ZBED2 is an antagonist of Interferon Regulatory Factor 1 and modulates cell identity in pancreatic cancer

Tim D.D. Somerville¹, Yali Xu¹, Xiaoli S. Wu¹, and Christopher R. Vakoc^{1,2*}

¹Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

²Lead contact

*Correspondence: vakoc@csHL.edu

Classification

BIOLOGICAL SCIENCES: Cell Biology.

Keywords

ZBED2, IRF1, interferon, lineage plasticity, pancreatic ductal adenocarcinoma

ABSTRACT

Lineage plasticity is a prominent feature of pancreatic ductal adenocarcinoma (PDA) cells, which can occur via deregulation of lineage-specifying transcription factors. Here, we show that the zinc finger protein ZBED2 is aberrantly expressed in PDA and alters tumor cell identity in this disease. Unexpectedly, our epigenomic experiments reveal that ZBED2 is a sequence-specific transcriptional repressor of interferon-stimulated genes, which occurs through antagonism of Interferon Regulatory Factor 1 (IRF1)-mediated transcriptional activation at co-occupied promoter elements. Consequently, ZBED2 attenuates the transcriptional output and growth arrest phenotypes downstream of interferon signaling in multiple PDA cell line models. We also found that ZBED2 is preferentially expressed in the squamous molecular subtype of human PDA, in association with inferior patient survival outcomes. Consistent with this observation, we show that ZBED2 can repress the pancreatic progenitor transcriptional program, enhance motility, and promote invasion in PDA cells. Collectively, our findings suggest that high ZBED2 expression is acquired during PDA progression to suppress the interferon response pathway and to promote lineage plasticity in this disease.

SIGNIFICANCE STATEMENT

Pancreatic ductal adenocarcinoma (PDA) is one of the most lethal human malignancies, attributed in part to lineage infidelity downstream of deregulated lineage-specifying transcription factors (TFs). Here we define the biological effects of a poorly understood TF ZBED2 in the most aggressive subtype of PDA, defined by the expression of squamous lineage markers. Our study reveals two molecular functions of ZBED2 in PDA cells: an inhibitor of interferon response genes and a modifier of epithelial lineage programs. Both functions can be explained by the ability of ZBED2 to antagonize the functional output of Interferon Regulatory Factor-1 (IRF1). Our study reinforces the concept of aberrant lineage identity in cancer and highlights an unexpected connection between interferon response pathways and squamous-subtype PDA.

INTRODUCTION

Cells are capable of adopting different identities along a phenotypic spectrum, a process referred to as lineage plasticity. While lineage plasticity is a critical feature of normal development and during wound-healing responses, it is also a powerful contributor to the pathogenesis of human cancer (1, 2). Through a multitude of genetic and non-genetic mechanisms, cancer cells gain access to diverse lineage and developmental transcriptional programs, resulting in heterogeneous cell identities emerging in a clonally-derived tumor. The consequences of lineage plasticity in human cancer are far-reaching, but include a prominent role in the acquisition of metastatic traits and in the evasion of targeted therapy (1).

Pancreatic ductal adenocarcinoma (PDA) is an emerging paradigm for the study of lineage plasticity in human cancer. While PDA is defined by its histopathological resemblance to ductal epithelial cells of the exocrine pancreas, studies in mice have shown that acinar cells can serve as a cell-of-origin of this disease, which trans-differentiate into the ductal fate following acquisition of a *Kras* activating mutation (3-5). At later stages of tumor development, aberrant up-regulation or silencing of master regulator transcription factors (TFs) in PDA can lead to reprogramming of ductal identity towards that of other cell lineages, including mesenchymal (6-8), foregut endodermal (9), or squamous epithelial fates (10-12). While each of these lineage transitions are capable of promoting disease progression in experimental systems, only the presence of squamous characteristics correlates with a shorter overall survival in human PDA patients (13, 14). For this reason, the identification of mechanisms that promote squamous trans-differentiation in PDA has become an active area of investigation in recent years (10-13, 15, 16).

The interferon (IFN) transcriptional response is a conserved pathway that protects organisms from infectious pathogens and malignancy (17, 18). IFN pathway activation occurs via autocrine or paracrine IFN signalling that can be triggered in response to the detection of foreign nucleic acids

as well as ectopically located self-DNA (18, 19). Whereas almost all cell types can produce type I IFNs (e.g. IFN- α and IFN- β), type II IFN (i.e. IFN- γ) production is restricted to a subset of activated immune cells (20). IFN pathway activation promotes the transcriptional induction of hundreds of IFN-stimulated genes (ISGs), which encode diverse proteins with anti-viral, anti-proliferative, and immunostimulatory functions (21). The key TFs that promote ISG induction belong to the Signal Transducer and Activator of Transcription (STAT) and IFN Regulatory Factor (IRF) families, which can bind in an IFN-inducible manner at the promoters of ISGs (22, 23). In the classical pathway, phosphorylation of STATs downstream of IFN receptor activation triggers a rapid ISG response (23). This primary response includes the STAT-dependent transcriptional activation of several genes encoding IRFs, which subsequently drive an amplifier circuit resulting in sustained ISG induction (22). Within this complex transcriptional response, IRF1 is a critical positive regulator required for the full range of overlapping target gene activation following type I or type II IFN pathway activation (22). IRF1 is a broadly acting antiviral effector and exhibits tumor suppressor functions in multiple cellular contexts (24, 25). With respect to PDA, prior studies have shown that IRF1 can promote a differentiated epithelial cell state and inhibit cell proliferation (26, 27).

The ZBED gene family encodes nine zinc finger-containing TFs in humans, which originated from a domesticated DNA transposase gene from an hAT transposable element (28). While lacking in transposase activity, human ZBED TFs instead retain their zinc finger domain to perform sequence-specific DNA binding and function as transcriptional regulators in a cell-type specific manner (29-31). Within this family, ZBED2 is one of the least understood members, owing in part to its recent evolution and lack of a mouse ortholog (28). A prior genomewide-association study identified *ZBED2* as a candidate locus influencing risk of smoking-induced pancreatic cancer (32). More recently, *ZBED2* was found to be highly expressed in the basal layer of the epidermis, where it plays a role in regulating keratinocyte differentiation (33). Another study identified *ZBED2* as a

marker of T cell exhaustion in human CD8 T cells, although the function of ZBED2 was not investigated in this context (34). We are unaware of any prior study characterizing a transcriptional function for ZBED2 or its role in cancer.

Here we identify ZBED2 as one of the most aberrantly up-regulated TFs in human PDA. This prompted our characterization of the transcriptional function of ZBED2, which we demonstrate to be a sequence-specific transcriptional repressor. We show that the repression targets of ZBED2 are highly enriched for genes within the IFN response pathway. By interacting with ISG promoters, ZBED2 blocks the transcriptional output and growth arrest phenotypes caused by IRF1 activation downstream of IFN stimulation. We also provide evidence that ZBED2 is preferentially expressed in squamous-subtype PDA tumors and promotes loss of pancreatic progenitor cell identity in this context. Collectively, our study suggests that aberrant ZBED2 expression in PDA cells blocks the IFN response and alters epithelial cell identity in this disease.

RESULTS

Aberrant *ZBED2* expression in pancreatic ductal adenocarcinoma correlates with inferior patient survival outcomes

In this study, we sought to identify novel TFs that deregulate cell identity in PDA. By evaluating previously published transcriptome data comparing normal human pancreas and PDA (35), we identified *ZBED2* as among the most aberrantly up-regulated TF genes in tumor and metastatic lesions (Fig. 1A and B). Notably, the fold-induction of *ZBED2* was comparable to other aberrantly expressed TFs in PDA, *FOXA1* and *TP63*, which have established roles in disease progression (Fig. 1A and Dataset S1) (9, 10). In two additional independent transcriptome studies of human PDA tumors (13, 36), we validated high levels of *ZBED2* expression in 16-20% of primary patient samples (Fig. 1C and D). We further corroborated the aberrant expression pattern of *ZBED2* in ~20% of established organoid cultures derived from human PDA tumors, while *ZBED2* was expressed at low levels in organoids derived from normal pancreatic epithelial cells (Fig. 1E and Dataset S2) (37). Considering the limited number of prior studies of *ZBED2*, these observations prompted us to investigate the molecular function of *ZBED2* in PDA.

To gain initial insight into *ZBED2* function, we examined its expression across a diverse collection of normal and malignant human tissues using publicly available transcriptome data (38, 39). This analysis revealed a highly tissue-specific pattern of *ZBED2* expression in normal tissues, with skin, lung, esophageal mucosa, and thyroid tissues expressing *ZBED2* at the highest levels (Fig. 1F). In accord with the findings above, *ZBED2* expression was not detected in the normal human pancreas (Fig. 1F). Transcriptome analysis of 32 tumors types profiled by the TCGA Pan-Cancer Atlas (39) revealed pervasive *ZBED2* expression in numerous human cancer types (Fig. 1G). In specific malignancies, like thyroid carcinomas (THCA) and squamous cell carcinomas (HNSC, LUSC, ESCA), high *ZBED2* expression can be explained by its expression pattern in the normal tissue counterpart (i.e. cell-of-origin) of these tumors. However, in other cancer types, like PDA and

ovarian cancer, *ZBED2* appears to be up-regulated in an aberrant manner (Fig. 1F and G). Remarkably, we found that patients with *ZBED2*^{high} tumors have a significantly shorter overall survival than patients with *ZBED2*^{low} tumors in the context of PDA, colorectal cancer, kidney cancer, glioblastoma/glioma, lung squamous cell carcinoma and lung adenocarcinoma (Fig. 1H and I and *SI Appendix*, Fig. S1A) (40). This highlights *ZBED2* as biomarker of disease aggressiveness across several tumor subtypes.

We further interrogated *ZBED2* expression in 1,156 cancer cell lines from the CCLE database (41) and identified high expression in ~24% of cell lines (Fig. 1J). In accord with the TCGA analysis, *ZBED2*^{high} cell lines were significantly enriched for lines derived from pancreatic, ovarian, intestinal, endometrial and upper aerodigestive tract tumors (Fig. 1J and *SI Appendix*, Fig. S1B). In addition, *ZBED2*^{high} cell lines were significantly enriched for genetic alterations in *KRAS*, *TP53*, *CDKN2A* and *SMAD4* (*SI Appendix*, Fig. S1C and D), which are also the most common recurrently mutated genes in PDA (42-44). Despite this high level expression, analysis of CRISPR screening data from 559 cancer cell lines revealed that *ZBED2* is not required to sustain cell growth in any cell line under standard tissue culture conditions (*SI Appendix*, Fig. S1E).

***ZBED2* inhibits the expression of genes in the interferon response pathway**

To understand the cancer-relevant function of *ZBED2*, we performed RNA-seq analysis following *ZBED2* knockout or overexpression in PDA cell lines. Western blot analysis revealed high level expression of endogenous *ZBED2* in three out of nine PDA cell lines examined (Fig. 2A). We used lentiviral CRISPR-Cas9 editing to target *ZBED2* in the highest expressing PDA cell line (PANC0403) with two independent sgRNAs and performed RNA-seq analysis. An unbiased gene set enrichment analysis (GSEA) of the *ZBED2* knockout RNA-seq data revealed induction of the IFN pathway as top-ranked gene signatures in this experiment (Fig. 2B and C and Dataset S3 and S4). To complement this approach, we performed RNA-seq after overexpressing a *ZBED2* cDNA

in 15 different PDA cell lines, which likewise revealed consistent suppression of IFN pathway gene signatures as a top-ranked alteration (Fig. 2B, D and E and *SI Appendix*, Fig. S2A and Dataset S3-5). We confirmed the down-regulation of the ISGs *STAT2*, *CMPK2* and *MX1* following ectopic ZBED2 expression by RT-qPCR and western blotting (Fig. 2F and *SI Appendix*, Fig. S2B). We also confirmed the up-regulation of the ISGs *STAT2* and *CMPK2* in response to ZBED2 knockdown using shRNAs as an alternative approach to CRISPR (*SI Appendix*, Fig. S2C and D). Taken together, this transcriptome analysis implicates an inhibitory effect of ZBED2 on the IFN transcriptional response in PDA cells.

Despite the inhibitory effect of ZBED2 on IFN pathway genes, our western blot and RNA-seq analysis revealed that *ZBED2*^{high} PDA lines tended to have higher baseline level of IFN pathway activation, as judged by levels of STAT2 phosphorylation and ISG expression (Fig. 2A and G). This high basal activity of the IFN pathway was not attributed to autocrine IFN production, since inactivating the IFN alpha receptor (IFNAR1) did not suppress ISG levels (*SI Appendix*, Fig. S2E and F). Moreover, the positive correlation between levels of *ZBED2* and IFN pathway genes was observed across all 1,156 cell lines from the CCLE, encompassing PDA lines as well other diverse cancer types (Fig. 2H and *SI Appendix*, Fig. S2G). Additionally, we did not observe an upregulation of *ZBED2* following stimulation with IFN- β or IFN- γ , indicating that *ZBED2* is not itself an ISG in this context (*SI Appendix*, Fig. S2H). Taken together, these findings suggest that ZBED2 expression may have been acquired in PDA cells to dampen a pre-existing activation state of the IFN pathway.

We next validated the inhibitory effect of ZBED2 on the IFN pathway by performing western blot analysis of IFN components following exposure of PDA cells to IFN- β or IFN- γ (Fig. 2I and J). In these experiments, ZBED2 had no effect on STAT phosphorylation, which suggests that ZBED2 does not function in the signaling pathway downstream of receptor activation. In addition, this

analysis revealed that specific ISG products were induced normally following IFN treatment (e.g. IRF1, IRF9), while others were attenuated by the presence of ZBED2 (STAT2, MX1, and CMPK2) (Fig. 2J). These findings suggest that ZBED2 functions as a repressor of a specific subset of ISGs.

ChIP-seq analysis implicates ZBED2 as a sequence-specific repressor of ISG promoters

In order to investigate the function of ZBED2 at the molecular level, we performed chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) analysis of lentivirally expressed FLAG-ZBED2 in AsPC1 and SUIT2 PDA cell lines. We observed high concordance between ZBED2 occupancy in both cell lines, enabling us to define a set of 2,451 high confidence binding sites (Fig. 3A and *SI Appendix*, Fig. S3A and Dataset S6), which was highly biased towards promoter regions (Fig. 3B). Notably, genomic occupancy of ZBED2 was globally correlated with its repressive effect on transcription when comparing ChIP-seq with the aforementioned RNA-seq analyses. This included several ISGs (e.g. *CMPK2* and *STAT2*) that are both downregulated and located near ZBED2 binding sites (Fig. 3C and Dataset S7). To further verify the repressor function of ZBED2, we made use of GAL4 fusion protein to artificially tether ZBED2 to a plasmid with luciferase expression downstream of a thymidine kinase promoter. Unlike the established activator protein IRF1, tethering ZBED2 to this promoter resulted in a significant decrease in luciferase expression (Fig. 3D). Together, these findings suggest that ZBED2 is a transcriptional repressor that occupies specific sites in the genome.

Since other ZBED zinc fingers function as sequence specific DNA binding domains (30, 31), we attempted to define a ZBED2 motif that correlates with its genomic occupancy. Using a *de novo* motif discovery analysis in the MEME-ChIP software (45), we derived an 8-nucleotide position weight matrix that closely correlated with ZBED2 enrichment observed by ChIP-seq (Fig. 3E). In addition, we found that the presence of two or more ZBED2 motifs correlated with strong genomic occupancy as well as with a stronger repressive effect on transcription than peaks displaying only a

single ZBED2 motif (Fig. 3*F* and *G* and *SI Appendix*, Fig. S3*B*). Notably, we found that less than 1% of all human promoters of protein coding genes contain two or more ZBED2 motifs, and these promoters are strongly enriched for those driving expression of genes in the IFN pathway (Fig. 3*H* and *I* and Dataset S8). *MXI*, *CMPK2* and *STAT2* are examples of ISGs that contain two ZBED2 motifs in their promoters (Fig. 3*J*). Collectively, these results suggest that ZBED2 functions as a sequence-specific DNA-binding repressor of select ISG promoters harboring multiple ZBED2 motifs.

Antagonistic regulation of ISG promoters by ZBED2 and IRF1

In our MEME analysis of ZBED2-enriched locations in the human genome, we noticed a strong correlation between ZBED2 occupancy and a motif recognized by the IRF family TFs (Fig. 4*A* and *SI Appendix*, Fig. S4*A*). RNA-seq analysis revealed variable expression of IRF family TFs in PDA cell lines, however the levels of *ZBED2* expression were closely correlated with that of *IRF1* (Fig. 4*B* and *SI Appendix*, Fig. S4*B* and *C*). A significant positive correlation between *ZBED2* and *IRF1* expression was also observed across all 1,156 cell lines within the CCLE database (*SI Appendix*, Fig. S4*D*). We next performed ChIP-seq analysis of IRF1 in AsPC1 cells, which confirmed its overlap with ZBED2 occupancy (*SI Appendix*, Fig. S4*E* and Dataset S9). For further analysis, we defined a high-confidence set of IRF1/ZBED2 co-occupied sites (Fig. 4*C* and Dataset S10). Notably the presence of IRF1 and ZBED2 co-occupancy was a feature that was highly enriched for ISG promoters, such as *CMPK2* and *STAT2* (Fig. 4*D* and *E* and *SI Appendix*, Fig. S4*F-I*).

Since ZBED2 is a transcriptional repressor and IRF1 is a transcriptional activator (Fig. 3*D*), we reasoned that these two TFs might function in an antagonistic manner to regulate ISG expression. Consistent with this hypothesis, RT-qPCR analysis showed that IRF1 led to potent activation of *CMPK2* while this effect was ablated in the presence of ZBED2 (Fig. 4*F*). RNA-seq analysis further verified that IRF1 and ZBED2 function in an antagonistic manner to regulate the entire

program of co-occupied sites (Fig. 4G and H and *SI Appendix*, Fig. S4J-M). Considering these observations, we next treated AsPC1-empty and AsPC1-ZBED2 cells with IFN- β , IFN- γ or control for 12 hours and performed RNA-seq analysis. This analysis revealed that genes associated with IRF1/ZBED2 co-occupied sites were significantly induced by IFN and this effect was attenuated by the presence of ZBED2 (Fig. 4I and *SI Appendix*, Fig. S4N). Taken together, these data indicate that ZBED2 co-localizes with IRF1 at a subset of ISG promoters that are potently modulated by IFN pathway activation.

We noticed that the IRF1 motif and ZBED2 motif bear similarity with one another, with both factors able to recognize a GAAA sequence (Fig. 4J) (46). This led us to hypothesize that IRF1 and ZBED2 might bind to specific promoter sequences in a competitive manner, and this may contribute to the antagonistic function of these two TFs. To evaluate this, we performed IRF1 ChIP-seq in AsPC1 cells that either express or lack ZBED2. This analysis revealed a two-fold reduction of IRF1 occupancy at ~19% of IRF1/ZBED2 co-bound sites (Fig. 4K and Dataset S11). This included sites at the promoters of ISGs *STAT2* and *CMPK2* (Fig. 4K and L), which were consistently observed to be among the most strongly downregulated genes upon ZBED2 expression (Fig. 2E and 3C). Notably, these two promoters both possess motifs with a common GAAA core that are capable of being recognized by both IRF1 and ZBED2 (Fig. 4L). Taken together, these findings reveal two distinct mechanisms by which ZBED2 can block the output of IRF1 at ISG promoters: through an intrinsic repressor function and via competitive DNA binding at specific motifs.

ZBED2 protects PDA cells from IRF1- and interferon- γ -induced growth arrest

In addition to coordinating an anti-viral cellular response, IRF1 leads to a powerful growth arrest when activated in tumor cells (47-49). Consistent with this observation, we found that forced IRF1 expression in several human and murine PDA cell line contexts resulted in a significant growth

arrest (Fig. 5A and *SI Appendix*, Fig. S5A). Remarkably, if we repeated this experiment in the presence of ZBED2, the IRF1-induced growth arrest was prevented (Fig. 5B-E). To verify this result in the setting of a more physiological context, we used exogenous IFN- γ to promote IRF1 activation in PDA cells. Notably, in PDA cell lines, IRF1 knockout only blocked IFN- γ -induced growth arrest but not growth arrest caused by IFN- β , which instead relied on IRF9 (Fig. 5F-H and *SI Appendix*, Fig. S5B-D). Given that the growth inhibition effects of IFN- γ stimulation are at least partially mediated through IRF1 induction, we reasoned that ZBED2 expression might confer a protective effect in this context. Consistent with this hypothesis, ZBED2 expression was sufficient to protect PDA cells from the effects of IFN- γ -mediated growth arrest, which was observed in AsPC1 cells, as well as in three murine PDA cell lines derived from the KPC (*Kras*^{+/*LSL*-*G12D*}; *Trp53*^{+/*LSL*-*R172H*}; *Pdx1-Cre*) mouse model (Fig. 5I and J and *SI Appendix*, Fig. S5E) (50, 51). Taken together, these data suggest that ZBED2 can antagonize both the transcriptional and phenotypic consequences of IRF1 activation in PDA cells.

ZBED2 represses pancreatic progenitor lineage identity in PDA

Prior transcriptome analyses of tumor samples have revealed two major molecular subtypes of PDA: a ‘pancreatic progenitor’ and a ‘squamous’ subtype (13, 35, 36, 52). The pancreatic progenitor subtype is also known as classical PDA, and expresses endodermal TFs (e.g. GATA6, HNF4, FOXA) at high levels. In contrast, the squamous subtype (also known as basal-like) silences the pancreatic progenitor gene signature and instead expresses markers of the squamous lineage (e.g. KRT5, TP63) in association with inferior patient survival outcomes (10, 13). Remarkably, we found in three independent human tumor transcriptome datasets (13, 35, 36), *ZBED2* expression was highly associated with squamous subtype PDA tumors (Fig. 6A-F and Dataset S12). This prompted us to investigate whether ZBED2 has causal role influencing the ‘pancreatic progenitor’ or ‘squamous’ transcriptional signatures in PDA cell lines. Using the aforementioned RNA-seq analysis, we found that the pancreatic progenitor signature was significantly repressed by ZBED2

expression in 13/15 PDA cell lines examined (Fig. 6G and H). The two exceptions were MIAPaca2 and PANC1 cells, which do not express pancreatic progenitor signature genes (10). The squamous signature was more variably affected by ZBED2 expression (*SI Appendix*, Fig. S6A). In accord with previous findings (27), we found that IRF1 activates the pancreatic progenitor signature, which is consistent with its antagonism with ZBED2 (Fig. 6I). We noticed that one of the IRF1/ZBED2 co-occupied sites was found at the promoter of the *GATA6* gene, which encodes an established marker of the pancreatic progenitor transcriptional program (Fig. 6J and K) (36, 52-54). Consistent with their antagonistic activities, we found ZBED2 or IRF1 expression resulted in the repression or activation of *GATA6* in this context, respectively (Fig. 6L and *SI Appendix*, Fig. S6B and C). Of note, out of the 15 PDA cell lines used in this study, we observed the lowest levels of *GATA6* and *IRF1* in MIAPaca2 and PANC1 cells (*SI Appendix*, Fig. S6D and E).

Consistent with the results above, patient tumor samples with high levels of ZBED2 expression were much more likely to present with poorly differentiated tumors compared to those patients with low ZBED2 expression (*SI Appendix*, Fig. S6F and G). Moreover, we found that ZBED2 expression in PDA cells was sufficient to promote loss of epithelial integrity and increased invasiveness when cells were plated in Matrigel (Fig. 6M), as well as enhanced migratory behavior in scratch assays (Fig. 6N). These findings are in support of ZBED2 regulating the lineage identity of PDA cells and can account for the biased expression of ZBED2 in squamous subtype PDA tumors.

DISCUSSION

Here, we have shown that a previously uncharacterized zinc finger protein ZBED2 is aberrantly expressed in a diverse collection of human tumors in a manner that correlates with poor clinical outcomes. Using PDA cells as an experimental system, we have defined two major molecular functions of ZBED2 that may underpin its role in promoting cancer: as an antagonist of the IFN response and as a modulator of epithelial cell identity. In addition, we provide evidence that both of these functions of ZBED2 can be accounted for by antagonism of IRF1-mediated transcriptional activation.

A key mechanistic advance in our study is in revealing the DNA sequence motif recognized by ZBED2, which we show is disproportionately enriched at ISG promoters in juxtaposition with IRF1 motifs. This observation suggests that the evolved function of ZBED2 in mammalian biology is as a tissue-specific attenuator of the IFN response. The tissues that express *ZBED2* at the highest levels include thyroid, esophagus, lung, and skin, which are organs that are often exposed to pathogens and may employ ZBED2 to define thresholds for IFN responsiveness. Notably, inflammation of the thyroid is one of the most common side effects of interferon- α therapy in humans (55), which leads us to speculate that ZBED2 may have evolved to protect vulnerable tissues from the damaging effects of chronic IFN stimulation. Other ZBED TFs have also been implicated in the regulation of immune responses. For example, ZBED1 has been described as a viral restriction factor that negatively regulates viral growth (56). In addition, a recent study in plants identified a ZBED protein as a novel regulator of immunity in this context (57). Thus, the function of ZBED2 defined in our study may have its origins in the early evolutionary functions of ZBED proteins as important regulators of immune responses.

We have defined ZBED2 as an attenuator of IFN responses through its ability to antagonize the transcriptional output of IRF1. Importantly, IRF1 is a validated tumor suppressor (47-49, 58) and

IFN pathways have a well-established role in the regulation of antitumor immunity (59). Thus, antagonism of IRF1 provides a plausible explanation for why high ZBED2 expression is selected for during tumorigenesis. Less clear are the initial triggers of IRF1/IFN pathway activation in PDA. A recent study demonstrated that cancer cells exhibit high ISG expression due to chronic IFN signaling sustained by the STING adaptor protein, which is activated by various intracellular nucleic acid sensors (60, 61). Thus, elevated ISG expression in cancer cells could be attributed to the accumulation of cytosolic double-stranded DNA, which may arise from the disruption of micronuclei as a result of DNA damage, chromosome mis-segregation, and chromothripsis (62, 63). Such features of high genomic instability have been strongly implicated in the development and progression of PDA (43, 64). However, sustained activation of ISG expression via the STING pathway was shown to be dependent on chronic production of type I IFNs (60). In our analysis of human PDA cell lines, we observed high basal levels of ISG expression in PDA cells in a manner that is independent of type I IFNs. While the triggers of the ISG pathway in PDA await further characterization, our results strongly suggest that the function of ZBED2 in this context is to block the anti-tumor effects of IRF1 within the IFN pathway.

Histopathological and transcriptome studies have shown that a subset of aggressive PDA tumors undergo trans-differentiation into the squamous lineage (13, 65-67). In addition to the aberrant upregulation of squamous lineage markers, a key attribute of these tumors is the silencing of genes associated with pancreatic progenitor cell identity. A key finding in our study is that ZBED2 is selectively upregulated in squamous-subtype PDA tumors, and in this context represses the pancreatic progenitor transcriptional program. Notably, we also observe high *ZBED2* expression in normal squamous epithelial tissues, such as the skin, lung and esophageal mucosa, suggesting a normal function for ZBED2 in the regulation of squamous epithelial cells. Interestingly, single-cell RNA-seq analysis of human foreskin keratinocytes recently identified ZBED2 as a key TF in the regulation of keratinocyte differentiation programs (33). ZBED2 was predicted to promote the

basal keratinocyte state and experimental depletion of ZBED2 was shown to induce keratinocyte differentiation, demonstrated by the upregulation of *KRT10* (33). It is noteworthy that IRF family members have also been ascribed functional roles in regulation of the skin epidermis. For example, both IRF2 and IRF6 are necessary for regulating proliferation and terminal differentiation of keratinocytes within the epidermis (68-71). Thus, it is possible that the ability of ZBED2 to modulate the transcriptional output of IRF TFs may extend to its ability to regulate stem cell function and cell identity within normal squamous epithelial tissues.

Similar to ZBED2, prior studies have implicated TP63 and GLI2 as additional determinants of squamous cell identity in PDA (10, 11, 13, 15). Our analysis has shown that ZBED2 and TP63 are among the most aberrantly expressed TFs in the squamous subtype of PDA, however these two TFs are not expressed in a mutually exclusive manner. These findings suggest the aberrant expression of TP63, GLI2, or ZBED2 offer distinct mechanisms by which pancreatic cancer cells with a ductal identity adopt squamous features during the pathogenesis and progression of this disease. An important area of future investigation will be to determine whether such factors can function in a collaborative manner to regulate squamous identity in this disease.

An important feature of squamous subtype tumors is that they display poorly differentiated histopathological features and an exceptionally poor prognosis (13, 35, 36, 52, 72, 73). Interestingly, a recent study identified IRF1 as one of several candidate TFs responsible for preserving differentiation and epithelial identity in human PDA (27). In this study, IRF1 was shown to associate with ‘low-grade’ enhancers controlling grade-specific transcriptional programs, which is in accord with our functional studies of IRF1. We have shown that a key IRF1/ZBED2 co-occupied target gene is *GATA6*, which is overexpressed and amplified in the pancreatic progenitor subtype of PDA (52, 54, 74). Indeed, *GATA6* expression is altered by multiple mechanisms in pancreatic tumors and was found to be a robust biomarker for PDA subtypes (36, 73). Thus, our

study further reinforces how cell identity in PDA is defined by antagonism among TFs important for maintaining the epithelial identity of this lineage.

MATERIALS AND METHODS

Cell Lines and Cell Culture

Human PDA cell lines were cultured in RPMI 1640 (Gibco) containing 10% FBS. Mouse PDA cell lines, HEK 293T cells and Plat-E cells were cultured in DMEM containing 10% FBS. Plat-E and HEK 293T cells were used for packing retrovirus and lentivirus, respectively. All cells were cultured with 1% L-glutamine and 1% Penicillin/Streptomycin at 37°C with 5% CO₂. All cell lines were routinely tested for mycoplasma. Interferon treatments of cell lines were performed with IFN- β (PeproTech; #300-02BC), human IFN- γ (PeproTech; #300-02), or mouse IFN- γ (PeproTech; #315-05).

Plasmid Construction

ZBED2 and IRF1 cDNA was PCR amplified from oligo-dT primed cDNA from gDNA harvested from CAPAN2 cells, gel purified and cloned into LentiV_Puro vector (addgene #111886) using In-Fusion following the manufacturer's instructions (Clontech: #638909). A 3*FLAG sequence was added to the N-terminus of ZBED2 cDNA to generate the FLAG-ZBED2 construct.

Lentiviral Production and Infection

Lentivirus was produced in HEK 293T cells by transfecting plasmid DNA and packaging plasmids (VSVG and psPAX2) using Polyethylenimine (PEI 25000; Polysciences; Cat# 23966-1). Media was replaced with target media 6-8 hours following transfection and lentivirus-containing supernatant was subsequently collected every 12 hours for 48 hours prior to filtration through a 0.45 μ m filter. For infection of human and mouse PDA cells, cell suspensions were mixed with lentiviral-containing supernatant supplemented with polybrene to a final concentration of 4 μ g/ml. Cells were plated in tissue culture plates of the appropriate size and lentiviral-containing supernatant was replaced with fresh media after an incubation period of 24 hours. At 48 hours post-

infection, transduced cells were selected with puromycin (3µg/ml) or G418 (1mg/ml) for five days prior to RNA extraction, lysate preparation and/or phenotypic analyses on day seven post infection.

Retroviral Production and Infection

Retrovirus was produced in Plat-E cells by transient transfection with PEI (Polysciences; Cat# 23966-1) according to standard protocols. The viral supernatants were harvested at 48 and 72 hours post-transfection prior to filtration through a 0.45µm filter. To maximize transduction efficiency, PANC0403 cells were infected with retrovirus by centrifugation at 2,000 rpm for 30 mins in the presence of 8µg/ml polybrene. Retroviral-containing supernatant was replaced with fresh media after an incubation period of 24 hours. At 48 hours post-infection, transduced cells were selected with G418 (1mg/ml) for five days prior to RNA extraction.

CRISPR-based Targeting

To generate cell lines in which ZBED2, IRF1, IRF9 or IFNAR1 had been stably knocked out, PDA cells expressing Cas9 in the LentiV-Cas9-puro vector (addgene # 108100) were infected in a pooled fashion with domain-targeting sgRNAs or a control sgRNA (sgNEG) in the LRG2.1_Neo vector (addgene #125593). sgRNA targeting sequences can be found in Dataset S13.

shRNA-based Targeting

shRNAs targeting ZBED2 or control were cloned into the miR30-based retroviral shRNA expression vector LENC (LTR-miR30-shRNA-PGK-neo-mCherry) (#111163) (75). Two days post infection with shRNAs, transduced cells were selected with 1mg/ml of G418 for five days prior to RNA extraction. shRNA sequences can be found in Dataset S13.

RNA extraction and RT-qPCR analysis

Total RNA was extracted using TRIzol reagent following the manufacturer's instructions. 1µg of total RNA was reverse transcribed using qScript cDNA SuperMix (Quanta bio; 95048-500), followed by RT-qPCR analysis with Power SYBR Green Master Mix (Thermo Fisher Scientific; 4368577) on an ABI 7900HT fast real-time PCR system. Gene expression was normalized to *GAPDH*. RT-qPCR primers used can be found in Dataset S13.

Cell Lysate Preparation and Western Blot Analysis

Cell cultures were collected and 1×10^6 cells were counted by trypan blue exclusion, washed with ice cold PBS, resuspended in 100µl PBS and lysed with 100µl of 2x Laemmli Sample Buffer supplemented with β-mercaptoethanol by boiling for 30 minutes. Samples were centrifuged at 4°C for 15 mins at 10,000xg and the supernatant was used for western blot analysis with standard SDS-PAGE-based procedures. Primary antibodies used were ZBED2 (Sigma-Aldrich; SAB1409809), IRF1 (Cell Signaling; #8478), IRF9 (Cell Signaling; #76684), MX1 (Cell Signaling; #37849), pSTAT2 (Cell Signaling; #88410), pSTAT1 (Cell Signaling; #9167), STAT2 (Cell Signaling; #72604), CMPK2 (abcam; ab103658), HSC70 (Santa Cruz; sc-7298) and ACTB (Sigma-Aldrich; A3854). Proteins were detected using HRP-conjugated secondary antibodies.

GAL4 Fusion Reporter Assays

To measure transcriptional activation activity of ZBED2 and IRF1, the GAL4 DNA-binding domain was fused to full length ZBED2 and IRF1 cDNA and these constructs were individually co-transfected into HEK 293T cells with the pGL4.35[luc2P/9XGAL4UAS/Hygro] vector (Promega, #E1370). 48 hours post transfection, an equal volume of Dual-Glo luciferase buffer and Dual-Glo Stop&Glo buffer were added sequentially and relative luciferase activity was measured according to manufacturer's instructions. Firefly luciferase activity was normalized to internal Renilla luciferase activity.

In Vitro Luciferase Imaging and 3D Matrigel Assays

To generate luciferase expressing PDA cell lines, cells were infected with a luciferase transgene in a Lenti-luciferase-blast vector (10) and stable cell lines were generated by selection with 10 μ g/ml blasticidin. Following CRISPR-based targeting or cDNA expression, cells were plated at a density of 5x10³-2x10⁴ cells per well of a black, clear-bottom, ultra-low attachment 96-well plate (10014-318; CELLSTAR) in 200 μ l of RPMI supplemented with 2% FBS and 5% Matrigel with or without the addition of human IFN- γ (PeproTech; #300-02), mouse IFN- γ (PeproTech; #315-05) or control. Cells were imaged on day seven post-plating using an IVIS Spectrum system (Caliper Life Sciences) six minutes post addition of D-Luciferin (150 μ g/ml) to each well. Bright field images were also captured on day seven post plating.

Scratch Assays

Cells were first plated to confluency in triplicate in wells of a standard 24-well plate. At day 0 of the assay, a wound was applied down the center of the well using a pipette tip. Media was subsequently removed and cells washed with PBS before addition of 1ml serum-free RPMI. Bright field images were captured using a 4x objective immediately (0 hours) and then at 6 hours, 24 hours and 48 hours post-plating. Area of the wound was quantified using ImageJ software (NIH).

RNA-seq Library Construction

RNA-seq libraries were constructed using the TruSeq sample Prep Kit V2 (Illumina) according to the manufacturer's instructions. Briefly, 2 μ g of purified RNA was poly-A selected and fragmented with fragmentation enzyme. cDNA was synthesized with Super Script II Reverse Transcriptase (Thermo Fisher; 18064014), followed by end repair, A-tailing and PCR amplification. RNA-seq libraries were single-end sequenced for 50bp using an Illumina NextSeq platform (Cold Spring Harbor Genome Center, Woodbury).

RNA-Seq Data Analysis

Single end 50bp sequencing reads were mapped to the hg19 or hg38 genomes using HISAT2 with standard parameters (76). Structural RNA was masked and differentially expressed genes were identified using Cuffdiff (77). All the following analysis was performed on genes with an RPKM value no less than 2 in either control or experimental samples. To generate ranked gene lists for Pre-ranked GSEA, genes were ranked by their mean $\log_2$ fold change between the two experimental groups of interest. For unbiased interrogation of the transcriptional consequences following ZBED2 knockout in PANC0403 cells or ZBED2 cDNA expression in HPAFII cells, ranked gene lists were ran against the H1, C2, C5 and C6 gene sets within the MSigDB v7.0 ($n = 15,736$) with those with ≤ 15 or ≥ 500 expressed genes filtered out, leaving 7,573 and 7,593 interrogated genes sets for PANC0403 and HPAFII cells, respectively. The Interferon Response gene signature was defined by merging the Broad Institute's Hallmark Interferon Alpha Response and Interferon Gamma Response gene signatures to generate a gene set comprising 222 unique gene symbols. The Progenitor-PDA and Squamous-PDA Identity signatures have been previously described (10). Heat maps of expression values were generated using Morpheus from the Broad Institute (software.broadinstitute.org/morpheus). Gene ontology analysis was performed using Metascape (78).

ChIP-Seq Library Construction

ChIP procedures were performed as previously described (10). Antibodies used for ChIP-seq in this study were IRF1 (Cell Signaling; #8478) and FLAG (Sigma-Aldrich; F1804). ChIP-seq libraries were constructed using Illumina TruSeq ChIP Sample Prep kit following manufacture's protocol. Briefly, ChIP DNA was end repaired, followed by A-tailing and size selection (300-500bp) by gel electrophoresis using a 2% gel. 15 PCR cycles were used for final library amplification that was analyzed on a Bioanalyzer using a high sensitivity DNA chip (Agilent). ChIP-seq libraries were

single-end sequenced for 50bp using an Illumina NextSeq platform (Cold Spring Harbor Genome Center, Woodbury).

ChIP-Seq Data Analysis

Single end 50bp sequencing reads were mapped to the hg19 genome using Bowtie2 with default settings (79). After removing duplicated mapped reads using SAM tools (80), MACS 1.4.2 was used to call peaks using input genomic DNA as control (81). Only peaks enriched greater than or equal to 10-fold over input samples were used for subsequent analyses. FLAG-ZBED2 peaks that were >10 fold over input with >5 tags per million in both the AsPC1 and SUIT2 datasets (overlapping intervals by at least 1bp) were taken forward as high confidence ZBED2 peaks (n = 2,451). De novo motif discovery was performed on DNA sequences extending 200bp upstream and downstream of the top 1,000 high confidence peak summits using MEME-ChIP (45) from the MEME Suite (82). The same DNA sequences were used to identify motifs enriched at ZBED2 binding sites using CentriMo via the MEME Suite with default settings (83). For this analysis, the non-redundant JASPAR core vertebrate database (v2018) containing 579 transcription factor motifs were interrogated. Annotation of ChIP-seq peaks was performed using HOMER v4.9 with default settings (84). To interrogate expression changes of ZBED2 bound genes in HPAFII cells, the top 500 unique, protein coding genes that were expressed and annotated to ZBED2 peaks (ranked by ZBED2 binding intensity) were defined as ZBED2 bound genes used for GSEA. To identify high confidence IRF1 peaks in AsPC1 cells, cells were first treated overnight with control or IFN- γ (0.2ng/ μ l; PeproTech; #300-02) prior to ChIP to increase IRF1 expression. Peaks that were >10 fold over input from both the control and IFN- γ treated datasets were merged giving 11,783 peaks. Tag counts were recounted and those peaks with >5 tags per million in either dataset were taken forward as high confidence IRF1 peaks (n = 1,946). Peaks that may be recognized by both ZBED2 and IRF1 were defined as those intervals in high confidence ZBED2 and IRF1 peaks that overlap by at least 1bp. TreeView software was used to generate the heat-map density plots and the contrast

was adjusted proportionally to the total uniquely mapped reads for visual comparison across samples (85). H3K27ac ChIP-seq in AsPC1, SUIT2 and HPAFII cells were from a previous study (10).

ZBED2 Motif Scanning

The ZBED2 motif output from *de novo* motif discovery analysis was used for this analysis. Genomic DNA and transcriptional start site annotation were downloaded from GENCODE: GRCh37 v19 for human (n=73,182). Promoters were defined as 500 bp sequences upstream and 100 bp downstream of annotated transcriptional start sites of protein coding transcripts. FIMO (86) was used to scan both strands of each promoter DNA sequence using a p-value cutoff of 10^{-4} and default settings using the MEME suite. For GSEA evaluating ZBED2 bound genes at promoter regions with ≥ 2 motif counts upon ZBED2 cDNA expression, this analysis was restricted to 139 protein coding genes with RPKM expression values ≥ 2 in at least 10/15 of the PDA cell lines interrogated; for comparative analysis, a random set of 139 protein coding genes meeting the same criteria with 0 or 1 ZBED2 motifs in their promoter was used.

Analysis of Publicly Available Datasets

For analysis of TF expression in primary PDA tumors versus normal pancreas tissue from the study by Moffitt et al (GSE71729), as well as Basal-like versus Classical tumors, TFs (87) with a mRNA array value ≥ 2 in at least 10% of samples (n=252 samples) were considered as expressed (n=1,079 TFs). For analysis of TF expression in Basal-like versus Classical tumors from the TCGA-PAAD study, TFs with a $\log_2$ RSEM value ≥ 6.1 in at least 10% of samples (n=150 samples) were considered as expressed (n=1,110 TFs). For analysis of TF expression in Squamous versus Progenitor tumors from the study by Bailey et al (ICGC-PACA-AU), TFs with a CPM value ≥ 4 in at least 10% of samples (n=96 samples) were considered as expressed (n=1,052 TFs). Patient survival analysis across the TCGA dataset was performed using a publically available web tool

(40). Patient survival data for PDA studies were obtained from the cBioPortal (TCGA-PAAD) (88) and ICGC Data Portal (PACA-AU) (13) and were downloaded in January 2018. For the TCGA-PAAD study, only the 150 confirmed PDA cases were used for the analyses in this study (36). In the two PDA studies, samples were designated as *ZBED2*^{high} or *ZBED2*^{low} based on z-score expression values >0.35 or <0, respectively. For analysis of *ZBED2* expression in the TCGA PanCancer Atlas studies (39) and the Broad Institute's CCLE (41), data was downloaded from the CBioPortal (88) in October 2019. Cell lines in the CCLE database with a *ZBED2* log₂(RPKM+1) expression value ≥ 1.5 , which approximates to cell lines in 75th percentile, were stratified as *ZBED2*^{high}; those less than this value were stratified as *ZBED2*^{low}. Analysis of genetic alterations in *ZBED2*^{high} versus *ZBED2*^{low} cell lines was performed using the CBioPortal and 127 significantly mutated genes identified from 12 tumor types were interrogated (89).

Statistical Analysis

For graphical representation of data and statistical analysis, GraphPad Prism was used and statistical significance was evaluated by *p* value using Prism software as indicated in the figure legends. Data are presented as mean with SEM and n refers to the number of biological repeats. For Kaplan-Meier survival curves, the log rank (Mantel-Cox) test was used to estimate median overall survival and statistical significance.

Data and Software Availability

ChIP-seq and RNA-seq data reported in this paper is available via the GEO: GSE141607 with reviewer token: gpatcqwtxihfgt. Previously deposited ChIP-seq data (10) can be found here GEO: GSE115463. PDA patient microarray data (35) can be found here GEO: GSE71729.

Acknowledgements

The authors would like to thank the Cold Spring Harbor Cancer Center Support Grant (CCSG) shared resources: Bioinformatics Shared Resource and Next Generation Sequencing Core Facility. We would also like to thank Kenneth Olive and Carlos Maurer (Columbia University, NY) for critical discussions during manuscript preparation and Noah Dukler (CSHL, NY) for phylogenetic analysis of ZBED family members. T.D.D.S. was supported by a grant from the State of New York, contract no. C150158. C.R.V. was supported by Pershing Square Sohn Cancer Research Alliance, the Cold Spring Harbor Laboratory and Northwell Health Affiliation, the National Cancer Institute (NCI) 5P01CA013106-Project 4 and 1RO1CA229699, the Thompson Family Foundation, the Simons Foundation and a Career Development Award from the Pancreatic Cancer Action Network-American Association for Cancer Research (AACR) 16-20-25- VAKO.

Disclosures

C.R.V. has received funding from Boehringer-Ingelheim and is an advisor to KSQ Therapeutics.

FIGURE TITLES AND LEGENDS

Fig. 1. Aberrant *ZBED2* expression in pancreatic ductal adenocarcinoma correlates with inferior patient survival outcomes.

(A) Expressed TFs ranked by their mean log₂ fold change in primary PDA versus normal pancreas tissue samples. Selected transcription factors are labeled along with their rank. (B) *ZBED2* expression in normal pancreas tissue and primary and metastatic PDA tumor samples. *p* value was determined by one-way ANOVA. (C and D) *ZBED2* expression across PDA patient samples. Scale bar indicates the standardized expression value. (E) *ZBED2* expression in human organoids derived from normal pancreatic tissue or neoplastic PDA tumor samples. (F and G) *ZBED2* expression in normal tissue samples from the GTEx portal (F) or profiled tumors from the TCGA Pan-Cancer Atlas (G). Arrow indicates pancreas tissue. (H and I) Survival curves of patients stratified according to high or low *ZBED2* expression. *p* value was calculated using the log rank (Mantel-Cox) test. (J) *ZBED2* expression in cancer cell lines from the CCLE database. Arrow indicates pancreas tissue. See also Fig. S1.

Fig. 2. *ZBED2* inhibits the expression of genes in the interferon response pathway.

(A) Western blot analysis for *ZBED2* and IFN pathway components in human PDA cell lines. (B) GSEA following *ZBED2* knockout in PANC0403 cells (left panel) or *ZBED2* cDNA expression in HPAFII cells (right panel) versus their respective controls. Normalized enrichment score (NES) and family-wise error rate (FWER) *p* value were ranked and plotted for 7,573 (PANC0403) and 7,593 (HPAFII) genes sets within MSigDB v7.0 (H1, C2, C5 and C6 collections). Each gene set is depicted as a single dot. (C and D) GSEA plots evaluating the Interferon Response signature upon *ZBED2* knockout in PANC0403 cells (C) or *ZBED2* cDNA expression in HPAFII cells (D). (E) Gene expression changes in HPAFII cells infected with *ZBED2* cDNA versus those infected with an empty vector control. *ZBED2* and the ISGs *STAT2*, *MX1* and *CMPK2* are labeled along with their rank with respect to down regulated genes. (F) Western blot analysis for HSC70, *ZBED2* and

the IFN pathway proteins STAT2 and MX1 following ZBED2 cDNA expression in HPAFII and PANC0403 cells. (G) ZBED2 expression versus the median expression value of Interferon Response genes across 15 human PDA cell lines. (H) Median expression values of Interferon Response genes across all cancer cell lines within the CCLE database (left panel) or in the indicated tissue lineages stratified according to high or low ZBED2 expression. Each cell line is depicted as a single dot. **p <0.01, ***p <0.001 by Student's t test. (I) Schematic representation of IFN signaling pathways and the IRF1 amplification loop. ISGs, interferon stimulated genes. Adapted from ref (22). (J) Western blot analysis for HSC70, ZBED2 and the indicated IFN pathway components following 12 hour stimulation with the increasing concentrations of IFN- β or IFN- γ in AsPC1 cells stably expressing ZBED2 or the empty vector. See also Fig. S2.

Fig. 3. ChIP-seq analysis implicates ZBED2 as a sequence-specific repressor of ISG promoters.

(A) Density plots showing FLAG-ZBED2 and H3K27ac enrichment surrounding a 2-kb interval centered on the summit of 2,451 high-confidence ZBED2 peaks in AsPC1 and SUIT2 cells, ranked by FLAG-ZBED2 peak intensity in AsPC1 cells. (B) Pie chart showing the distribution of high-confidence FLAG-ZBED2 peaks in AsPC1 cells. TTS, transcription termination site; TSS, transcription start site; UTR, untranslated region. (C) GSEA plot evaluating ZBED2 bound genes upon ZBED2 cDNA expression in HPAFII cells. Leading edge (indicated by the red line), Interferon Response genes are listed. (D) GAL4 fusion reporter assay testing full length ZBED2 and IRF1 transactivation activity normalized to Renilla luciferase internal control. Mean+SEM is shown. n=3. **p <0.01, ***p <0.001 by Student's t test. (E) ZBED2 ChIP-seq derived *de novo* motif logo, distribution, and E value for the ZBED2 binding motif derived using MEME-ChIP. (F) Meta-profile comparing ZBED2 occupancy in AsPC1 cells around the center of peaks with 0, 1 or ≥ 2 motif counts. ZBED2 binding intensity is shown as sequencing depth normalized tag count. (G) Summary of GSEA evaluating ZBED2 bound genes at promoter regions with 0, 1 or ≥ 2 motif

counts upon ZBED2 cDNA expression in 15 PDA cell lines. (H) Summary of ZBED2 motif density frequency in the human genome. The ZBED2 motif frequencies were calculated in all annotated human promoter regions (n=73,182). Considering the presence of multiple promoter sequences per gene, one representative promoter with the highest count per gene was selected, leaving n=20,094 unique gene promoters to calculate the frequencies. (I) Gene ontology (GO) analysis with Metascape for genes annotated by HOMER to promoter regions with ≥ 2 ZBED2 motif counts. Terms are ranked by their significance (p value) and the most significant term is highlighted. (J) ChIP-seq profiles of FLAG-ZBED2 in AsPC1 cells and H3K27ac in HPAFII cells at the promoters of the interferon genes *MX1*, *CMPK2* and *STAT2*. See also Fig. S3.

Fig. 4. Antagonistic regulation of ISG promoters by ZBED2 and IRF1.

(A) CentriMo motif enrichment analysis for JASPAR motifs at ZBED2 binding sites. (B) ZBED2 expression versus the *IRF1* expression across 15 human PDA cell lines. (C) Density plot showing IRF1 and FLAG-ZBED2 enrichment surrounding a 2-kb interval centered on the summit of 140 intersecting IRF1 and FLAG-ZBED2 sites in AsPC1 cells, ranked by IRF1 peak intensity. (D) ChIP-seq profiles of IRF1 and FLAG-ZBED2 in AsPC1 cells at the promoters of *CMPK2* and *STAT2*. (E) Gene ontology (GO) analysis with Metascape of genes annotated by HOMER to IRF1/ZBED2 sites. Terms are ranked by their significance (p value) and the most significant terms ($-\log_{10} p$ value >12) are shown. (F) RT-qPCR analysis of *CMPK2* in AsPC1-empty and AsPC1-ZBED2 cells following IRF1 cDNA expression. Mean+SEM is shown. n=3. ** $p < 0.01$ by Student's t test. (G and H) GSEA plots evaluating protein coding genes annotated by HOMER to IRF1/ZBED2 sites upon IRF1 (G) or ZBED2 (H) cDNA expression in AsPC1 cells. (I) Expression levels of protein coding genes annotated to IRF1/ZBED2 sites following 12-hour treatment with 0.2ng/ μ l of IFN- β , IFN- γ or control. **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$ * $p < 0.05$ by one-way ANOVA. (J) IRF1 motif logo from the JASPAR database (top panel) and the ZBED2 motif logo (bottom panel). (K) ChIP-seq analysis showing the $\log_2$ fold change in IRF1 occupancy at IRF1

binding sites in AsPC1-ZBED2 versus AsPC1-empty cells. (L) ChIP-seq profiles of FLAG-ZBED2 and IRF1 at the promoters of *CMPK2* (left panel) and *STAT2* (right panel) in AsPC1-empty or AsPC1-ZBED2 cells. ZBED2 and IRF1 motifs recovered by FIMO ($p < 0.001$) are highlighted in purple and yellow, respectively. See also Fig. S4.

Fig. 5. ZBED2 protects PDA cells from IRF1- and interferon- γ -induced growth arrest.

(A) Luciferase-based quantification of cell viability of AsPC1 cells grown in Matrigel on day 7 post infection with IRF1 cDNA or the empty vector. Representative bright field images (right panel) are shown. Scale bar indicates 200 μ m. (B) Luciferase-based quantification of cell viability of AsPC1-empty and AsPC1-ZBED2 cells grown in Matrigel following co-expression of IRF1 cDNA or empty vector for 7 days. (C) Representative bright field images from (B). Scale bar indicates 500 μ m. (D) Luciferase-based quantification of cell viability of KPC-derived FC1199 cells stably expressing ZBED2 or the empty vector grown in Matrigel following co-expression of IRF1 cDNA or empty vector for 7 days. (E) Representative bright field images from (D). Scale bar indicates 200 μ m. (F-H) AsPC1 cells infected with sgRNAs targeting IRF1, IRF9 or a control sgRNA (sgNEG) were plated in Matrigel and grown for 7 days in the presence of 20ng/ μ l of IFN- γ , IFN- β or control. Representative western blots following overnight stimulation with 20ng/ μ l of IFN- γ (F), luciferase-based quantification (G) and representative bright field images on day 7 of the assay (H) are shown. Scale bar indicates 500 μ m. (I and J) AsPC1 cells (I) or the indicated mouse KPC cell lines (J) were infected with the ZBED2 cDNA or an empty vector and grown in Matrigel with the increasing concentrations of IFN- γ . Bar charts show luciferase-based quantification on day 7. Mean+SEM is shown. n=3. For (A), (B), (D) and (G), *** $p < 0.001$ by Student's t test. For (I) and (J), *** $p < 0.001$, ** $p < 0.01$ * $p < 0.05$ by one-way ANOVA. See also Fig. S5.

Fig. 6. ZBED2 represses pancreatic progenitor lineage identity in PDA.

(A-C) TF expression in Basal-like/Squamous and Classical/Progenitor subtypes of PDA. TFs are ranked by their mean log₂ fold change in Basal-like versus Classical (A and C) or Squamous versus Progenitor (B) patient samples from the indicated studies. (D-F) ZBED2 expression in PDA patient samples stratified according to molecular subtype. Each dot represents one patient sample. *p* value was calculated using Student's *t* test. (G) Summary of GSEA evaluating the Progenitor PDA Identity signature upon ZBED2 cDNA expression in 15 PDA cell lines. (H and I) GSEA plots evaluating the Progenitor-PDA Identity signature following expression of ZBED2 (H) or IRF1 (I) in AsPC1 cells. (J) Overlap of Progenitor PDA Identity genes with protein coding genes associated with IRF1/ZBED2 sites. (K) ChIP-seq profiles of IRF1 and FLAG-ZBED2 at the promoter of *GATA6* in AsPC1 cells. (L) Overlap of leading edge (LE) genes associated with IRF1/ZBED2 sites that are repressed by ZBED2 or activated by IRF1 in AsPC1 cells. (M) Quantification of colony size in three-dimensional (3D) Matrigel colony formation assays. Means+SEM are shown. n=3. Representative images at day 7 are shown (right panel). (N) Quantification of scratch assays at the indicated time points post-seeding and representative images (right panel). Scale bar indicates 500µm. Means+SEM are shown. n=3. ****p*<0.001, ***p*<0.01 **p* <0.05 by Student's *t* test. See also Fig. S6.

REFERENCES

1. Yuan S, Norgard RJ, & Stanger BZ (2019) Cellular Plasticity in Cancer. *Cancer Discovery* 9(7):837-851.
2. Merrell AJ & Stanger BZ (2016) Adult cell plasticity in vivo: de-differentiation and transdifferentiation are back in style. *Nature Reviews Molecular Cell Biology* 17(7):413-425.
3. Grippo PJ, Nowlin PS, Demeure MJ, Longnecker DS, & Sandgren EP (2003) Preinvasive Pancreatic Neoplasia of Ductal Phenotype Induced by Acinar Cell Targeting of Mutant Kras in Transgenic Mice. *Cancer Research* 63(9):2016-2019.
4. Kopp Janel L, *et al.* (2012) Identification of Sox9-Dependent Acinar-to-Ductal Reprogramming as the Principal Mechanism for Initiation of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* 22(6):737-750.
5. De La O J-P, *et al.* (2008) Notch and Kras reprogram pancreatic acinar cells to ductal intraepithelial neoplasia. *Proceedings of the National Academy of Sciences* 105(48):18907.
6. Rhim AD, *et al.* (2012) EMT and Dissemination Precede Pancreatic Tumor Formation. *Cell* 148(1):349-361.
7. Krebs AM, *et al.* (2017) The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. *Nature Cell Biology* 19:518.
8. David CJ, *et al.* (2016) TGF- β Tumor Suppression through a Lethal EMT. *Cell* 164(5):1015-1030.
9. Roe J-S, *et al.* (2017) Enhancer Reprogramming Promotes Pancreatic Cancer Metastasis. *Cell* 170(5):875-888.e820.
10. Somerville TDD, *et al.* (2018) TP63-Mediated Enhancer Reprogramming Drives the Squamous Subtype of Pancreatic Ductal Adenocarcinoma. *Cell Reports* 25(7):1741-1755.e1747.
11. Hamdan FH & Johnsen SA (2018) DeltaNp63-dependent super enhancers define molecular identity in pancreatic cancer by an interconnected transcription factor network. *Proceedings of the National Academy of Sciences* 115(52):E12343-E12352.
12. Andricovich J, *et al.* (2018) Loss of KDM6A Activates Super-Enhancers to Induce Gender-Specific Squamous-like Pancreatic Cancer and Confers Sensitivity to BET Inhibitors. *Cancer Cell* 33(3):512-526.e518.
13. Bailey P, *et al.* (2016) Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* 531:47.
14. Collisson EA, Bailey P, Chang DK, & Biankin AV (2019) Molecular subtypes of pancreatic cancer. *Nature Reviews Gastroenterology & Hepatology*.
15. Adams CR, *et al.* (2019) Transcriptional control of subtype switching ensures adaptation and growth of pancreatic cancer. *Elife* 8.

16. Candido JB, *et al.* (2018) CSF1R+ Macrophages Sustain Pancreatic Tumor Growth through T Cell Suppression and Maintenance of Key Gene Programs that Define the Squamous Subtype. *Cell Reports* 23(5):1448-1460.
17. Zitvogel L, Galluzzi L, Kepp O, Smyth MJ, & Kroemer G (2015) Type I interferons in anticancer immunity. *Nature Reviews Immunology* 15(7):405-414.
18. McNab F, Mayer-Barber K, Sher A, Wack A, & O'Garra A (2015) Type I interferons in infectious disease. *Nature Reviews Immunology* 15(2):87-103.
19. Motwani M, Pesiridis S, & Fitzgerald KA (2019) DNA sensing by the cGAS–STING pathway in health and disease. *Nature Reviews Genetics* 20(11):657-674.
20. Pestka S, Krause CD, & Walter MR (2004) Interferons, interferon-like cytokines, and their receptors. *Immunological Reviews* 202(1):8-32.
21. Schneider WM, Chevillotte MD, & Rice CM (2014) Interferon-Stimulated Genes: A Complex Web of Host Defenses. *Annual Review of Immunology* 32(1):513-545.
22. Michalska A, Blaszczyk K, Wesoly J, & Bluysen HAR (2018) A Positive Feedback Amplifier Circuit That Regulates Interferon (IFN)-Stimulated Gene Expression and Controls Type I and Type II IFN Responses. *Frontiers in Immunology* 9(1135).
23. Platanias LC (2005) Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nature Reviews Immunology* 5(5):375-386.
24. Schoggins JW, *et al.* (2011) A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* 472(7344):481-485.
25. Alsamman K & El-Masry Omar S (2018) Interferon regulatory factor 1 inactivation in human cancer. *Bioscience Reports* 38(3).
26. Sakai T, *et al.* (2014) The Roles of Interferon Regulatory Factors 1 and 2 in the Progression of Human Pancreatic Cancer. *Pancreas* 43(6):909-916.
27. Diaferia GR, *et al.* (2016) Dissection of transcriptional and *cis* - regulatory control of differentiation in human pancreatic cancer. *The EMBO Journal* 35(6):595-617.
28. Hayward A, Ghazal A, Andersson G, Andersson L, & Jern P (2013) ZBED Evolution: Repeated Utilization of DNA Transposons as Regulators of Diverse Host Functions. *PLOS ONE* 8(3):e59940.
29. Yamashita D, *et al.* (2007) hDREF Regulates Cell Proliferation and Expression of Ribosomal Protein Genes. *Molecular and Cellular Biology* 27(6):2003.
30. Mokhonov VV, Theendakara VP, Griбанова YE, Ahmedli NB, & Farber DB (2012) Sequence-Specific Binding of Recombinant Zbed4 to DNA: Insights into Zbed4 Participation in Gene Transcription and Its Association with Other Proteins. *PLOS ONE* 7(5):e35317.
31. Markljung E, *et al.* (2009) ZBED6, a Novel Transcription Factor Derived from a Domesticated DNA Transposon Regulates IGF2 Expression and Muscle Growth. *PLOS Biology* 7(12):e1000256.

32. Tang H, *et al.* (2014) Axonal guidance signaling pathway interacting with smoking in modifying the risk of pancreatic cancer: a gene- and pathway-based interaction analysis of GWAS data. *Carcinogenesis* 35(5):1039-1045.
33. Finnegan A, *et al.* (2019) Single-Cell Transcriptomics Reveals Spatial and Temporal Turnover of Keratinocyte Differentiation Regulators. *Frontiers in Genetics* 10(775).
34. Li H, *et al.* (2019) Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated Compartment within Human Melanoma. *Cell* 176(4):775-789.e718.
35. Moffitt RA, *et al.* (2015) Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nature Genetics* 47:1168.
36. Cancer Genome Atlas Research Network (2017) Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* 32(2):185-203.e113.
37. Tiriac H, *et al.* (2018) Organoid profiling identifies common responders to chemotherapy in pancreatic cancer. *Cancer Discovery*.
38. GTEx Consortium (2015) The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* 348(6235):648-660.
39. Hoadley KA, *et al.* (2018) Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* 173(2):291-304.e296.
40. Smith JC & Sheltzer JM (2018) Systematic identification of mutations and copy number alterations associated with cancer patient prognosis. *eLife* 7:e39217.
41. Ghandi M, *et al.* (2019) Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* 569(7757):503-508.
42. Biankin AV, *et al.* (2012) Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* 491(7424):399-405.
43. Waddell N, *et al.* (2015) Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* 518(7540):495-501.
44. Jones S, *et al.* (2008) Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses. *Science* 321(5897):1801-1806.
45. Machanick P & Bailey TL (2011) MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* 27(12):1696-1697.
46. Escalante CR, Yie J, Thanos D, & Aggarwal AK (1998) Structure of IRF-1 with bound DNA reveals determinants of interferon regulation. *Nature* 391(6662):103-106.
47. Tanaka N, *et al.* (1994) Cellular commitment to oncogene-induced transformation or apoptosis is dependent on the transcription factor IRF-1. *Cell* 77(6):829-839.
48. Kirchhoff S, Schaper F, & Hauser H (1993) Interferon regulatory factor 1 (IRF-1) mediates cell growth inhibition by transactivation of downstream target genes. *Nucleic Acids Research* 21(12):2881-2889.

49. Harada H, *et al.* (1993) Anti-oncogenic and oncogenic potentials of interferon regulatory factors-1 and -2. *Science* 259(5097):971-974.
50. Engle DD, *et al.* (2019) The glycan CA19-9 promotes pancreatitis and pancreatic cancer in mice. *Science* 364(6446):1156-1162.
51. Hingorani SR, *et al.* (2005) Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* 7(5):469-483.
52. Collisson EA, *et al.* (2011) Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nature Medicine* 17:500.
53. Decker K, Goldman DC, L. Grash C, & Sussel L (2006) Gata6 is an important regulator of mouse pancreas development. *Developmental Biology* 298(2):415-429.
54. Fu B, Luo M, Lakkur S, Lucito R, & Iacobuzio-Donahue CA (2008) Frequent genomic copy number gain and overexpression of GATA-6 in pancreatic carcinoma. *Cancer Biology & Therapy* 7(10):1593-1601.
55. Tomer Y, Blackard JT, & Akeno N (2007) Interferon Alpha Treatment and Thyroid Dysfunction. *Endocrinology and Metabolism Clinics of North America* 36(4):1051-1066.
56. Radko S, *et al.* (2014) Adenovirus E1A Targets the DREF Nuclear Factor To Regulate Virus Gene Expression, DNA Replication, and Growth. *Journal of Virology* 88(22):13469-13481.
57. Kroj T, Chancelud E, Michel-Romiti C, Grand X, & Morel J-B (2016) Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. *New Phytologist* 210(2):618-626.
58. Nozawa H, *et al.* (1999) Loss of transcription factor IRF-1 affects tumor susceptibility in mice carrying the Ha-ras transgene or nullizygosity for p53. *Genes & Development* 13(10):1240-1245.
59. Parker BS, Rautela J, & Hertzog PJ (2016) Antitumour actions of interferons: implications for cancer therapy. *Nature Reviews Cancer* 16:131.
60. Liu H, *et al.* (2019) Tumor-derived IFN triggers chronic pathway agonism and sensitivity to ADAR loss. *Nature Medicine* 25(1):95-102.
61. Ng KW, Marshall EA, Bell JC, & Lam WL (2018) cGAS–STING and Cancer: Dichotomous Roles in Tumor Immunity and Development. *Trends in Immunology* 39(1):44-54.
62. Mackenzie KJ, *et al.* (2017) cGAS surveillance of micronuclei links genome instability to innate immunity. *Nature* 548:461.
63. Zhang C-Z, *et al.* (2015) Chromothripsis from DNA damage in micronuclei. *Nature* 522(7555):179-184.
64. Notta F, *et al.* (2016) A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* 538(7625):378-382.

65. Ishikawa O, *et al.* (1980) Adenosquamous carcinoma of the pancreas: A clinicopathologic study and report of three cases. *Cancer* 46(5):1192-1196.
66. Morohoshi T, Held G, & Klöppel G (1983) Exocrine pancreatic tumours and their histological classification. A study based on 167 autopsy and 97 surgical cases. *Histopathology* 7(5):645-661.
67. Motojima K, Tomioka T, Kohara N, Tsunoda T, & Kanematsu T (1992) Immunohistochemical characteristics of adenosquamous carcinoma of the pancreas. *Journal of Surgical Oncology* 49(1):58-62.
68. Ingraham CR, *et al.* (2006) Abnormal skin, limb and craniofacial morphogenesis in mice deficient for interferon regulatory factor 6 (Irf6). *Nature Genetics* 38(11):1335-1340.
69. Mercado N, *et al.* (2019) IRF2 is a master regulator of human keratinocyte stem cell fate. *Nature Communications* 10(1):4676.
70. Hida S, *et al.* (2000) CD8⁺ T Cell-Mediated Skin Disease in Mice Lacking IRF-2, the Transcriptional Attenuator of Interferon- α / β Signaling. *Immunity* 13(5):643-655.
71. Oberbeck N, *et al.* (2019) The RIPK4-IRF6 signalling axis safeguards epidermal differentiation and barrier function. *Nature* 574(7777):249-253.
72. Maurer C, *et al.* (2019) Experimental microdissection enables functional harmonisation of pancreatic cancer subtypes. *Gut* 68(6):1034-1043.
73. Aung KL, *et al.* (2018) Genomics-Driven Precision Medicine for Advanced Pancreatic Cancer: Early Results from the COMPASS Trial. *Clinical Cancer Research* 24(6):1344-1354.
74. Martinelli P, *et al.* (2017) GATA6 regulates EMT and tumour dissemination, and is a marker of response to adjuvant chemotherapy in pancreatic cancer. *Gut* 66(9):1665-1676.
75. Fellmann C, *et al.* (2013) An Optimized microRNA Backbone for Effective Single-Copy RNAi. *Cell Reports* 5(6):1704-1713.
76. Kim D, Langmead B, & Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* 12:357.
77. Trapnell C, *et al.* (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* 28:511.
78. Zhou Y, *et al.* (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications* 10(1):1523.
79. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9:357.
80. Li H, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.

81. Feng J, Liu T, Qin B, Zhang Y, & Liu XS (2012) Identifying ChIP-seq enrichment using MACS. *Nature Protocols* 7:1728.
82. Bailey TL, *et al.* (2009) MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research* 37(suppl_2):W202-W208.
83. Bailey TL & Machanick P (2012) Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Research* 40(17):e128-e128.
84. Heinz S, *et al.* (2010) Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* 38(4):576-589.
85. Saldanha AJ (2004) Java Treeview—extensible visualization of microarray data. *Bioinformatics* 20(17):3246-3248.
86. Grant CE, Bailey TL, & Noble WS (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27(7):1017-1018.
87. Vaquerizas JM, Kummerfeld SK, Teichmann SA, & Luscombe NM (2009) A census of human transcription factors: function, expression and evolution. *Nature Reviews Genetics* 10(4):252-263.
88. Cerami E, *et al.* (2012) The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery* 2(5):401-404.
89. Kandoth C, *et al.* (2013) Mutational landscape and significance across 12 major cancer types. *Nature* 502(7471):333-339.

Fig. 1.

Fig. 2.

Fig. 3.

Fig. 4.

Fig. 5.

Fig. 6.

Supplementary Information for

ZBED2 is an antagonist of Interferon Regulatory Factor 1 and modulates cell identity in pancreatic cancer

Tim D.D. Somerville¹, Yali Xu¹, Xiaoli S. Wu¹, and Christopher R. Vakoc^{1,2*}

¹ Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

²Lead contact

*Correspondence: vakoc@csHL.edu

This PDF file includes:

Figures S1 to S6

Legends for Datasets S1 to S13 (Datasets provided as separate Excel files)

Supplementary Figures

Overview of supplementary figures

Fig. S1. *ZBED2* is a poor prognosis marker in numerous cancers, is highly expressed in a subset of cancer cell lines and is dispensable for cell growth *in vitro*.

Fig. S2. *ZBED2* inhibits the expression of genes in the interferon response pathway.

Fig. S3. Epigenomic analysis of *ZBED2*.

Fig. S4. Antagonistic regulation of ISG promoters by *ZBED2* and IRF1.

Fig. S5. *ZBED2* protects PDA cells from IRF1- and interferon- γ -induced growth arrest.

Fig. S6. *ZBED2* represses pancreatic progenitor lineage identity in PDA.

A

TCGA survival analysis

B

Tissue type	<i>ZBED2</i> ^{low} (n=879)	<i>ZBED2</i> ^{high} (n=277)	p value
AUTONOMIC GANGLIA	24 (2.7%)	0 (0%)	0.0025
BILIARY TRACT	6 (0.7%)	5 (1.8%)	ns
BONE	33 (3.8%)	0 (0%)	0.0002
BREAST	47 (5.3%)	6 (2.2%)	0.0309
CENTRAL NERVOUS SYSTEM	71 (8.1%)	4 (1.4%)	0.0001
CERVIX	6 (0.7%)	2 (0.7%)	ns
ENDOMETRIUM	11 (1.3%)	17 (6.1%)	0.0001
FIBROBLAST	35 (4.0%)	0 (0%)	0.0052
HEMATOPOIETIC	171 (19.5%)	17 (6.1%)	0.0001
INTESTINE	36 (4.1%)	22 (7.9%)	0.0168
KIDNEY	23 (2.6%)	11 (4.0%)	ns
LIVER	22 (2.5%)	3 (1.1%)	ns
LUNG	150 (17.1%)	48 (17.3%)	ns
ESOPHAGUS	20 (2.3%)	11 (4.0%)	ns
OTHER	3 (0.3%)	3 (1.1%)	ns
OVARY	17 (1.9%)	39 (14.1%)	0.0001
PANCREAS	18 (2.0%)	31 (11.2%)	0.0001
PLEURA	8 (0.9%)	1 (0.4%)	ns
PROSTATE	6 (0.7%)	2 (0.7%)	ns
SKIN	53 (6.0%)	2 (0.7%)	0.0001
SOFT TISSUE	53 (6.0%)	1 (0.4%)	0.0001
STOMACH	30 (3.4%)	7 (2.5%)	ns
THYROID	9 (1.0%)	3 (1.1%)	ns
UPPER AERODIGESTIVE TRACT	8 (0.9%)	25 (9.0%)	0.0001
URINARY TRACT	19 (2.2%)	17 (6.1%)	0.0022

C

Mutated gene	<i>ZBED2</i> ^{low} (n=597)	<i>ZBED2</i> ^{high} (n=221)	p value
KRAS	90 (15.1%)	64 (29.0%)	0.0001
SMAD4	28 (4.7%)	28 (12.7%)	0.0001
CDKN2A	14 (2.3%)	14 (6.3%)	0.0085
TP53	356 (59.6%)	150 (67.9%)	0.035
VHL	5 (0.8%)	7 (3.2%)	0.0211
NSD1	34 (5.7%)	23 (10.4%)	0.0292
RB1	67 (11.2%)	14 (6.3%)	0.0471

D

Altered gene	<i>ZBED2</i> ^{low} (n=597)	<i>ZBED2</i> ^{high} (n=221)	p value
KRAS	118 (19.8%)	77 (34.8%)	0.0001
SMAD4	62 (10.4%)	56 (25.3%)	0.0001
TP53	365 (61.1%)	152 (68.8%)	0.05
CDKN2A	234 (39.2%)	91 (41.2%)	ns

E

Fig. S1. *ZBED2* is a poor prognosis marker in numerous cancers, is highly expressed in a subset of cancer cell lines and is dispensable for cell growth *in vitro*. Related to Fig. 1.

(A) Survival analysis based on *ZBED2* expression levels across TCGA profiled tumors. Z-score values ≤ -2 and ≥ 2 are considered significant. See Smith and Sheltzer (2018). KIPAN, pan kidney cancer; PAAD, pancreatic adenocarcinoma; GBMLGG, Glioblastoma/Glioma; COADREAD, colorectal adenocarcinoma; LUSC, Lung squamous cell carcinoma; LUAD, Lung adenocarcinoma; STES, Esophagus-Stomach Cancers; HNSC, Head and Neck squamous cell carcinoma; BRCA, Breast invasive carcinoma; UCEC, Uterine Corpus Endometrial Carcinoma; SARC, Sarcoma; LIHC, Liver hepatocellular carcinoma; BLCA, Bladder Urothelial Carcinoma; OV, Ovarian serous cystadenocarcinoma; PRAD, Prostate adenocarcinoma; SKCM, Skin Cutaneous Melanoma. (B) Proportion of cell lines from the CCLE stratified as *ZBED2*^{low} or *ZBED2*^{high} classified based on their tissue of origin. (C and D) Association of *ZBED2* expression with genetic features in CCLE cell lines. 127 significantly mutated genes in human cancer (Kandoth et al., 2013) were analyzed. For (C) only genetic mutations were considered and only significantly mutated genes are shown. For (D) both genetic mutation and copy number alteration (i.e. loss of a tumor suppressor or gain of an oncogene) were considered as altered genes. The table displays the top four most significantly altered genes in PDA. Statistical significance for the indicated comparisons was assessed using Fisher's Exact Test, ns = not significant. (E) Summary of *ZBED2* dependency score (CERES) and expression in 559 cancer cell lines from the depmap portal (depmap.org). The number of cell lines dependent on *ZBED2* for cell growth across all cell lineages is indicated, as well as the 26 pancreatic cancer cell lines (PDA) included in this dataset. Cell lines with a CERES ≤ -0.5 were considered as dependent.

Fig. S2. ZBED2 inhibits the expression of genes in the interferon response pathway. Related to Fig. 2.

(A) Summary of GSEA evaluating Hallmark gene sets within the MSigDB v7.0 database upon ZBED2 cDNA expression in 15 PDA cell lines. NES scores for each gene set were averaged across each cell line and this value was used to rank each gene set. The top three most repressed (top panel) and activated (bottom panel) gene sets are shown. “#” indicates the number of cell lines in which the indicated gene set was determined to be repressed or activated with a false discovery rate (FDR) or <0.1 . NES, normalized enrichment score; SD, standard deviation. (B) RT-qPCR analysis of the ISGs *STAT2*, *MX1* and *CMPK2* in HPAFII cells (left panel) and PANC0403 cells (right panel) following infection with ZBED2 cDNA or the empty vector. (C and D) Western blot analysis of HSC70 and ZBED2 (C) and RT-qPCR analysis of *ZBED2*, *STAT2* and *CMPK2* (D) in PANC0403 cells following infection with ZBED2 shRNAs or a control shRNA (shNEG). (E and F) RT-qPCR analysis of *IRF1*, *MX1* and *ZBED2* in PANC0403 cells following CRISPR-based targeting of IFNAR1 with three independent sgRNAs or a control sgRNA (E) and subsequent treatment with 20ng/ μ L of IFN β for 12 hours (F). (G) *ZBED2* expression across all cancer cell lines within the CCLE database (left panel) or in the indicated tissue lineages stratified according to high or low *ZBED2* expression. Each cell line is depicted as a single dot. (H) RT-qPCR analysis of *IRF1* and *ZBED2* following treatment with 20ng/ μ L of IFN- β or IFN- γ for 12 hours in PDA cell lines. Mean+SEM is shown. n=3. ns, not significant, * $p<0.05$, ** $p<0.01$, *** $p<0.001$ by Student's t test.

A **B**

Fig. S3. Epigenomic analysis of ZBED2. Related to Fig. 3.

(A) Overlap of FLAG-ZBED2 ChIP-seq peaks in AsPC1 and SUIT2 cells. (B) Meta-profile comparing ZBED2 occupancy in SUIT2 cells around the center of peaks with 0, 1 or ≥ 2 motif counts. ZBED2 binding intensity is shown as sequencing depth normalized tag count.

A Motif analysis at ZBED2 binding sites

Motif	Motif family	E-value	adj p-value
IRF5	IRF	1.10E-53	1.90E-56
IRF4	IRF	4.60E-11	7.90E-14
ELF4	ETS	7.20E-10	1.30E-12
SIX1	SIX	1.60E-09	2.80E-12
ELF1	ETS	3.50E-09	6.10E-12
EHF	ETS	3.80E-09	6.50E-12
HSF4	HSF	1.50E-08	2.50E-11
IRF8	IRF	1.20E-07	2.20E-10
IRF7	IRF	3.30E-07	5.70E-10
ELF3	ETS	4.00E-07	6.90E-10

D Correlation in 1,156 cancer cell lines (CCLE, Broad 2019)

Correlated gene	Spearman's correlation	p-value
IRF6	0.408	1.20e-47
IRF1	0.363	2.23e-37
IRF7	0.250	7.20e-18
IRF9	0.235	5.07e-16
IRF5	0.235	5.08e-16
IRF3	0.216	1.22e-13
IRF8	0.120	4.10e-5
IRF2	0.075	0.0107
IRF4	-0.021	0.478

C Expression in 15 PDA cell lines

Fig. S4. Antagonistic regulation of ISG promoters by ZBED2 and IRF1. Related to Fig. 4.

(A) Summary of CentriMo motif enrichment analysis for JASPAR motifs at ZBED2 binding sites. The top 10 transcription factor (TF) motifs ranked by E value are shown. The nucleotide sequence 200bp up- or downstream of the peak summit of the top 1000 ZBED2 peaks in AsPC1 cells was used for this analysis. (B) Expression of *ZBED2* and IRF family TF genes in 15 human PDA cell lines. (C) *ZBED2* expression versus *IRF2*, *IRF3*, *IRF5*, *IRF6*, *IRF7* and *IRF9* expression across 15 human PDA cell lines. (D) *ZBED2* expression correlation with IRF family TF genes in 1,156 cancer cell lines from the CCLE database analyzed using CBioPortal. (E and F) Density plots showing IRF1 and FLAG-ZBED2 enrichment surrounding a 2-kb interval centered on the summit of all IRF1 peaks (E) or 140 random IRF1 peaks that do not intersect with FLAG-ZBED2 sites in AsPC1 cells (F), ranked by IRF1 peak intensity. (G) ChIP-seq profiles of IRF1 and FLAG-ZBED2 in AsPC1 cells at the promoters of *BTN3A3* and *SAMD9*. (H) Gene ontology (GO) analysis with Metascape of genes annotated by HOMER to IRF1 only sites. Terms are ranked by their significance (*p* value) and no terms reached the significant threshold ($-\log_{10} p$ value >12). (I) Pie chart showing the distribution of 140 IRF1 only peaks (left) or IRF1/ZBED2 peaks (right) in AsPC1 cells. TTS, transcription termination site; TSS, transcription start site; UTR, untranslated region. (J and K) GSEA plots evaluating protein coding genes annotated by HOMER to IRF1 only sites upon IRF1 (J) or ZBED2 (K) cDNA expression in AsPC1 cells. (L and M) GSEA plots evaluating the Interferon Response signature upon IRF1 cDNA expression in AsPC1 cells (L) or IRF1 knockout in PANC0403 cells (M). (N) Expression levels of protein coding genes annotated to IRF1 only sites following 12-hour treatment with 0.2ng/μl of IFN-β, IFN-γ or control. ****p*<0.001, ns = not significant, by one-way ANOVA.

Fig. S5. ZBED2 protects PDA cells from IRF1- and interferon- γ -induced growth arrest. Related to Fig. 5.

(A) Luciferase-based quantification of cell viability of SUI2 cells grown in Matrigel on day 7 post infection with IRF1 cDNA or the empty vector. Representative bright field images (right panel) are shown. Scale bar indicates 200 μm . (B-D) SUI2 cells infected with sgRNAs targeting IRF1, IRF9 or a control sgRNA (sgNEG) were plated in Matrigel and grown for 7 days in the presence of 20ng/ μl of IFN- γ , IFN- β or control. Representative western blots following overnight stimulation with 20ng/ μl of IFN- γ (B), luciferase-based quantification (C) and representative bright field images on day 7 of the assay (D) are shown. Scale bar indicates 500 μm . (E) FC1242 KPC cells were infected with the ZBED2 cDNA or an empty vector and grown in Matrigel with the increasing concentrations of IFN- γ . Bar chart shows luciferase-based quantification on day 7. Mean+SEM is shown. n=3. For (A) and (C), ***p < 0.001 by Student's t test. For (E), ***p < 0.001, **p < 0.01 by one-way ANOVA.

Supplementary Datasets

Overview of supplementary datasets

- Dataset. S1. Ranked list of expressed transcription factors in primary PDA tumor samples versus normal pancreas tissue.
- Dataset. S2. *ZBED2* expression values as normalized counts in normal and neoplastic pancreatic organoids.
- Dataset. S3. Results from GSEA for gene sets within the MSigDBv7.0 following *ZBED2* knock out in PANC0403 or *ZBED2* cDNA expression in HPAFII cells.
- Dataset. S4. Interferon Response signature genes.
- Dataset. S5. Results from GSEA for Hallmark gene sets following *ZBED2* cDNA expression in 15 PDA cell lines.
- Dataset. S6. 2,451 high confidence *ZBED2* peaks and their HOMER annotations.
- Dataset. S7. GSEA of *ZBED2* bound genes following *ZBED2* cDNA expression in HPAFII cells.
- Dataset. S8. Results from *ZBED2* motif scanning at human promoters.
- Dataset. S9. 1,946 high confidence IRF1 peaks and their HOMER annotations.
- Dataset. S10. 140 IRF1/*ZBED2* sites and the control set of IRF1 only sites with their HOMER annotations.
- Dataset. S11. Fold change in IRF1 tag counts in AsPC1-*ZBED2* versus AsPC1-empty cells at high confidence IRF1 sites with their HOMER annotations.
- Dataset. S12. Ranked lists of expressed TFs in Basal-like versus Classical or Squamous versus Progenitor PDA tumor samples from three independent studies.
- Dataset. S13. Sequences for RT-qPCR primers, sgRNAs and shRNAs used in this study.

Additional information is provided as a header in the first page of each Excel workbook.