Longitudinal Analysis of Human Pancreatic Adenocarcinoma Development Reveals Transient Gene Expression Signatures

Jungsun Kim1,2,3, Taenel Ekstrom2, Wenli Yang4, Greg Donahue5, Dmytro Grygoryev7, Thuy T.M. Ngo1,2,6, John L. Muschler3,6, Terry Morgan2,7, and Kenneth S. Zaret5

ABSTRACT

Previous transcriptome studies of human pancreatic ductal adenocarcinoma (PDAC) compare non-cancerous pancreatic intraepithelial neoplasias (PanIN) with late-stage PDAC obtained from different patients, thus have limited ability to discern network dynamics that contribute to the disease progression. We demonstrated previously that the 10-22 cell line, an induced pluripotent stem cell–like line reprogrammed from late-stage human PDAC cells, recapitulated the progression from PanINs to PDAC upon transplantation into NOD/LtSz-scid/IL2R-gamma-/- mice. Herein, we investigated the transition from precursor to PDAC using the isogenic model. We analyzed transcriptomes of genetically tagged 10-22 cells progressing from PanINs to PDAC in mice and validated the results using The Cancer Genome Atlas PDAC dataset, human clinical PanIN and PDAC tissues, and a well-established murine PDAC model. We functionally studied candidate proteins using human normal (H6C7) and cancerous (Miapaca2, Aspc1) pancreatic ductal epithelial cell lines. 10-22 cell–derived PDAC displayed the molecular signature of clinical human PDAC. Expression changes of many genes were transient during PDAC progression. Pathways for extracellular vesicle transport and neuronal cell differentiation were derepressed in the progression of PanINs to PDAC. HMG-box transcription factor 1 (HBPI) and BTB domain and CNC homolog 1 (BACH1) were implicated in regulating dynamically expressed genes during PDAC progression, and their expressions inversely correlated with PDAC patients’ prognosis. Ectopic expression of HBPI increased proliferation and migration of normal and cancerous pancreatic cells, indicating that HBPI may confer the cell dissemination capacity in early PDAC progression. This unique longitudinal analysis provides insights into networks underlying human PDAC progression and pathogenesis.

Implications: Manipulation of HBPI, BACH1, and RUN3 networks during PDAC progression can be harnessed to develop new targets for treating PDAC.

Introduction

A better understanding of the development and progression of pancreatic ductal adenocarcinoma (PDAC) will inform better diagnoses and therapies (1). Pancreatic intraepithelial neoplasias (PanIN) are major PDAC precursors (2) and are classified as low grade (PanIN1 and 2) and high grade (PanIN3) based on cytologic atypia severity. Low-grade PanINs are considered clinically benign, whereas high-grade PanINs, usually found in pancreata with PDAC, are regarded as carcinoma in situ (2). PanIN2 is the earliest lesions that harbor genetic alterations in KRAS, CDKN2A, and TP53 typically seen in PDAC (3, 4). PanIN3 and PDAC share common mutations, whereas PDAC-specific mutations other than those in SMAD4 appear to be passengers (5, 6).

Gene expression in human PanINs (7–9), PDAC (10–19), and PDAC cell lines (20) has been investigated. The transcriptome studies revealed two major PDAC subtypes, a basal (or squamous) subtype with a worse prognosis and a classical subtype with a better prognosis (12–15). Different cell populations in PDAC have been characterized using an intraductal transplantation model (21). However, most works with human samples did not longitudinally compare gene expression in individual isogenic series. Although genetic mutational profiling can reveal cancer evolution (5, 19, 22, 23) and the pseudotemporal ordering of single cells can suggest lineages (24), the pseudotemporal order of cancer cells can be confounded by undocumented genetic changes. Thus, an analysis of transitions from human PanINs to their corresponding PDAC is currently lacking. Genetically engineered mouse models (25) have provided the basis for studying the phenotypic manifestation of defined genetic mutations. However, most transcriptome studies of these models compare normal with cancer (26) or acinar to ductal conversion (27) and not revealing the dynamic process.

We previously created an induced pluripotent stem cell (iPS)-like cell line (designated 10–22) from a recurrent advanced stage of PDAC. Despite having mutations, including an activating KRAS mutation, heterozygous deletion of CDKN2A, and decreased PTEN and DPC4 copy number as well as 20 of the 23 chromosomal rearrangements...
presented in the original tumor (30), 10-22 cells develop into PanIN2/3 by 3 months and then into PDAC by 9 months when injected into NOD/LtSz-scid IL2R-gammanull (NSG) mice (30). Furthermore, ductal lesions are surrounded by human stromal cells that are also differentiated from 10-22 cells. The system revealed an HNF4a-driven network specific to the early to intermediate stages of PDAC (30) and a biomarker, THBS2, that can discriminate patients with resectable or advanced stages of PDAC from healthy controls (31).

In this study, we sought to gain insights into gene networks during the progression of human PanINs to PDAC. We tagged 10-22 cells with the KIR-EG reporter vector that expresses GFP from a ubiquitous promoter and a near-infrared fluorescent protein (iRFP; ref. 32) from a duct-selective K19 promoter (33) then injected tagged cells into NSG mice. We monitored lesions in vivo over time by iRFP fluorescence and used GFPþiRFPþ cells of the pancreatic ductal epithelial lineage isolated at different timepoints for RNA sequencing (RNA-seq) analysis. Our unique ability to capture events in isogenic, longitudinal tumor development allowed us to identify effectors and networks associated with the progression of early PanINs to PDAC.

**Materials and Methods**

**PanIN and PDAC mouse model**

All animal works were performed with the Institutional Animal Use and Care Committee (IACUC) approval of the Oregon Health & Science University School of Medicine (OHSU). Ages 2–2.5 and 4–6 months old PDAC mice (p66-Cre; LSL-KrasG12D/þ; Trp53R172H; ref. 25) were analyzed for HBPI expression at the PanINs and tumor stages, respectively. All transgenic lines used were backcrossed at least five generations onto the C57Bl/6 background, and female and male mice are used.

**Culturing 10-22 cells, generation of KIR-EG vector, labeling 10-22 cells, and gating strategy**

The KIR-EG lentiviral vector was constructed to express iRFP (32) driven by the keratin 19 (K19) gene promoter (33) in the backbone of GFP driven by a ubiquitous EF1a promoter and a near-infrared fluorescent protein (iRFP) from a duct-selective K19 promoter (33) then injected tagged cells into NSG mice. We monitored lesions in vivo over time by iRFP fluorescence and used GFPþiRFPþ cells of the pancreatic ductal epithelial lineage isolated at different timepoints for RNA sequencing (RNA-seq) analysis. Our unique ability to capture events in isogenic, longitudinal tumor development allowed us to identify effectors and networks associated with the progression of early PanINs to PDAC.

**RNA-seq and data processing**

Pooled cDNA libraries, generated from total RNA, were sequenced using NextSeq500 sequencer and the trimmed reads were aligned to the hg 19 genomes using STAR (34). HTSeq count was used to quantify tags over genes (Supplementary Table S2), then normalized and called differentially expressed (DE) genes (FDR Benjamini–Hochberg adjusted q < 0.05) between two pairwise samples by Wald test or among all the samples across each time-point by likelihood-ratio test (LRT) using DESeq2 package (35). Homofermentation analysis (36) identified de novo Transcription factor (TFs) bind to promoters (1,000 bp upstream of transcription start site) of genes within selected Ward cluster groups. RNA-seq data are available on NCBI GSE173286 and GSE173489. RNA-seq and data processing codes are available on Code Ocean (Capsule #6902043) and Github (https://github.com/gdonahue/Kim_MCR_2021). Detailed analysis is available in the Supplementary Data. Additional results are available in supplementary figures and tables as resource datasets.

**Comparison of 10-22 cells gene expression datasets to other datasets**

The Cancer Genome Atlas (TCGA) PDAC RNA-seq dataset (n = 150; ref. 37), which ranked genes based upon RSEM values from 0% to 100% ("IlluminaHiSeq percentile" dataset), was downloaded (38). Nonlinear dimensionality reduction method, multidimensional scaling (MDS) was performed with cmdscale in R package. To compare with human pluripotent embryonic stem cells (ESC), rank-ordered FQ values of human ESC (39), the third quartile SAGE tags of human ESCs tags (40), and Progenitor Cell Biology Consortium (PCBC) RNA-seq dataset, which contains gene expression data from 48 pluripotent stem cells (7 ESC lines, 41 iPSC lines) and their differentiated cell populations (41), were used. The PCBC data code is publicly available on GitHub.

**Kaplan–Meier survival analysis of TCGA PDAC dataset**

The cutoff for HBPI expression was determined by equally segregating TCGA PDAC cohorts (37) into two groups at 10.57 (log2 median) for each cohort, respectively. For HBPI expression level was evenly spread across the PDAC cases. To determine the cutoff for BACH1 expression, we examined the risk table with three different BACH1 log2-transformed normalized RSEM levels (from the first quartile to the third quartile) because BACH1 is expressed abundantly in many PDAC cases and then stratified the survival curve with 10.27, the value in the 75th percentile.

**HBPI CRISPR-cas9 targeting**

AsPC-1 cells were transfected with two CRISPR-Cas9 vectors targeting the RB-binding domain of HBPI exon 2 (gRNA#1, gRNA#2). Transfected cells were selected using 5 μg/mL puromycin for 48 hours, cultured for 2 more weeks without puromycin, then examined for the deletion of HBPI using Sanger sequencing.

**Results**

**PanINs derived from 10-22 cells resemble human disease**

We confirmed the stepwise PDAC progression of 10-22 cells after transplantation into NSG mice. Six of the 7 NSG mice injected with 10-22 cells developed PanINs that express K19, MUC5AC, PDX1, and SOX9 by 1.5 to 3 months (Supplementary Table S1A; Fig. 1A–F). PanINs possessed an intact collagen IV+ basement membrane (Fig. 1D; ref. 42). Tumors grown in vivo for 13 months showed highly
dysmorphic (Fig. 1G) K19+, PDX1+, and mesothelin+ structures with collagen IV+ basement membrane breakdown (n = 3/4; Fig. 1H–K). Mesothelin is a marker for advanced PDAC (Fig. 1L; ref. 43). After 13 months, 3 of 4 mice exhibited additional tumors that migrated beyond the injection sites (Supplementary Table S1B). These findings demonstrate that 10-22 cells can recapitulate PanINs, invasive, and migratory phenotypes with features common to those documented for clinical human PDAC.

PDACs derived from 10-22 reporter subclones are histologically heterogeneous

PanINs and PDAC express K19, 10-22 cells express K19 upon ductal differentiation but they also differentiate into other lineages (30). To isolate ductal and non-ductal 10-22 cells from mouse cells, we developed the lentivirus KiR-EG vector that expresses iRFP driven by a K19 promoter (33) and GFP driven by an EF1α promoter (Fig. 2A: Supplementary Fig. S1A). We labeled 10-22 cells with KiR-EG lentivirus and isolated clones that gave rise to iRFP and GFP double-positive cells during in vitro differentiation (Supplementary Fig. S1B). Four stable KiR-EG clones were injected into NSG mice individually and their growth was followed for 9 months. Clones #6 and #9 generated PanINs followed by PDAC like the parental 10-22 cells, and thus they were chosen for detailed analysis (Fig. 2B–D; Supplementary S1C–S1E, S2, and S3A). While various cell types were seen in lesions at 1 week after injection, PanIN-like ductal epithelial structures were enriched by 1 to 3 months after injection and became increasingly heterogeneous over time (Supplementary Fig. S2 and S3A). For example, after 6 months, mouse 8560 RT had a well-differentiated tumor with a relatively intact basement membrane, whereas mouse 5074T had a poorly differentiated tumor (Supplementary Fig. S2). Tumors greater than 4 mm in diameter in vivo are detectable by their iRFP signal (Fig. 2F; Supplementary Fig. S3B). In sum, we validated...
that 10-22 KiR-EG clones recapitulate the PDAC progression of their parental 10-22 cells.

**Identify PDAC progression-associated gene signatures**

To characterize the global gene expression during the transition from iPS-like cells to PanINs and PDAC, we performed RNA-seq on iRFP+ cells, representing the ductal lineage, from 10-22 KiR-EG clone 6 at 1 week, 2 weeks, 3 months, 6 months, and clones 6, 9, and 10 at 9 months after transplantation into NSG mice (Fig. 2G). GFP+ cells at day 0 were used as controls.

On the basis of histology of the lesions (Supplementary Fig. S2), we defined each stage with RNA-seq replicates as follows: undifferentiated
IPS stage (day 0, triplicates), released from near-pluripotency (1 week after transplantation, triplicates; and 2 weeks after transplantation, duplicates). PanINs (3 months after transplantation, triplicates), early invasive PDAC (6 months after transplantation, triplicates), and PDAC (9 months after transplantation, duplicates). Spearman rank correlation coefficient with log-transformed normalized DESeq2 counts indicated a strong correlation between biological replicates in each timepoint, including across different clones at the 9-month timepoint (Supplementary Fig. S3C).

To identify stage-specific gene signatures in the 10-22 PDAC progression model, we discovered 1,075 DE genes across all the timepoints using the LRT (FDR < 0.05; Fig. 2H; Supplementary Table S2). PCA and Euclidian clustering showed that the earliest differentiated population (1 week) was located the farthest from day 0, indicating rapid differentiation of 10-22 cells after transplantation (Fig. 2I; Supplementary Fig. S3D). Moreover, undifferentiated 10-22 cells were located most closely to late-stage tumors (9 months), consistent with reported similarities between pluripotent and cancer cells (Fig. 2I; Supplementary Fig. S3D; ref. 44). Similar results were observed when we percentile ranked the gene expression values of 10-22 Kir-EG clone 6 and its differentiated counterparts and compared them with the percentile ranked-ordered gene expression values of 150 TCGA PDAC RNA-seq datasets (37) by MDS (n = 150; Fig. 2J). To avoid ubiquitously expressed genes and focus on pathologically relevant genes, we studied genes enriched between the 75th and 92nd percentile of expression (38), which varied the most highly across samples. PDAC can be classified into classical and basal subtypes based on gene expression (12–15). Clustering normalized counts with the PDAC-subtype gene signatures (15) showed that 10-22 cells derived from PDAC at 9 months had a classical and basal hybrid subtype (Supplementary Fig. S4A and S4B). Altogether, we validated that the 10-22 PDAC progression model reflected the heterogeneity observed in natural PDAC (14, 45).

We then assessed the expression of genes associated with pluripotency, endoderm, oncogenesis, and tumor suppression. The majority of Muller pluripotent genes (44) were expressed in the starting 10-22 cells and downregulated by 1 week after transplantation, as expected (Supplementary Fig. S4C). Endodermal genes such as KRT19, SOX9, HNF1B, and FOXA3 were upregulated by 2 weeks of transplantation (Supplementary Fig. S4D). While KRAS was consistently expressed across timepoints, with transient downregulation 3 months after transplantation, CDX2A, TP53, and BRCA1/2, known to contribute to PDAC development (3, 4), variedly expressed during PDAC development (Supplementary Fig. S4E).

10-22 cells retain the memory of the pancreatic epithelial lineage

We rank-ordered transcripts by their expression levels in undifferentiated 10-22 Kir-EG clone 6 and compared them with the ranked-ordered transcripts of normal ESCs (39). The Spearman rank-order correlation coefficient (0.60) and the coefficients of linear regression (0.79; Supplementary Fig. S4F) were consistent with our original observation that 10-22 cells were different from ESCs (30). Unexpectedly, early endodermal genes FOXA1, GATA4, and GATA6 were clearly expressed in undifferentiated 10-22 cells (Supplementary Fig. S4D), indicating that 10-22 cells retain endodermal gene signatures in the IPS-like stage (30). To further investigate genes aberrantly upregulated in 10-22 cells compared with normal ESCs, we compared the top 20% of rank-ordered genes in 10-22 cells with the bottom 20% rank-ordered genes in human ESCs. These 129 genes aberrantly expressed in undifferentiated 10-22 cells were enriched with the Gene Ontology (GO) associated with epithelium development, regulation of microvillus organization, endoderm origin, small airway epithelium, and PDGF pathway (Supplementary Table S3C). The signature was enriched for malignant neoplasia, including pancreatic cancer (ref. 46; q < 0.05; Supplementary Table S3C), indicating the memory of the pancreatic cancer epithelial state in 10-22 cells. In summary, 10-22 cells express both pluripotency networks and polarized endodermal epithelial cell networks, suggesting how they tend to develop into the pancreatic epithelial lineage (30).

Genes dynamically expressed in the 10-22 PDAC progression model show distinct gene signatures at different stages

We clustered DE genes from 10-22 derivatives into five stage-specific signatures based on the expression pattern across timepoints using Ward method (Fig. 3A; ref. 47). It showed a gene signature specific to the 10-22 Kir-EG clone 6 cultured in pluripotent conditions (group 1; Fig. 3A). While most group 1 genes (Supplementary Fig. S5A and S5B) expressed higher in pluripotent ESCs and iPSCs, some (n = 51) expressed less in pluripotent ESCs compared with their differentiated counterparts (Supplementary Fig. S5B and S5C). Comparing group 1 genes with the published RNA-seq (39) and third quartile-SAGE tags (40) from human ESC confirmed that group 1 genes significantly overlap with genes explicitly expressed in pluripotent ESCs (Supplementary Fig. S6). The expressed genes in group 1 were enriched with GO terms associated with cell-cycle regulation (q < 0.17; Supplementary Fig. S5C), suggesting how they tend to develop into the pancreatic epithelial lineage (30).

The expression of group 2 genes gradually increased over time and stayed up at 9 months (Fig. 3A), reflecting the late PDAC stages. GO terms showed that group 2 genes are enriched to cell adhesion (n = 13/44; q < 0.05); including extracellular matrix (ECM) features (q = 1.69E-04), consistent with ECM remodeling in later stages of PDAC progression (Supplementary Table S3E). Validating this finding, 30 of 44 group 3 genes that are highly expressed in TCGA PDAC are enriched in malignant neoplasms (q < 0.05; Fig. 3C; ref. 46). Thus, we confirmed that 10-22-derived tumors display adenocarcinoma gene signatures by 9 months, particularly with ECM remodelers. These findings agree with the breakdown in collagen IV basement membrane structures in late-stage 10-22 tissue (compare Fig. 1D and J) and further validate the model system.

The KRAS pathway and extracellular vesicle transports are reactivated early in the 10-22 PDAC progression model

Upon differentiation, 10-22 cells upregulated a small portion of basal subtype signature (Supplementary Fig. S4B), which correlates with the amplification of KRAS-mutant alleles (15). Nevertheless, upon differentiation, 10-22 cells developed PanIN-like lesions that undergo PDAC progression and profoundly increase their classical subtype signatures in 9-month samples (Supplementary Fig. S4A). We thus asked how the classical subtype gene program can override the default basal cell fate.

We examined groups 4 and 5 genes that covered PanINs and early invasive stages during 10-22 PDAC progression. Group 4 genes include genes increased at the earliest differentiation upon 10-22 cells being released from pluripotency (1 week) and remained until early invasive stages (6 months; Fig. 3A). As expected, GO terms of group 4 genes enriched signal pathways activated in response to external cues.
upon differentiation and developmental processes (n = 368; Fig. 3D; Supplementary Table S3F). Consistently, gene set enrichment analysis (GSEA) showed activation of KRAS pathways as early as 1 week (Supplementary Table S5A) and throughout the entire PDAC progression after transplantation, compared with day 0 (Fig. 3E; Supplementary Table S5B). This agrees with the well-established role of mutant KRAS, such as the G12D mutation in 10-22 cells, as an early driver of PDAC progression (30). Genes involved in lysosomal membrane protein and exosomes were rapidly upregulated upon differentiation (Fig. 3F). Although lysosomal activity (49–51) and exosome pathways (52) play key roles in human PDAC, it was unclear whether they occur early in the disease, that is, higher in PanINs than in PDAC. Our data suggest that such cell-intrinsic networks can establish a microenvironment that is favorable to cancer progression early. On the
other hand, we found that genes in neural stem cell proliferation, neuronal differentiation, and neural invasion-driving chemokines (53, 54) were expressed the highest in early invasion stages (6 months) during PDAC progression of 10-22 cells (Fig. 3F).

The extracellular vesicle pathway was also enriched in group 5 genes, which arise slower than group 4 genes during the 10-22 PDAC progression (Fig. 3A–G). Products of most group 5 genes were associated with membrane-bound organelles (80%, n = 62), and half of them (39 of 77 genes) are related to vesicle transport (q = 3.60E-06, fold enrichment 2.51; Fig. 3G; Supplementary Table S3G), including extracellular vesicles (n = 30, q = 4.36E-02).

Altogether, neuronal development, lysosomal membrane, and exosome gene programs were derepressed during the transition from precursor to invasive PDAC in the 10-22 cell model.

**HBP1 is involved in cell migration pathways, and high HBP1 mRNA level in PDAC is associated with poor prognosis of patients**

To identify transcriptional regulators associated with the progression of PanIN to early PDAC, we performed motif analysis of the promoters of groups 4 and 5 genes. The HBP1 (55) motif ranked the highest among the group 4 genes (P = 1e-91; Fig. 4A and B; Table 4; Supplementary Table S4).

Figure 4. Enrichment of HBP1 early in PDAC progression. A, HBP1 motif. B, HBP1 mRNA expression during PDAC progression of 10-22 cells. Representative GO categories for HBP1 target genes (C) and genes correlated with HBP1 in TCGA PDAC (D). H&E staining of PanINs and benign ductal mucosa (E), HBP1-IHC in PanINs and benign ducts (F), and HBP1-IHC in PDAC, benign ducts (inset, G), islet cells (‘’), and acinar units (‘’), in a murine PDAC model. The arrowhead indicates an HBP1 positive material in the lumens of glandular cancer groups (G). Size bar is 100 μm.
Supplementary Table S4). Apparent HBP1 targets included genes involved in cell migration, including neuron differentiation (Fig. 4C; FDR < 0.05). Indeed, HBP1 is known to regulate neuronal lineage differentiation (56). Consistently, cell migration–related pathways were ranked top among genes whose expression correlated with the HBP1 mRNA level in TCGA PDAC RNA-seq dataset (57) and cellular response to DNA damage as expected (Fig. 4D; FDR < 0.01; ref. 55).

Next, we examined the expression of HBP1 protein in the pancreata of mice (25) bearing PanINs and PDAC and in human pancreatic cancer tissues (Supplementary Fig. S7A). HBP1 was barely detectable in the benign ductal epithelium (Fig. 4E and F), acinar cells (Fig. 4G, *), and islet cells (Fig. 4G, *). In contrast, HBP1 was strongly labeled in both nucleus and cytoplasm/membranes of PanINs (Fig. 4F) and slightly diffusely labeled in PDAC cells (Fig. 4G), except for budding cancers within the lumens of PDAC (Fig. 4G–G', arrowheads; n = 3).

In humans, HBP1 protein was detected mainly in islet cells of the healthy pancreas (Fig. 5A, arrow) and the acinar compartments of chronic pancreatitis tissues adjacent to PDAC (n = 22/22; Fig. 5B, arrow). The HBP1 antibody labeled the ductal region of PanINs (n = 22/22).

**Figure 5.** HBP1 in human pancreata (see Supplementary Table S7B). HBP1-IHC in normal human pancreata (A), chronic pancreatitis at the margin of PDAC (B), PanIN2 (C), stage 1 PDAC (D), stage 2 PDAC (E), and another stage 2 PDAC (F). Boxes show a magnified view, arrows indicate positive staining, and dashed arrows indicate cytoplasmic diffused/weaker signal. Scale bars indicate 50 μm. G, Semi-quantification of HBP1-IHC in chronic pancreatitis (n = 22), PanIN1 (n = 9) and PanIN2 (n = 8), and various stages of PDAC (stage 1, n = 29; stage 2, n = 26; stage 4, n = 3). Left bars indicate % of samples with indicated staining intensity, 0; barely, 1: weakly, 2: moderately, and 3: intensely. Middle bars indicate % of samples with indicated staining patterns. Right bars indicate % of samples with indicated % area of investigation stained positively. H, Overall survival curve of TCGA PDAC patients (n = 150) stratified by the HBP1 mRNA level in their PDAC.

For the HBP1 immunostaining data, the intensity of staining was classified as follows: 0 = no label, 1 = weak, 2 = moderate, and 3 = intense.
14/17; Fig. 5C) and a subset of PDAC with various degrees of intensity yet slightly diffused patterns (Fig. 5D–G). Notably, high HBP1 mRNA levels in PDAC were correlated with poor patient prognosis (n = 150, TCGA PDAC; Fig. 5H). Altogether, we identified HBP1 as a feature of the cell migration phenotype and transient neural gene expression pattern during PDAC progression and a potential biomarker for prognosis.

**HBP1 derepresses cell migratory genes and drives cell proliferation**

To understand how elevated HBP1 expression impacts pancreatic cell functions, we performed RNA-seq to define DE genes in a normal pancreatic epithelial cell line (H6C7) before (parental) and after overexpressing HBP1 (HBP1-OE) or with the backbone control vector (CV; Supplementary Fig S7B). Consistent with prior studies (55), mRNAs for cell-cycle regulators were significantly increased in HBP1-OE compared with controls (Fig. 6A and B, yellow cluster). Genes encoding growth factors, including receptor tyrosine kinases (RTK), were also upregulated in HBP1-OE, suggesting that aberrant HBP1 can trigger cell proliferation (Fig. 6A and B, blue cluster).

In support of results obtained from 10-22 cells, a group of genes enriched in ECM disassembly and axon guidance was significantly upregulated in HBP1-OE (Fig. 6A and B, blue cluster). In contrast, genes enriched in type I IFN pathways or ncRNA processing were significantly downregulated in HBP1-OE (Supplementary Fig. S6A, purple and red clusters). Altogether, the aberrant expression of HBP1 can induce cell migratory and proliferation pathways in normal pancreas cells.

Among human PDAC cell lines examined, the AsPC1 line expresses a high level of HBP1 mRNA (Fig. 6C). Thus, AsPC-1 cells were introduced with two CRISPR-Cas9 gRNAs to disrupt HBP1 or with CRISPR-Cas9 control vectors. As HBP1 requires RB binding for its function (58), we designed two gRNAs targeting its domain that binds to RB1 (Fig. 6D; ref. 55). gRNA#1 almost completely disrupted RB1-binding sites, but gRNA#2 did not (Fig. 6D). We then determined the proliferation rate of AsPC-1 cells with the ATP levels that are proportional to the number of cells and the invasion rate with the number of cells migrated through matrigel. HBP1 knockout cells (gRNA#1) had significantly reduced proliferation by 30% and invasiveness by 50%, compared with control AsPC-1 cells (Cas9; Fig. 6E and F). As expected, AsPC-1 gRNA#2 cells had less reduced proliferation and invasiveness (Fig. 6E and F). Conversely, ectopic expression of HBP1 (OE) in MiaPaCa2 cells, which express a low level of the endogenous HBP1, slightly increased invasion (Fig. 6G and H) and moderately increased proliferation (P < 0.05; Fig. 6I) compared with control cells with the backbone vector (CV). Altogether, these results suggest that aberrant expression of HBP1 increases cell proliferation and migration by upregulating the RTK pathway, axon guidance, and ECM remodeling genes during PDAC progression.

**BACH1 and RUNX3 associate with PDAC progression features**

BACH1/MAFK and RUNX motifs were the highest-ranked at promoters of group 5 genes (Fig. 7A and B; Supplementary Table S4). BACH1 forms a heterodimer with small MAF proteins such as MAFK to repress transcription (59). While MAFK was constitutively expressed across timepoints after differentiation (Supplementary Fig. S7C), the expression of BACH1 was variable throughout the time course and increased at the 9 months stage (Fig. 7C). Genes with a BACH1 binding motif were involved in degranulation/intracellular vesicle tracking (Fig. 7E). Proteins encoded by one-third of BACH1 target genes were identified in urine exosomes (FDR < 0.05; ref. 60). Patients with a high level of BACH1 mRNA in PDAC also had a worse prognosis in TCGA dataset (Fig. 7F), indicating that BACH1 may impact PDAC prognosis. The BACH1 protein was expressed mainly in islet cells of the normal human pancreas (Fig. 7G, arrows) and in acinar and ductal cells in chronic pancreatitis (n = 14/21) in the cytoplasm; Fig. 7H), but not in stromal cells in either. BACH1 was detected in both ductal and stromal cells in PanINs (n = 17) and PDAC of all stages (n = 8; Fig. 7I–M).

Among proteins that bind to the RUNX motif, RUNX1 and RUNX3 were expressed late in PDAC progression (Supplementary Fig. S7C) and the expression of RUNX3, a transcriptional repressor (61), was inversely associated with the expression of group 5 genes (Fig. 7D). Most promoters harboring the RUNX motif were of genes involved in vesicle transport (q < 0.01) and encoding proteins found in cancer-specific exosomes (n = 7/9; asterisks; Fig. 7N; ref. 62). Indeed, genes whose expression negatively correlate with the RUNX3 mRNA level in the PDAC TCGA (Pearson Correlation=0.2) were most enriched in the endosome transport with multivesicular body assembly and Endosomal Sorting Complex Required for Transport (ESCRT; ref. 63; FDR < 0.05; Supplementary Table S3H).

**Discussion**

Herein, we use the 10-22 cell model to provide the first transcriptional characterization of an isogenic trajectory of ductal cells in human PDAC development. We showed that 10-22 cells exhibit ESC gene signatures while retaining signatures of the polarized epithelial lineage from which they were derived under pluripotency culture conditions, explaining their preference to undergo PDAC development upon differentiation in mice. Reflecting their endogenous genetic alterations in KRA,s TP53, and CDKN2A, 10-22 cells exhibit basal subtype gene signatures soon after being released from pluripotency, yet classical ductal cell programs eventually take over.

Neuronal differentiation and neural stem cell proliferation genes were upregulated during early PDAC progression, implying that a neuronal differentiation program may trigger cell migration early during PDAC progression. Indeed, genetic alterations in axon guidance are observed in human PDAC (64). Perineural invasion, an unfavorable prognostic marker, is common in patients (65, 66). Perineural invasions did not correlate with tumor size and 75% of stage I PDAC cases were accompanied by neural invasion (65). These clinical observations support our finding that perineural invasion may contribute to early invasion or dissemination.

Extracellular vesicles function as mediators in cellular contact (67), and PDAC-derived exosomes induce a pre-metastatic niche formation in naive mice (68). We found that the gene signature for extracellular vesicles was transiently upregulated in PanIN and early PDAC and significant enrichment of Notch and Hedgehog pathways, which are implicated in exosome formation (63), in early PDAC progression of the 10-22 model (Supplementary Table S5). However, genes encoding extracellular vesicles are downregulated with the concomitant increase of the repressor RUNX3 (69). Indeed, most genes harboring a RUNX motif relate to cancer-specific exosomes (Fig. 7N; ref. 62). RUNX3 expression is negatively correlated with endosome transport genes via ESCRT, which is required to form a subset of exosomes (Supplementary Table S3H; ref. 63). Thus, our longitudinal transcriptome analysis uncovered transient gene signatures that could be masked at later stages by a regulator needed for PDAC. In contrast, BACH1 is a negative regulator of oxidative stress-induced response (59). The BACH1 level correlates with PDAC stages and
Figure 6. HBP1 increases cell migration and proliferation. A, Ward clustering of DE genes across parental, control GFP vector (CV), and HBP1-overexpressing H6C7 cells (OE; FDR < 0.05 by LRT). The heatmap indicates Z-score gene expression. Representative top GO in each cluster is described. B, Relative expression of top GO genes of the yellow and blue clusters, which are upregulated DE genes in HBP1-overexpressing H6C7 cells. Cell-cycle regulator genes (n = 107) are enriched in the yellow cluster, and RTK (n = 85), axonal guidance (n = 117), and ECM disassembly (n = 10) genes are enriched in the blue cluster. C, HBP1 and BACH1 mRNA levels in human PDAC cell lines shown as Transcripts Per Million (TPM) obtained from Expression Atlas. D, Validation of indel mutations in HBP1 exon 2 in AsPC-1 cells targeted with CRISPR-Cas9 with two guide RNA (gRNA)s or CRISPR-Cas9 controls (Cas9 CV) by T7 Endonuclease I treatment. Sequences of guide RNA #1 and #2, cutting the neighboring sequences to the RB-binding domain of HBP1 exon 2. E, Invasion assay at 36 hours post-plating AsPC-1 cells targeted with CRISPR-Cas9 control (Cas9 CV) or the indicated HBP1 gRNAs. Y-axis indicates the number of invading cells (technical replicates n = 6; *P = 0.04). F, Proliferation assay at 48 hours post-plating AsPC-1 cells targeted with CRISPR-Cas9 controls or the indicated HBP1 gRNAs. Data are shown as a box-and-whisker plot with a median from six independent technical replicates. Y-axis shows the proliferation rate of gRNAs cells over the control cells. *P < 0.005. G, Western blot image of HBP1 expression in MiaPaca2 cell expressing HBP1 (OE) and control vector (CV). Bar graph shows normalized HBP1 levels over GAPDH (P = 0.0226, n = 3 paired t test). Invasion assay (H, one-way ANOVA with multiple comparisons, n = 3; *P = 0.0384) and proliferation assay (I, unpaired t test with Welch correction, n = 8; **P = 0.0081) of parental (P) and MiaPaca2 cells transduced with lentiviruses expressing control GFP vector (CV) or HBP1 vectors ("OE").
BACH1 and RUNX3 as upstream regulators of genes associated with the early lesions to invasive stages. BACH1/MAFK (A) and RUNX (B) motifs. BACH1 (C) and RUNX3 (D) expressions during PDAC progression of 10-22 K-RasG12V cells. E, Venn diagram of group 5 genes with BACH1 target genes. F, Kaplan–Meier survival curve of TCGA PDAC patients stratified by BACH1 mRNA level (n = 150). BACH1-IHC in normal human pancreata (G), chronic pancreatitis at the margin of PDAC (H), ductal compartment in PanIN1 (I) and PanIN2 (J), and stage 1 (K) and stage 2B (L) PDAC. Boxes show a magnified view, arrows indicate positive staining, and dashed arrows indicate cytoplasmic diffused/weaker signal. Scale bars indicate 50 μm. M, Semi-quantification of BACH1-IHC in chronic pancreatitis (n = 21), PanIN1 (n = 9) and PanIN2 (n = 8), and PDAC (n = 6 in stage 1, n = 2 in stage 2B). Left bars indicate % of samples with indicated staining intensity: 0; barely, 1: weakly, 2: moderately, and 3: intensely. Middle bars indicate % of samples with indicated staining patterns. Right bars indicate % of samples with indicated % area of investigation stained positively (see Supplementary Table S7C for IHC quantification). N, Venn diagram of vesicle transporter genes and RUNX target genes. Asterisks denote RUNX target genes identified in cancer-specific exosomes.

High BACH1 mRNA levels associate with human PDAC patients’ poor prognosis. Ectopic BACH1 expression in MiaPaca 2 cells showed increased invasiveness and decreased proliferation (Supplementary Fig. S7D–S7G). A polymorphism in the BACH1 locus has been associated with increased pancreatic cancer risk (70), and a chimeric BACH1 transcript has been identified in a patient’s PDAC (16). BACH1 was reported to promote epithelial-to-mesenchymal transition and metastasis of PDAC (71), implying PDAC promoting roles.
Intriguingly, HBP1 is transiently expressed in early PDAC in both the 10-22 model and human clinical samples (Figs. 4B, 5A–G). Survival analyses (Fig. 5H) and our functional tests (Fig. 6) indicate a negative role of HBP1 in patients’ prognosis. Further studies are needed to unveil how HBP1 regulates pancreatic cell migration and confers neuronal migratory phenotypes and how these signaling influence PDAC progression. On the other hand, ectopic expression of HBP1 delays cells entering S-phases by prolonging the G1-phase, where cells respond to extracellular growth pathways. Indeed, we found that various RTK genes were simultaneously upregulated with cell-cycle regulators in the HBP1-overexpressing H6C7 cells (Fig. 6B). However, the effect of HBP1 on cell proliferation appears to depend on cell types (72–74). Deleterious mutations in HBP1 are observed frequently in breast cancers (72) but rarely in PDAC. Instead, focal HBP1 amplification is observed in 2.7% TCGA PDAC [57], implying a tumor-promoting role for HBP1 in PDAC.

In summary, our longitudinal transcriptome study using the 10-22 cell model shows that HBP1, RUNX3, and BACH1 may regulate cell-cycle regulators in the HBP1-overexpressing H6C7 cells (Fig. 6B).

Authors’ Disclosures

J. Kim reports grants from Medical Research Foundation New Investigator Grant and CRUK-OHSU Project Award during the conduct of the study. No disclosures were reported by the other authors.

References

22. Badea I, Herlea V, Dima SO, Dumitrascu T, Popescu I. Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adeno...
25. Hingorani SR, Petricoin EF, Maitra A, Rajapakse V, King C, Jacobetz MA, et al. Preinvasive and invasive ductal pancreatic cancer and its early detection in
27. Krah NM, De La OJ, Swift GH, Hoang CQ, Willet SG, Pan FC, et al. The acinar
32. Filonov GS, Piatkevich KD, Ting LM, Zhang J, Kim K, Verkhusha VV. Bright and
34. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR:
47. Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering meth-
48. Smith CL, Goldsmith CA, Eppig JT. The Mammalian Phenotype Ontology as a
tool for annotating, analyzing and comparing phenotypic information. Genome
49. Commisso C, Davidson SM, Suynder-Azenlogu RG, Parker SJ, Kamphorst JJ, Hackett S, et al. Macropinocytosis of protein is a amino acid supply route in
55. Tevosian SG, Shih HH, Mendelson KG, Sheppard KA, Paulson KE, Yee AS. HBPl: a HMG box transcriptional repressor that is targeted by the retinoblas-
56. Watanabe N, Koyagura R, Ohtsuka T. HBPl regulates the timing of neuronal
63. Mathivanan S, Simpson RJ. ExoCarta: a compendium of exosomal proteins and
65. Pour PM, Bell RH, Batra SK. Neural invasion in the staging of pancreatic cancer. Pancreas 2003;26:322
67. Tkach M, Thery C. Communication by extracellular vesicles: where we are and
68. Mathivanan S, Simpson RJ. ExoCarta: a compendium of exosomal proteins and
71. Tkach M, Thery C. Communication by extracellular vesicles: where we are and
72. Mathivanan S, Simpson RJ. ExoCarta: a compendium of exosomal proteins and
Longitudinal Analysis of Human Pancreatic Adenocarcinoma Development Reveals Transient Gene Expression Signatures

Jungsun Kim, Taelor Ekstrom, Wenli Yang, et al.