Cancer Cells Retrace a Stepwise Differentiation Program during Malignant Progression

Sadegh Saghafinia1,2,3, Krisztian Homicsko1, Annunziata Di Domenico4, Stephan Wullschleger1, Aurel Perren6, Ilaria Marinoni4, Giovanni Ciriello2,3, Iacovos P. Michael1, and Douglas Hanahan1,5

Cancer Research. on November 15, 2021. © 2021 American Association for Cancer Discovery.
Pancreatic neuroendocrine tumors (PanNET) comprise two molecular subtypes, relatively benign islet tumors (IT) and invasive, metastasis-like primary (MLP) tumors. Until now, the origin of aggressive MLP tumors has been obscure. Herein, using multi-omics approaches, we revealed that MLP tumors arise from IT via dedifferentiation following a reverse trajectory along the developmental pathway of islet β cells, which results in the acquisition of a progenitor-like molecular phenotype. Functionally, the miR-181cd cluster induces the IT-to-MLP transition by suppressing expression of the MEIS2 transcription factor, leading to upregulation of a developmental transcription factor, HMG3. Notably, the IT-to-MLP transition constitutes a distinct step of tumorigenesis and is separable from the classic proliferation-associated hallmark, temporally preceding accelerated proliferation of cancer cells. Furthermore, patients with PanNET with elevated HMG3 expression and an MLP transcriptional signature are associated with higher-grade tumors and worse survival. Overall, our results unveil a new mechanism that modulates cancer cell plasticity to enable malignant progression.

**SIGNIFICANCE:** Dedifferentiation has long been observed as a histopathologic characteristic of many cancers, albeit inseparable from concurrent increases in cell proliferation. Herein, we demonstrate that dedifferentiation is a mechanistically and temporally separable step in the multistage tumorigenesis of pancreatic islet cells, retractiong the developmental lineage of islet β cells.

**INTRODUCTION**

Pancreatic neuroendocrine tumors (PanNET) are rare human cancers that arise from the endocrine cells of the pancreas, the islets of Langerhans, and are broadly classified as “functional” (secreting islet cell hormones, e.g., insulin) and “nonfunctional” (NF-PanNET; ref. 1). According to the WHO, PanNETs are subdivided into three grades as G1, G2, and G3 based on the mitotic index and proliferation index (by Ki-67; refs. 2, 3). The majority of PanNETs are sporadic and typically exhibit alterations in the MEN1 (~40%) and DAXX/ATRX (~35%) genes (4). Although novel therapies have improved the prognosis of low-grade PanNETs, high-grade tumors are invariably lethal (1), which motivates further efforts to better understand PanNET biology and the molecular programs that characterize high-grade tumors.

Genetically engineered mouse models of cancer have been instrumental in understanding the underlying mechanisms of tumor progression. RIP1-Tag2 (RT2) is a prototypical mouse model of PanNET, in which the expression of SV40 T antigen in the insulin-producing pancreatic islet β cells incapacitates p53 and Rb family tumor suppressors, and leads to stepwise tumor development and progression (5). During tumorigenesis, a subset of hyperplastic islets, which appear between 4 and 6 weeks of age, undergo an angiogenic switch between 6 and 9 weeks, and a subset of these “angiogenic islets” subsequently progress to form overt tumors. Histologically, the tumors are heterogeneous, representing either encapsulated solid tumors (adenomas) or invasive carcinomas, and a subset of cancer cells are capable of metastasizing to the lymph nodes and less frequently to the liver (5–7).

Using various transcriptomic data, we have previously reported three main molecular subtypes of human PanNETs, namely well-differentiated islet/insulinoma tumors (IT), “intermediate” tumors, and poorly differentiated tumors associated with liver metastasis, called metastasis-like primary (MLP; refs. 4, 6, 8). The IT subtype represents well-differentiated G1/2 PanNETs in humans, which are typically insulin-secreting, noninvasive, and poorly metastatic. The human intermediate subtype also consists of well-differentiated tumors, both functional and nonfunctional, typically enriched for loss of the MEN1 tumor suppressor. The human MLP subtype is the most heterogeneous, and consists of nonfunctional and poorly differentiated PanNETs (G3), and around half have associated liver metastasis (8). Molecular subtyping analysis of the RT2 model revealed that a majority of tumors belong to the insulinoma-like (IT) subtype, whereas a smaller subset presents as the more aggressive MLP subtype, which has the capability to metastasize to lymph nodes and liver (8, 9). Despite lacking the mutational alterations that are present at varying frequencies in human PanNET (4), the IT and MLP tumors that arise in RT2 model have highly similar mRNA and miRNA transcriptome.
profiles to the human counterparts, arguing that it is a valid model of human cancer (8). The human intermediate subtype is not phenocopied in the RT2 model, but rather in an engineered mouse model with β-cell–specific inactivation of Men1 tumor suppressor (10).

Previous studies have suggested two possible pathways for the development of the MLP subtype, based on descriptive analyses. The first proposed endocrine progenitors as the cell-of-origin for a separate tumorigenesis pathway leading to MLP tumors (6, 8, 11), whereas the second suggested dedifferentiation to the MLP subtype from preexisting cancer cells as the mechanism of malignant progression (7, 12). In lieu of definitive functional studies, there has remained a lack of clarity about the molecular mechanism of malignant progression to G3/MLP tumors. The integration of high-throughput multi-omics data has proved to be a powerful resource for investigating molecular determinants of tumor development (13). In this study, we used multi-omics approaches to characterize the two main PanNET subtypes (i.e., insulinoma/IT and MLP), aiming to investigate the origin of the aggressive MLP subtype. To this end, we profiled primary and metastatic lesions of the RT2 mouse model of PanNET, integrating single-cell transcriptomics with bulk mRNA and miRNA sequencing and proteomic analysis, followed by functionally perturbing elements of a regulatory pathway implicated in specifying the development of the aggressive MLP subtype of PanNET.

RESULTS
MLP PanNETs Reactivate Pancreatic β-cell Progenitor Signaling

To dissect the underlying genetic program of primary and metastatic PanNETs, we collected samples via laser-captured microdissection from primary tumors (31 samples) and liver metastases (6 samples) from the RIP1-TAG2 (RT2) genetically engineered mouse model in three different genetic backgrounds, as well as normal mouse pancreatic islets (4 samples), and normal mouse liver (3 samples; Fig. 1A; Supplementary Fig. S1A–S1D; Supplementary Table S1). The collected samples—based on sufficient availability—were profiled for mRNA, miRNA, and proteomic analysis (Supplementary Fig. S1D). First, to assess the purity of the samples for cancer cells, we applied a cell-type deconvolution method—xCell (14)—to the samples that were subjected to RNA-sequencing (RNA-seq) analysis. This revealed a low level of immune and stroma scores, indicative of the high proportion of cancer cells in the bulk samples (Supplementary Fig. S1D). To evaluate the overall heterogeneity of the samples, we separately performed clustering analysis on all the samples from each platform, using the nonnegative matrix factorization (NMF) method (15). Subsequently, a multi-omics clustering method—“Similarity Network Fusion” (SNF; ref. 16)—was applied to integrate information from the distinctive datasets—mRNA, miRNA, and protein. Because more samples were profiled for both mRNA and miRNA, in a separate analysis, we also applied SNF on the paired samples from these two datasets. The multi-omics clusters defined by SNF were highly concordant with the NMF clusters derived from a separate analysis of mRNA, miRNA, and proteomics profiles (Supplementary Fig. S1D).

In agreement with our previous reports (6, 8), the tumors segregated into two broad subtypes with highly distinctive mRNA, miRNA, and proteomic profiles (Fig. 1A; Supplementary Fig. S1D). The first cluster, denoted as IT, contained only primary tumors, which maintained high expression of well-defined markers of mature β cells, including Ins1/2, Insm1, Iapp, Nkx6, Pax6, Pdx1, and Chga (17, 18), as detected in both transcriptomic (Supplementary Fig. S1E) and proteomic analyses (Fig. 1B). The second cluster, dubbed by Olson and colleagues (6) as “metastasis-like primary” tumors (MLP), contained both primary and metastatic samples and was characterized by low expression of mature β-cell markers, along with an elevated expression of a number of endocrine progenitor markers (Fig. 1C; Supplementary Fig. S1F), including Hmgbl2/3, Fev, and Myc (19). The Fisher exact test revealed no association between the IT and MLP phenotypes with the mouse strains from which the tumors arose (BL6 vs. AJ in all samples; P = 0.32, BL6 vs. AJ in RNA-seq samples; P = 0.13).

To further characterize the MLP tumors, a miRNA signature and an mRNA signature for the MLP cluster were developed using differential expression analysis and NMF-selected features (see Methods). We identified a 28-member miRNA signature and a 203-member mRNA signature that each distinguished the IT and MLP subtype samples (Supplementary Table S1). Next, functional enrichment analyses [Gene Ontology (GO) terms] for upregulated and downregulated genes in the MLP mRNA signature were performed. Downregulated genes were enriched in GO terms associated with hormone secretion, pancreas development, cell–cell adhesion, and regulation of insulin secretion (Fig. 1D). The majority of the genes in these categories are involved in the homeostasis of mature β cells, consistent with the observation that MLP tumors suppress mature β-cell markers.

Notably, repressors of cell differentiation and developmental processes were among the most significantly enriched GO terms for upregulated genes in the MLP mRNA signature (Fig. 1E). The MLP tumors had high expression of genes previously reported to be involved in maintaining stem-like features, playing essential roles in embryonic development and regeneration and in the epithelial-to-mesenchymal transition, such as Sox11 (20, 21), Sox6 (22), Cited1 (23), Id1 (24, 25), and Zfp536 (ref. 26; Supplementary Table S1). These results are in agreement with our previous study by Sadanandam and colleagues (8), in which we observed higher expression of pancreatic progenitor-specific markers in MLP tumors (Fig. 1C; Supplementary Fig. S1F). Following up on this observation, we hypothesized that MLP tumors might have undergone dedifferentiation, thereby enabling the transition from IT to MLP subtype.

To begin assessing the dedifferentiation hypothesis, we leveraged preexisting knowledge about islet β-cell differentiation from pancreatic progenitors to fully mature islet cells. Both human and mouse β cells develop through three sequential phases of differentiation. Specifically, in mice, a primary transition takes place from E9.5 to E12.5, a secondary transition from E12.5 to birth, and finally, a postnatal maturation phase from birth to weaning (Fig. 1F; refs. 27–29). Querying the presence of the MLP mRNA signature in two separate datasets profiling the secondary transition (GSE8070) and...
Retracing Stepwise Differentiation during Tumor Progression

Figure 1. The aggressive MLP subtype of PanNET has a similar expression profile to islet β-cell progenitors. A, Multi-omics clustering of samples isolated via LCM from primary tumors and liver metastasis from the RT2 mice, as well as pancreatic islets and liver from wild-type mice, in the C57BL6/N, C57BL6/J, and A/J genetic backgrounds. Samples of the IT and MLP subtype are colored as green and red, respectively. B and C, Heat maps of protein levels, detected by mass spectrometry from LCM samples, of mature β-cell markers (B) and endocrine progenitor markers (C) in IT and MLP PanNET samples.

D and E, GO categories significantly enriched (FDR q-value on the y-axis) for downregulated (D) and upregulated (E) genes in the MLP transcriptome signature. F, Schematic representation of mouse endocrine pancreas development depicting the stages from early progenitors (E9.5) to fully mature β cells (P60). G, Score of the MLP mRNA signature in different embryonic stages during the secondary transition of mouse pancreatic development (left) and postnatal maturation (right). In both datasets there MLP score declines in the curse of differentiation. H, The MLP mRNA signature scores are high in endocrine pancreatic progenitor cells (E10), but not in mature β cells. I, PCA of samples from the temporal mouse pancreatic islet postnatal maturation phase compared with primary tumors and liver metastasis from the RT2 mouse model along with normal (nontransgenic) islets. The x-axis shows PC2, which represents a surrogate for dedifferentiation time points. The samples from each cohort are separated in different rows (see also Supplementary Fig. S1K).
the postnatal maturation of β cells (19) revealed that the less differentiated pancreatic/β-cell samples had a higher MLP mRNA signature score (Fig. 1G). These results indicated that the MLP transcriptome profile is active in normal β-cell progenitor cells. This observation was further substantiated by applying the MLP mRNA signature onto a single-cell profile of β-cell postnatal maturation (19), which showed a high correlation between the MLP gene signature and less-differentiated states, independent of Mki67 proliferation marker (ANOVA P = 6.52e-23; Supplementary Fig. S1G–S1I). Furthermore, by applying the MLP miRNA signature as a filter in an independent dataset profiling miRNA expression of pancreatic progenitor cells and mature β cells (30), we again found enrichment of MLP miRNA expression in pancreatic progenitor cells (Fig. 1H).

In addition, the enrichment analysis of upregulated genes revealed, among others, “neurogenesis/cetal nervous system (CNS) development” and “metabolic processes” as highly enriched categories (Fig. 1E). Notably, these categories largely overlapped with the enrichment analysis of upregulated genes in embryonic E17.5 progenitors versus mature β cells (Supplementary Fig. S1J). Concordantly, we have previously reported that PanNETs exploit neuronal synaptic signaling pathways to acquire invasive capabilities (31), and that altered cell metabolism is a feature of aggressive MLP tumors (8).

Next, we sought to evaluate the overall transcriptomic program of PanNETs further and assess the potential similarity between the two subtypes and the distinctive stages of β-cell differentiation. To this end, the ComBat algorithm (Combating Batch Effects When Combining Batches of Gene Expression Microarray Data; ref. 32) was used to merge RNA-seq datasets describing the stages of mouse postnatal maturation (19) with those of tumor samples. Principal component analysis (PCA) on the merged dataset (Supplementary Fig. S1K) found PC2 as a highly correlated component with postnatal maturation time points (ANOVA P = 0.00224; Fig. 1I); therefore, PC2 was considered as a proxy of the differentiation timeline trajectory. As expected, RNA samples from normal adult islets collected along with the PanNETs grouped closer to the dataset of mature β cells (P00) β cells. Notably, all samples from IT tumors grouped with intermediate-mature P9–P18 β cells, and MLP samples clustered with P0–P3 immature β cells, whereas (presumably more aggressive) metastatic samples clustered closer to the inception point of the β-cell differentiation trajectory (Fig. 1I).

Collectively, these results demonstrate that MLP tumors share characteristics with pancreatic progenitor cells and pose the intriguing hypothesis that MLP tumors emerge from the progression of IT cells via dedifferentiation and reactivation of progenitor-like signaling pathways.

**miR-181cd Induces the Activation of the Progenitor-Like Program in IT-Like Cell Lines**

To evaluate the hypothesis that dedifferentiation is a discrete step in tumorigenesis that gives rise to MLP tumors, we sought to identify signaling pathways that induce the MLP phenotype in IT-like cancer cells. miRNAs have been shown to regulate diverse signaling pathways and biological processes, including cellular reprogramming (33, 34), and we have previously shown that a set of miRNAs is differentially expressed in MLP tumors compared with IT tumors (6, 8). To identify candidate miRNAs with the potential to activate the MLP program in IT-like cell lines, the most highly upregulated MLP-associated miRNAs were sorted according to their correlation with the MLP-mRNA signature score in the RT2 tumor samples dataset (Supplementary Fig. S2A). In addition, we segregated miRNAs that were differentially expressed between pancreatic progenitor cells and mature β cells (Fig. 2A). This analysis led us to focus on two members of the miR-181 family, namely miR-181c and miR-181d, which compose the miR-181cd cluster (Fig. 2B).

To assess a possible functional role, the miR-181cd cluster was conditionally overexpressed in the βTC3 cell line, which showed IT-like phenotype (Supplementary Fig. S2B). To this end, we used a ppgyBac transposon system enabling doxycycline-inducible miRNA expression (35). RNA-seq analysis was performed for collected samples at 24 hours and seven days after miR-181cd overexpression (Supplementary Fig. S2C and S2D), and activation of the progenitor-like program was evaluated by applying the MLP mRNA signature onto the transcriptome profiles of the samples. Intriguingly, seven days of miR-181cd expression in the βTC3 IT-like cancer cells resulted in the transition of these cells toward the MLP subtype (Fig. 2C).

Congruently, IT marker genes, such as Ins1, Ins2, and Iapp, were downregulated upon miR-181cd expression, whereas various markers of MLP, including Peg10, members of the ID family of transcriptional regulators, and the Miat long noncoding RNA (lncRNA), were among the upregulated genes (Fig. 2D). Moreover, functional enrichment analysis of significantly upregulated genes after seven days of miR-181cd expression (Supplementary Table S2) revealed multiple GO terms associated with neurogenesis and cell differentiation regulation as the most enriched categories (Fig. 2E).

In addition to a significant shift in the transcriptomic program and upregulation of genes involved in dedifferentiation, βTC3 cells underwent a morphologic change upon induction of miR-181cd expression, characterized by the development of neuronal-like structures, similar to the morphology that is exhibited by the MLP-like cell line (A1-5257–1; Fig. 2F). These morphologic changes were concordant with the enrichment of neurogenesis GO term categories for upregulated genes both in the MLP subtype (Fig. 1E) and in β-cell progenitors (Supplementary Fig. S1G), as well as upregulated genes after seven days of miR-181cd expression (Fig. 2E). In addition, we also evaluated the effect of miR-181cd expression in the 99-3o IT-like cell line (Supplementary Fig. S2B). Similar to the βTC3 cells, upon expression of miR-181cd in the 99-3o cells (Supplementary Fig. S2C and S2D), we observed downregulation of Ins2, an IT marker (Supplementary Fig. S2E), as well as morphologic changes, which were accompanied by the appearance of neuronal-like structures (Supplementary Fig. S2F).

Although the βTC3 parental cells expressed higher endogenous levels of miR-181cd compared with the 99-3o cells (Supplementary Fig. S2C and S2D), overexpression of the miR-181cd transgene in both IT-like cell lines was necessary to induce gene expression and morphology changes that resemble the MLP subtype. We suspect that the comparatively higher levels of endogenous miR-181cd in the βTC3 versus 99-3o cells might be due to sequestration by competitive endogenous
Figure 2. miR-181cd induces dedifferentiation when upregulated in IT cancer cells. 

A. Heat map of differentially expressed MLP-signature miRNAs comparing mature β cells and pancreatic progenitor cells from wild-type mice. B, Scatter plot showing the correlation between miR-181c (top) and miR-181d (bottom) expression (x-axis) and the MLP mRNA signature score (y-axis) in RT2 PanNET samples. C, MLP mRNA signature scores in cell cultures collected 24 hours and 7 days following induction of miR-181cd expression in the βTC3 IT-like cancer cell line, which reveals a transcriptional shift toward MLP after 7 days. D, Heat map of MLP mRNA signature genes in miR-181cd-transfected βTC3 cells, which indicates that the majority of upregulated MLP genes (colored coded as red in the y-axis) were upregulated after 7 days of miR-181cd expression; conversely, the genes downregulated in the MLP signature (colored coded as green in the y-axis) were congruently downregulated after 7 days of miR-181cd expression. E, GO categories that were significantly enriched (FDR q-value on the y-axis) for upregulated genes after 7 days of miR-181cd cluster overexpression in βTC3 cells. F, Brightfield images of cancer cells: (i) βTC3 control; (ii) βTC3 miR-181cd 2 weeks; (iii) AJ-5257-1 MLP-like cells. Bottom, magnified image illustrating miR-181cd-induced neuronal-like structures (in dashed lines) in βTC3 cells.
Prospective negative regulators might be potential targets of the miR-181cd cluster. To this end, a recently developed algorithm that enables the identification and ranking of biologically relevant miRNA targets (Bio-miRTa; ref. 35) was used to query two datasets: (i) downregulated genes in the MLP signature and (ii) downregulated genes in the miR-181cd overexpression experiment (Supplementary Table S2). *Mei2*, which encodes for a homeobox protein that belongs to the three amino acid loop extension (TALE) family of homeodomain-containing proteins, ranked as the top miR-181cd gene target (Supplementary Fig. S3A). *Mei2* has been implicated in prostate cancer and neuroblastoma (41, 42), and VIPER analysis revealed that the genes downstream of *Mei2* are also altered significantly between IT and MLP samples (Fig. 3A). Congruently, *Mei2* was downregulated upon forced miR-181cd expression in both βTC3 and 99-3e IT-like cell lines (Fig. 3B; Supplementary Fig. S3B), as well as in MLP compared with IT tumors (Supplementary Fig. S3C). Using reporter assays, we verified the ability of miR-181cd to bind to the miRNA response element (MRE) in the *Mei2* UTR of *Hmgb3* (Fig. 3C; Supplementary Fig. S3D).
To then identify and prioritize potential inducers of the IT-to-MLP dedifferentiation, we queried candidate TFs from the VIPER analysis in four mRNA datasets: (i) the secondary-transition phase of pancreatic development, (ii) the β-cell postnatal maturation phase, (iii) the dedifferentiation trajectory of primary cancer cells (PC2 in Fig. 1I), and (iv) one resultant to miR-181cd overexpression in the βT3C IT-like cells (Supplementary Table S2). Hmgb3, an X-linked member of the high-mobility group box (HMGB) family, was the only gene that scored significantly high in all four datasets (Fig. 3D; Supplementary Fig. S3E–S3G), and the VIPER-inferred downstream genes of Hmgb3 were also significantly altered comparing MLP with IT samples (Fig. 3A). Hmgb3 has been shown to modulate transcription, replication, recombination, DNA repair, and genomic stability (43), and its expression is associated with invasion, metastasis, and poor prognosis of a number of human cancers (44). Concordantly, with these implications, Hmgb3 was highly expressed in MLP samples at the mRNA and protein level (Supplementary Fig. S3H and S3I), and it was upregulated upon forced miR-181cd expression in the βT3C and 99–30 IT-like cell lines (Fig. 3D and E; Supplementary Fig. S3J and S3K).

**Meis2 and Hmgb3 TFs Regulate Dedifferentiation to the MLP Subtype**

To assess the suspected regulatory pathway involving miR-181cd, Meis2, and Hmgb3 in the transition from the IT to the MLP phenotype, an inducible shRNA knockdown system (as we have described previously in ref. 45), was used to downregulate the miR-181cd target, Meis2, in the IT-like βT3C cell line (Supplementary Fig. S3L). Meis2 downregulation led to the upregulation of Hmgb3 at the mRNA and protein level (Supplementary Fig. S3M and S3N), concomitant with downregulation of the IT marker Ins2 (Supplementary Fig. S3O).

Next, to ascertain the role of Hmgb3, we used the piggyBac (doxycycline)-inducible system to overexpress Hmgb3 in the βT3C cell line (Fig. 3F; Supplementary S3P). To characterize the effect of Hmgb3 at the transcriptomic level in this IT-like cell line, RNA-seq analysis was performed for samples collected seven days after Hmgb3 overexpression, as well as for control cells. Subsequently, by applying the MLP mRNA signature to the transcriptome profiles, we observed that the βT3C cell line transitioned toward the MLP subtype upon seven days of forced Hmgb3 expression (Fig. 3G). To assess the similarity in MLP gene network activation between the overexpression of miR-181cd and Hmgb3, we merged the two datasets (see Methods) and ran hierarchical clustering on the merged MLP–mRNA gene expression matrix. Intriguingly, the result showed two main clusters, with the Hmgb3 overexpression dataset clustering closest to the transcriptome profile of miR-181cd overexpression in βT3C cell line (Fig. 3H).

Similar to miR-181cd, seven days of Hmgb3 overexpression caused downregulation of IT markers such as Ins2 and Nkx6–1 (Fig. 3I). Differential expression analysis of Hmgb3-overexpressing and control cells revealed increased expression of multiple genes involved in regulating neuronal programs, pluripotency, and cellular morphogenesis and migration (Fig. 3J; Supplementary Table S2). Consistently, the majority of these genes were similarly upregulated by miR-181cd overexpression (Supplementary Fig. S3Q–S3U). Overall, these results establish Hmgb3 as a key downstream effector of the miR-181cd–induced transition from IT cells to the dedifferentiated MLP subtype.

**Hmgb3 Is a Marker of Dedifferentiation and Metastasis-Like Cells**

Seeking to substantiate the evidence for dedifferentiation as a discrete regulatory event, we assessed the proposed miR-181cd/Meis2/Hmgb3 axis in vivo. First, we immunostained tumor lesions for Insulin and HMG3, as well as for the oncoprotein T-antigen (to identify cancer cells) from two RT2 strains, that is, AB6F1 and C57BL6/N, at early (7–9 weeks), middle (10–11 weeks), and advanced (14–16 weeks) stages of tumorigenesis. As expected, cancer cells (T-antigen+ cells) with Ins+/HMG3+ expression marked the IT tumors (Fig. 4A; Supplementary Fig. S4A). Notably, the initial down-regulation of Insulin was coincident with HMG3 upregulation, as indicated by the presence of Ins+/HMG3+ cancer cells (Fig. 4A), which marked the transition from IT to MLP subtype emerging at the early stages of tumorigenesis. The IT-to-MLP transition was followed by complete loss of Insulin and high expression of HMGB3 in MLP lesions, indicated by Ins+/HMGB3+ cancer cells, in both early- and late-stage primary tumors (Fig. 4A; Supplementary Fig. S4B). These results indicate that the induction of HMGB3 is an early event in the dedifferentiation process of IT cancer cells into MLP subtype.

Examining the tumor sections from three different temporal stages of the disease showed an increase in the incidence of the IT-to-MLP and MLP lesions for both models during tumor progression (Supplementary Fig. S4C and S4D). Furthermore, when lymph node and liver metastases in the highly metastatic RT2;AB6F1 model were examined, all lesions proved to be negative for the IT marker Insulin, and uniformly positive for HMG3, providing further evidence that HMGB3 marks the aggressive and metastatic cancer cells of the MLP subtype, which consistently express this MLP marker, even after metastasizing and colonizing secondary parenchyma (Fig. 4B; Supplementary Fig. S4B and S4E).

Next, we sought to investigate the intratumoral heterogeneity at the single-cell level for IT and MLP subtypes in advanced lesions from the prototypical RT2;C57BL6/N mouse model. Accordingly, two individual tumors were collected from two 14-week-old RT2 mice. The samples were processed separately, and single-cell RNA-seq (scRNA-seq) was performed for 2,375 individual cells in total. The clustering analysis was performed to identify and visualize the cells with similar expression patterns (see Methods; Supplementary Fig. S5A and S5B). Using known gene markers (Supplementary Fig. S5C), we annotated different cell populations of the tumor microenvironment in primary PanNETs (Supplementary Fig. S5D), which revealed a similar ratio of various immune (infiltrating B- and T leukocytes, macrophages, and neutrophils) and stromal cell types [endothelial cells, pericytes, and cancer-associated fibroblasts (CAF)] in both primary tumor samples (Supplementary Fig. S5E).

Cancer cells were identified by expression of the SV40 large T-antigen oncogene and by a PanNET mRNA signature score comprising a 62-gene set diagnostic for transformed islet β cells that was developed from the bulk RNA-seq data (Supplementary Fig. S5F and S5G; Supplementary Table S1). Clusters...
**Figure 4.** Hmgb3 upregulation is an early event in the dedifferentiation of IT into MLP. **A** and **B**, Representative images of Insulin and HMG3 immunostaining, alongside DAPI, of early pancreatic tumors (**A**) and liver metastasis (**B**) in the RT2;AB6F1 mouse model of PanNETs. **A**, Representative images of early lesions showing: IT cells; identified as Ins$^\text{hi}$/Hmgb3$^\text{neg}$ (left), cells undergoing IT-to-MLP transition; identified as Ins$^\text{hi}$/Hmgb3$^\text{lo}$ and Ins$^\text{lo}$/Hmgb3$^\text{hi}$ (middle), and fully dedifferentiated MLP cells; identified as Ins$^\text{neg}$/Hmgb3$^\text{hi}$ (right; see also supplementary Fig. 4A–4D). **B**, Metastatic cancer cells exhibit Ins$^\text{neg}$/Hmgb3$^\text{hi}$ expression pattern (see also Supplementary Fig. 4E). **C**, t-SNE analysis for all cancer cells based on scRNA-seq analysis, color coded according to the seven distinct subclusters. **D**, Individual cancer cells (points) in two-dimensional t-SNE plots, color-coded according to the MLP signature score. **E**, mRNA expression levels of Ins2 (left), Meis2 (middle), and Hmgb3 (right) in scRNA-seq data from primary tumor samples. **F**, Schematic representation of dedifferentiation process during PanNET tumor progression: The IT cancer cells (Ins$^\text{hi}$/HMBG3$^\text{neg}$) go through dedifferentiation and transition to MLP subtype (Ins$^\text{neg}$/HMBG3$^\text{hi}$), enabling them to disseminate to the liver. This transition is induced by upregulation of miR-181cd, which directly inhibits Meis2 expression, and indirectly effects the upregulation of Hmgb3 expression.
with average signature score above –0.4 were selected as cancer cells. Cluster 2 was excluded from the follow-up analysis due to low RNA-seq reads, and cluster 13 was excluded due to the low number of SV40-expressing cells. Although cluster 6 also scored highly for the PanNET signature, this cluster is comprised of macrophages (Csf1r-positive cells; T-antigen negative, Supplementary Fig. S5H), and we envisage that phagocytosis incorporated the so-called “passenger” transcripts originating from engulfed apoptotic cancer cells (46). Accordingly, the cells within clusters 0, 1, 3, 4, and 5 were classified as cancer cells.

Next, clustering analysis was performed specifically for the selected cancer cells, which revealed seven distinct populations of cancer cells (cancer subclusters i–vi) with different transcriptome profiles (Fig. 4C; Supplementary Fig. S5I); both of the analyzed tumors showed a similar distribution in these subclusters (Supplementary Fig. S5J). Although cancer cells had variable levels of SV40 expression, all subclusters demonstrated a similar level of the PanNET signature score (Supplementary Fig. S5K). Seeking to delineate cancer cells into IT versus MLP subtypes, the MLP mRNA signature was overlaid onto the single-cell transcriptome data. Subcluster i demonstrated the lowest MLP score, with high expression of mature β-cell markers such as Ins2 (Fig. 4D and E; Supplementary Fig. S5L). Relative to subcluster i, the rest of the six subclusters in the primary tumor demonstrated high activity of the MLP transcriptomic program (Fig. 4D). Therefore, we concluded that subcluster i represents the IT cancer cells, and subclusters ii–vi are MLP cancer cells within primary tumors. Notably, we observed an inverse correlation of the Meis2 and Hmgb3 expression in the IT (subcluster i) and MLP cells (subclusters ii–vi; Fig. 4E; Supplementary Fig. S5L), supporting our previous data implicating the Meis2/Hmgb3 axis in the IT-to-MLP transition.

Collectively, these results establish a miRNA/TF regulatory pathway, wherein miR-181cd induces downregulation of the Meis2, which leads to upregulation of Hmgb3 expression. This pathway evidently plays an important role in the induction of the dedifferentiation process and the consequent transition of IT into MLP cancer cells, contributing thereby to the acquisition of invasive and metastatic capabilities (Fig. 4F).

**scRNA-seq Reveals a High Proliferation Rate in Late-Stage MLP Tumors**

Next, we used the PCA on the transcriptomic data from the scRNA-seq to investigate the gene regulatory pathways contributing to the heterogeneity of the late-stage tumors. Considering the first two components of PCA (Fig. 5A; Supplementary Fig. S6A), PC1 clearly showed a high correlation with the MLP mRNA signature score, separating IT from MLP cancer cells (Fig. 5B; Supplementary Fig. S6B). Congruently, functional enrichment analysis on differentially expressed genes showed that the categories associated with the regulation of hormone and insulin secretion in normal β cells were enriched in IT cancer cells (subcluster i, Fig. 5C; Supplementary Table S3). Conversely, categories related to cellular homeostasis, regulation of cell differentiation, and CNS development were associated with MLP cancer cells (subclusters ii–vi, Fig. 5D; Supplementary Table S3).

Intriguingly, in addition to GO terms representing MLP phenotype, the upregulated genes in MLP clusters exhibited high enrichment of functional categories associated with regulation of cell division and cell cycle (Fig. 5D; Supplementary Table S3). Therefore, we hypothesized that proliferation might be another major pathway contributing to the phenotypic state of cancer cells in the primary tumor. Accordingly, the second most variable component in the PCA (PC2) revealed a high correlation with cellular proliferation rate (Fig. 5E), with the highest significance for MLP cancer cells (subclusters ii–vi, p = 0.96, P < 2.2 e-16). Concordantly, the MLP clusters within the primary tumors exhibited different levels of proliferation capability (Fig. 5F) and Mek67 expression (Supplementary Fig. S6C), showing the significant contribution of proliferation regulatory network in cellular heterogeneity of late-stage MLP cancer cells. Thus, the initial MLP subclusters ii and iii had a low proliferation score, whereas the progressive MLP subclusters iv, v, and vi had a high score, with the intermediate subcluster iv highly variable, suggestive of a transition phase. Notably, the proliferation status of the cancer cells in these seven subclusters was independent of the variable expression of the driving SV40 large T-antigen oncogene (Supplementary Fig. S6D and S6E).

To further investigate the subpopulations of MLP cancer cells, differential expression analysis was performed on each subcluster, followed by GO term enrichment analysis (Supplementary Table S3). As expected, subcluster i showed a high expression of β-cell markers, including Ins1/2, Nkx-6-1, Iapp, and Mafa (Fig. 5G). On the other hand, subcluster ii, the first cluster with high expression of MLP signature genes (Supplementary Fig. S6B), was enriched in GO terms representing negative regulation of differentiation (Supplementary Table S3). Interestingly, we observed upregulation of the islet α cell gene markers Gip (encoding for glucagon) and Mafb in subcluster ii. The switch from Mafb to Mafa expression is vital for islet β-cell maturation (28); hence, the transition from Mafa expression in subcluster i to Mabf expression in subcluster ii can be envisaged to signify the activation of dedifferentiation in cancer cells toward an endocrine progenitor-like state (Fig. 5G). Genes distinguishing subcluster iii proved to be involved in the G- to S-phase transition, as well as DNA repair and chromosomal stability; while subcluster iv had the highest expression of genes associated with metabolism and hypoxia (Fig. 5G; Supplementary Table S3). This result suggests the presence of unfavorable conditions in the tumor microenvironment, such as hypoxia, experienced by a subset of MLP cancer cells. Finally, functional enrichment analysis for the differentially expressed genes in the last three subclusters (v, vi, and vii) revealed cell division and cell-cycle regulation as the most enriched categories (Fig. 5G; Supplementary Table S3). Most of the upregulated genes in these clusters were involved in the last phases of mitosis and G2–M phase of the cell cycle, showing the hyperproliferative status of the cancer cells in these three subclusters (Fig. 5G).

These results collectively establish dedifferentiation and proliferation as the two qualitatively distinct cellular programs shaping the heterogeneity of late-stage primary tumors in the mouse model of PanNET.

**Dedifferentiation Precedes the Activation of Hyperproliferation during Tumorigenesis**

The hyperproliferative state of MLP cancer cells at advanced stages prompted us to investigate whether induction of dedifferentiation is temporally associated with substantive increases in the rate of cancer cell proliferation. Therefore, we examined the proliferation index of IT and MLP lesions in the early
**Figure 5.** Dedifferentiation and proliferation are two separate molecular pathways contributing to the heterogeneity of primary tumors. A, PCA of all cancer cells based on scRNA-seq analysis, color-coded according to the cell subclusters. B, Scatter plot showing a high correlation of PC1 from the PCA analysis and the MLP mRNA-signature score. C and D, GO categories that are significantly enriched (FDR q-value on the y-axis) for upregulated genes in subcluster I (C; IT cancer cells), and in subclusters II–VII (D; MLP cancer cells). E, Scatter plot showing a high correlation of PC2 from PCA with the mRNA proliferation signature score for both IT (subcluster I) and MLP clusters (subclusters II–VII). F, Violin plot showing the mRNA proliferation signature score for each cancer cell subcluster. G, Heat map of top 20 differentially expressed genes for each cancer cell subcluster in the primary tumor. H, Images of Insulin, HMGBl, and EdU immunostaining, along with DAPI, for early lesions from 7- to 8-week-old RT2.86F1 mice, illustrating representative IT, MLP, and transitional histologic stages. I, Quantification of EdU immunostaining to reveal proliferation of cancer cells in IT and MLP tumor lesions isolated from early, 7- to 8-week-old, RT2.86F1 mice. J, Cell-cycle analysis of BTC3 cells before [control] and after 7 days of miR-181c doxycycline-induced expression.
stages of tumorigenesis by immunostaining the small, well-separated tumor lesions that are typical in young (8–10-week-old) RT2AB6F1 mice. To this end, Insulin and HMGB3 were used as biomarkers to distinguish IT and MLP subtypes, respectively, while EdU immunostaining marked proliferating cells. Notably, a similar number of proliferating cells was observed in both IT and MLP lesions in this early phase of tumor growth (Fig. S5H and I), indicating that dedifferentiation does not lead to an immediate increase in the proliferative capacity of cancer cells.

To further validate the immunostaining data, the effect of the miR-181c expression on proliferation was examined by performing cell-cycle analysis of βTC3 IT-like cancer cells seven days after miR-181c overexpression. While overexpression of miR-181c induced dedifferentiation in βTC3 cell line, we found no difference in proliferation status of the cells before and after miR-181c induction (Fig. 5j; Supplementary Fig. S5F). Concordantly, the GO terms enrichment analysis of differentially expressed genes after seven days of miR-181c expression did not show any categories associated with the cell cycle (see Fig. 3D). Overall, these results suggest that dedifferentiation is an early event in tumor progression, one that does not directly affect the proliferation capability of the cancer cells. This observation agrees with the previous data from the scRNA-seq analysis of islet β cells during postnatal maturation that similar revealed a dissociation between the β-cell differentiation program and proliferation (see Supplementary Fig. S1G–S1I).

Finally, we evaluated the proliferation rate in late-stage IT and MLP lesions from older (14- to 15-week-old) RT2 mice, again via immunostaining for Insulin, HMGB3, and EdU. Consistent with the scRNA-seq data, late-stage MLP lesions showed a higher proliferation rate than late-stage IT lesions (Supplementary Fig. S6G–S6I). However, late-stage IT tumors exhibited a higher rate of proliferation compared with IT lesions in the early stage (Fig. 5i; Supplementary Fig. S6H). Collectively, these data reveal that induction of the progenitor-like program is separable from and temporally precedes activation of the hyperproliferative signature in MLP cancer cells.

The MLP Cluster in Human PanNET Is Associated with Higher Grades, More Frequent Metastasis, and Poor Prognosis

To investigate the possibility that the described dedifferentiation pathway and its regulators might be operative in human PanNET, we applied the MLP mRNA signature onto a cohort of 110 human samples from both primary and metastatic tumors (47). Mirroring the approach used above to analyze mouse PanNET, RNA-seq datasets from mouse postnatal maturation (19) and human PanNETs (47) were merged. PCA (Supplementary Fig. S7A) revealed that PC2 correlated with the postnatal maturation timeline trajectory of developing β cells (Fig. 6A; \( \rho = 0.53; P = 2.09e-09 \)). Accordingly, human MLP tumors, identified by applying the mouse MLP mRNA signature, were most similar to immature β cells (Fig. 6A), suggesting that human cancer cells follow the same reverse developmental trajectory to acquire a progenitor-like phenotype that ostensibly contributes to increased malignancy.

Then we queried the MLP mRNA signature in a cohort of 29 human tumor samples with available clinical follow-up data (4, 48). The MLP signature had a significantly higher score in more aggressive human PanNETs, based on grade and T stage (Fig. 6B and C). The MLP mRNA signature was also assessed in a cohort of primary PanNETs and liver metastasis (47), which revealed a significantly higher signature score in metastatic samples (Supplementary Fig. S7B). Finally, the MLP mRNA signature was associated with overall survival: patients with a high MLP score had poorer prognosis (Fig. 6D); in notable contrast, tumor grade and stage in the same cohort of patients did not have any predictive value (Supplementary Fig. S7C and S7D).

To assess the role of miR-181c in tumor progression in the human disease, miRNA profiling of a cohort of human PanNETs (49) was used to examine the correlation of miR-181c expression with clinicopathologic features. Intriguingly, we found that high miR-181c and miR-181d expression was associated with more aggressive and nonfunctional PanNETs (Supplementary Fig. S7E–S7G).

Then, to explore the predicted anticorrelation of MEIS2 and HMGB3 in human PanNETs, their expression was assessed in the human PanNET datasets (4, 47, 48). Consistent with the data from the PanNET mouse model, we observed a significant negative correlation between the mRNA expression of these two TFs in human patients (Supplementary Fig. S7H). Interestingly, MEIS2 and HMGB3 expression exhibited negative and positive correlation with more aggressive tumors, respectively (Fig. 6E and F; Supplementary Fig. S7I and S7J). This result is in agreement with a previous report showing downregulation of MEIS2 in metastatic pancreatic neuroendocrine neoplasm (50).

Furthermore, we used a cohort of archival human PanNET tissue microarrays (TMA), which was immune-stained for HMGB3 protein. Congruently, HMGB3-positive sections were associated with higher grade, T stage, and metastasis (Fig. 6G–I), suggesting that this TF is also operative in specifying the human MLP PanNET subtype. Subsequently, we performed double immunostaining of the human PanNET TMA for the IT and MLP markers Insulin and HMGB3, respectively. The results revealed an anticorrelation for the two markers; high insulin-positive cells were HMGB3-negative, while insulin-negative cells were HMGB3-positive (Fig. 6J). Notably, similar to the mouse model, a small portion of cells had low expression of both markers, suggestive of an ongoing transition from IT to MLP phenotype (Fig. 6J; green arrows). Furthermore, we also observed interspersed IT (insulin-positive), and MLP (HMGB3-positive) cells in the same specimen, revealing intratumoral heterogeneity of human tumors with respect to IT and MLP subtypes (Supplementary Fig. S7K).

Finally, the majority of liver metastases were negative for insulin while expressing high levels of HMGB3 (Fig. 6K).

These results implicate miR-181/MEIS2/HMGB3-mediated dedifferentiation during malignant progression of human PanNETs to the more aggressive and metastatic MLP subtypes.

**DISCUSSION**

PanNETs are histologically diverse and oligo-mutated tumors lacking prevalent driving mutations in oncogenic signaling (1, 4). Studies in mouse PanNET models have suggested two alternative tumorigenesis pathways leading to the predominant molecular and histopathologic subtypes, dubbed IT and MLP. One schematic envisioned parallel pathways, wherein β-cell progenitors were the cell-of-origin for
the aggressive MLP subtype, and mature islet β-cell spawned benign ITs (6, 8). The alternative scenario suggested dedifferentiation of relatively benign ITs into the invasive and metastatic MLP subtype (7). These alternative pathways were proposed on the basis of comparative analyses of the molecular profiles of the IT and MLP tumors, but neither has been functionally validated so as to establish which mechanism is actually operative.

Although we cannot exclude the possibility that the aggressive MLP subtype can also arise directly from β-cell progenitor cells, herein we present compelling lines of evidence and functional validation that dedifferentiation and reactivation of a β-cell progenitor transcriptional program dictates the transition from the IT to the MLP subtype. First, our computational analysis of single-cell and bulk mRNA transcriptome profiling, and miRNA transcriptomic data, along with proteomic data, revealed a transition phase from IT to MLP tumors concordant with the activation of a β-cell progenitor-like program. Second, the miR-181cd cluster was identified as a regulatory factor that functionally contributes to the IT-to-MLP transition. Expression of the miR-181cd cluster was enriched both in normal pancreatic islet cell progenitors and in the MLP tumor subtype, and overexpression of miR-181cd in IT-like cancer cells activated a progenitor-like molecular phenotype and led to distinct morphologic changes similar to the MLP subtype. Third, we uncovered elements of a regulatory circuit mediating dedifferentiation of PanNETs,
in which the transcription factors Meis2 and Hmgb3 act as negative and positive regulators, respectively. Notably, Meis2 is a direct target of miR-181c/d, and its downregulation leads to the induction of Hmgb3 expression. Hmgb3 expression correlated both with β-cell progenitor stages and with the MLP signature, and its overexpression in an IT-like cell line similarly induced the IT-to-MLP transition. Concordantly, IHC revealed a progressive increase in HMG3 expression correlated with decreasing expression of the IT marker Insulin, demonstrating a gradual phase transition from IT to MLP subtypes. Finally, lymph node and liver metastases were characterized by a complete loss of insulin expression and uniformly elevated expression of HMG3, consistent with the dependence of metastatic growth on the activation of this progenitor-like signaling pathway. The results establish the miR-181c/d/Meis2/Hmgb3 axis as a key regulatory mechanism orchestrating the dedifferentiation program in PanNETs.

Furthermore, we validated our findings in different cohorts of human PanNETs by demonstrating that the MLP signature, upregulation of miR-181c/d and HMG3, as well as downregulation of MEIS2, were all associated with aggressive tumors. Similar to the mouse PanNETs that develop de novo in RT2 mice, the MLP signature score in human PanNETs also correlated with pancreatic progenitors. In addition, metastatic samples had higher expression of MLP-specific genes, and patients bearing PanNET with high MLP signature scores had worse prognosis. Notably, we also observed an anticorrelation of MEIS2 and HMG3 expression, further substantiating the role of miR-181c/d/MEIS2/HMG3 axis in the dedifferentiation of human PanNETs.

Cellular plasticity enables cancer cells to evolve and activate ectopic programs to facilitate malignant progression (51). Dedifferentiation has been implicated, but not functionally established, as a discrete step in the progression of various cancers, such as melanoma (52), non–small cell lung cancer (53, 54), colorectal cancer (55, 56), breast cancer (57, 58), and pancreatic ductal adenocarcinoma (59, 60). Herein we have revealed dedifferentiation to a progenitor-like state is a separable step in multistage tumorigenesis, wherein activation of the dedifferentiation transition to the MLP subtype does not directly alter cancer cell proliferation, which establishes dedifferentiation as an independent hallmark of aggressive tumors, one that is mechanistically separate from proliferation.

The human PanNET grading system currently used in the clinic is based on the proliferation index, characterized by the Ki-67 histologic marker, where poorly differentiated MLP-like neuroendocrine tumors show the highest Ki-67 labeling index (>20%; ref. 61). Our data reveal that initial dedifferentiation from IT-like cancer cells produces progenitor-like cells that are unaltered in their low proliferation index. Subsequently, the dedifferentiated cancer cells undergo dynamic evolution to a highly proliferative state that becomes predominant in MLP tumors. As such, some bona fide MLP lesions in human PanNET may not show a high proliferative index, despite the likely possibility of progressing in the future to the well-described highly proliferative PanNET G3 stage, indicating, therefore, that proliferation index may not be a fully informative marker of tumor grade. As such, incorporation of biomarkers of dedifferentiation, in particular HMG3, might offer added value to the stratification of patients and better prediction of clinical outcomes.

Cancer cell progression coevolves with the tumor microenvironment (TME). This includes microvascular density (MVD), immune infiltration, and CAFs. In our previous study, we reported that the IT and MLP tumors have distinct MVD (8), while a recent study has shown that at least a subset of MLP in human patients with PanNET exhibits high immune infiltration (62). The distinctive molecular profile of IT and MLP tumors warrants future studies to investigate reciprocal interaction between cancer cells and different cells within the TME and the roles that they play in the process of dedifferentiation and development of the MLP subtype.

Characterization of the miRNA transcriptome during multistage tumorigenesis of PanNET has previously revealed signature sets of differentially expressed miRNAs in the distinctive stages of PanNET tumorigenesis (6). We have recently shown that two members of the MLP signature set, that is, miR-137 and the miR-23b cluster, contribute to the invasive and metastatic capabilities of PanNETs (35). Specifically, the miR-23b cluster modulates the expression of ALK7, a receptor of the TGFβ superfamily, and a metastasis suppressor in PanNETs (45). Notably, ALK7 is one of the downregulated genes in MLP samples, which substantiates the link between the MLP phenotype and the evasion of homeostatic tissue barriers to tumor metastasis. In this study, we have demonstrated that another member of the MLP signature set, the miR-181c/d cluster, regulates a distinctive parameter of the invasive and metastatic phenotype, namely dedifferentiation. The transcription factor Hmgb3 was identified as an indirect downstream effector, demonstrably involved in the transition from IT to MLP subtype. Hmgb3 has been previously implicated in maintaining the proper balance between hematopoietic stem cell (HSC) self-renewal and differentiation, wherein its expression inhibits the terminal differentiation of hematopoietic progenitor cells (63). These results motivate future investigations aimed to further chart the regulatory pathway, to identify upstream inducers of miR-181c/d expression as well as downstream effectors of the miR-181c/d/Hmgb3-mediated progenitor state operative in invasive and metastatic neuroendocrine pancreatic tumors.

The dedifferentiation process and activation of progenitor-like molecular phenotype in the MLP PanNETs were accompanied by the enrichment of neuronal gene sets that are also elevated in β-cell progenitors. Congruently, expression of miR-181c/d in an IT-like cancer cell line led to morphologic changes in the form of neuronal-like structures and activation of genes involved in neurogenesis. Interestingly, neurite outgrowth has been previously implicated in PanNET progression (64), and we have shown that PanNETs activate neuronal NMDAR signaling to acquire invasive capabilities (31, 65), and that basal breast cancers activate the same pathway during brain metastasis (66). Neuronal pathways have also been implicated in neuroendocrine transdifferentiation in resistant tumors across different cancer types (67), and neuronal-like morphology such as axon-like protrusions linked to the malignant phenotype of small-cell lung cancer (SCLC; ref. 68). Thus, ectopic activation of neuronal signaling pathways may prove to be a broader mechanism involved in acquiring invasive and metastatic capabilities.

In summary, this work has established dedifferentiation as a discrete stepwise transition via which neuroendocrine cancer cells acquire progenitor-like features, contributing...
to malignant progression. The dedifferentiation pathway is demonstrably orchestrated by the upregulation of the miR-181c cluster, which inhibits Meis2 expression, leading in turn to the expression of Hmgb3. Identifying this network of progenitor-associated genes fills a conceptual gap in our understanding of the factors contributing to the specification of invasive and metastatic neuroendocrine pancreatic tumors, knowledge that could lay a foundation for therapeutic strategies aimed to hinder tumor progression and metastasis.

METHODS

Animal Studies

All animals used in this study were maintained as a colony in a pathogen-free animal facility and all animal studies were approved by the Veterinary Authorities of the Canton Vaud according to Swiss Law. The following four RIP1-Tag2 (RT2) strains were used: RT2;C57B6/J (obtained by G. Christofori, University of Basel, Basel, Switzerland), RT2/A/J (RT2;C57BL6/N were backcrossed to A/J mice for 10 generations), and RT2;A6F1 (F1 generation of A/J female mice crossed with male RT2;C57BL6/J; ref. 45).

Laser Capture Microdissection Sample Collection

Tumor frozen sections were mounted onto PET membrane slides (MMA). Sections were incubated with 70% ethanol and stained with cresyl violet. Each sample was microdissected using a PALM laser dissection microscope (Zeiss). Serial sections from each lesion were used for the RNA or protein isolation:

- Total RNA was isolated using the QiAzoL Lysis Reagent according to the instructions (QiAGEN; catalog no. 217084).
- Total protein was extracted in 0.1% RapiGest SF Surfactant (Waters). Protein extracts were In-Solution digested as described previously (69). The reaction was stopped, and RapiGest cleaved by the addition of pure trifluoroacetic acid (TFA) during a final 1-hour incubation at 37°C.

EdU Cell Proliferation Assays

Mice were injected intraperitoneally with 100 μg of EdU in PBS. One hour postinjection, the mice were anesthetized using pentobarbital sodium and followed with transcardial perfusion with 20 mL of 2% PFA, and then with 20 mL of 1% PFA solution.

Cell Culture

All cell lines used in this study were maintained in a 5% CO2 incubator at 37°C. The βTC3 cell line (RRID:CVCL_0172) was cultured in DMEM containing 10% (v/v) FBS, and the AJ-5257–1 (ref. 45) βTC3 cell line (RRID:CVCL_0172) was cultured in DMEM-F12 media containing 10% (v/v) FBS, 1% (v/v) Insulin/Selenium, 4 mg/mL hydrocortisone, and 5 mg/mL mouse transferrin/Selenium, 4 mg/mL hydrocortisone, and 5 mg/mL mouse erythropoietin. oxycycline (2 mg/mL) was added to the media for the induction of the siRNA or protein isolation.

DNA Synthesis

The genomic area encompassing the miR-181c and miR-181d cluster was PCR amplified from genomic DNA isolated from C57BL6/N wild-type mice. The forward and reverse primers (below) contained restriction sites to facilitate cloning in the pDNR221 vector, and subsequent subcloning to the PB-31 destination vector (45).

```
TTCATCTTCCTCCTCTTCCTCCTCC
```

miRNA Cloning

The genomic area encompassing the miR-181c and miR-181d cluster (GRCm38/mm10 chr8:84,179,184–84,178,505) was PCR amplified from genomic DNA isolated from C57BL6/N wild-type mice. The forward and reverse primers (below) contained AtrB1 and AtrP1 sites to facilitate cloning in the pDRNR221 vector, and subsequent subcloning to the PB-33 destination vector (45).

```
FWd-attB1-miR-181c: GGGGACCACTTTGTACAAGAAAGCTGGGTTCA
GACAGATGTGACACTTAATATGATACATCTGTGGC
Gacttcaggggtag

Rev-attB2-miR-181c: GGGGACCACTTTGTACAAGAAAGCTGGGTTCA
GACAGATGTGACACTTAATATGACACATCTGTGGC
Gacttcaggggtag
```

Transfections

One million cells were plated in 6-well plates 24 hours before transfection and the media were changed 2 hours before transfections. For the transfection, the following liposomal complexes were set up:

- Single piggyBac vector: 4.5 μg PB-1 + 1.5 μg PBase; 12 μL of Lipofectamine 2000
- Two piggyBac vectors: 2.25 μg PB-1 + 2.25 μg PB-2 + 1.5 μg PBase; 12 μL of Lipofectamine 2000

The complexes were incubated for 20 minutes at room temperature and subsequently applied dropwise to the cells. The media was changed the next day, and stable cell lines were selected for resistance to genetic (G418; 1 mg/mL) and/or puromycin (2 μg/mL).

DNA Synthesis

The coding area of mouse Hmgb3 transcript variant 2 (Gene Accession NM_008253) was purchased from OriGene (Hmgb3 Mouse Tagged ORF Clone; catalog no. MR202042). The Hmgb3 open reading frame (ORF) was amplified using the primers shown below. The Kozak sequence GCCACCATG was included before the start codon in the Forward primer. AtrB1 and AtrP1 sites were also included to facilitate cloning in the pDNR221 vector, and subsequent subcloning to the PB-33 destination vector (45).

```
Forward primer: GGGGACCAAGAGTTGTACAAAAAAGCAGGCTGCAGCGACACGAGGCAGTAGG
GCTGAGATGTGACACTTAATATGACACATCTGTGGC
Gacttcaggggtag

Reverse primer: GGGGACCACTTTGTACAAGAAAGCTGGGTTCA
GACAGATGTGACACTTAATATGACACATCTGTGGC
Gacttcaggggtag
```

Knockdown Experiments

For the doxycycline-inducible gene knockdown experiments, we used previously described PB-31 piggyBac transposon vector (45), which allows for the expression of three tandem miR-E based shRNAs. The oligos for miR-E based shRNA are listed below:

- **Meis2-shRNA-1:**
  - 1. Meis2_17536_313_b: tcagaaaggatatGCTGTTGACAGTGGAGC
    GCCACACGAAAAATTTTATATACATGGTGACACAGAT
  - 2. Meis2_17536_313_a: aattctgactatatGCTGTTGACAGTGGAC
    GAACAGTGGACACTTAAATAGAATGAGGCAACAGAT
  - 3. Meis2_17536_313_c: gaggaggtatatGCTGTTGACAGTGGAGC
    GAACAGTGGACACTTAAATAGAATGAGGCAACAGAT

- **Hmgb3-shRNA-2:**
  - 1. Hmgb3_17536_353_b: aattctgactatatGCTGTTGACAGTGGAC
    GAACAGTGGACACTTAAATAGAATGAGGCAACAGAT
  - 2. Hmgb3_17536_353_a: aattctgactatatGCTGTTGACAGTGGAC
    GAACAGTGGACACTTAAATAGAATGAGGCAACAGAT

- **Accession NM_008253**
  - 1. Meis2_17536_313_b: tcagaaaggatatGCTGTTGACAGTGGAGC
    GCCACACGAAAAATTTTATATACATGGTGACACAGAT
  - 2. Meis2_17536_313_a: aattctgactatatGCTGTTGACAGTGGAC
    GAACAGTGGACACTTAAATAGAATGAGGCAACAGAT
  - 3. Meis2_17536_313_c: gaggaggtatatGCTGTTGACAGTGGAGC
    GAACAGTGGACACTTAAATAGAATGAGGCAACAGAT

- **Accession NM_008253**
  - 1. Meis2_17536_353_b: aattctgactatatGCTGTTGACAGTGGAC
    GAACAGTGGACACTTAAATAGAATGAGGCAACAGAT
  - 2. Meis2_17536_353_a: aattctgactatatGCTGTTGACAGTGGAC
    GAACAGTGGACACTTAAATAGAATGAGGCAACAGAT

A single shRNA cloned from genomic DNA was used to knock down Meis2 expression. The oligos for this strategy are given above.

DNA Synthesis

The coding area of mouse Hmgb3 transcript variant 2 (Gene Accession NM_008253) was purchased from OriGene (Hmgb3 Mouse Tagged ORF Clone; catalog no. MR202042). The Hmgb3 open reading frame (ORF) was amplified using the primers shown below. The Kozak sequence GCCACCATG was included before the start codon in the Forward primer. AtrB1 and AtrP1 sites were also included to facilitate cloning in the pDNR221 vector, and subsequent subcloning to the PB-33 destination vector (45).

```
Forward primer: GGGGACCAAGAGTTGTACAAAAAAGCAGGCTGCAGCGACACGAGGCAGTAGG
GCTGAGATGTGACACTTAATATGACACATCTGTGGC
Gacttcaggggtag

Reverse primer: GGGGACCACTTTGTACAAGAAAGCTGGGTTCA
GACAGATGTGACACTTAATATGACACATCTGTGGC
Gacttcaggggtag
```
Retracing Stepwise Differentiation during Tumor Progression

**Reporter Assay**

The Meis2 3′UTR (GRCM39/mm39 chr12:115,693,545–115,694,018) containing either the wild-type or mutated miR-181 MRE were synthesized and subsequently cloned in the destination vector pCDA-hLuc-RFA. The HEK293T cells (RRID:CVCL_0063) were stably transfected with the vector PB-13/PB-RB, allowing constitutive expression of the miR-181cid cluster. The reporter assays were performed as previously described in ref. 35.

**Cell-Cycle Analysis**

For this assay, the cells in suspension as well as adherent cells were collected. The total number of 0.2 × 10^6 cells were centrifuged and resuspended in 500 μL of 70% EtOH and incubated for 30 minutes in −20°C. Then, the cells were centrifuged and washed with FACS buffer (PBS 2% FBS). Subsequently, the cells were centrifuged again and then incubated staining buffer [FACS buffer, 20 μg/mL RNase A (Invitrogen #12091-021), 40 μg/mL PI (Sigma #P4864-10ML)] for 15 minutes at room temperature.

**mRNA Extraction and cDNA Synthesis**

mRNA extractions from cultured cell lines were performed with the QIAGEN RNeasy Kit (74106). Reverse transcription into cDNA was performed with PrimeScript RT Master Mix (Takara Bio Europe SAS) using 300 ng of total RNA.

**RT-PCR For mRNA Quantification**

RT-PCR was carried out with primers designed according to the PrimerBank database (https://pga.mgh.harvard.edu/primerbank/index.html; RRID:SCR_008689) and synthesized by Microsynth AG. All RT-PCR reactions were performed in triplicates using the 7900HT Fast RT-QPCT System (Applied BiosystemSTM). • In12 forward primer: GCTTCCTTCACACCCATGTC • In12 reverse primer: AGCAGTCTCATACATGCA • Mei2 forward primer: CAGGGTGGTCCAATGGGAATG • Mei2 reverse primer: GGGGGTCATGCTTATACAGT • Rpl13 forward primer: AGCCGGATGTCGATGACTG • Rpl13 reverse primer: ATCTCAGTGTAGGACCTCA

**RT-PCR For miRNA Quantification**

MystiCq MicroRNA SYBR Green qPCR ReadyMix (catalog no.: MIRCP00001) was used to quantify miRNA expression levels in different cell lines. All the primers used for this analysis were ordered from MystiCq microRNA qPCR kit as the following:

- miR-181c: mmu-mir-181c-3p (catalog no.: MIRAP01276)
- miR-181d: has-mir-181d (catalog no.: MIRAP00210)
- U6 (housekeeping miRNA): RNU6-1 (catalog no.: MIRCP00001)

**mRNA Sequencing**

RNA-seq libraries were prepared by first generating double-stranded cDNA from 10 ng total RNA with the NuGEN Ovation RNA-Seq System V2 (NuGEN Technologies). One hundred nanograms of the resulting double-stranded cDNA was fragmented to 350 bp using Covaris S2 (Covaris). Sequencing libraries were prepared from the fragmented cDNA with the Illumina TruSeq Nano DNA Library Prep Kit (Illumina) according to the protocol supplied by the manufacturer. Cluster generation was performed with the libraries using the Illumina TruSeq SR Cluster Kit v4 reagents and sequenced on the Illumina HiSeq 2500 with TruSeq SBS Kit v4 reagents. Sequencing data were processed using the Illumina Pipeline Software version 1.82. Purity-filtered reads were adapted and quality trimmed with Cutadapt (v. 1.2.1; RRID:SCR_011841). Reads matching to ribosomal RNA sequences were removed with fastq_screen (v. 0.11.1). The remaining reads were further filtered for low complexity with reaper (v. 15-065; RRID:SCR_009354). Reads were aligned against Mus musculus. GRCm38.86 genome using STAR (v. 2.5.3a; RRID:SCR_004463). The number of read counts per gene locus was summarized with htsq-count (v. 0.9.1; RRID:SCR_011867) using Mus musculus. GRCm38.86 gene annotation. The quality of the RNA-seq data alignment was assessed using RSeQC (v. 2.6.4; RRID:SCR_005275). Raw counts were further processed for normalization using DESeq2 pipeline (RRID:SCR_015687), where the data is normalized with the size factor to bring the count values to a common scale. This method is implemented in the R Bioconductor (RRID:SCR_006442) package DESeq2 (RRID:SCR_015687).

**miRNA Profile**

The miRNA expression profiles were evaluated using the Agilent miRNA microarrays. Fluorescence was scanned with an Agilent G2566AA scanner and analyzed using the Feature Extraction 10.7.3.1 software. Normalization was performed on the Total Gene Signal from Agilent “GeneView” data files in R. Data were log2 transformed after adding a small constant (here: 4). Quantile normalization was performed using the “normalize.quantiles” function from R package “preprocessCore” from the Bioconductor project (ref. 70).

**Proteomics**

Protein extracts from LCM samples were digested using the FASP procedure as described previously (ref. 71). Peptides were desalted using StageTips (ref. 72) and dried using a vacuum concentrator. For LC/MS-MS analysis, resuspended peptides were separated by reverse-phase chromatography on a Dionex Ultimate 3000 RSLC nano UPLC system connected in-line with an Orbitrap Fusion (Thermo Fisher Scientific). Raw data analysis was processed using MaxQuant 1.5.1.2 (RRID:SCR_014485), and database searches were performed against a human UniProt protein database.

**scRNA-seq**

Single-cell tumor samples were freshly prepared. Tissues were cut with surgical scissors into 1 to 3 mm³ cubes. The tissue was digested in a digestion mix with collagenase-I (25 mg) and collagenase-F (25 mg) and DNAse-I (5 mg) in 10 mL HBSS. Tissue was digested for 30 to 40 minutes at 37°C in 5 mL digestion mix. After digestion, 35 mL of FACS buffer (2% BSA in 1x PBS) was added to the 5 mL digestion mix. This solution was passed through a 70-μm cell strainer and then centrifuged at 600 × g for 5 minutes. The single-cell pellet was resuspended in 5 mL of 1x lysis buffer for 3 minutes. Thirty-five milliliters of FACS buffer (2% BSA in 1x PBS) was added to the lysis buffer and centrifuged again at 600 × g for 5 minutes. Cells were counted and diluted to 1 million cells/mL concentration. We used 10x Chromium standard 5′ seq assay. We aimed at 3,000, to 5,000 cells final. The standard library preparation was done following 10x Chromium protocol. Sequencing was done at a HiSeq4000 machine with the standard sequencing parameters (Read 1: 26 cycles, i7 index: 8 cycles, i5 index: 0 cycles, Read 2: 98 cycles).

**Western Blots**

For Western blotting, 10 to 20 micrograms of total protein extract was separated using SDS-PAGE and transferred onto polyvinylidene difluoride blotting membrane. Subsequently, the membranes were blocked with 5% BSA Fraction V (BSA) in TBS-T (pH 7.6, 0.5% Tween 20), and incubated with the primary and secondary antibodies diluted in 5% BSA in TBS-T. Finally, the membranes were visualized using ECL and imaged with Fusion FX7. The HMG83 antibody (Abcam, ab75782; RRID:AB_1310317) and HSP90 antibody (Santa Cruz Bio-technology, SC-13119; RRID:AB_675659) were used in this study.
**Immunostaining (Mouse Samples)**

For immunostaining, the tissue sections were first dewaxed and rehydrated, and heat-induced epitope retrieval was done with 10 mmol/L sodium citrate (pH 6.0), followed by washing with PBS. The samples were incubated overnight at 4°C with antibodies diluted in 1% BSA. The next day, the samples were incubated for 40 minutes with the secondary Alexa Fluor antibodies (Alexa Fluor 488, Alexa Fluor 568, and Alexa Fluor 647; diluted according to the manufacturer’s instruction). When needed, Tyramide Signal Amplification (TSA) revelation was done with Alexa Fluor 488. DAPI was used to stain the nuclei.

The immunofluorescence EdU was performed manually. After dewashing and rehydration, sections were pretreated with heat in 10 mmol/L sodium citrate buffer pH 6 using the PT module (Thermo Fisher Scientific) for 20 minutes at 95°C. The primary antibodies were incubated overnight at 4°C under agitation. After three washes, the slides were incubated for 45 minutes with the secondary antibodies at room temperature. Tissue was treated with 0.5% Triton in PBS for 20 minutes at room temperature before incubation with the Azide reaction cocktail (0.1 mol/L TBS pH 7.4, 4 mmol/L CuSO4, 100 mmol/L sodium ascorbate, and 10 mmol/L azide Alexa 594) to reveal the presence of EdU. The HMGB3 antibody (Abcam, ab75782; RRID:AB_1310317; dilution: 1:200), Insulin antibody (Dako A0564; dilution: 1:500), and SV40 T-antigen (in-house; dilution 1:1,000) were used in this study.

**Immunostaining (TMA Cohort from Human PanNETs)**

IHC was performed on a TMA including 110 primary PanNETs and 62 metastases from patients who underwent surgery at the Inselspital Bern between 1990 and 2017. The study on this cohort was approved by the local ethics committees (KEK number 105/2015).

Sections of 2.5 μm prepared from a TMA were used for staining with HMGB3 antibody (Abcam, ab75782; RRID:AB_1310317). Immunostaining was performed by automated staining using Bond RX (Leica Biosystems) immunostainer. All slides were dewaxed in Bond dewax solution (product code AR9222, Leica Biosystems) Heat-induced epitope retrieval at pH 9 in Tris buffer based (code AR9640, Leica Biosystems) for 30 minutes at 95°C. HMGB3 was diluted 1:200 and incubated for 30 minutes. Samples were then incubated horseradish peroxidase (HRP) polymer for 15 minutes and subsequently visualized using 3,3-Diaminobenzidine (DAB) as brown chromogen (Bond polymer refine detection, Leica Biosystems, Ref DS9800) for 10 minutes. HMGB3 was scored by a pathologist (A. Perren) as negative, positive, and heterogeneous.

For the double HMGB3 and Insulin staining, the antibodies were incubated sequentially. Following on HMGB3 staining as described above, mouse Insulin antibody (Sigma Aldrich, I2018) was diluted 1:1,200, incubated for 15 minutes. Secondary antibody, AP (Alkaline phosphatase)-polymer for 8 minutes, and visualized using fast red as red chromogen (Red polymer refine Detection, Leica Biosystems) for 30 minutes at 95°C. Induced epitope retrieval at pH 9 in Tris buffer based (code AR9640, Leica Biosystems) Heat-induced epitope retrieval at pH 9 in Tris buffer based (code AR9640, Leica Biosystems) Heat-induced epitope retrieval at pH 9 in Tris buffer based (code AR9640, Leica Biosystems) Heat-induced epitope retrieval at pH 9 in Tris buffer based (code AR9640, Leica Biosystems). Slides were scanned and photographed using Pannoramic 250 (3-DHistech).

**Total RNA Extraction—Including Small mRNA—from Formalin-Fixed Human PanNETs**

Total RNA extraction, including miRNAs, has been performed using the RecoverAll Total Nucleic Acid Isolation Kit for FFPE (Ambion; AM1975) according to the provider’s instructions. DNA digestion via DNase has been performed during RNA isolation according to the manufacturer’s instructions.

**miRNA Expression Profiling on Human PanNETs**

We have analyzed miRNA expression from 24 PanNETs using the NanoString nCounter Human v3 miRNA Panel (NanoString Technologies) according to the manufacturer’s instructions, 150 ng of total RNA have been used for running the assay. Normalization has been performed using the quantile normalization method (70). All the analyses have been performed within the R environment.

**Image Data Acquisition and Analysis**

The entire tissue area of all slides was scanned using an Olympus slide scanner at 20x magnification, and the images were analyzed using the open-source software QuPath (RRID:SCR_018257). For quantifying the cells stained positive for a marker, we selected sections from different regions used “CellIntensityClassifications” method, as described in ref. 73.

**Clustering Analysis**

For subtype identification, we applied the NMF algorithm (“NMF” package in R, ref. 15) on each profiling platform (i.e., mRNA, lncRNA, miRNA, and protein), separately. For clustering, we selected the top 25% most variable genes (features) from the normalized datasets. The Cophenetic coefficient was used to select the optimized number of clusters for each analysis. The features contributing the most to each cluster were also extracted using “extractFeatures” function in the R package.

SNF algorithm (ref. 16) was applied for multi-omics clustering. The top 25% of the most variable features in mRNA, lncRNA, miRNA, and proteomics were used as an input for the algorithm. We used “SnFtool” package in R for running this analysis, and we applied the default setting in the package. Estimating the most optimal number of clusters was done using “estimateNumberOfClustersGivenGrowth” function. For a better resolution on the final clustering, we excluded the normal pancreatic and liver samples from this analysis. The final clusters were manually curated and annotated, based on single platform NMF clustering in addition to SNF results.

**Differential Expression Analysis**

For mRNA differential expression analysis, we used DESeq pipeline (DESeq2 R package; RRID:SCR_015687), where the data is modeled with a negative binomial distribution to estimate the dispersion and fold changes. For extracting the differential expressed genes, we consider the samples’ mouse strain as covariates to remove its effect on the downstream analysis. The cutoffs for gene selection were $P_{adj} < 0.01$ and fold change > 1.5.

Differential miRNA expression was assessed using the limma package (RRID:SCR_010943). $P$ values were adjusted using the Benjamini–Hochberg method, and significance cutoff set at 0.05. We selected miRNA with greater than 1.5 fold change.

**Subtype Signature Development**

The MLP mRNA signature was generated on the basis of the result of the differential expression analysis of mRNA and lncRNA datasets (IT samples vs. MLP samples). We also included the genes that were identified as the contributing features for NMF clustering in both mRNA and lncRNA datasets. Furthermore, we also checked in the protein analysis to see which of the genes differentially expressed have a similar trend in the proteomics data and a control dataset.

Likewise, the miRNA signature was generated by the union of the differential expressed miRNA (IT vs. MLP samples), and features contributing to the NMF clustering in the miRNA dataset.

**PanNET Signature Development**

The 62-miRNA signature for transformed islet β cells (PanNET signature) was generated from the bulk RNA-seq data following these criteria:

1. Differentially upregulated genes in IT and MLP tumors versus normal islet β-cells (fold change > 4, $P_{adj} < 0.01$)

**Image Data Acquisition and Analysis**

The entire tissue area of all slides was scanned using an Olympus slide scanner at 20x magnification, and the images were analyzed using the open-source software QuPath (RRID:SCR_018257). For quantifying the cells stained positive for a marker, we selected sections from different regions used “CellIntensityClassifications” method, as described in ref. 73.

**Clustering Analysis**

For subtype identification, we applied the NMF algorithm (“NMF” package in R, ref. 15) on each profiling platform (i.e., mRNA, lncRNA, miRNA, and protein), separately. For clustering, we selected the top 25% most variable genes (features) from the normalized datasets. The Cophenetic coefficient was used to select the optimized number of clusters for each analysis. The features contributing the most to each cluster were also extracted using “extractFeatures” function in the R package.

SNF algorithm (ref. 16) was applied for multi-omics clustering. The top 25% of the most variable features in mRNA, lncRNA, miRNA, and proteomics were used as an input for the algorithm. We used “SnFtool” package in R for running this analysis, and we applied the default setting in the package. Estimating the most optimal number of clusters was done using “estimateNumberOfClustersGivenGrowth” function. For a better resolution on the final clustering, we excluded the normal pancreatic and liver samples from this analysis. The final clusters were manually curated and annotated, based on single platform NMF clustering in addition to SNF results.

**Differential Expression Analysis**

For mRNA differential expression analysis, we used DESeq pipeline (DESeq2 R package; RRID:SCR_015687), where the data is modeled with a negative binomial distribution to estimate the dispersion and fold changes. For extracting the differential expressed genes, we consider the samples’ mouse strain as covariates to remove its effect on the downstream analysis. The cutoffs for gene selection were $P_{adj} < 0.01$ and fold change > 1.5.

Differential miRNA expression was assessed using the limma package (RRID:SCR_010943). $P$ values were adjusted using the Benjamini–Hochberg method, and significance cutoff set at 0.05. We selected miRNA with greater than 1.5 fold change.

**Subtype Signature Development**

The MLP mRNA signature was generated on the basis of the result of the differential expression analysis of mRNA and lncRNA datasets (IT samples vs. MLP samples). We also included the genes that were identified as the contributing features for NMF clustering in both mRNA and lncRNA datasets. Furthermore, we also checked in the protein analysis to see which of the genes differentially expressed have a similar trend in the proteomics data and a control dataset.

Likewise, the miRNA signature was generated by the union of the differential expressed miRNA (IT vs. MLP samples), and features contributing to the NMF clustering in the miRNA dataset.

**PanNET Signature Development**

The 62-miRNA signature for transformed islet β cells (PanNET signature) was generated from the bulk RNA-seq data following these criteria:

1. Differentially upregulated genes in IT and MLP tumors versus normal islet β-cells (fold change > 4, $P_{adj} < 0.01$)
2. Excluding the genes that were also differentially expressed between IT and MLP tumor samples.
3. Excluding the genes that are involved in cell-cycle regulation (genes in Cell-Cycle GO term).

Cross-Species Analysis
To merge different datasets and correct for the batch effects, we employed ComBat (“sva” package in R; RRID:SCR_010974). For merging human and mouse datasets, the HUGO gene nomenclature was used as the reference.

scRNA-seq Data Analysis
We used the Cell Ranger pipeline (version 3.1; RRID:SCR_017344) for processing the raw data. The downstream analysis was performed using Seurat package in R (RRID:SCR_007322). Cells were filtered on the basis of the number of features from 200 to 7,500, percentage of mitochondrial genes <10%. We used all standard Seurat setting for normalization, PCA, t-distributed Stochastic Neighbor Embedding (tSNE) analysis, and clustering.

Single-Cell Cluster Annotation
Gene expression levels of different markers used to identify the cell-type population in the tumor microenvironment: SV-40 infected cells, IM2; β cells and cancer cells, CD3e: T-cell leukocytes, CD19 and MHC1: B-cell leukocytes, Pcam1: endothelial cells, Csf3: neutrophils, Csf1r: macrophages, Mrc1: M2 macrophages, Marp: CAFs, Acta2: pericytes. We also used the total number of features for the cells to filter out the cancer cells having a significantly lower number of detected features (<2,500). This cluster was called “Tumor low-reads” and was excluded from the downstream analysis of cancer cells.

Differential Expression Analysis in scRNA-seq Data
Differential expression analysis between clusters was performed using the Wilcoxon rank-sum test, using 0.05 as the cutoff for P adj value, and 1.5 for the average fold change in expression level. Then, the genes were sorted on the basis of average fold change and the top 20 were selected for clusters’ marker genes. In the case of IT versus MLP analysis, the top 100 genes were kept and reported.

Proliferation Signature Score for scRNA-seq Data
To estimate the proliferation capability of the single cells, we used a signature of G 2–M marker genes provided by Seurat R package (V.3.2; RRID:SCR_007322).

PanNET Regulator Network
The regulatory network was constructed using the ARACNe package (RRID:SCR_002180) from 54 samples from the RT2 mouse model. We ran ARACNe with 100 bootstrap iterations using all the genes in the dataset, and the parameters were set to 0.15 DPI (Data Processing Inequality) tolerance and MI (Mutual Information) P value threshold of 10−7.

TF Regulators of IT-to-MLP Transition
The candidate TFs regulating IT-to-MLP transition were inferred using VIPER analysis as described in ref. 40. The differential expression analysis for all the genes in the datasets was performed according to the package manual at Bioconductor, by comparing IT to MLP samples identified from the clustering analysis. For selecting the TF from the list of VIPER-inferred regulators we used a list of 1636 Mus-musculus TFs reported at ref. 74.

GO Terms Enrichment Analysis
GO terms enrichment analysis was performed using the online web service based on the Molecular Signatures Database (MSigDB; RRID:SCR_016863). Enriched GO terms were defined as GO biological process (BP) obtaining an FDR-adjusted P < 0.05, retrieving a maximum of 50 terms.

Gene Signature Enrichment Analysis
Single-sample Gene Set Enrichment Analysis (ssGSEA; ref. 75), implemented in the R package GSVA, was used to calculate an expression score for each gene expression signature and each sample. The method that was used to estimate the gene set enrichment scores were specified to “ssgsea” in GSVA package.

Correlation Analysis
To evaluate the degree of correlation between two continuous variables, we employed “cor” function in R to retrieve the Pearson correlation coefficient. For the significance estimation of the extracted correlation, we used “lm” function in R for linear modeling the data and retrieved FDR-adjusted P values.

Survival Analysis
Kaplan–Meier survival analysis was used to assess the relationship of the signature scores with overall survival. We applied the large-sample P2 test (log-rank test) to determine the associations between predictor variables and to obtain adjusted HRs. These analyses were performed with the R package “survival.”

Quantification and Statistical Analysis
All statistical analyses described above were performed with R.

Data Availability
RNA-seq data and regulon analysis presented in this study are deposited at ZENODO database (RRID:SCR_004129), and are accessible via https://doi.org/10.5281/zenodo.4160441

Authors’ Disclosures
A. Perren reports personal fees from Ipsen and personal fees from Novartis outside the submitted work. D. Hanahan reports grants from Swiss National Science Foundation and grants from Human Frontier Science Program during the conduct of the study. No disclosures were reported by the other authors.

Authors’ Contributions
S. Saghaﬁnia: Conceptualization, resources, data curation, formal analysis, validation, investigation, visualization, methodology, writing-original draft, writing-review and editing. K. Homicsek: Formal analysis. A. Di Domenico: Formal analysis. S. Wullschleger: Formal analysis. A. Perren: Resources, formal analysis. I. Marinoni: Resources, formal analysis. G. Ciriello: Supervision. I.P. Michael: Conceptualization, resources, data curation, formal analysis, supervision, validation, investigation, visualization, methodology, writing-original draft, writing-review and editing. D. Hanahan: Conceptualization, supervision, funding acquisition, writing-original draft, project administration, writing-review and editing.

Acknowledgments
We thank G. Christofori (University of Basel, Basel, Switzerland), and E.F. Wagner and M. Farlik-Födinger (Medical University of Vienna, Vienna, Austria) for insightful feedback and comments on the manuscript; S. Lamy, A. Etienne, M. Gaveta, and L. Dori for technical support; M. Tichet for FACs analysis; J. Sordet-Dessimoz (Histology Core Facility–EPFL) for performing IHC; N. Zanger (Swiss Institute of Bioinformatics), S. Calderon (LGTF-UNIL), S. Pradervand (LGTF-UNIL), and B. Mangeat (Gene Expression Core Facility–EPFL) for assistance with RNA-seq analysis; and the EPFL

Downloaded from cancerdiscovery.aacrjournals.org on November 15, 2021. © 2021 American Association for Cancer Research.
School of Life Sciences BIOP, FCCF and HCF cores and animal care facility; and members of the Hanahan laboratory. This research was principally supported by a grant from the Swiss National Science Foundation (SNSF; to D. Hanahan). I.P. Michael was supported in part by a Human Frontier Science Program Organization (HFSPO) fellowship.

Received November 10, 2020; revised March 6, 2021; accepted April 8, 2021; published first April 28, 2021.

REFERENCES


Retracing Stepwise Differentiation during Tumor Progression


Cancer Cells Retrace a Stepwise Differentiation Program during Malignant Progression

Sadegh Saghafinia, Krisztian Homicsko, Annunziata Di Domenico, et al.

Cancer Discov 2021;11:2638-2657. Published OnlineFirst April 28, 2021.

Updated version
Access the most recent version of this article at:
doi:10.1158/2159-8290.CD-20-1637

Supplementary Material
Access the most recent supplemental material at:
http://cancerdiscovery.aacrjournals.org/content/suppl/2021/04/24/2159-8290.CD-20-1637.DC1

Cited articles
This article cites 74 articles, 17 of which you can access for free at:
http://cancerdiscovery.aacrjournals.org/content/11/10/2638.full#ref-list-1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://cancerdiscovery.aacrjournals.org/content/11/10/2638.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.