ESCAPE FROM TGFβ TUMOR SUPPRESSION IN PANCREATIC CANCER

By

Yun-Han Huang

A Dissertation

Presented to the Faculty of the Louis V. Gerstner, Jr.

Graduate School of Biomedical Sciences,

Memorial Sloan-Kettering Cancer Center

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

New York, NY

July, 2018

_____ _____

Joan Massagué Date
Dissertation Mentor

DEDICATION

This thesis is dedicated to my family and friends:

Mei-Yu Chen, my mother and guide to the beauty of the natural world

Yue-wern Huang, my father whose support never fails

Yu-Hsien Chiu, my stepmother and a constant source of positive energy

Helen Branson & Nancy Hopkins, mentors with invaluable guidance

Kevin, Lenka, Oscar, Ieva & Igor, my dance family and inspiration

Dan, Karishma, Ming, and Robert, my friendly scientific sounding boards

Arthur, Dan, HH, Nick, Rui & Will, my consistent moral support

The encouragement and support of these individuals and many others over the years makes everything possible.

ABSTRACT

Transforming growth factor beta (TGFβ) signaling can be pro-tumorigenic or tumor suppressive. The TGFβ pathway is recurrently inactivated in pancreatic ductal adenocarcinomas (PDAs), suggesting that it is tumor suppressive; however, alterations in its core components affect only half of cases, most frequently via inactivation of the transcription factor SMAD4. The goal of this study was to understand the role of TGFβ in preventing the growth of PDAs and the commonalities of the tumors that emerge.

The first part of this work describes a TGFβ-induced epithelial-mesenchymal transition (EMT), generally considered a pro-tumorigenic event, that becomes lethal by converting TGFβ-induced SOX4 from an enforcer of tumorigenesis into a promoter of apoptosis. This is the result of an EMT-linked remodeling of the cellular transcription factor landscape, including the repression of the gastrointestinal lineage-master regulator KLF5. KLF5 cooperates with SOX4 in oncogenesis and prevents SOX4-induced apoptosis. SMAD4 is required for EMT but dispensable for SOX4 induction by TGFβ. TGFβ-induced SOX4 is thus geared to bolster progenitor identity, whereas simultaneous SMAD4-dependent EMT strips SOX4 of an essential partner in oncogenesis. This work demonstrates that TGFβ tumor suppression functions through an EMT-mediated disruption of a lineage-specific transcriptional network.

Given that the components identified in the first phase of the project were unable to explain the large numbers of TGFβ pathway competent PDAs, the question remained of how tumors that retain TGFβ components escape TGFβ tumor

suppression. Gene expression and chromatin profiling coupled with genetic and chemical manipulation of PDA cell lines, organoids, and mouse models demonstrated that PDAs that develop with this pathway intact avert its apoptotic effect via Inhibitor of Differentiation 1 (ID1). *ID1* is primed for expression in the normal pancreas and is amongst a set of core transcriptional regulators shared by PDAs. ID1 uncouples EMT from apoptosis in KRAS-mutant pancreatic progenitors. PDA development selects against TGFβ-mediated *ID1* repression, and different mechanisms linked to low-frequency genetic events converge on ID1 as a common node during tumor progression. These results identify ID1 as a potential therapeutic target in multiple PDA subtypes.

Altogether, the work presented here demonstrates the roles of several transcriptional modules that are perturbed by TGFβ tumor suppression during the course of PDA development. It highlights the importance of the common core transcriptional regulators within PDAs, and shows that different genetic paths can lead to different phenotypic outputs while promoting a common core of highly expressed transcriptional regulators important to cancer progression.

BIOGRAPHICAL SKETCH

Yun-Han (Hannah) Huang was born to science and art teachers Yue-wern Huang and Mei-Yu Chen on 18 March 1989 in Taoyuan, Taiwan. The family immigrated to the United States in 1992, living in Madison, Wisconsin, for six years and East Lansing, Michigan, for two years before moving to Rolla, Missouri. After her mother died of aplastic anemia, a rare and poorly understood disease, during her first year at Rolla Senior High School, Hannah developed a strong interest in the intersection of science and medicine.

At the Massachusetts Institute of Technology (MIT), Hannah studied Biology with a minor in Applied International Studies. While at MIT, she performed research under the mentorship of Dr. Jeffrey Karp (Harvard-MIT Health Sciences & Technology) in stem cell homing, Dr. Amy Fleischman (Massachusetts General Hospital) in pediatric metabolism, Dr. Janet Hall (Marie Curie Institute) in DNA damage signaling, and Dr. Paul Chang (Koch Institute for Integrative Cancer Research) in poly(ADP-ribose) biology. She also became a certified Emergency Medical Technician with the MIT ambulance and a tutor in physics and chemistry.

After graduating from MIT in 2011, Hannah enrolled in the Weill Cornell/Rockefeller/Sloan Kettering Tri-Institutional MD-PhD Program. She completed laboratory rotations with Dr. Sean Brady in functional metagenomics, Dr. Hermann Steller in protein degradation, and Dr. Michael Overholtzer in cell engulfment. She joined the Louis V. Gerstner Jr. Graduate School of Biomedical Sciences at Memorial Sloan Kettering Cancer Center and the laboratory of Dr. Joan Massagué in 2013 to study TGFβ signaling.

ACKNOWLEDGEMENTS

First and foremost, I thank my mentor, Dr. Joan Massagué, for the training and resources to complete my dissertation work. I thank my fellow lab members for their assistance and advice--Dr. Charles David for his collaboration and guidance during my first years in the laboratory, Dr. Ekrem Emrah Er for critical scientific discussions, and Drs. Jie Su, Yilong Zou, and Mo Chen for their collaboration.

I am grateful to my thesis committee members Drs. Lewis Cantley, Christine Iacobuzio-Donahue, Steven Leach, and Andrea Ventura for insightful commentary; I similarly thank my thesis proposal and dissertation defense committees, Drs. Robert Benezra, Sarat Chandarlapaty, Ross Levine, and David Tuveson. The expertise and knowledge of these individuals guided the concepts and techniques throughout this project. In addition, the human tissue samples provided by Drs. Leach and Iacobuzio-Donahue were invaluable, and I thank my collaborators who collected the samples, including Dr. Nicolas Lecomte, Dr. Peter Allen, and Rajya Kappagantula.

To my collaborators in the MSKCC Center for Molecular Oncology, Molecular Diagnostics Service, Integrated Genomics Core, Flow Cytometry Core, and Center for Epigenetics Research, thank you for your technical assistance and scientific advice. In particular, Dr. Ross Levine, Dr. Richard Koche, and Matthew Witkin contributed heavily to the development of the epigenetic regulation described in this project.

Finally, I thank the Tri-Institutional MD-PhD Program and the Gerstner Sloan Kettering Graduate School for their support. The leadership of Dr. Olaf Andersen, Dr. Ruth Gotian, and Dr. Kenneth Marians truly enables students to develop the skills and knowledge necessary to become independent physician-scientists.

Table of Contents

List of Tables

TABLE OF FIGURES

LIST OF ABBREVIATIONS

ADM	Acinar-to-ductal metaplasia
AKT	AKR-transforming
ATAC-seq	Assay for transposase-accessible chromatin with sequencing
bHLH	Basic helix-loop-helix
BMP	Bone morphogenetic protein
CC3	Cleaved caspase 3
ChIP-seq	Chromatin Immunoprecipitation with sequencing
CRISPR	Clustered regularly interspaced short palindromic repeats
Dox/doxy	Doxycycline
EMT	Epithelial-to-mesenchymal transition
FACS	Fluorescence-activated cell sorting
GFP	Green fluorescent protein
ID1	Inhibitor of Differentiation 1
KC	*LSL-Kras^{G12D};Pdx1-Cre*
KIC	*LSL-Kras^{G12D};Pdx1-Cre;Cdkn2a$^{fl/fl}$*
KLF5	Kruppel-like family 5
KSC	*LSL-Kras^{G12D};Pdx1-Cre;Smad4$^{fl/fl}$*
KSIC	*LSL-Kras^{G12D};Pdx1-Cre;Cdkn2a$^{fl/fl}$;Smad4$^{fl/fl}$*
MK2206	AKT inhibitor
PanIN	Pancreatic intraepithelial neoplasia
PCA	Principal component analysis
PDA	Pancreatic ductal adenocarcinoma
PNET	Pancreatic neuroendocrine tumor
RNA-seq	Messenger ribonucleic acid sequencing
SB505124 (SB)	TGFBRI kinase inhibitor
SMAD	SMA (small) + Mothers Against Decapentaplegic
SOX4	SRY-related HMG box 4
TF	Transcription factor
TGFβ (Tβ)	Transforming Growth Factor Beta
UGT	UDP-glucuronosyltransferase

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDA) is a cancer of the exocrine compartment of the pancreas and currently the third leading cause of cancer deaths in the United States due to late detection and lack of effective treatment (SEER, 2000). A better understanding of the development of PDA has the potential to yield new, more effective therapeutic strategies. Of note, the endogenous mechanisms present in the pancreas to prevent PDA formation are remarkable. At a first approximation, of the 170 billion acinar cells in human pancreas, from which PDAs are primarily predicted to arise (Kopp et al., 2012), about 4.2 billion divide yearly (Tomasetti and Vogelstein, 2015). Given that each cell division is associated with the accumulation of three genetic mutations and the human genome size is three billion base pairs, the acinar compartment of the pancreas of an average person potentially mutates the entire genome approximately 330 times during an average lifetime of 79 years even in the absence of exogenous mutagens. Yet, the lifetime risk of developing pancreatic ductal adenocarcinoma cancer is only 1.6% (SEER, 2000). Many mechanisms exist to prevent the accumulation of cancer cells bearing oncogenic mutations, including signals from the microenvironment, nutrient availability, and immune surveillance (Hanahan and Weinberg, 2011). The goal of the following project is to understand the role of one tumor suppressive signal, transforming growth factor beta (TGFβ), in preventing the growth of PDAs and the commonalities of the tumors that emerge.

Cancer genetics

Multiple genetic alterations are required for cancer

Cancer arises from the accumulation of multiple mutations, each of which provides a selective advantage to the cell harboring it, compared to its neighbors. Statistician Carl Nordling first proposed the multiple-hit hypothesis in the British Journal of Cancer in 1953 based on the observation that cancer incidence increases according to the sixth exponential with age (Nordling, 1953). He posited that cell proliferation over time correlates with the emergence of mutations and that tumor incidence over time could be modeled by an n-1 order equation, where n is the number of mutations needed to form a tumor. While certain pediatric tumors appear to be an exception, as these can be driven by single events (Greaves, 2015), the requirement of multiple genetic alterations in adult human cancers is now well-supported by several lines of evidence, including human tumor genome sequencing data as well as mouse models where combinations of oncogenic mutations accelerate tumor growth. Interestingly, the earliest transgenic mouse models of cancer were created by introducing single virus-derived transcriptional modulators like transgenes for the SV40 T-antigen and Myc into a range of developmentally mature and immature cells (Adams et al., 1985; Brinster et al., 1984); however, subsequent tissue-specific models have demonstrated that multiple genetic alterations cooperate to accelerate tumor development. For example, in the pancreas, *Kras* activation alone can cooperate with occasional spontaneous loss of *Cdkn2a* and result in progression to PDA, but this process is greatly accelerated with

combination of *Kras* activation with *Cdkn2a*, *Trp53*, or *Smad4* loss of function (Aguirre et al., 2003; Bardeesy et al., 2006b; Hingorani et al., 2005).

The number of genetic alterations required for transformation of a normal cell to an invasive cancer cell, and the specific nature of the alterations has been a subject of debate. Metastatic carcinoma can be produced in genetically engineered mouse models with only one or two oncogenic mutations (Hingorani et al., 2005; Muller et al., 1988), and even combinations of mutations not commonly observed in human tumors appear to have this potential (Hill et al., 2010). Furthermore, contrary to the early hypothesis of Carl Nordling that six to seven mutations are required for the formation of invasive carcinoma, Bert Vogelstein's group recently proposed that, based on mutation rates in whole genome sequencing data and the incidence of lung and colorectal cancers, only three mutations are required for the development of cancer (Tomasetti et al., 2015). While the exact number of mutations necessary for tumor formation is therefore debatable, the fact that tumors engineered with multiple genetic alterations can often be reversed by simply removing one of the alterations (Dow et al., 2015; Ying et al., 2012) and human tumors with multiple mutations can be treated with some efficacy with single targeted therapies (Chapman et al., 2011; Druker et al., 2001; Hodis et al., 2012) suggests that a full repertoire of functionally significant mutations may be required for tumor progression.

Cancer genome landscapes

Over the past ten years, thousands of cancer genomes have been sequenced (Ding et al., 2018). These sequencing studies provide a wealth of information--

amongst others, they have further characterized the mutation spectrum of previously characterized oncogenes and tumor suppressors and identified interactions with the properties of the tissue of origin (Hoadley et al., 2018). They have also highlighted genetic complexity of cancers (Ding et al., 2018), revealed that some genetic alterations like KRAS activation and TP53 inactivation happen across cancer types while others like SMAD4 and SHH mutations are private to particular tumor types (Sanchez-Vega et al., 2018), and furthered the dilemma of distinguishing driver genetic alterations and passengers (Bailey et al., 2018). The study of pancreatic cancer has followed a similar route.

Even before the earliest whole genome and whole exome sequencing studies of PDA, frequent activation of the oncogene *KRAS*^{G12D} along with inactivation of the tumor suppressor genes *CDKN2A*, *TP53*, and *SMAD4* were reported in pancreatic cancer (Almoguera et al., 1988; DiGiuseppe et al., 1994; Hahn et al., 1996; Wilentz et al., 1998). Tumors from individuals with germline genetic alterations that predispose them to PDA such as *BRCA2*, *ATM*, *PRSS1*, or *PALB2*, also often bear mutations in these genes (TCGA, 2017; Waddell et al., 2015), supporting that these genetic changes are key to the transformation of a normal pancreatic epithelial cell into an invasive carcinoma cell. Subsequent genome-scale sequencing studies have confirmed that *KRAS, CDKN2A, TP53,* and *SMAD4* are the most frequently altered genes in PDA, followed by a long tail of additional, less frequent alterations (Figure 1-1). Amongst this tail are genes related to the four major alterations including MAPK activation (related to KRAS) and *TGFBR2* inactivation (related to SMAD4), previously-described oncogenes and tumor suppressors in PDA like *ARID1A* and

RNF43, oncogenes and tumor suppressors with no previously described role in PDA like *PREX2* and *KDM6A*, and a variety of not yet functionally characterized alterations (TCGA, 2017; Waddell et al., 2015).

Efforts to subcategorize PDAs into different clinical subclasses based on genetic alterations have yielded varying results. One study, for example, describes SMAD4-inactivated PDAs as more widely metastatic than SMAD4-expressing PDAs (Iacobuzio-Donahue et al., 2009); yet another reports that SMAD4 loss is associated with more resectable tumors and better survival after resection (Biankin et al., 2002). More recent studies have subtyped PDAs based on the gene expression profiles of tumors (Bailey et al., 2016; Collisson et al., 2011; Moffitt et al., 2015). Across these studies, two subtypes recurrently emerge: a basal/squamous subtype and a classical/pancreatic progenitor subtype. Subsequent analysis of the additional subtypes that have been identified—such as the exocrine-like, quasimesenchymal, and immunogenic subtypes—have suggested that these appear to reflect the presence of stromal cells such as stellate, immune, and normal acinar cells (TCGA, 2017). PDAs are notoriously desmoplastic with low cancer cell content and high stromal infiltration, and it is possible that the differences in infiltration are clinically significant (Heid et al., 2017); however, in general, these subtypes have yet to become clinically useful. In part, this is because PDA has a universally poor outcome if detected past the point at which surgical removal is possible, with an average five-year survival of 1-14% depending on stage at diagnosis as well as the lack of therapeutic options corresponding to these subtypes.

Inter-tumor genetic heterogeneity

At the core of the debate around the number of alterations necessary for tumorigenesis is the fact that tumors often contain hundreds of mutations coupled with extensive copy number variations, many of which may or may not contribute to the tumorigenic potential. Genetic alterations occur as a result of exposure to exogenous mutagens like ultraviolet irradiation or cigarette smoke, as well as normal cellular processes like DNA replication and transcription. Cancer cells therefore carry both driver mutations, which provide a selective advantage to a cell, and passenger mutations, which may provide either negative or neutral selective advantage (Bozic et al., 2010).

Collectively, the mutational repertoire of a tissue and the selective pressures of its microenvironment determine the genetic alterations in a tumor resected from a cancer patient. Because tumors arising from the same tissue are often exposed to the same mutagenic processes and selective pressures, driver mutations are often recurrent across tumors and have therefore been traditionally identifiable as genetic alterations that are recurrently found in many tumors and highly represented within a tumor. With the advent of high throughput sequencing of tumor genomes, many computational algorithms have been developed to attempt to identify driver mutations. The most commonly employed algorithms leverage large datasets to determine whether certain alterations occur more frequently than predicted by chance or whether there is predicted functional impact on the gene of interest (Tokheim et al., 2016). By relying heavily on statistical analysis with limited biological insight, however, these methods are restricted in their ability to detect

low frequency genetic alterations that may nonetheless be biologically and clinically significant.

Cancer rewires developmental signaling pathways

Signaling pathways in cancer

Functionally, genetic alterations allow cells to escape the constraints of normal tissue homeostasis by altering the way in which cells interpret signals from the microenvironment. Within normal adult epithelial tissues, the majority of cells are in a quiescent, non-dividing state. Adult stem cells, which have the capability to both self-renew and produce differentiated progeny, respond to cues from the microenvironment to re-enter the cell cycle and proliferate (Holley, 1975). Given their seemingly unlimited self-renewal capacity, one model of cancer suggests cells within a tumor have either originated from an aberrant tissue stem cell or reacquired the characteristics of stem cells such that a small population of such cells can indefinitely repopulate a tumor (Passegue et al., 2003).

In support of this hypothesis, genetic alterations found in cancer disrupt the signaling pathways that regulate cell division, differentiation, and survival (Vogelstein et al., 2013). Interestingly, genetic alterations in pathways that emerge later in metazoan evolution, such as TGFβ and WNT, tend to occur in tissue-specific patterns. In comparison, the pathways conserved even amongst unicellular organisms, such as MAPK and PI3K signaling, tend to occur across cancer types (Kandoth et al., 2013; Sanchez-Vega et al., 2018) (Figure 1-2).

The complexity of cellular and tissue homeostasis, however, often prevents a single perturbation from leading to cancer. While cancer genes are typically

classified as oncogenes or tumor-suppressor genes, a growing number play a dual role and defy this classification (Stepanenko et al., 2013). Feedback mechanisms often lead to negative selection against oncogenic mutations--KRAS activation, for example, can lead to cellular senescence and clearance by immunity (Collado et al., 2005). Additional genetic alterations can overcome the new selection barriers presented by the transforming cell and its evolving microenvironment, resulting in multiplicative rather than simply additive effects of driver mutations. The developing tumor is therefore a dynamic landscape in which natural selection occurs in the setting of a changing microenvironment.

Transcriptional dysregulation of cancer cells

Signaling pathways coalesce in the nucleus to regulate the transcriptional output of a cell, and transcriptional dysregulation is a common feature of emerging tumors. Profiling of chromatin status has demonstrated that cancer cells often harbor aberrantly activated enhancers that support the tumorigenic properties of the cell (Sengupta and George, 2017; Sur and Taipale, 2016). Whereas promoters, associated with H3K4me3, are located near the transcription start site, enhancers associated with H3K27ac are flexible as to orientation and location (Bulger and Groudine, 2011). These enhancers often contain clusters of transcription factor binding sites, allowing the integration of multiple signaling pathways, and studies have shown that these enhancers can be activated by genetic alterations or by signals from the microenvironment and in turn regulate gene expression that promotes tumorigenesis (Hnisz et al., 2013). The altered transcriptional state of a cancer cell therefore reflects adaptation to genetic alterations in cancer cells and to

inputs from the tumor microenvironment (Christophorou et al., 2006; Garraway and Sellers, 2006; Lee and Young, 2013; Ventura et al., 2007).

Similar to the concept of oncogene-addiction, in which cancer cells become dependent on an oncogenic alteration for survival, lineage-dependency has emerged as one in which tumor cells are dependent for survival on transcriptional regulatory networks that may pre-exist during the development of the lineage or emerge during tumor formation (Garraway and Sellers, 2006). In melanoma, MITF, which is necessary for melanocyte differentiation and survival, is an example of one such gene, and in prostate cancer, the androgen receptor is another example. These genes are critical in the normal development of the cell types from which these tumors arise, are expressed highly in tumors of that lineage, and are required for the progression of those tumors (Garraway and Sellers, 2006). Their high expression may therefore provide both clues regarding the signaling pathways active in the tumor as well as vulnerabilities for directed therapeutic targeting.

The TGFβ pathway

Amongst the developmental signaling pathways co-opted in cancer is that of transforming growth factor beta (TGFβ). TGFβ emerged as a signaling pathway in metazoans, where it regulates various aspects of cell fate and state in a cell-type specific manner. The earliest metazoan known to harbor TGFβ ligands, receptors, and SMAD transcription factors is the sponge genus *Amphimedon*; however, the *Trichoplax* TGFβ pathway more closely resembles that found in most metazoans outside of Porifera (Richards and Degnan, 2009). In Eumetazoa, including mammals, the TGFβ superfamily has expanded to include TGFβs, Activins, Nodal, GDFs and

BMPs. These ligands are expressed in every tissue type and function through paracrine and autocrine signaling to play critical roles in development, immunomodulation, and epithelial tissue homeostasis.

Transcriptional regulation via the SMAD transcription factors

Generally, TGFβ ligands are secreted as pre-pro-peptides that are then cleaved to yield the mature factor trapped by the pro-domain and other proteins as a latent complex. This inactive form resides in the extracellular matrix until it is activated by interactions with integrins or proteases, providing precise spatio-temporal control of the ligand (Dong et al., 2017; Qin et al., 2018). Dimers of the peptide then bind to constitutively phosphorylated type II serine-threonine receptors, which recruit and phosphorylate type I receptors (Wrana et al., 1994). A heterotetrameric complex of two type I and two type II receptors then phosphorylates receptor SMADs, which classically associate with SMAD4 and translocate to the nucleus (Lagna et al., 1996; Massague, 1998). SMAD complexes, commonly trimers cooperate with co-activators and co-repressors to modulate the expression of cell-type specific target genes (Figure 1-3). Each step of the signal transduction process--ligand activation, receptor availability, SMAD recycling--is intricately regulated (David and Massague, 2018a).

The domain structures of the SMADs are highly conserved: all SMADs share an MH1 domain with a β hairpin loop that is primarily responsible for binding to DNA and an MH2 domain that is critical for many protein-protein interactions, including interfacing with the receptors and oligomerization between SMADs (Macias et al., 2015). The major signaling output of TGFβs is via the type I receptors

TGFBR1/ACVR1B/ACVR1C to SMAD2/3, whereas BMPs operate mainly via ALK1/ACVR1A/BMPR1A/BMPR1B and SMAD1/5/8 (David and Massague, 2018a). While it was originally thought that BMP SMADs favor GC-rich regions while TGFβ SMADs favor CAGAC motifs, it was recently demonstrated that both favor clusters of CAGAC and 5GC-rich regions of DNA (Martin-Malpartida et al., 2017; Shi et al., 1998). Therefore, signaling specificity is likely due to the cofactors that cooperate with the various receptor SMADs.

Like the receptor SMADs, SMAD4 contains MH1 and MH2 domains, but it differs from the receptor SMADs in several of its structural loops, in its linker region, and in the lack of the C-terminal Ser-X-Ser motif that is necessary for receptor phosphorylation (Macias et al., 2015). Despite the seemingly minor differences between SMAD4 and the receptor SMADs, it has a unique role in TGFβ signal transduction. Not only does it promote SMAD complex binding to DNA and activate transcription (Liu et al., 1997), it is also necessary in both TGFβ and BMP signaling and uniquely the most commonly disabled member of the TGFβ pathway in cancer (Waddell et al., 2015). While one possibility is that this latter may be due to its roles in both the TGFβ and BMP pathways, the dispensability of SMAD4 in pancreatic development where SMAD2/3 are required points to a more specific role of SMAD4 in comparison to SMAD2/3 (Bardeesy et al., 2006b; Goto et al., 2007).

Context-specificity of TGFβ signaling

Even the earliest reports about TGFβ hinted at the context-specificity of its effects. Initial efforts to purify TGFβ resulted in the purification of TGFα, a member of the EGF family, along with TGFβ, and it was found that TGFβ could only promote

cell growth in colony formation assays in the presence of TGFα (Anzano et al., 1982). Later, it was observed that TGFβ could also inhibit cell growth (Tucker et al., 1984), and that the combination of growth factors present with TGFβ could alter the effect of TGFβ (Roberts et al., 1985). Together with the placement of SMAD4, independently identified in humans as deleted in pancreatic cancer 4 (DPC4) (Hahn et al., 1996), in the TGFβ pathway (Sekelsky et al., 1995), the opposing observations regarding the effect of TGFβ on cell growth solidified the context-dependency of TGFβ signaling.

Beyond a few key target genes, most notably the negative feedback regulator, *SMAD7*, the genes regulated by TGFβ vary greatly amongst cell types (David and Massague, 2018a). A major factor in determining the genes that are regulated in a particular cell is the chromatin landscape. SMAD transcription factors bind primarily to 5GC SBE motifs in regions of open chromatin (Martin-Malpartida et al., 2017) and cooperate with lineage- and signal-dependent transcription factors, which vary between cell types (David and Massague, 2018a). In embryonic stem cells, myotubes, and pro-B cells, for example, SMAD complexes co-localize with the cell type-specific master transcription factors Oct4, Myod1, and PU.1 genome-wide, respectively (Mullen et al., 2011); similarly, SMADs co-localize with other signal-regulated transcription factors like the PI3K/AKT-modulated forkhead family of transcription factors (Seoane et al., 2004). In this way, the gene expression response to TGFβ integrates both the cell's developmental history and its current signaling status.

TGFβ regulation of ID1

Amongst the best characterized of the context-dependent genes regulated by TGFβ are the Inhibitor of DNA-binding/Inhibitor of Differentiation (ID) genes. Mammals have four genes that encode ID proteins, ID1-4, which are members of the basic helix-loop-helix (bHLH) family of transcription factors that lack the basic region necessary for DNA binding (Jen et al., 1996). bHLH proteins primarily operate via dimerization of a ubiquitous E-protein member of the bHLH family with a tissue-specific bHLH protein, and ID proteins act as dominant-negative regulators by binding to E-proteins and blocking the DNA-binding capability of the bHLH complexes (Benezra et al., 1990). In doing so, the ID proteins promote cellular differentiation and cell cycle arrest (Benezra et al., 1990; Ruzinova and Benezra, 2003).

The effects of ID proteins on cellular differentiation have been observed in both normal tissues like embryonic stem cells, hematopoietic lineages, epithelial tissues, and neurons as well as cancer, particularly in gliomas (Anido et al., 2010; Barrett et al., 2012; Lasorella et al., 2014; Ruzinova and Benezra, 2003). Whereas ID1-3 are widely expressed across different normal tissue types, ID4 expression is generally more restricted to the nervous system (Jen et al., 1996). Concordantly, knockout of *Id4* results in perinatal lethality in mice with defects in neural differentiation. Knockout of *Id1* on the other hand, results in no major abnormalities, suggesting functional redundancy from the other ID proteins. Combined knockout of *Id1* with *Id2* or *Id3*, however, results in embryonic lethality with defects in the hematopoietic and neuronal lineages (Lasorella et al., 2014).

Multiple inputs regulate the expression of the *ID* genes. Receptor tyrosine kinases via MEK-ERK-EGR, Notch signaling, MYC, WNT activation, ETS proteins, BMPs, and TGFβ are amongst the signals that can induce *ID* expression. Forkhead box proteins, p53, various microRNAs, and TGFβ, on the other hand, can repress its expression. In general, tumor-promoting signals appear to correlate with higher *ID* expression whereas tumor-suppressive signals correlate with lower *ID* expression (Lasorella et al., 2014; Perk et al., 2005), and elevated levels of *ID* expression have been reported in various cancers of endodermal, ectodermal, and mesodermal origin (Perk et al., 2005).

Of note, both BMPs and TGFβ regulate the expression of *ID1-3* genes. The mouse and human *ID1* promoters are highly conserved and are the best described, but similar regulatory elements exist in the promoters of *ID2/3*. *ID1-3* are therefore co-regulated in many contexts (Korchynskyi and ten Dijke, 2002; Ruzinova and Benezra, 2003). Whereas BMPs induce the expression of *ID1-3*, TGFβ can either induce or repress *ID1-3* expression (Kang et al., 2003; Korchynskyi and ten Dijke, 2002; Padua et al., 2008). Both BMPs and TGFβ regulate *ID1* via a series of GC-rich and CAGAC elements in the 2 kilobases upstream of *ID1* (Kang et al., 2003; Korchynskyi and ten Dijke, 2002). The repression of *ID1* by TGFβ additionally requires an ATF3 binding element in the *ID1* promoter, and induction of ATF3 by TGFβ but not BMP results in repression of *ID1* by TGFβ (Kang et al., 2003). While cells derived from normal epithelial tissues respond to TGFβ by repression of *ID1* whereas cells derived from malignant tissue induce *ID1* in response to TGFβ (Padua et al., 2008), it is unknown what regulates the difference in response to TGFβ.

Developmental roles of TGFβ in pancreas specification

Signaling by the TGFβ superfamily plays important roles throughout the specification and development of the gastrointestinal tract and pancreas. Nodal signaling is a key part of gastrulation, and *Smad2* knockout mice are embryonic lethal at this stage (Brennan et al., 2001; Nomura and Li, 1998). Interestingly, *Smad3* knockout mice are viable but harbor skeletal abnormalities and develop metastatic colorectal cancer (Zhu et al., 1998). Patterning of the mesendoderm is also influenced by Nodal: high levels of Nodal signaling specify endoderm, whereas lower levels result in mesoderm (Zorn and Wells, 2007). Furthermore, Nodal and BMPs play key roles in left-right patterning of the embryo and organ specification (Wu and Hill, 2009). In particular, Activin signaling via SMAD2/3 is necessary for pancreatic bud formation in the foregut, and deletion of *Acvr2b*, *Smad2*, or *Smad3* result in hypoplastic pancreata in mice (Kim et al., 2000). Remarkably, however, mice with *Smad4* knockout develop normal pancreata (Bardeesy et al., 2006b), highlighting a distinct role for SMAD4 in TGFβ signaling in the pancreas.

TGFβ in tumor suppression and progression

TGFβ and its signaling pathway are a paradigm of duality in cancer (Massague, 2008)(Figure 1-4). Its initial isolation as part of sarcoma growth factor (SGF) and ability to promote colony formation in soft agar suggested a tumor-promoting role for this cytokine (Anzano et al., 1982). Soon, however, it was demonstrated that TGFβ was present in normal tissues as well and could also act as a growth inhibitor (Tucker et al., 1984). Further studies in mice demonstrated that keratinocytes expressing elevated amounts of TGFβ1 were less susceptible to

benign skin tumors but the tumors that developed were spindle-like and associated with a more aggressive phenotype (Cui et al., 1996). Similarly, mouse mammary epithelial cells with activated receptor signaling were less likely to form tumors but displayed enhanced metastatic potential (Siegel et al., 2003). And in humans, administration of TGFβ inhibitors and blocking antibodies has been associated with decreased growth of aggressive tumors but also the appearance of benign, reversible keratoacanthomas and squamous cell carcinomas (Connolly et al., 2012; Morris et al., 2014).

At a cellular level, depending on the context of tumor progression, TGFβ can result in cell cycle arrest, apoptosis, epithelial-to-mesenchymal transition (EMT), and microenvironment reprogramming. TGFβ-mediated cell cycle arrest occurs in G1 via repression of c-Myc and induction of cyclin-dependent kinase (CDK) inhibitors including p15Ink4b (*Cdkn2b*) and p21Cip1 (*Cdkn1a*) (Massague, 2008). While TGFβ from the inflammatory tumor microenvironment has been observed to cause apoptosis in several settings, the mechanism by which this occurs is less well characterized, though it is likely to involve pro-apoptotic BCL2 family proteins (Guasch et al., 2007; Siegel et al., 2003). Non-cell-autonomous modes of tumor suppression also exist—TGFβ can, for example, also promote the secretion of growth factors by cells within the tumor microenvironment (Bhowmick et al., 2004). On the other end of the spectrum, TGFβ can also induce an epithelial-mesenchymal transition (EMT) along with invasive features associated with cancer cell metastasis (Heldin et al., 2012; Kang et al., 2005; Padua et al., 2008), promote cancer stem cell heterogeneity and drug resistance (Oshimori et al., 2015), and favor

an immunosuppressive, tumor-promoting microenvironment (Calon et al., 2012; Mariathasan et al., 2018; Tauriello et al., 2018). The mechanistic basis for this duality is an unsolved question.

In support of the tumor suppressive effect of TGFβ, somatic and germline *SMAD4* mutations are found in human cancers (Hahn et al., 1996; TCGA, 2017). The genetic evidence for this is particularly strong in endoderm-derived tissues like the gastrointestinal system (Goggins et al., 1998; Hahn et al., 1996), and germline mutations result in familial juvenile polyposis syndrome, which predisposes individuals to adenocarcinomas of the colon, stomach, and pancreas (Howe et al., 1998). Both somatic and germline mutations frequently perturb SMAD activation and oligomerization (Hata et al., 1997; Shi et al., 1997). It is unknown whether somatic alterations in SMAD2, SMAD3, TGFBR1, and TGFBR2 are much less common due to functional differences in silencing these genes or the mutability of these genes, although mutations in these genes do occur and tend to be mutually exclusive to SMAD4 alterations, suggesting that they achieve a similar selective advantage.

The genetics of human cancers indicate that TGFβ is tumor suppressive, particularly in the gastrointestinal tract; at the same time, given that TGFβ can promote cancer progression through both cell autonomous and non-cell autonomous means, it is clear that malignant cells can be rewired to co-opt TGFβ signaling in a manner that is no longer tumor suppressive (Massague, 2008). However, it is unclear *how* TGFβ is tumor suppressive, particularly at the premalignant stage, and *what* determines whether it is tumor suppressive (Figure 1-4). Using PDA as a model, the following work describes how TGFβ regulates central

transcriptional networks within the premalignant epithelial cell of the pancreas and how these networks are rewired in pancreatic cancer.

MATERIALS AND METHODS

Experimental model and subject details

Overview of experimental models

The presence of inactivating genetic alterations in the TGFβ pathway suggests that premalignant pancreatic epithelial cells are sensitive to the tumor suppressive effects of TGFβ and that PDA cells have lost this sensitivity. To simulate the TGFβ-sensitive state, mouse PDA cells were derived from *Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$* tumors and transduced with SMAD4, or organoids were derived from *Kras^{G12D}; Cdkn2a$^{-/-}$* mouse pancreata prior to tumor development (David et al., 2016). To model the TGFβ tumor suppression-resistant state of PDA cells, the SMAD4-restored cells were cultured in TGFβ for three weeks to generate cells resistant to the pro-apoptotic effects of TGFβ, or cells were isolated from human or mouse PDAs (Figure 2-1). Together, these models provide a platform to query the genetic, epigenetic, transcriptional, and phenotypic status of the sensitive and resistant pancreatic cells and the response to TGFβ.

Human subjects research

The study was conducted under Memorial Sloan-Kettering Cancer Center Institutional Review Board approval (MSKCC IRB 15-149 or 06-107) and all patients provided informed consent prior to tissue acquisition. Normal pancreas and PDA samples were collected from patients undergoing routine surgical resection and as part of MSKCC's rapid autopsy program for paraffin embedding or organoid generation. Normal pancreatic samples were from patients with tumors at distant sites (lung and brain) with no observed pathology in the pancreas.

Animal models

Animal experiments were performed as approved by the MSKCC Institutional Animal Care and Use Committee. Acute pancreatitis was induced by caerulein treatment, performed as previously described: mice received IP injections of 50 mg/kg caerulein eight times per day for two days (Morris et al., 2010). For orthotopic tumor models, 500 dissociated cells were implanted in 25 μL of Matrigel (Corning Matrigel GFR Membrane Matrix, #356231), injected through 31G insulin syringes into the pancreata of 4-week old female athymic nude mice (ENVIGO, Hsd: Athymic Nude-Foxn1nu). Mice were started on 2500 mg/kg doxycycline diet (ENVIGO, TD.07383 2014-2500-B, irradiated) after surgery for Tet-On shRNA experiments. Tumor growth was tracked weekly using bioluminescence (Goldbio Firefly D-Luciferin, potassium salt). Tumors were collected from genetic mouse models of PDA (*LSL-Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$* & *LSL-Kras^{G12D}; Cdkn2a$^{-/-}$*) for cell lines, organoid lines, and immunohistochemistry.

Cell based models

All cell lines and organoids were maintained at 37ºC and 5% CO2. Pancreatic cell lines were maintained in high glucose DMEM supplemented 10% FBS and 2 mM L-Glutamine. The 806 *LSL-Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$*mouse PDA cell line was provided by Nabeel Bardeesy (Bardeesy et al., 2006b). Pancreatic organoids were generated as previously described (Boj et al., 2015). Briefly, tumor samples were dissociated mechanically and with collagenase (Sigma C9263) and TrypLE (Fisher 12563029) and embedded in Matrigel (Corning Matrigel GFR Membrane Matrix, #356231).

Pancreatic organoids were maintained embedded in Matrigel (Corning Matrigel GFR Membrane Matrix, #356231) with Advanced DMEM/F12 (Gibco, 12634-028) supplemented with B-27 (Life Technologies, 12587-010), HEPES (10 mM), 50% Wnt/R-spondin/Noggin-conditioned medium (ATCC, CRL-3276), Glutamax (Invitrogen, 2 mM), N-acetyl-cysteine (Sigma, 1 mM), Nicotinamide (Sigma, 10 mM), EGF (Peprotech, 50 ng/mL), Gastrin (Sigma, 10 nM), FGF-10 (Peprotech, 100 ng/mL), A83-01 (Tocris, 0.5 μM) as previously described (Boj et al, Cell, 2015). Pancreatic oncospheres were grown in Ultra Low Attachment Culture plates (Corning) in DMEM supplemented with Glutamax (2 mM) and heparin (5 μg/mL).

Method Details

CRISPR/Cas9-mediated Genetic Knockins & Knockouts

CRISPR-mediated knockouts were obtained by cloning sgRNAs into the lentiGuide-Puro (Addgene plasmid # 52963) or pSpCas9(BB)-2A-Puro (PX459) (Addgene plasmid # 48139) vectors as previously described (Ran et al., 2013).

For CRISPR-mediated knock-in of the GFP tag to the 3' end of the endogenous Id1 locus, Id1-cterm sgRNA was cloned into the pSpCas9(BB)-2A-Puro vector, and HA-L, GFP, HA-R were cloned into the pUC19 vector (Norrander et al., 1983) using the KpnI, BamHI, EcoRI, and XbaI restriction sites. The donor plasmid was cleaved at KpnI and XbaI, and the fragment containing the tag and flanking homology arms was isolated by gel purification.

Short Hairpin-mediated Genetic Knockdowns

shRNA-mediated knockdowns were performed with mir-E-based shRNAs cloned into the doxycycline-inducible lentiviral LT3GEP vector (Fellmann et al., 2013). Except for tumor and oncosphere formation assays with ID protein depletion, all shRNAs were induced for 3 days in doxycycline before the beginning of the assay. To maintain cell viability, shRNAs to *ID* mRNAs were induced at the start of the experiment.

Complementary DNA (cDNA) Expression

The pLVX-Tight-Puro vector (Clontech) was used for doxycycline-inducible cDNA expression, and the pLVX-IRES-Hyg vector (Clontech) was used for constitutive expression. *Id1* was cloned into the pLVX-Tight-Puro site using the BamHI/EcoRI sites. Doxycycline (1 μg/mL, Fisher) was added 12 hours prior to experiments to induce gene expression.

Transient Transfections

For mammalian expression vectors, transient transfections were performed using 2.5 μg of DNA and 7.5 μL of Lipofectamine 3000 in 6-well plates seeded overnight with 100,000 cells.

For the CRISPR-mediated knock-in of the GFP tag to the 3' end of the endogenous *Id1* locus, 1.25 μg of the Cas9 vector and 1.25 μg of the cleaved donor fragment were transiently transfected using Lipofectamine 3000 in 6-well plates seeded overnight with 100,000 cells. Cells were selected in 10 μg/mL puromycin (Sigma) for 48h, recovered for 48h, and GFP-positive clones were identified by FACS.

Viral Transductions

Lentivirus was produced by transfection of lentiviral vector with second generation packaging constructs psPAX2 and pMD2.G (Didier Trono, Addgene plasmids 12260 &12259) into 70% confluent 293T cells using Lipofectamine 2000. Viral particles were collected, filtered through 0.45 μm sterile filters, and incubated with the cells of interest for 5 hours with 8 μg/mL polybrene. Cells were recovered in growth medium overnight before addition of selection media--500 μg/mL hygromycin (Life Technologies, 10687-010), 10 μg/mL puromycin (Sigma), 500 μg/mL G418 (GIBCO, 10-131-035), 10 μg/mL blasticidin (Thermo Fisher Scientific, R21001).

Immunodetection

Western immunoblot analysis. 100k cells were lysed in 100 μL of 1.5x NuPAGE LDS Sample Buffer supplemented with 1.5x NuPAGE Sample Reducing Agent, heated at 95ºC for 10 min and sonicated in a water bath for 120 seconds. 10 μL protein per sample were separated on NuPAGE Novex 4-12% gels in MOPS running buffer at 200V for 50 min and transferred to nitrocellulose membranes at 400 mA for 30-60 min. Membranes were blocked with 1:1 Odyssey Blocking Buffer and PBS + 0.1% Tween (PBST) for 1h, incubated overnight in designated antibody, washed 3x in PBST, and incubated 1h in secondary antibody in PBST + 5% milk. Detection using the 680 and 800 channels of the Odyssey CLx imager. Primary antibodies were used at 1:1000 dilution except for anti-Id1 antibodies, which were used at 1:2500.

Immunohistochemistry (IHC) & Immunofluorescence (IF). Tissue blocks were prepared from tissues fixed in 4% paraformaldehyde overnight and dehydrated in 70% alcohol. Paraffin-embedded sections were rehydrated using Histo-Clear (National Diagnostics) followed by 100-70% EtOH and dH2O. Endogenous peroxidase activity was quenched with H_2O_2, antigen retrieval was performed in a steamer for 30 min in citrate antigen retrieval solution. Cell cultures plated on glass coverslips were fixed for 10 min a room temperature in 3% paraformaldehyde. For IHC, Avidin and Biotin blocking solutions (Vector Labs) were followed by normal horse serum (Vector Labs) and overnight incubation in the designated primary antibody. ImmPRESS HRP Anti-Rabbit Ig and ImmPACT DAB Peroxidase (Vector Labs) were used for detection. Detection was followed by dehydration of tissue in 70-100% EtOH and HistoClear, followed by mounting with Vectashield mounting medium. For IF, permeabilization in PBS + 0.1% Triton X (PBSTr) was performed, followed by a 1h block in PBSTr supplemented with 1% normal goat serum and 1% bovine serum albumin. Primary antibody and secondary antibody incubation of 45 min is separated by 3 washes with PBSTr and finished in 3 washes of PBSTr, 5 min incubation in 1x Hoechst, and mounting with Invitrogen ProLong Gold. Primary antibodies were used at 1:100 except for anti-ID antibodies, which were used at 1:500.

Drug Treatments

Cells were plated in D10F + 2.5 μM MK2206 (Chemietek) 12h prior to SB505124 (Sigma, 2.5 μM) or TGFβ (R&D, 100 pM) treatment for the indicated times.

Cell Viability Measurement

CellTiter-Glo and Caspase-Glo 3/7 (Promega) was used according to manufacturer's protocol with incubation for 1 hour. Caspase 3/7 signal was normalized to CellTiter-Glo signal to account for increased apoptosis in larger cell populations.

Ribonucleic acid (RNA) Isolation

Total RNA was purified using Qiagen RNeasy Mini Kit. For RNA-seq, RNA quality and quantity was checked by Agilent BioAnalyzer 2000. 500 ng RNA per sample with RIN > 9.5 was used for library construction with TruSeq RNA Sample Prep Kit v2 (Illumina) according to manufacturer's instructions. Libraries were multiplex sequenced libraries on a Hiseq2500 platform, and more than 20 million raw paired-end reads were generated for each sample.

Quantitative Real Time Polymerase Chain Reaction (qRT-PCR)

cDNA was synthesized from 100-1000 ng RNA using random hexamer oligos and the Transcriptor First Strand cDNA Synthesis Kit (Roche). qPCR was performed on a ViiA 7 Real-time PCR System (Applied Biosystems).

Electromobility Shift Assay (EMSA)

SMAD4-restored cells were treated with 2.5 μM SB505124 or 100 pM TGFβ for 6h. Samples were prepared for EMSA using a canonical E-box motif as previously described (Kreider et al., 1992) and DNA was detected using a 2nd generation DIG Gel Shift Kit (Roche).

Chromatin Immunoprecipitation (ChIP)

10 million cells were collected for each ChIP sample. Cells were crosslinked at 37ºC for 10 min with 1% formaldehyde, quenched with 125 mM glycine, washed with PBS, and sonicated in lysis buffer: 50 mM HEPES/KOH pH7.5, 140 mM NaCl, 0.1% Na-deoxycholate, 1% Triton X-100, 1 mM EDTA, complete protease inhibitor cocktail (Roche). Samples were incubated with 5 µg of anti-RNAPII antibody (Abcam) overnight and washed 7 times with 20 mM Tris, pH7.9, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS. After one wash with Tris-EDTA (TE), DNA was eluted in TE+1% SDS for 1 hour at 65ºC, and reverse-crosslinked with RNAse A for 4 hours and Proteinase K for 1 hour at 65ºC. DNA was purified using a PCR Purification Kit (Qiagen).

Libraries were prepared using the NEBNext ChIP-seq Library Prep Master Mix Set for Illumina (NEB, E6240L) and quality checked using Agilent Technologies 2200 TapeStation to determine fragment size and PicoGreen (Life Technologies/Invitrogen, P7589) to quantify concentration. Samples were pooled and submitted to New York Genome Center for single-end 50 bp sequencing using a HiSeq 2500.

Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)

50,000 cells were collected and washed with 1mL of cold PBS, then 1mL of ice-cold ATAC Buffer (10mM Tris pH 7.4, 10mM NaCl, 3mM MgCl2). Cells were suspended in 50µL of ATAC Lysis Buffer (10mM Tris pH 7.4, 10mM NaCl, 3mM MgCl2, 0.1% NP-40 or IGEPAL-Ca630), incubated on ice for 2 min. 1mL of cold ATAC Buffer was added and nuclei were pelleted at 1500rpm for 10 min at 4ºC in a bucket

centrifuge. Nuclei were resuspended in 22.5μL of the supernatant and transferred to 2.5μL Tagmentation Enzyme (transposase) and 25μL of Tagmentation Buffer (Illumina Nextera DNA Sample Preparation Kit). Reaction was incubated at 37ºC for 30 min. After tagmentation, SDS (final concentration of 0.2%) was added and sample was incubated at room temperature for 5 min before purifying with 2X Agencourt AMPure XP beads (Beckman Coulter A63881). Purified samples were eluted in 50μL of 0.1X Tris-EDTA (Buenrostro et al., 2013).

Libraries were prepared with 50 μL sample + 55μL of NEBNext Q5 Hot Start HiFi PCR Master Mix (NEB, catalogue M0543L) and 5μL of primer mix using 25 μM of the Nextera primers from Buenrostro et al 2013. PCR amplification was performed as previously described for 12 cycles. Samples were purified with 1.5X AMPure XP beads. Concentration was measured using PicoGreen and median fragment size measured using the Agilent D1000 screentape on an Agilent Technologies 2200 TapeStation. Samples were sequenced (paired-end 50 bp) on a HiSeq 2500.

Screens

CRISPR screens were performed as described in Sanjana et al, Nat Methods, 2014. For the whole genome screen, the GeCKO v2 mouse whole genome library (Addgene Pooled Library 1000000049) (Joung et al., 2017; Sanjana et al.) was introduced into an 806-SMAD4pLVX-ID1GFP-Cas9Blast clone at 10% infection efficiency and 200x representation. Cells were plated at 2e6 per plate in 2.5 μM MK2206 for 12h and then treated with 2.5 μM SB505124 or 100 pM TGFβ for 36h

prior to collection for FACS. Between rounds of selection, cells were expanded in 15 cm plates and representation was kept at >200x.

For the targeted CRISPR/Cas9 screens, sgRNA oligos (IDT) were annealed and phosphorylated separately, then pooled and cloned as previously described (Ran et al., 2013) into the LentiGuide-puro vector. Screening proceeded as with the whole genomes screen except that only 2 rounds of selection were performed and representation was maintained at >1000x.

Analysis

General. GNU parallel (Tange, 2011) was used where possible to increase efficiency of analysis.

RNA-seq. Reads were quality checked using FastQC v0.11.5 (Andrews, 2010) and mapped to the human/mouse genome (hg19/mm10) with STAR2.5.2b (Dobin et al., 2013) using standard settings for paired reads. On average, 85% of reads were uniquely mapped. Uniquely mapped reads were assigned to annotated genes with HTSeq v0.6.1p1 (Anders and Huber, 2010) with default settings. Read counts were normalized by library size, and differential gene expression analysis based on a negative binomial distribution was performed using DESeq2 v3.4 (Love et al., 2014). Unless otherwise indicated, thresholds for differential expression were set as follows: adjusted p-value<0.05, fold change > 2.0 or < 0.5, and average normalized read count > 10. Gene set enrichment analysis was performed using GSVA v3.4 (Hanzelmann et al., 2013) or GSEA and previously curated gene sets (Subramanian et al., 2005).

ChIP-seq & ATAC-seq. Reads were quality checked using FastQC v0.11.5 and mapped to the human/mouse genome (hg19/mm10) with Bowtie2 (Langmead and Salzberg, 2012). Sam Tools was used to manipulate .sam and .bam files (Li et al., 2009). Tag directories, visualization in UCSC genome browser, and downstream analyses were performed using the HOMER suite (Heinz et al., 2010).

Screens. Analysis of CRISPR screens was performed as previously described (Sanjana et al., 2014). For the whole genome screen, after count table creation and filtering for expressed genes, guides were collapsed and genes tested for enrichment using CAMERA (Wu and Smyth, 2012). For the targeted screens RIGER was used as a rank-based test for enrichment (Luo et al., 2008)

Analysis of public datasets. Count data were downloaded from the ICGC, GTEx, and Broad GDAC Firehose portals, respectively. For the cross-study comparisons, transcription factors were curated and ranked as previously described (David et al., 2016) and expression values were simulated based on rank-expression correlation. The top 5 factors in each case were selected where indicated, representing ~30% of the reads in each case. Principal components were calculated and plotted. Briefly, simulated expression levels were log-transformed and centered on the mean, and eigenvectors were calculated based on the correlation matrix. All data manipulation and analyses were performed in R using base R packages unless otherwise specified. Code for the described analyses are available upon request.

ChIP-seq, RNAseq, and microarray were downloaded from UCSC Genome Browser, the Epigenome Roadmap Browser, or GEO when available (Bernstein et al., 2010; Diaferia et al., 2016; Kent et al., 2002; Moffitt et al., 2015; Rosenbloom et al., 2013). In all other cases, fastq files were generated from the Sequence Read Archive using fastq-dump and re-analyzed using our in-house ChIPseq pipeline.

MSK-IMPACT (Cheng et al., 2015; Zehir et al., 2017), TCGA ((TCGA, 2017)), and ICGC ((Biankin et al., 2012)) oncogenomic data were accessed via cBioPortal (Cerami et al., 2012; Gao et al., 2013). To enrich for possible driver mutations and

reduce confounders introduced by high levels of passenger mutations, only pancreatic adenocarcinoma cases where <10% of the assayed genes were mutated and <50% contained copy number mutations (representing 95% of the cases) were included in the mutual exclusivity analysis. Mutation impact prediction was based on OncoKB (Chakravarty et al., 2017), hotspot annotation, and recurrence of mutation in COSMIC and cBioportal databases.

Statistical tests. Statistical tests were performed in Prism 7 or R as appropriate. P values are reported based on two-sided, unpaired t tests, $\alpha=0.05$, unless otherwise specified. Bar graphs represent mean ± standard deviation unless otherwise specified.

TGFβ TUMOR SUPPRESSION THROUGH A LETHAL EMT

Introduction

TGFβ is a major tumor-suppressive signal in the gastrointestinal tract and the pancreas (Goggins et al., 1998; Hahn et al., 1996), but how does it achieve tumor suppression? More than 95% of PDAs exhibit *KRAS* activation, most commonly via $KRAS^{G12D}$ (Almoguera et al, Cell, 1998). Given its high frequency across PDA cases and the high clonal representation within tumors, *KRAS* activation is thought to be an initiating event in pancreatic cancer. Interestingly, *RAS* mutant premalignant cells are particularly susceptible to TGFβ-induced apoptosis (Guasch et al., 2007), and in mouse models carrying a $Kras^{G12D}$ allele in the pancreatic epithelium, loss of *Smad4* accelerates progression to PDA (Bardeesy et al., 2006b). Restoration of *Smad4* expression in *Smad4*-defective cancer cells inhibits tumorigenic activity (Duda et al., 2003) and results in apoptosis (Bardeesy et al., 2006b).

Cancer cells avoid the tumor-suppressive action of TGFβ through inactivation of TGFβ receptors or *SMAD* genes or through selective silencing of TGFβ effects (Massague, 2008). Carcinoma cells that are resistant to the tumor suppressive effects of TGFβ can undergo a TGFβ-induced EMT (Heldin et al., 2012). EMT is a developmental plasticity process involving a loss of epithelial features, such as expression of the cell-junction molecule E-cadherin, and a gain in mesenchymal features (Thiery et al., 2009). In cancer cells, EMT promotes invasiveness and stem cell-like features (Valastyan and Weinberg, 2011).

Apoptosis and EMT are generally viewed as separate fates for TGFβ-stimulated cancer cells and opposite poles of the duality of TGFβ in cancer. TGFβ

triggers EMT through induction of *Snail* and *Zeb1/2* (Thiery et al., 2009), but little is known about how the TGFβ pathway triggers apoptosis (Massague, 2012). To address this question, the tumor-suppressive action of TGFβ was dissected in *Ras* mutant pancreatic cancer cells, demonstrating that TGFβ induces PDA cells to undergo a SMAD4-dependent EMT, and EMT then triggers apoptosis. The mechanism involves a conversion of the transcription factor (TF) SOX4 from pro-tumorigenic to pro-apoptotic. This occurs as the result of Snail-mediated repression of KLF5, an essential master regulator of endodermal progenitors. These results illustrate a paradigm in which TGFβ tumor-suppressive action revolves around an EMT-associated disruption of a pro-tumorigenic transcriptional network.[1]

Results

SMAD4-Dependent EMT and Apoptosis in PDA Cells

Activation of the $Kras^{G12D}$ oncogene in the pancreatic epithelium using *Pdx1-Cre; lox-stop-lox (LSL)-*$Kras^{G12D}$ alleles (KC) results in the formation of pancreatic intraepithelial neoplasia (PanIN) lesions that can spontaneously progress to pancreatic ductal adenocarcinoma (PDA) (Hingorani et al., 2003). Pancreatic injury or the presence of additional spontaneous or engineered epigenetic or genetic alterations facilitates the conversion from premalignant lesions to invasive carcinoma (Bardeesy et al., 2006a; Guerra et al., 2007; Hingorani et al., 2003; Hingorani et al., 2005; Ijichi et al., 2006). When combined with $Cdkn2a^{fl/fl}$ and $Smad4^{fl/fl}$ alleles, KC mice rapidly develop well-differentiated PDA with a median survival of under 8 weeks (Bardeesy et al., 2006b) (Figure 3-1). Treatment of mice

[1] This project was a collaboration, primarily with Dr. Charles David (CD). For additional information on this work, see David et al, Cell, 2016.

with caerulein, a decapeptide that mimics cholecystokinin and causes pancreatic acinar cells to degranulate, induces acute pancreatitis and widespread PanIN formation (Morris et al., 2010) and is accompanied by an inflammatory response that includes recruitment of TGFβ-secreting cells (Konturek et al., 1998). To observe the effect of SMAD4 loss in an inflammatory microenvironment that predisposes to PDA, KC and KC mice bearing *Smad4*$^{fl/fl}$ alleles (KSC mice) were treated with caerulein. Three days later, SMAD4-wild-type pancreata exhibited a higher rate of apoptosis than the SMAD4 mutant pancreata (Figure 3-2), suggesting that SMAD4 loss protects cells from apoptosis in this setting and consistent with evidence from squamous cell carcinomas that the main outcome of tumor-suppressive TGFβ signaling is apoptosis (Guasch et al., 2007).

To recapitulate the TGFβ-sensitive cells of the premalignant pancreas and examine the mechanism of apoptosis, SMAD4 was reintroduced into early-passage primary PDA cells isolated from KSC; *Cdkn2a*$^{fl/fl}$ (KSIC) mice (Figure 3-1). The cells, having been selected during tumorigenesis in the absence of SMAD4 signaling, should be sensitive to the tumor-suppressive effect of restored SMAD4 expression. When cultured under low-mitogen cell suspension conditions, cancer cells form "oncospheres" that are enriched for tumor-initiating cells (Lonardo et al., 2011). Treatment of PDA oncospheres with TGFβ resulted in a Smad4-dependent an increase in apoptosis (Figure 3-3). When pancreatitis was induced with caerulein and the cells were orthotopically implanted, mutant PDA cells efficiently formed tumors (4/6 mice), whereas no tumors (0/7 mice) were detected after 2 months

with Smad4+ cells, supporting that the restoration of SMAD4 in this setting effectively re-sensitizes PDA cells to the tumor suppressive effects of TGFβ.

TGFβ or the TGFβ receptor kinase inhibitor SB505124 (DaCosta Byfield et al., 2004) had no effect on the viability of SMAD4+ cells in monolayer culture with serum growth factors (Figure 3-4). Serum starvation sensitizes PDA cells to TGFβ-induced apoptosis (Bardeesy et al., 2006b), and growth factor-activated AKT mediates anti-apoptotic effects through Bcl-xL (Clem et al., 1998). Treatment of SMAD4-expressing PDA cells with the AKT inhibitor MK2206 (Figure 3-4) synergized with TGFβ to promote apoptosis, and this effect was inhibited by overexpression of Bcl-xL (Figure 3-5). Similarly, orthotopic implantation of SMAD4-restored PDA cells with expression of Bcl-xL decreased the number of cells with cleaved caspase 3 detected shortly after implantation (Figure 3-6). These results show that TGFβ/SMAD4 signaling triggers AKT- and Bcl-xL-sensitive apoptosis in SMAD4+ PDA cells.

SMAD4-Dependent EMT Precedes Apoptosis in PDA Cells

Time-lapse imaging of PDA cells for 3 days after the addition of TGFβ revealed that Smad4 mutant cells exhibited little cell death throughout this period and retained epithelial morphology (Figure 3-7); however, SMAD4+ cells started with epithelial morphology but within 24 hr lost cellular polarity and cell-cell contacts and assumed a spindle-like appearance. From 24–72 hr, most spindle-like cells began to shrink, formed membrane blebs, and dissociated into apoptotic bodies. Consistent with this sequence of events, TGFβ treatment in SMAD4+ cells decreased E-cadherin levels at 18 hr, followed by CC3 accumulation (Figure 3-7).

Organoids derived from the premalignant pancreata of genetic mouse models of PDA exhibited similar behaviors (Boj et al., 2015). Organoids from KSIC tumors were little affected by TGFβ addition. In contrast, KIC organoids underwent extensive apoptosis after TGFβ addition, accompanied by a loss of structural integrity, loss of E-cadherin, and increased CC3 staining (Figure 3-8). These observations provided additional evidence that TGFβ lethal EMT occurs in premalignant cells of the pancreas.

EMT TFs Prime Cells for TGFβ-Induced Apoptosis

RNA-seq of the SMAD4-restored PDA cells after 90 min of TGFβ treatment revealed that *Snail* was prominently induced by TGFβ in the observed EMT (Figure 3-9). SMAD2/3 ChIP-seq analysis at 90 min post-TGFβ treatment showed binding of SMAD2/3 to the *Snai1* locus in SMAD4+ cells but not SMAD4 mutant cells (Figure 3-9). Importantly, shRNAs targeting *Snai1* blocked TGFβ/SMAD4-induced EMT-associated E-cadherin down-regulation and apoptosis by TGFβ (Figure 3-10). Further in support of a role for SNAIL in tumor suppression, deletion of *Snai1* in a mouse model accelerated PDA formation. At 4-5 months, in contrast to *Pdx1-Cre; LSL-Kras^{G12D}* mice, which display various grades of acinar-to-ductal metaplasia (ADM) and PanIN but no invasive PDA (Aguirre et al., 2003; Hingorani et al., 2003; Hingorani et al., 2005), 7/9 *Pdx1-Cre; LSL-Kras^{G12D}; Snai1$^{flox/flox}$* mice developed multifocal regions of atypical ADM, with confluent and highly disorganized CK19+ cells displaying nuclear atypia, consistent with carcinoma in situ (Figure 3-11). These results suggested that Snai1 deletion promotes formation of carcinoma *in situ* directly from ADM without progression through a PanIN intermediary.

Although SNAIL was necessary in promoting TGFβ-mediated apoptosis, it was not sufficient. *Snai1* overexpression using a TetOn-*Snai1* construct in SMAD4+ cells did not induce apoptosis in the absence of TGFβ but synergized with TGFβ in the induction of apoptosis (Figure 3-12). These results indicated that SNAIL induction by TGFβ is necessary but not sufficient to cause apoptosis in PDA cells, and induction of apoptosis must involve an additional TGFβ-dependent input.

SOX4 Is Required for EMT-Associated Apoptosis

In search of additional TGFβ-induced factors that cooperate with SNAIL to induce cell death, miR30-based shRNA libraries (Zuber et al., 2011) were constructed to target the mRNAs induced by TGFβ over 90 min in SMAD4+ PDA cells. These pools were transduced into SMAD4+ PDA cells, which were then treated with SB505124 or TGFβ for 5 days. shRNA representation was then determined by deep sequencing. shRNAs targeting the TGFβ receptor subunits *Tgfbr1* and *Tgfbr2*, included as positive controls, were enriched following cell treatment with TGFβ (Figure 3-13). Snai1.1197, which caused a potent knockdown of *Snai1* and rescue of cell viability, was among the top 10% of enriched shRNAs in each of two replicates in the screen. *Sox4* scored as a candidate factor required for TGFβ-induced apoptosis. SOX4 (SRY-related HMG box 4) is widely expressed during mouse embryogenesis and functions in the development of many tissues (Vervoort et al., 2013a). After birth, Sox4 expression is restricted to progenitor cells in pancreatic, intestinal, uterine, and mammary epithelia and maintains epithelial progenitor identity in a range of tissues (Doulatov et al., 2013 2009, Zhang , 2013b), including

pancreatic epithelium (Wilson et al., 2005). SOX4 is up-regulated in various types of carcinoma and leukemia (Vervoort et al., 2013b).

RNAi-mediated knockdown of *Sox4* in SMAD4+ PDA cells abrogated TGFβ-mediated apoptosis (Figure 3-14) but had little effect on EMT morphology or E-cadherin expression (Figure 3-15). Conversely, shRNA-mediated depletion of *Snail* had no effect on SOX4 expression (Figure 3-16). Like SNAIL, TetOn-driven *Sox4* expression in SMAD4+ PDA cells synergized with TGFβ to trigger extensive apoptosis (Figure 3-17). These results show that neither SNAIL nor SOX4 are pro-apoptotic separately, but the simultaneous induction of EMT and SOX4 by TGFβ is lethal to PDA cells.

The SMAD4-Independent TGFβ Program and SOX4

The involvement of SOX4 in EMT-associated apoptosis was surprising given its role as a progenitor identity factor in epithelial tissues, including the pancreas. An additional parallel between SOX4 and pancreatic development emerged when from examining the requirement of SMAD4 for *Sox4* expression. Although SMADs 2 and 3 play a key role in the development and patterning of the foregut and pancreas (Wiater and Vale, 2012), SMAD4 is dispensable for normal pancreas development (Bardeesy et al., 2006b). Notably, *Sox4* induction by TGFβ was SMAD4 independent in SMAD4 mutant mouse and human PDA lines (Figure 3-18).

RNA-seq transcriptomic analysis revealed 165 genes that were differentially expressed >2-fold after addition of TGFβ for 90 min to SMAD4+ PDA cells. Of these responses, 145 were SMAD4 dependent (SMAD4 WT:mutant fold-change ratio > 1.5), including well-known SMAD4-dependent genes such as *Smad7*, *Interleukin-11*,

and *Snai1* (Figure 3-19). An additional set, including *Sox4*, *Junb*, *Sgk1*, and *Fn1*, were equally induced by TGFβ regardless of SMAD4 status, with induction occurring over a longer timescale (Figure 3-19, Figure 3-20). Interestingly, assessment of gene responses to TGFβ at longer timescales revealed that many SMAD2/3/4 target genes could be induced independently of SMAD4 at longer timescales, albeit with lower efficiency (Figure 3-20). SMAD2/3 bound similarly to the *Sox4* locus in the presence and the absence of SMAD4 by ChIP-seq (Figure 3-21). Thus, *Sox4* is in a small group of genes whose response to TGFβ involves SMAD2/3 but is independent of SMAD4.

Role of SOX4 in PDA Tumor-Initiating Activity

The observations that SOX4 is pro-apoptotic in pancreatic epithelial cells that undergo a TGFβ-induced EMT but regulated independently of SMAD4 and amplified rather than deleted in PDAs suggested that SOX4 may play a second role important for the tumor-initiating capacity of these cells (Figure 3-22). In line with this hypothesis, cells grown as oncospheres exhibited higher SOX4 levels than cells grown in adherent conditions and retreatment of SMAD4 mutant oncospheres with TGFβ resulted in increased capacity for secondary oncosphere formation (Figure 3-23). Functionally, PDA cells required SOX4 for growth as oncospheres (Figure 3-24), and *Sox4* depletion in PDA cells inhibited tumor formation in both subcutaneous (Figure 3-25) and orthotopic implantations of PDA cells (Figure 3-26).

Collectively, these results show that TGFβ induces SOX4 expression in PDA cells independently of SMAD4, and SOX4 supports tumor-initiating properties of PDA cells, consistent with the role of SOX4 in the epithelial progenitor state. Yet,

SOX4 triggers apoptosis in cells that undergo a TGFβ/SMAD4-induced EMT, implying that EMT switches SOX4 function from pro-tumorigenic in epithelial PDA cells to pro-apoptotic.

KLF5 as a Lineage-Survival Gene that Cooperates with SOX4 in PDA

To profile the SOX4 transcriptional program in epithelial PDA cells, SOX4 ChIP-seq was performed in these cells and approximately 3,000 SOX4-bound peaks were identified. Analysis of sequences within the SOX4 peaks revealed enrichment of consensus binding motifs for other TFs (Figure 3-27), including the related but often functionally opposed TFs KLF4 and KLF5 (Ghaleb et al., 2005). KLF5 was of particular interest because it enforces epithelial identity (Zhang et al., 2013), is required for the integrity and oncogenicity of intestinal stem cells (Nakaya et al., 2014), and is a lineage-survival oncogene in gastric cancer (Chia et al., 2015). Genome-wide SOX4 ChIP-seq confirmed that approximately 80% of SOX4 peaks were associated with an overlapping KLF5 peak (Figure 3-28), and like SOX4, KLF5 is required for the formation of oncospheres (Figure 3-29) and orthotopic tumors by PDA cells (Figure 3-30).

High, lineage-specific expression is a hallmark of lineage-survival oncogenes such as MITF in melanoma. Similarly, KLF5 expression in human tumors is lineage restricted, with KLF5 ranking among the most highly expressed TFs in GI and squamous carcinomas (Figure 3-31), tumors that exhibit high rates of loss of TGFβ signaling. Taken together, these data suggest that KLF5 functions as a lineage-survival gene in pancreatic and other GI cancers. Given the EMT phenotype induced by TGFβ, and the repression of TFs that determine the epithelial state by EMT TFs

(Cicchini et al., 2006), RNA-seq data of the SMAD4-restored PDA cells after TGFβ treatment for 24 hours were queried for *Klf5* repression. The analysis revealed a loss in the expression of 20 abundant PDA TFs, including *Klf5*, during EMT (Figure 3-32). Knockdown of *Snail* blunted the down-regulation of *Klf5* by TGFβ (Figure 3-33), further supporting a loss of KLF5 as part of the lethal EMT program.

Repression of KLF5 Turns SOX4 into a Pro-apoptotic TF

These data indicated that KLF5 may be an important determinant of SOX4 function in PDA. Consistent with this, *Klf5* knockdown in SMAD4 mutant PDA cells was accompanied by increased apoptosis, which was further accentuated by TGFβ addition. Knockdown of *Foxa2*, another highly expressed TF, did not promote TGFβ-induced apoptosis (Figure 3-34). In a reciprocal experiment, enforced expression of KLF5 in SMAD4+ cells prevented TGFβ-induced apoptosis (Figure 3-35). These results suggested that a low KLF5:SOX4 ratio might trigger PDA cell death. Tandem knockdown of SOX4 and KLF5 by shRNA confirmed this hypothesis (Figure 3-36), thus linking apoptosis in PDA to an imbalance between SOX4 and KLF5.

To identify SOX4-dependent mediators of apoptosis in the post-EMT context, RNA-seq was performed on SMAD4+ cells with and without *Sox4* knockdown. After 24 hr of TGFβ treatment, 1,169 genes were induced by TGFβ. Of these, 290 showed a >2-fold difference in their TGFβ responses in SOX4-depleted cells compared to controls. The set of TGFβ gene responses that were lost in SOX4-depleted PDA cells included BIM (encoded by *Bcl2l11*) and BMF (Figure 3-37). BIM and BMF are BH3-only pro-apoptotic proteins, which are sequestered by BCL-XL (Pinon et al., 2008). These observations were consistent with the ability of BCL-XL to blunt TGFβ-

induced apoptosis in this context (refer to Figures 3-5 and 3-6). Increased expression of *Bim* and *Bmf* in response to TGFβ in PDA cells was confirmed to require SOX4 (Figure 3-37). Thus, repression of KLF5 by TGFβ-induced EMT-TFs switches SOX4 from a pro-tumorigenic enforcer in malignant endodermal progenitors to an inducer of tumor-suppressive, pro-apoptotic genes (Figure 3-38).

Discussion

TGFβ Control of Cellular Identity Drives Lineage-Specific Tumor Suppression

TGFβ superfamily members are powerful orchestrators of cell-identity changes throughout development, and the regulation of KLF5 and SOX4 by TGFβ implicate this property in tumor suppression. Pre-neoplastic lesions are dependent on TFs that maintain progenitor populations endemic to the tissue of origin (Garraway and Sellers, 2006). In tumor types most sensitive to TGFβ, KLF5 appears central to the identity of these progenitors. KLF5 marks proliferative progenitors in the developing GI tract (Ohnishi et al., 2000). KLF5 deletion in mouse intestinal stem cells results in apoptosis and depletion of LGR5+ cells (Nakaya et al., 2014 2015) and abrogates tumor formation (Nakaya et al., 2014). KLF5 displays amplifications, stabilizing mutations (Bialkowska et al., 2014 2015), and high-level expression; all of these are restricted to a set of GI and squamous tumor types most prone to genetic inactivation of the TGFβ pathway.

SOX4 is associated with stem/progenitor cell properties in many contexts (Ikushima et al., 2009 2009, Zhang , 2013b). By repressing endodermal TFs, TGFβ drives the lethal collapse of the progenitor-like state of the SOX4-KLF5 network, imposing a cell-identity-specific bottleneck. Cells could emerge from this TGFβ-

imposed bottleneck either by (a) losing TGFβ signaling components, (b) attaining a self-renewing epigenetic state compatible with TGFβ signaling, or (c) activating signaling pathways that blunt the effects of TGFβ. SMAD4-wild-type mouse models often give rise to tumors with loss of E-cadherin and sarcomatoid differentiation (Rhim et al., 2012 2012); cells derived from these tumors resist death when treated with TGFβ (Bardeesy et al., 2006b), a likely example of (b). Such tumors are rare in human pancreas cancer (Kane et al., 2014), indicating that different mechanisms of escape from TGFβ suppression predominate in human tumors.

A Tumor-Suppressive EMT

An unexpected finding is that tumor suppression is coupled to an EMT, a developmental program thought to promote the acquisition of malignant traits in advanced carcinomas (Valastyan and Weinberg, 2011). This coupling occurs through the function of the EMT-TFs SNAIL and ZEB1, which are induced by and cooperate with SMAD4 to bring about the repression of KLF5. *KRAS* mutations sensitize PDA cells to TGFβ-induced EMT (Horiguchi et al., 2009) and sensitize keratinocytes to TGFβ-induced apoptosis (Guasch et al., 2007). Taken together, these observations suggest that *KRAS* mutations sensitize incipient carcinoma cells to lethal EMT.

One implication of this work is that EMT plays a role in tumor suppression. An association of EMT with apoptosis is not unprecedented. Serum-starved canine kidney epithelial cells undergo EMT followed by cell death in response to TGFβ (Peinado et al., 2003). EMT can also precede apoptosis in response to TGFβ in murine mammary epithelial cells (Gal et al., 2008 2013a). In the presence of

sustained TGFβ treatment, the emergence of proliferative cells that survive EMT requires 2–3 weeks (Gal et al., 2008), which is consistent with our findings that lethal EMT imposes a narrow bottleneck on naive epithelial cells. Although the bulk of naive PDA cells are susceptible to EMT-linked death, EMT-permissive cells can emerge from the TGFβ-imposed bottleneck (Bardeesy et al., 2006b; Rhim et al., 2012). Moreover, lethal EMT suggests that tumorigenicity linked to an EMT is not broadly endowed on incipient carcinoma cells but is rather a selectable trait during carcinoma progression.

The Dual Nature of TGFβ and Core TF Regulation

TGFβ has a long history of duality in cancer. Despite its clear tumor-suppressive properties, a high level of TGFβ signaling in PDA tumors is associated with poor prognosis (Friess et al., 1993 1999). The observation that *Sox4* is induced in a SMAD4-independent manner provides one potential mechanism by which TGFβ may promote aggressiveness in SMAD4 null PDA and could account for the predominance of SMAD4 loss over TGFβ receptor inactivation in human PDA. Approximately half of PDA harbor SMAD4 mutations, but few harbor TGFBR mutations (Goggins et al., 1998), underscoring the requirement to bypass SMAD4-dependent lethal EMT but not TGFβ-mediated tumor-promoting effects. PDA tumors with wild-type SMAD4 status tend to have fewer metastases (Iacobuzio-Donahue et al., 2009) perhaps reflecting an increased susceptibility of these disseminated cells to suppression by stromal TGFβ. The mouse PDAs with *Snai1* knockout that retain metastatic competency are consistent with a recent report that EMT is not always connected with effective metastatic dissemination in PDA (Zheng et al., 2015).

SMAD4-independent induction of SOX4, a known component of pancreatic progenitors (Wilson et al., 2005), also provides a potential resolution to the apparently paradoxical observation that Activin receptor mutant mice develop a severely hypoplastic pancreas, whereas SMAD4 is dispensable for normal pancreas development (Bardeesy et al., 2006b 2000).

SOX4 both mirrors and contributes to the context-dependent dual nature of TGFβ. SOX4 is an established stem/progenitor factor that functions as a driver of apoptosis in some contexts (Vervoort et al., 2013b). The major determinant of the context for SOX4 activity is the cellular repertoire of transcriptional master regulators. Why would a stemness factor be endowed with the ability to induce apoptosis in the absence of appropriate transcriptional partners? This property may have evolved as a mechanism to eliminate stem cells with developmentally untenable TF configurations.

ID1 MEDIATES ESCAPE FROM TGFβ TUMOR SUPPRESSION

Introduction

Despite the possibility of escaping the tumor suppressive effects of TGFβ by mutational inactivation of the core pathway components, the majority of tumors retain an active TGFβ pathway (Yachida et al., 2012). Even in pancreatic ductal adenocarcinoma (PDA), which presents the highest frequency of TGFβ pathway mutations in cancer, these mutations occur only in one half of cases, leaving the other half of PDAs with a functional TGFβ pathway.

Selective pressure for the inactivation of TGFβ signaling in PDA results from a lethal EMT triggered by TGFβ via SMAD2/3/4 (David et al., 2016). In normal pancreatic progenitors, the lineage determining transcription factors SOX4 and KLF5 cooperatively impose a pancreatic epithelial progenitor phenotype. Mutational activation of KRAS, a nearly universal tumor-initiating event in PDA, causes TGFβ to induce expression of the master EMT transcription factor SNAIL, which represses KLF5 expression. As cells undergo an EMT in this dysregulated, KLF5-depleted context, SOX4 is directed to activate pro-apoptotic genes, resulting in cell death (David et al., 2016). However, TGFβ can also promote progression by enhancing invasive and pro-metastatic effects in established tumors (Cui et al., 1996; Padua et al., 2008; Tauriello et al., 2018), and cancer cells that retain an active TGFβ pathway while evading its tumor suppressive action can benefit from TGFβ-driven invasive and metastatic functions (David and Massague, 2018b).

How PDAs with an intact TGFβ pathway avert the apoptotic effect has remained an open question. Genomic analysis of PDAs has revealed three other

frequently mutated genes –*KRAS, CDKN2A* and *TP53*– in addition to *SMAD4* (TCGA, 2017). Mutations in these three genes are not mutually exclusive with SMAD4 mutations. Beyond these genetic alterations, PDAs present a long series of low frequency mutated genes, none of which is an obvious inhibitor of cellular responses to TGFβ signaling (Bailey et al., 2016; Jones et al., 2008; TCGA, 2017; Witkiewicz et al., 2015). Given that no single, high-frequency genetic alteration has emerged as mutually exclusive to TGFβ pathway inactivation, multiple alterations may converge on a common regulatory node that is critical to escape from tumor suppression in PDAs with an intact TGFβ pathway. Identifying this regulatory node would provide a potential therapeutic target in PDA.

Transcriptional dysregulation is a common feature of emerging tumors, reflecting adaptation to genetic alterations in cancer cells and to inputs from the tumor microenvironment (Christophorou et al., 2006; Garraway and Sellers, 2006; Lee and Young, 2013; Ventura et al., 2007). A focus on the analysis of dominant transcriptional networks in PDA revealed that PDA cells with an active TGFβ pathway avert apoptosis by transcriptional dysregulation of ID1, and this dysregulation results from a diverse set of infrequently mutated pathways, including PI3K-AKT and DNA damage response pathways. ID1 thus emerges as a target of interest in pancreatic cancer.

Results

TGFβ signaling is active in half of pancreatic cancers

SMAD4 is inactivated in 38-43% of human PDAs, and the full set of TGFβ pathway core components (*TGFBR1*, *TGFBR2*, *SMAD2*, *SMAD3* and *SMAD4*)

collectively are inactivated in approximately 53% of PDAs (Figure 4-1). To determine whether PDAs lacking mutations in these core components retain a functional TGFβ pathway, human PDA organoids were assayed for responsiveness to TGFβ. Since the functional transcriptional unit of TGFβ signaling is a trimer of phospho-SMAD2/3 with SMAD4 (Figure 4-2), SMAD4 expression and receptor-phosphorylated SMAD2 (pSMAD2) in response to TGFβ were measured. Six of the 13 organoids showed low levels of pSMAD2 or SMAD4 by Western blot, indicative of functional loss of TGFβ receptors or SMADs (Figure 4-2). With induction of the common TGFβ target gene *SMAD7* as readout, the six PDA organoids with defective SMAD levels were found to have low (<2-fold) or no increase in SMAD7 mRNA levels in response to TGFβ whereas the other seven showed a 4- to 9-fold increase (Figure 4-2).

SMAD2 is phosphorylated by TGFBR1/2 regardless of SMAD4 status, and staining of a tissue microarray of 130 human PDAs for pSMAD2 (Winter et al., 2012) revealed that 5% of samples were negative for pSMAD2 in the tumor cells (Figure 4-3), congruent with the frequency of reported genetic TGFβ receptor inactivation. These samples retained pSMAD2 staining in the tumor stroma. The remainder of the samples was positive for pSMAD2, suggesting the presence of active TGFβ receptors in the cancer cell compartment of these tumors (Figure 4-3). 69% of the tumors in this TMA stain positive for SMAD4 (Winter et al., 2012). These results suggest that a large subset of PDA retains a functional TGFBR-pSMAD2-SMAD4 pathway and are exposed to TGFβ in the tumor microenvironment.

PDAs share a dominant transcriptional network

Cell lineages are characterized by the dominant transcription factors that the cells specifically express (Davis et al., 1987; Takahashi and Yamanaka, 2006; Yamamizu et al., 2013). To explore the hypothesis that PDAs with either active or inactive TGFβ/SMAD pathway develop under similar selective pressures and achieve certain common end-point characteristics, the transcriptional networks that are selected for in pancreatic cancer were compared to those in the normal pancreas. Principal component analysis (PCA) of 225 cases of PDA, normal pancreas, and pancreatic neuroendocrine tumor (PNET) with gene expression datasets (Bailey et al., 2016; GTEx, 2013), based on the top 5 most highly expressed transcription factors within each sample (Figure 4-4), yielded separation of normal, PDA, and PNET samples. This separation was similar to that of PCA using a full list of expressed transcription factors (Figure 4-5). Similar results were obtained when the same analysis was applied to a second, independent PDA gene expression dataset (Figure 4-5)(TCGA, 2017). PDA cases with wild type or mutant TGFβ pathway components clustered together (Figure 4-4), suggesting that PDA evolution selects for similar transcription factor networks regardless of the genetic integrity of the TGFβ pathway.

Examination of the principal components of this analysis revealed transcription factors highly expressed specifically in normal pancreas, PNET and PDA, respectively. Of the transcription factors highly expressed in normal tissue, RBPJL, BHLHA15/MIST1 and NR5A2 are crucial in pancreatic specification and development and important to the acinar lineage, which is the putative cell of origin

of PDA (Masui et al., 2010; Pin et al., 2001; von Figura et al., 2014). In PNET, the highly expressed transcription factors correspond to the endocrine lineage: ISL1, MAFB, NKX2.2, PAX6 and ST18 (Artner et al., 2007; Henry et al., 2014; Karlsson et al., 1990; Sander et al., 1997; Sussel et al., 1998). Those that are highly expressed in PDA include AHR, HMGB2, ID1, KLF5 and PPARG, which play roles in epithelial progenitor specification and transformation (Fusco and Fedele, 2007; Lasorella et al., 2014; Murray et al., 2014; Tetreault et al., 2013). Analysis of other cancer types demonstrated a similar enrichment for transcription factors with tissue-specific traits (Figure 4-6), highlighting the generality of this principle. The ten genes that contribute the most to principal component 1 in this analysis were sufficient to separate PDA and normal tissue samples by unsupervised hierarchical clustering (Figure 4-7). In an independent dataset (Moffitt et al., 2015), these factors are enriched in PDA regardless of whether they fall into the recently defined classical or basal subtypes (Figure 4-8), as well as in both primary and metastatic tumors (Figure 4-9).

To investigate the effect of TGFβ/SMAD signaling on the expression of the PDA-associated transcription factors, a SMAD4 lentiviral expression vector was transduced into cancer cells derived from in mouse *Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$* PDA tumors. This rescued the sensitivity of the cells to the tumor suppressive effects of TGFβ, recapitulating the exposure of premalignant pancreatic cells to TGFβ (David et al., 2016). Notably, treatment of these SMAD4-restored PDA cells with TGFβ for 12 h, compared to treatment with SB505124 (a TGFBR1 kinase inhibitor), decreased the expression of the five transcription factors enriched in PDA, as determined by

RNA-seq (Figure 4-10). *ID1* and *PPARG* in particular were strongly down-regulated by TGFβ in PDA cells. This effect was accompanied by the induction of *Snai1*, encoding the master EMT transcription factor SNAIL (Figure 4-10), and by an enrichment for EMT and apoptosis gene signatures (Figure 4-11), which are typical responses of SMAD4-restored PDA cells to TGFβ under these conditions (David et al., 2016).

ID1 expression in pancreatic progenitors

Of the factors enriched in PDA and down-regulated by TGFβ, ID1 was of particular interest because the high expression of this factor is also correlated with the tumor types that most frequently lose TGFβ pathway components (David et al., 2016), and because of the known role of the ID family members (ID1~4) in the negative regulation of differentiation in embryonic and adult stem and progenitor cells (Anido et al., 2010; Ying et al., 2003; Zhang et al., 2014). ID1~4 function as negative regulators of bHLH transcription factors and promote self-renewal and proliferation in stem and progenitor cells, and their expression is low or null in more differentiated cells. ID proteins lack a DNA binding domain and primarily interact with bHLH E-proteins to prevent their dimerization with tissue-specific bHLH lineage differentiation transcription factors, sequestering these proteins away from their target chromatin domains (Benezra et al., 1990; Lasorella et al., 2014; O'Toole et al., 2003). The 5' promoter region of *ID1* contains four canonical CAGAC SMAD-binding elements, three clustered at -1 kb and one at -38 bp relative to the transcription start site. The CAGAC cluster at -1 kb and a GC-rich region at -1.1 kb have both been implicated in SMAD-mediated induction of *ID1* by BMPs, whereas

only the CAGAC cluster at -1 kb was found to be necessary for the repression of *ID1* by TGFβ (Kang et al.; Lopez-Rovira et al., 2002). In keratinocyte, mammary, and bronchial epithelial progenitor cells, TGFβ represses ID1∼3 (Kang et al., 2003), whereas in breast cancer cells and glioblastoma cells, TGFβ induces ID1 expression (Anido et al., 2010; Barrett et al., 2012; Padua et al., 2008).

To better understand the role of ID1 in PDA, ID1 expression was analyzed in a panel of human and mouse pancreatic tissues. In both human and mouse tissues, a proportion of epithelial cells within the PDA samples, inclusive of areas of both adenomatous and squamous trans-differentiation, stained positive for ID1 relative to normal pancreatic tissue, where staining was restricted to endothelial cells (Figure 4-12). Of 30 cases of human PDA, all of which are KRAS-mutant and 67% of which harbor p53 mutations, 26 (87%) exhibited cancer cells with nuclear ID1 staining; on average, 5-10% of cancer cells in these tumors exhibited positive ID1 staining, regardless of the SMAD4 status of the tumor (Figure 4-13). In comparison, ID1-positive epithelial cells were not observed within normal human pancreatic tissues collected via rapid autopsy from two patients with lung adenocarcinoma or anaplastic glioma who had no signs of pancreatic disease or inflammation (Figure 4-12, 4-13).

Given the role of the ID proteins as regulators of differentiation, one hypothesis is that the minority of ID1 positive cells in heterogeneous PDA cell populations may act as progenitors that give rise to the ID1 negative cells. In support of this hypothesis, ID1+ cells constitute a subset of PDX1+ cells (Figure 4-14). PDX1 (pancreatic and duodenal homeobox 1) is a transcription factor that is

necessary for pancreatic development and associated with a pancreatic progenitor population in PDA (Ischenko et al., 2014; Jonsson et al., 1994). In contrast, immunofluorescence detection of amylase, a marker of mature pancreatic acinar cells, is mutually exclusive with ID1 (Figure 4-14).

To recapitulate the heterogeneity in ID1 expression observed in pancreatic tumors, a green fluorescent protein (GFP) reporter was knocked-in at the endogenous *Id1* locus in mouse *Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$* PDA cells (Figure 4-15) (Lopes et al., 2016; Perez et al., 2017). ID1-GFP^{high} cells isolated by FACS from these populations grew better as oncospheres, a surrogate growth assay for stem-like cancer cells in which the cells are plated in low-attachment plates with limited growth factors (Hermann et al., 2007)(Figure 4-16). ID1-GFP^{high} cells were able to generate both ID1-GFP^{high} and ID1-GFP^{low} progeny, whereas ID1-GFP^{low} cells only generated ID1-GFP^{low} progeny (Figure 4-16), a pattern that has been observed previously in glioma (Barrett et al., 2012). The transcriptional repertoire of pancreatic stem cells is not well defined, however, the ID1-GFP^{high} cells expressed transcriptomic signatures associated with BMI1+ breast cancer progenitors and HOPX+ intestinal stem cells (Figure 4-17). Interestingly, both BMI1 and HOPX markers are associated with the +4 position of the intestinal crypt, a subpopulation of intestinal stem cells that express ID1 (Takeda et al., 2011; Zhang et al., 2008; Zhang et al., 2014). In contrast, ID1-GFP^{low} cells expressed transcriptomic signatures of pancreatic epithelial differentiation, as determined by mRNA sequencing (RNA-seq) and gene set enrichment analysis (Figure 4-17).

Along with ID1, ID2 and ID3 are expressed in PDAs, and could play redundant roles to ID1 (Lyden et al., 1999; Yan et al., 1997)(Figure 4-18); therefore, combinations of knockdowns of these factors were tested. shRNA-mediated depletion of ID1 in mouse PDA cells decreased the oncosphere-forming ability of these cells, and additional depletion of ID2 and ID3, which have been reported to have redundant activities to ID1, further decreased oncosphere growth (Figure 4-19). Together, these results suggest that, as in other contexts (Niola et al., 2012; Romero-Lanman et al., 2012; Zhang et al., 2014), ID1 and, to a lesser extent, ID2 and ID3 are selectively expressed in PDA progenitor cells and support the progenitor phenotype.

In order to determine the role of ID proteins in the tumorigenic potential of PDA cells, 500 *Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$* PDA cells were orthotopically implanted in mice and doxycycline was administered to the mice to induce expression of *Id1~3* shRNAs immediately after implantation (Figure 4-20, 4-21). Tumor growth was tracked by bioluminescence of a transduced firefly luciferase gene in the cells. Like with the oncosphere assays, ID1 depletion alone led to a decrease in tumor burden (Figure 4-20); however, the effect of depleting ID1~3 was 100x greater—it decreased tumor burden and prolonged the survival of the mice by 25% (Figure 4-21). Positive ID1 immunostaining was present in the late-emerging ID1~3 shRNA tumors, indicating that these tumors formed from cells that escaped shRNA-mediated depletion of ID1 (Figure 4-21). To determine if established tumors retain sensitivity to Id depletion, the same cells were implanted and allowed to grow for 3 weeks. The mice were then randomized into luminescence-matched groups, and

administered doxycycline or not. In this setting, tumor growth was again blunted by ID depletion and mouse survival was similarly prolonged by 25% (Figure 4-22).

Together, these data suggest a model in which PDA cells express high levels of ID1 relative to normal cells, and that this is associated with a progenitor cell phenotype. TGFβ-mediated repression of ID1 in this context results in lethality. To progress to invasive PDA, tumor cells must therefore escape this repression, either through TGFβ pathway inactivation or other means (Figure 4-23).

$ID1^{high}$ vs $ID1^{low}$ states are associated with enhancer status

Given that the $ID1^{high}$ and $ID1^{low}$ populations in both the genetic mouse models and the ID1-GFP reporter cells possess the same oncogenic alterations, epigenetic changes were likely to explain the difference in ID1 expression. A query of the status of the ID1 locus using the Epigenome Browser, a database of epigenetic characterization of a variety of tissues, revealed that the *ID1* promoter in normal human pancreas contains H3K4me3, a histone modification that denotes promoter activation and priming for expression (Figure 4-24) (Bernstein et al., 2010). In comparison, in differentiated neurons, in which ID1 is not expressed (Ying et al., 2003), this locus is silenced, as indicated by the presence of H3K27me3 (Figure 4-24). In the small intestine, where ID1-expressing cells contribute to normal regeneration, H3K27ac, a mark of activated enhancers, is enriched in a region near ID1 (Figure 4-24) (Bernstein et al., 2010; Creyghton et al., 2010; Zhang et al., 2014). Like the small intestine, human pancreatic cancer cells have H3K27ac enrichment near ID1, indicating an activated ID1 enhancer (Figure 4-25). As a control, the RBPJL locus, which is necessary for normal pancreas development, was analyzed (Masui et

al., 2010). RBPJL is activated in normal pancreas but silenced in neurons, the small intestine, and pancreatic cancer cells (Figure 4-24, 4-25). These observations suggest that accumulation of H3K27ac at the ID1 locus is associated with high expression of this gene.

To characterize the functional consequence of this enhancer, ATAC-seq was used to detect regions of increased accessibility, which are associated with gene regulatory regions (Buenrostro et al., 2013). Several peaks of increased accessibility were detected within the *ID1* H3K27ac enhancer in these cells; of these, a subset is conserved between human and mouse and therefore more likely to be regulatory loci (Figure 4-26). To test this functionally, an sgRNA library was designed to tile the seven conserved regions at the *Id1* locus in mouse PDA cells. By restricting the screen to regions with conserved increased accessibility, potential regulatory regions were enriched while maintaining the technical advantages of a relatively small library size. This library was transduced into the cells with the ID1-GFP reporter and sorted for GFP^{high} and GFP^{low} cells (Figure 4-27). Analyzing sgRNAs enriched in the GFP^{low} population compared to the GFP^{high} population, regions within the 3' H3K27ac locus were found to score most highly as required for *Id1* expression, particularly Region 5 (Figure 4-27).

To functionally validate the findings of the screen, two reciprocal experiments were performed. First, three pairs of sgRNAs were designed to flank and remove region 5. CRISPR/Cas9 removal of this region resulted in loss of ID1-GFP expression (Figure 4-28). Second, the highest scoring sgRNA in the screen (Figure 4-27) was introduced into the ID1-GFP reporter cells, the cells were sorted

for high versus low ID1-GFP reporter expression, and a 198 bp region surrounding the predicted cut site was PCR amplified and sequenced. The ID1-GFP^{low} population was highly enriched for editing at this site (Figure 4-28). Together, these results support a functional role for the 3' enhancer in modulating high versus low ID1 expression in pancreatic epithelial cells.

The Transcription Factor ChIP-seq track from ENCODE in the UCSC genome browser suggested that multiple factors bind to Region 5 (Kent et al., 2002; Rosenbloom et al., 2013), so a small-scale screen of candidate factors was performed using the same scheme used to scan the *Id1* locus. sgRNAs that targeted to three transcriptional regulators –*Rxra*, *Tbl1xr1*, and *Tcf7l2*– were enriched in the cells that lost ID1-GFP expression (Figure 4-29). Retinoid X receptor alpha (RXRα), which can function as a homodimer or as a heterodimeric partner of other nuclear hormone receptors (Evans and Mangelsdorf, 2014), was of particular interest. Not only is there a nuclear receptor motif in Region 5 of the *Id1* locus, there is also RXRA ChIP signal at the corresponding locus in human liver cells (Kent et al., 2002; Rosenbloom et al., 2013), suggesting conservation across species, as would be expected for a regulatory element (Figure 4-30). ChIP-PCR confirmed RXRA binding to Region 5 of murine PDA cells, and inducible shRNA depletion of RXRα decreased the transcript level of *Id1* in mouse PDA cells (Figure 4-30), suggesting that RXRα is one of potentially many factors (Lasorella et al., 2014) that can increase *ID1* expression. These results suggest that the *ID1* locus is primed for expression in pancreatic progenitors and receives additional inputs during oncogenesis for enhancer activation and increased expression.

ID1 down-regulation in association with the TGFβ lethal EMT

The emergence of ID1 as a top PDA transcriptional network component, its role in supporting the growth and tumorigenic activity of PDA progenitor cells, and its down-regulation by TGFβ in the context of a lethal EMT prompted an investigation of the interrelationship between these various phenomena. In mouse *Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$* PDA cells with restored SMAD4 expression, treatment with TGFβ decreased the expression of *Id1*, *Id2* and *Id3* in conjunction with the previously described EMT and apoptosis (Figure 4-31, 4-10, 4-11, Chapter 3). ID proteins function primarily by sequestering E-proteins from binding to and activating gene transcription, and in support of this, a gel-shift assay demonstrated that TGFβ treatment of the SMAD4-restored PDA cells, which depletes the ID proteins, promotes protein binding to DNA containing an E-box motif (Figure 4-32).

Moreover, in SMAD4-restored PDA cells containing an endogenous ID1-GFP reporter, the sorted ID1-GFP^{high} progenitor cells underwent *Id1* down-regulation by TGFβ and apoptosis (Figure 4-33). In contrast, the more differentiated ID1-GFP^{low} counterparts did not undergo apoptosis in response to TGFβ (Figure 4-33). Collectively, these results suggested that TGFβ-induced ID1 down-regulation in association with a lethal EMT are specific responses of ID1^{high} pancreatic cancer progenitors but not the differentiated, ID1^{low} progeny of these cells.

Dysregulated ID1 expression in PDAs with a functional SMAD pathway

Many PDA tumors develop with a functional TGFβ/SMAD pathway and are exposed to TGFβ in the tumor microenvironment, as indicated by the presence of pSMAD2 in these tumors (refer to Figure 4-3). These tumors therefore must have

acquired an ability to avert the tumor suppressive effect of TGFβ. Indeed, cells derived from human and mouse PDAs that developed with a functional TGFβ/SMAD pathway were resistant to TGFβ-induced apoptosis, whereas those that developed with SMAD4 inactivation suffered TGFβ-induced apoptosis when SMAD4 expression was restored (Figure 4-34). Notably, ID1 was not repressed by TGFβ in human or mouse organoids derived from PDAs (Figures 4-35). In fact, treatment with TGFβ increased ID1 expression in a cohort of human PDA organoids, the extent of ID1 induction being proportional to that of SMAD7 induction across the cohort (Figures 4-35, 4-2).

To recapitulate this observation in cells derived from the same genetic background, SMAD4-restored mouse PDA cells (S4 cells) were cultured with TGFβ for 3 weeks to select for cells resistant to the pro-apoptotic effect of TGFβ (Figure 4-36). Pharmacologic inhibition of AKT with MK2206, an allosteric AKT inhibitor, strongly synergizes with TGFβ in the induction of a lethal EMT (David et al., 2016). Therefore, all lethal EMT experiments are performed in the presence of MK2206 in the media unless otherwise indicated. The cells that were derived (S4.1-3 cells) were resistant to the pro-apoptotic effects of TGFβ (Figures 4-36, 4-37), undergo a reversible EMT in response to TGFβ (Figure 4-37), and induce *Smad7* in response to TGFβ like the population from which they were selected (Figure 4-38). However, these cells increased *Id1* expression in response to TGFβ, instead of repressing it as in the TGFβ-sensitive population (Figures 4-38). Besides *Id1*, only four other genes were differentially regulated by TGFβ by more than 4-fold: *Id3*, *Fam167a*, *Trib3*, and *Chac1*. These results indicate that human and mouse PDA cells that retain an active

TGFβ pathway have a selective dysregulation of ID1 expression; TGFβ up-regulates ID1 in these cells in contrast to the down-regulation observed in premalignant pancreatic progenitors and related models of TGFβ-induced lethal EMT.

To understand the basis for the resistance to TGFβ-induced apoptosis in the S4.1 cells, the genes differentially expressed between the S4 and S4.1 cells (average expression >10 readcounts, fold change >2, $p<0.05$) were functionally annotated. The top three enriched pathways were all related to small molecule metabolism (Figure 4-39). The common genes between these pathways were all UDP glucuronosyltransferases (UGTs), which are key enzymes for the inactivation and excretion of polycyclic compounds, including MK2206 (Ahn et al., 2015). Several UGTs were up-regulated in the S4.1 cells relative to the S4 cells (Figure 4-39). Furthermore, gene set enrichment analysis based on an MK2206 gene signature--generated from the differentially expressed genes (average expression > 10 readcounts, fold change > 4, $p<0.05$) of S4 cells treated with or without 2.5 μM MK2206 for 16 h--showed that the MK2206 signature was decreased in the S4.1 cells relative to the S4 cells, suggesting less complete inhibition of AKT in the S4.1 cells (Figure 4-40).

A recent study reports that TGFβ phosphorylation of SMAD1/5 induces ID1 (Ramachandran et al., 2018). To test whether a switch from SMAD2/3 phosphorylation to SMAD1/5 phosphorylation could explain the transition from ID1 repression by TGFβ to ID1 induction, SMAD4-restored PDA cells were treated with TGFβ and BMP4. TGFβ treatment results in phosphorylation of both SMAD2/3 and SMAD1/5 in both cell lines that induce and repress ID1 (Figure 4-41), suggesting

that the difference in regulation occurs downstream of TGFβ mediated SMAD phosphorylation.

ID1 uncouples TGFβ-induced EMT from cell death

The association between resistance to TGFβ-induced apoptosis and retention of high levels of ID1 expression in the presence of TGFβ suggested that a loss of TGFβ-mediated ID1 repression could protect PDA progenitors from the pro-apoptotic effects of TGFβ. ID1 was expressed from a doxycycline-inducible promoter in SMAD4-restored mouse PDA cells that are normally sensitive to the apoptotic effects of TGFβ. The expression of exogenous ID1 in these cells in the absence of TGFβ, when endogenous ID1 is already highly expressed, resulted in very few significant changes in gene expression (average readcounts >10, fold change >2, $p<0.05$) (Figure 4-42). Strikingly, the enforced expression of ID1 in the presence of TGFβ protected the cells from apoptosis (Figure 4-43). Implantation of these cells into the TGFβ-rich pancreata of mice with caerulein-induced pancreatitis (David et al., 2016) resulted in increased tumor formation, supportive of ID1-mediated protection from TGFβ tumor suppression (Figure 4-44). Associated with the protection from apoptosis, the enforcement of ID1 expression altered a subset of TGFβ gene responses (Figure 4-45), five of which have known pro-apoptotic functions: *Fbxo32*, *Bmf*, *Rnf152*, *Ndrg1*, and *Errfi1* (Hopkins et al., 2012; Puthalakath et al., 2001; Stein et al., 2004; Tan et al., 2007; Zhang et al., 2010). Within the accessible chromatin regions detected by ATAC-seq near these genes, there was a strong enrichment of bHLH and SMAD motifs (Figure 4-46), and SMAD2/3 bind to these accessible chromatin regions in the presence of TGFβ (Figure 4-46). These

results suggest that ID1 enables TGFβ-driven SMADs to activate a set of transcriptional targets necessary for the pro-apoptotic effect of TGFβ.

Of note, enforced expression of ID1 in SMAD4-restored mouse PDA cells left the majority of TGFβ gene responses intact (Figure 4-45) and did not prevent cells from undergoing an EMT. Gene set variation analysis of the RNA-seq profiles of the cells 24 h after treatment with TGFβ revealed similar enrichment for EMT gene signatures in the TGFβ-treated cells regardless of ID1 enforcement (Figure 4-47). Among specific EMT transcriptional effectors and regulated genes, there was similar up-regulation of *Snai1*, *Zeb1*, *Zeb2* and *Cdh2* (N-cadherin) and similar down-regulation of *Cdh1* (E-cadherin), *Klf5*, and *Krt19* (cytokeratin 19) in response to TGFβ in the two conditions (Figure 4-47). Under prolonged (72 h) TGFβ treatment, SMAD4-null PDA cells retained an epithelial morphology, whereas SMAD4-restored cells with enforced ID1 expression survived in the presence of TGFβ with a mesenchymal morphology (Figure 4-48). Further supporting the independence of high ID1 expression and the epithelial state, there is no correlation between ID1 and KLF5 transcript levels in TCGA PDA RNAseq datasets (Figure 4-49). These results suggest that TGFβ repression of ID1 is necessary for the apoptotic effects of TGFβ and that a sustained expression of ID1 decouples TGFβ-induced EMT from apoptosis (Figure 4-50).

Dysregulation of ID1 repression as a nodal point

The switch of the ID1 response to TGFβ from repression to activation provides a common node for tumor progression by PDAs in the presence of TGFβ. Given that the majority of PDAs contain progenitor cells that express high levels of

ID1, and yet the highly recurrent genetic alterations in PDA (i.e. KRAS, TP53, CDKN2A) are not mutually exclusive to TGFβ pathway alterations (refer to Figure S1A), low-frequency alterations in multiple factors may achieve the same common end of preventing the repression of ID1 by TGFβ. What are some examples of such mechanisms?

RNA polymerase II (RNAPII) ChIP-seq analysis in S4 (*Id1* repressing) and S4.1 (*Id1* inducing) cells showed that RNAPII is already present at the promoter of *Id1* under basal conditions (Figure 4-51). TGFβ-mediated repression of *Id1* may therefore be due to the recruitment of a SMAD co-repressor that facilitates RNAPII pausing and turnover (Adelman and Lis, 2012), and that loss of a SMAD co-repressor could result in release of paused RNAPII and transcriptional elongation. Indeed, TGFβ decreased RNAPII occupancy within the *Id1* coding region in apoptosis-sensitive cells but increased RNAPII occupancy in the *Id1* coding region in apoptosis-resistant cells (Figure 4-51).

To identify some potential SMAD co-repressors, a genome-wide CRISPR/Cas9 screen was performed to identify factors in which a loss of function would cause persistent expression of ID1 in the presence of TGFβ. SMAD4-restored cells containing the endogenous ID1-GFP reporter were transduced with the Gecko_v2 genome-wide library containing 123,411 sgRNAs (Joung et al., 2017; Sanjana et al., 2014) and treated with TGFβ or SB505124 for 36 h. To select cells that bypassed ID1 repression by TGFβ, ID1-GFP^{high} cells were selected by FACS, and this procedure was repeated three times to improve the stringency of the screen.

sgRNAs targeting TGFβ pathway components were enriched in the samples treated with TGFβ (Figure 4-52), providing a positive control for the screen.

In addition to enrichment of TGFβ pathway components, Biocarta pathway analysis identified the ATM/ATR/BRCA DNA damage response-related pathways (ATRBRCA & ATM) as being enriched in the TGFβ-treated samples and PI3K/AKT-related pathways (AKT, EGF, IGF1MTOR, IGF1R, INSULIN, PTEN, HER2, GLEEVEC, MTOR, NGF) as depleted in those samples (Figure 4-53). These results suggested that loss of DNA damage pathway components or activation of the PI3K/AKT pathway may protect cells from TGFβ-mediated repression of ID1. Further supporting this hypothesis, analysis of genetic events in the MSK-IMPACT dataset (the largest PDA cohort to date) (Zehir et al., 2017) for alterations occurring with mutual exclusivity tendencies to TGFβ pathway inactivation concordantly enriched and depleted for these same pathways (Figure 4-53).

PI3K/AKT regulation of ID1

In line with these results and the observation that AKT inhibition with MK2206 synergizes with TGFβ in the induction of a lethal EMT (Figure 4-40) (David et al., 2016), several PI3K/AKT activating events including insulin receptor (INSR) and AKT2 amplification and PIK3CA and PTEN mutation occur in a subset of PDA with intact TGFβ pathway (Figure 4-54). Furthermore, expression of AKT1(W80A) and AKT2(W80A), which are forms of AKT that are growth-factor regulated but resistant to MK2206 (Kajno et al., 2015), rendered SMAD4-restored cells resistant to both TGFβ-mediated repression of ID1 and TGFβ-mediated apoptosis (Figure 4-55, 4-56).

PI3K/AKT signaling has previously been demonstrated to up-regulate ID1 expression via nuclear exclusion of the transcription factor FOXO3a in K562 cells with a BCR-ABL fusion (Birkenkamp et al., 2007). By RNA-seq, mouse PDA cells express FOXO1 10x higher than FOXO3, and by ChIP-PCR, both FOXO1 and FOXO3a bind to the two 5' regions (Region 1 and Region 2) of accessible chromatin at the mouse Id1 locus (Figure 4-57). Supportive of the hypothesis that these two regions are necessary for the negative regulation of ID1 by TGFβ, Regions 1 and 2 were enriched in a screen using the sgRNA library tiling the seven conserved regions at the *Id1* locus in mouse PDA cells and comparing the Id1GFP^{high} populations of TGFβ and SB505124-treated SMAD4-restored cells, (Figure 4-58). Furthermore, the highest scoring guide in this screen, which was validated by transfection of the individual sgRNA and measuring ID1GFP expression in response to TGFβ (Figure 4-58), overlaps with a FOXO binding motif. These results are consistent with the observation that the SMAD4-restored cells that repress ID1 lose RNAPII from the ID1 promoter whereas the ID1-inducing S4.1 derivatives exhibit dysregulation of RNAPII at the promoter (Figure 4-51).

Furthermore, using FOXO1 exclusion from the nucleus as a marker of AKT signaling, analysis of mouse PDA tissues demonstrated that AKT signaling is higher in SMAD4-wildtype than SMAD4-null PDA tissues (Figure 4-59). In fact, when cells are isolated from *Kras^{G12D}; Cdkn2a$^{-/-}$* PDAs, some are sensitive to TGFβ-induced apoptosis in the presence of low serum (Bardeesy et al, 2006) or AKT inhibition (Figure 4-60). This suggested that PI3K/AKT activation in PDA provides one mechanism to prevent TGFβ-induced ID1 repression and apoptosis.

Additional regulators of ID1

The unexpected enrichment of DNA damage response factors in both the screen for regulators of TGFβ-mediated repression of *ID1* and the MSK-IMPACT mutual exclusivity analysis led to a query of whether these factors provide another mechanism for transcriptional regulation of *ID1* which is altered in PDAs. The DNA damage repair genes enriched in at least one of the two analyses included ATM, BRCA1, ERCC4, CHEK2, RAD51C, CHEK1, and RECQL. All are known or proposed genetic cancer risk factors, with the first five reported specifically in pancreatic cancer (Bartsch et al., 2006; Cybulski et al., 2015; Li et al., 2006; McWilliams et al., 2008; Roberts et al., 2012). Of note, in addition to their roles in the DNA damage response, these proteins are implicated in transcriptional regulation (Hatchi et al., 2015; Shivji et al., 2018; Tresini et al., 2015; Zhu et al., 2011).

Given the presence of BRCA1 mutations in PDA (Figure 4-54) and its known roles in transcriptional regulation, the hypothesis emerged that BRCA1 may cooperate with SMADs in *Id1* regulation. At the Id1 locus, three regions of SMAD2/3 binding are detected (Figure 4-61), including a region 1 kb upstream of the promoter previously shown to be important for both the down-regulation of Id1 by TGFβ and up-regulation by the TGFβ family members bone morphogenetic protein 2 (BMP2) and BMP4 (Kang et al., 2003; Korchynskyi and ten Dijke, 2002; Lopez-Rovira et al.). BRCA1, which is found at the promoters of certain genes in ENCODE ChIP-seq datasets, binds this 1 kb upstream region of ID1 in various cell types, including the SMAD4-restored mouse PDA cells (Figure 4-62). Furthermore, BRCA1 depletion by shRNA prevented *Id1* repression by TGFβ with no effect on *Smad7*

regulation (Figure 4-63), and implantation of the BRCA1-depleted cells in mice with caerulein-induced acute pancreatitis demonstrated that these cells were more resistant to the pro-apoptotic effects of a TGFβ-rich microenvironment (Figure 4-64). Together, these results suggest that BRCA1 cooperates with TGFβ-activated SMADs in mediating ID1 repression, and that loss of BRCA1 is one alternative mechanism to SMAD4 loss or PI3K/AKT activation for the dysregulation of ID1 repression to evade tumor suppression.

Discussion

A key finding of this work is that transcriptional dysregulation of ID1 is a selected, common feature of PDAs, which imparts both stemness characteristics to PDA cells and protects them from TGFβ-induced apoptosis (Figure 4-65). Suppression of ID1 expression is a necessary component of TGFβ-induced lethal EMT in PDA (David et al., 2016). These findings implicate E-protein-bHLH complexes in a pro-apoptotic subset of TGFβ gene responses in premalignant pancreatic cells. ID1 is expressed in pancreatic development (Loh et al., 2014; Zhou et al., 2007) and remains primed for expression in the adult pancreas. Several other bHLH factors necessary for the specification and maintenance of mature pancreatic cell types, including MIST1 and PTF1A, have tumor suppressive roles in the pancreas (Krah et al., 2015; Shi et al., 2009), suggesting that there is a critical balance between the inhibitory and activating bHLH factors in pancreatic homeostasis.

Uncoupling lethal EMT

That enforced ID1 expression can uncouple EMT from the apoptotic responses to TGFβ, and that EMT can be a selected feature of cells that survive TGFβ-mediated apoptosis with high ID1, suggest that the association of stemness with EMT features may be linked to the selection for ID1 dysregulation. The observation here that ID1 prevents the induction of a subset of TGFβ genes without affecting the EMT initiated by TGFβ differs from a previous study in breast cancer metastasis where it was found that ID1 opposes the bHLH EMT factor TWIST1 to promote an epithelial state (Stankic et al., 2013). Interestingly, the previous report describes that TGFβ can only induce ID1 in cells that have previously undergone an EMT, which is consistent with the current observation that ID1 induction is a selected feature of the TGFβ response. In addition to its role in tumor suppression, TGFβ can play a variety of tumor-promoting roles, including evasion of immunity, local invasion, and metastasis (David and Massague, 2018b). The ability to uncouple these effects provides a framework for targeting specific features of this pathway to mitigate the complications of direct therapeutic inhibition of the TGFβ pathway (Smith et al., 2012).

Rare genetic alterations in PDAs and TGFβ tumor suppression

Although common PI3K/AKT activating mutations are relatively rare in PDA, the observation that *INSR* and *AKT2* are amplified in PDA and that AKT signaling opposes lethal EMT has both biological and clinical implications. AKT2 knockout mice have particular defects in insulin-mediated glucose metabolism compared to AKT1 knockout mice, which are stunted in growth but have normal glucose

metabolism (Cho et al., 2001a; Cho et al., 2001b; Pin et al., 2001). Insulin signaling is particularly relevant in the pancreas, where blood flows through the islets of Langerhans, where insulin is synthesized, to the capillaries supplying the pancreatic acini and ducts, where it regulates key functions in these cells, including the excretion of pancreatic enzymes (Williams and Goldfine, 1985). Furthermore, several studies have reported a correlation between insulin resistance, which results in high levels of circulating insulin, with increased incidence of PDA (Wang et al., 2003). Therefore, while *AKT2* amplification may not be sufficient to activate the pathway, its presence in a microenvironment rich in insulin suggests that this may be one way in which PI3K/AKT signaling is amplified in the pancreas and may contribute to the observation that receptor tyrosine kinase activation is necessary for PDA formation even in the presence of mutant KRAS in the majority of PDAs (Ardito et al., 2012; Navas et al., 2012).

The identification of several genetic alterations that contribute to dysregulation of TGFβ-mediated repression of ID1 suggests that multiple factors participate in this repression. Of interest, DNA damage proteins, including BRCA1, ATM, and ERCC4 have recently been implicated in transcriptional regulation via the formation and resolution of R-loops and regulation of heterochromatin (Grunseich et al., 2018; Hatchi et al., 2015; Zhang et al., 2017; Zhu et al., 2011). One possibility is that these proteins play a role in assembly of a repressive complex at the -1kb region upstream of the ID1 promoter, which is brought into contact with the promoter by TGFβ-driven SMADs to repress ID1 in premalignant pancreatic cells. Loss of components of this repressive complex would therefore contribute to

altered TGFβ-mediated regulation of ID1. The tail of low-frequency genetic alterations that are exclusive of SMAD4 mutations in genome-wide studies of PDA may contain additional mediators of ID1 dysregulation besides AKT2 and BRCA1. Of interest, the observations here may be relevant to a subset of epithelial tumors beyond PDA—the triple negative subtype of breast cancer, for example, is associated with both ID1 expression and BRCA1 mutation (Atchley et al., 2008; Gupta et al., 2007).

Commonalities of PDAs

Whereas recent studies of the PDA transcriptome have focused on subtyping pancreatic adenocarcinomas (Bailey et al., 2016; Collisson et al., 2011; Moffitt et al., 2015; TCGA, 2017), the analysis focused on highly expressed transcriptional regulators yields commonalities in PDA. Using a rank-based, transcription factor-centered algorithm internally normalizes datasets and makes cross-study comparisons possible. The focus on factors that are highly and differentially expressed enriches for factors key to integrating microenvironmental signals and genetic alterations to orchestrate cell fates and states that may represent lineage dependencies and constitute promising therapeutic targets.

Altogether, the results surrounding ID1 in PDA point at a critical role for ID1 transcriptional activation and maintenance in PDA development and identify ID1 as a therapeutic target in PDA. Despite the historical difficulty of targeting transcription factors, small molecule inhibitors of ID proteins have been developed (Chaudhary et al., 2009). Based on the dependence of PDA cells on ID1, PDAs would be predicted to be particularly sensitive to these inhibitors. PDAs of different

subtypes possess common transcriptional dependencies, such as on ID1, and therefore an effective treatment strategy may be to simultaneously target subtype-specific characteristics (such as BRCA-mutant tumors with PARP inhibitors or AKT-activated tumors with AKT inhibitors) along with these core, shared dependencies.

DISCUSSION

Switching the role of TGFβ in tumors

The mechanism for TGFβ-mediated tumor suppression delineated in these studies of pancreatic adenocarcinoma is as follows: SMAD4-dependent down-regulation of transcription factors key in pancreatic adenocarcinoma, including KLF5 and ID1, perturbs a pro-tumorigenic transcriptional network; in combination with SMAD4-independent up-regulation of SOX4, this results in apoptosis. The presence of multiple TGFβ-regulated factors that are necessary for this tumor suppression mechanism suggests that these factors operate in balance with one another and that perturbation of each shifts the balance towards either cell death or survival. Enforcement of SOX4, which appears to reinforce an endodermal progenitor identity of PDA cells, presents an opportunity for pancreatic cells to move towards a more tumorigenic state while presenting a liability towards tumor suppression. SNAIL-mediated EMT represses a set of endodermal TFs that would normally partner with SOX4 to establish this progenitor state, and repression of these endodermal factors, most notably KLF5, switches the activity of SOX4 from pro-tumorigenic in the presence of cooperating endodermal TFs to pro-apoptotic in their absence. In parallel, SMAD-mediated repression of the ID proteins frees E-proteins towards promotion of the pro-apoptotic state (Figure 4-65).

The simultaneous high expression of both KLF5 and ID1 in human PDAs, each of which individually can protect from TGFβ-mediated apoptosis suggests either safety in a double protection from TGFβ tumor suppression or gene co-regulation (or both). Given that KLF5 is tied to the epithelial state of PDA cells, it is

possible that such a configuration of redundant protection allows cells to switch between epithelial and mesenchymal states without the liability of undergoing lethal EMT.

TGFβ induces many gene responses in any given cell type, either as direct SMAD-mediated transcriptional effects or indirectly through SMAD effects on the expression of other TFs. The observation that simultaneous regulation of multiple of these genes is necessary in the process of tumor suppression is both intriguing from an evolutionary perspective of how the co-regulation of these genetic loci developed as well as important in understanding the consequences of a single signal that potently rewires the gene expression of epithelial cells in their progression towards cancer. The ability to alter single branches of the pathway and observe different phenotypes--for example, perturbation of KLF5 suppression resulting in epithelial cells with tumorigenic potential versus perturbation of ID1 suppression resulting in cells with mesenchymal and tumorigenic potential--suggests that these branches operate as modules that are interconnected but often separable.

Transcriptional configurations in pancreatic cancer

The observation that SOX4-, KLF5-, and ID1-high cells represent a unique population of cells that are enriched in the premalignant and malignant tissues of the pancreas suggests a unique state of cells that are both more likely to form tumors and susceptible to tumor suppressive signals. It may be that pediatric tumors, for example, which are more prone to result from a single genetic alteration, arise from cells that are already in such a progenitor-like, undecided state, whereas

adult epithelial tumors require the presence of a gateway mutation and signals from an inflammatory microenvironment to achieve such a state.

The dominant transcriptional network including both KLF5 and ID1 shared by many PDAs reflects a confluence of the spatiotemporally local signals from the PDA microenvironment and the developmental history of the PDA--its cell of origin, its exposure to tumor suppressive signals like TGFβ, and the genetic alterations that accumulate as a result. It is interesting to note that several studies report that transcriptional networks depend highly on the developmental origin of the tumor (Hoadley et al., 2018), yet the most highly expressed transcription factors in PDA are different from those expressed in the abundant acinar cells of the pancreas. Several studies have shown acinar cells to be the most susceptible to transformation to PDA using cell-type specific Cre recombinase in genetic mouse models (Habbe et al., 2008; Kopp et al., 2012; Yamaguchi et al., 2018). Intriguingly, SOX9, one of the few endodermal transcription factors induced by TGFβ, is expressed by normal ductal cells, and has been shown to synergize with KRAS activation in acinar cells to promote acinar-to-ductal metaplasia and subsequent transformation to PDA (Kopp et al., 2012). Whether this suggests that PDAs originate from a reprogramming event early in pancreatic tumorigenesis or the expansion of a rare population of cells primed with this particular transcriptional network is unknown.

Meanwhile, one possible clinical utility of profiling the dominant network of a particular tumor, beyond identifying drug targets, may be to predict outcome. The PCA analysis of PDAs and PNETs suggests that pancreatic neuroendocrine tumors differ in the major transcription factors expressed, and these diseases have

drastically different outcomes. The five-year survival of PDA is 1-14% depending on stage, whereas the five-year survival of PNET is 16-61% (SEER, 2000). However, undifferentiated PNETs, which are associated with poor prognosis, have been described to express high levels of ID1 (Hunter et al., 2013). These tumors may result from a different cell of origin than the differentiated variety or arrive at this more aggressive transcriptional configuration via a unique developmental trajectory. In either case, the suggestion is that an $ID1^{high}$ configuration may be a poor prognostic factor.

Unresolved questions

The investigation of TGFβ as a tumor suppressor and the commonalities of the tumors that develop in its presence has highlighted many interesting and important aspects of TGFβ signaling and raised additional questions. The unexpected finding that SMAD4 is not required for all gene responses to TGFβ, for example, highlights the limited knowledge about the unique function of SMAD4 in comparison to the receptor SMADs and raises the question of how the requirement for various SMADs is determined. This encompasses both the differentiation of SMAD2/3 gene responses vs SMAD2/3/4 responses, as well as the those of the BMP SMADs (1/5/8) vs the TGFβ SMADs (2/3), all of which were recently demonstrated to bind GC-rich regions, and which sometimes have the capacity to bind the exact same GC-rich motifs, as in the case of ID1. Some possibilities include the relative availability of each factor at a given spatio-temporal region, unique cofactors at specific regulatory loci, and the genomic structure of a particular chromatin region. While the signal transduction from membrane to nucleus of TGFβ signaling has been

relatively well-studied, the mechanisms of transcriptional activation and repression and the interaction of the SMADs with the core transcriptional machinery are less well-described.

From the perspective of understanding the genetics of pancreatic cancer, the concept that proteins like SOX4 can play dual roles in tumor progression highlights the potential of genome-focused analyses in missing important players in cancer progression. Investigating the mechanism by which TGFβ signaling is perturbed by alterations in PI3K/AKT, DNA damage proteins, and additional candidate genes identified in the PDA genetic analysis and ID1-repressive screen may therefore yield additional important insights about tumor progression. Similarly, further study of the mechanisms by which the E-proteins and SOX4 promote apoptosis in response to TGFβ could provide insight into the vulnerabilities of PDA cells. Finally, the role of SOX4/KLF5/ID1 in a subpopulation of functionally important cells in PDA beyond their role in TGFβ tumor suppression and escape from TGFβ is intriguing. Does this population pre-exist in the pancreas? What properties do cells gain from the high expression of these factors? How are these properties conserved and what are the equivalents in other tumor types? And how do tumor suppressive signals, TGFβ and otherwise, interact with these important factors?

Table 2- 1. Resources & Reagents

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit mAb Anti-Phospho-Smad2 (Ser465/467) (138D4)	CST	Cat #3108
Rabbit mAb Anti-Smad2/3 (D7G7)	CST	Cat #8585
Rabbit mAb Anti-Mouse/Human Id1	Biocheck	Cat #BCH-1/195-14
Rabbit mAb Anti-Mouse Id1 (BCH-1/37-2)	Biocheck	Cat #BCH-1/37-2
Rabbit mAb Anti-E-Cadherin (24E10)	CST	Cat #3195
Rabbit pAb Anti-Klf5 (ab137676)	Abcam	Cat #ab137676
Rabbit mAb Anti-Smad4 (ab40759)	Abcam	Cat #ab40759
Rabbit mAb Anti- Phospho-Akt (Thr308) (D25E6)	CST	Cat #13038
Rabbit mAb Anti- Phospho-Akt (Ser473) (D9E)	CST	Cat #4060
Mouse mAb Anti-Akt (pan) (40D4) Mouse	CST	Cat #2920
Rabbit mAb Anti- Phospho-Smad1/5 (Ser463/465) (41D10)	CST	Cat #9516
Mouse mAb Anti-N-Cadherin (13A9)	CST	Cat #14215
Rabbit mAb Anti-Vimentin (D21H3)	CST	Cat #5741
Mouse mAb Anti-α-Tubulin (DM1A)	Sigma	Cat #T6199
Mouse mAb Anti-Anti-RNA polymerase II CTD repeat YSPTSPS (8WG16)	Abcam	Cat #ab817
Rabbit pAb Anti-Sox4 (EVI16)	Diagenode	Cat # C15310129
Rat Anti-Ck19	DSHB	Cat #TROMAIII
Mouse Anti-Flag M2	Millipore	Cat #F1804
Rabbit Anti-Foxo1	CST	Cat #2880
Rabbit Anti-Foxo3a	Abcam	Cat #ab12162
Biological Samples		
Human PDA paraffin blocks	MSKCC CPCR	https://www.mskcc.org
Chemicals, Peptides, and Recombinant Proteins		
MK2206	ChemieTek	Cat #CT-MK2206
SB505124	Sigma	Cat #S4696
TGFβ	R&D	Cat #101-B1
Doxycycline Hydrochloride	Fisher	BP2653-5
Critical Commercial Assays		
Caspase-Glo 3/7	Promega	Cat #G8093
CellTiter-Glo	Promega	Cat #G7572

Experimental Models: Cell Lines		
H. sapiens: hPDA organoids	MSKCC CPCR	https://www.mskcc.org/
H. sapiens: MiaPaca2	ATCC	CRL-1420
H. sapiens: Panc1	ATCC	CRL-1469
H. sapiens: BxPC3	ATCC	CRL-1687
M. musculus: Cell line 806: $Kras^{G12D}$;$^{Cdkn2a-/-}$;$Smad4^{-/-}$	Bardeesy et al 2006	806
M. musculus: Cell line NB44: $Kras^{G12D}$;$Cdkn2a^{-/-}$	Bardeesy et al 2006	NB44
M. musculus: Cell line 4279: $Kras^{G12D}$;$Cdkn2a^{-/-}$	David et al, 2016	4279
M. musculus: Organoid 281: $Kras^{G12D}$;$Cdkn2a^{-/-}$	David et al, 2016	281
M. musculus: Organoid 1896: $Kras^{G12D}$;$Cdkn2a^{+/-}$;$Smad4^{+/-}$		1896
M. musculus: Organoid line T7: $Kras^{G12D}$; $Trp53^{R172H}$	Boj et al 2015	N/A
Experimental Models: Organisms/Strains		
M. musculus: Hsd:Athymic Nude-Foxn1nu	ENVIGO	Cat #069
M. musculus: LSL-$Kras^{G12D}$; $Cdkn2a^{fl/fl}$; $Smad4^{fl/fl}$	Bardeesy et al 2006	N/A
M. musculus: LSL-$Kras^{G12D}$; $Cdkn2a^{fl/fl}$	Bardeesy et al 2006	N/A
M. musculus: FVB/NJ	Jax	Cat #001800
Recombinant DNA		
pLVX-IRES-Hyg	Clontech	Cat #632185
pLVX-Tight-Puro	Clontech	Cat #632163
pLVX-IRES Hyg SMAD4	David et al, 2016	Addgene #107128
pLVX-Tight-Puro FLAG-SMAD4 TetON	David et al, 2016	Addgene #107129
pBABEpuro HA-BRCA1	Cortez et al, 1999	Addgene #14999
pLVX-IRES-tdTomato-FlagAkt1	Kajno et al, 2015	Addgene #64831
pLVX-IRES-tdTomato-FlagAkt1W80A	Kajno et al, 2015	Addgene #64833
pLVX-IRES-tdTomato-FlagAkt2W80A	Kajno et al, 2015	Addgene #64834
pLVX-IRES-tdTomato-FlagAkt2	Kajno et al, 2015	Addgene #64832
Mouse CRISPR Knockout Pooled Library (GeCKO v2)	Sanjana et al, 2014	Addgene #1000000053
pSpCas9(BB)-2A-Puro (PX459)	Ran et al, 2013	Addgene #48139
LT3GEP	Fellmann et al, 2013	N/A
pUC19	Norrander et al, 1983	Addgene ##50005

Deposited Data		
GTEx	Consortium, 2013	https://www.gtexportal.org/home/
ICGC	Bailey et al, 2016	https://dcc.icgc.org/
TCGA	Consortium, 2017	https://gdac.broadinstitute.org/
Pancreatic microarray data	Moffitt et al, 2015	GSE71729
Roadmap Epigenomics	Bernstein et al, 2010	http://www.roadmapepigenomics.org/
MSK-IMPACT	Zehir et al, 2017	http://www.cbioportal.org/
PDA ChIPseq	Diaferia et al, 2016	GSE64557
UCSC Genome Browser	Kent et al, 2002	http://genome.ucsc.edu/
ENCODE (SYDH, HAIB)	Consortium, 2012	https://www.encodeproject.org/
Smad2/3 ChIPseq, shSox4 RNAseq	David et al, 2016	GSE72069
RNAseq		GSE112940
ATACseq		GSE112940
ChIPseq		GSE112940
Software and Algorithms		
FastQC v0.11.5	Andrews, 2010	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
GNUparallel v2.5.2b	Tange et al, 2011	https://www.gnu.org/software/parallel/
STAR v0.6.1p1	Dobin et al, 2013	https://github.com/alexdobin/STAR
HTSeq v3.4	Anders & Huber, 2010	https://htseq.readthedocs.io/en/release_0.9.1/
DESeq2 v3.4	Love et al, 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
GSVA	Hanzelmann et al, 2013	https://bioconductor.org/packages/release/bioc/html/GSVA.html
Bowtie2	Langmead & Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Sam tools	Li et al, 2009	http://samtools.sourceforge.net/
RIGER	Luo et al, 2008	https://software.broadinstitute.org/GENE-E/extensions.html
Camera	Wu & Smyth, 2012	https://bioconductor.org/packages/release/bioc/html/limma.html
GSEA	Subramanian et al, 2005	http://software.broadinstitute.org/gsea/index.jsp

Table 2- 2. mirE-based shRNA sequences

Snai1.1197	TGCTGTTGACAGTGAGCGCGAGGTACAACAGACTATGCAATAGTGAAGC CACAGATGTATTGCATAGTCTGTTGTACCTCATGCCTACTGCCTCGGA
Sox4.965	TGCTGTTGACAGTGAGCGCCAGCGACAAGATTCCGTTCATTAGTGAAGC CACAGATGTAATGAACGGAATCTTGTCGCTGTTGCCTACTGCCTCGGA
Sox4.2137	TGCTGTTGACAGTGAGCGCTAGATGGAGAGTAGAAGGAGATAGTGAAG CCACAGATGTATCTCCTTCTACTCTCCATCTATTGCCTACTGCCTCGGA
Sox4.2091	TGCTGTTGACAGTGAGCGCAAAGAAGAAGAAAGAAAGAAATAGTGAAG CCACAGATGTATTTCTTTCTTTCTTCTTCTTTTTGCCTACTGCCTCGGA
Renilla.713	TGCTGTTGACAGTGAGCGCAGGAATTATAATGCTTATCTATAGTGAAGC CACAGATGTATAGATAAGCATTATAATTCCTATGCCTACTGCCTCGGA
Klf5.1343	TGCTGTTGACAGTGAGCGCCAGATACAACAGAAGGAGTAATAGTGAAGC CACAGATGTATTACTCCTTCTGTTGTATCTGATGCCTACTGCCTCGGA
Klf5.636	TGCTGTTGACAGTGAGCGACAGTTCTTCACTGACACTGAATAGTGAAGC CACAGATGTATTCAGTGTCAGTGAAGAACTGGTGCCTACTGCCTCGGA
Klf5.1704	TGCTGTTGACAGTGAGCGCACCGCAGACCTAACTTCATAATAGTGAAGC CACAGATGTATTATGAAGTTAGGTCTGCGGTTTGCCTACTGCCTCGGA
Foxa2.a649	TGCTGTTGACAGTGAGCGACACAGTGATCTGTCATTCTAATAGTGAAGC CACAGATGTATTAGAATGACAGATCACTGTGGTGCCTACTGCCTCGGA
Brca1.3153	TGCTGTTGACAGTGAGCGAAGGGACGATTTGAGAGACATATAGTGAAG CCACAGATGTATATGTCTCTCAAATCGTCCCTCTGCCTACTGCCTCGGA
Brca1.1324	TGCTGTTGACAGTGAGCGCTAGCAGCGTTCAGAAAGTTAATAGTGAAGC CACAGATGTATTAACTTTCTGAACGCTGCTATTGCCTACTGCCTCGGA
Rxra.5178	TGCTGTTGACAGTGAGCGCCAGATACGTGATGTCATCTAATAGTGAAGC CACAGATGTATTAGATGACATCACGTATCTGATGCCTACTGCCTCGGA
Rxra.2108	TGCTGTTGACAGTGAGCGATGGGTCATAGCTAACCTATAATAGTGAAGC CACAGATGTATTATAGGTTAGCTATGACCCAGTGCCTACTGCCTCGGA
Id1.840	TGCTGTTGACAGTGAGCGATGAAAATATTGTTTTACAATATAGTGAAGC CACAGATGTATATTGTAAAACAATATTTTCAGTGCCTACTGCCTCGGA
Id3.812	TGCTGTTGACAGTGAGCGACCCTGATTATGAACTCTATAATAGTGAAGC CACAGATGTATTATAGAGTTCATAATCAGGGCTGCCTACTGCCTCGGA
Id2.91	TGCTGTTGACAGTGAGCGCCCCGATGAGTCTGCTCTACAATAGTGAAGC CACAGATGTATTGTAGAGCAGACTCATCGGGTTGCCTACTGCCTCGGA
Sox4.965	TGCTGTTGACAGTGAGCGCCAGCGACAAGATTCCGTTCATTAGTGAAGC CACAGATGTAATGAACGGAATCTTGTCGCTGTTGCCTACTGCCTCGGA
Sox4.2509	TGCTGTTGACAGTGAGCGACCTTGGTGATTTCTTGTTTGATAGTGAAGC CACAGATGTATCAAACAAGAAATCACCAAGGCTGCCTACTGCCTCGGA

Table 2- 3. sgRNAs

	Top_oligo	Bottom_oligo
Id1_region5_sg1	caccgaggaccaggcgctactttcc	caccgctggctgccgtagtgaaagg
Id1_region5_sg2	caccgttccgatcagaagtggcact	caccggagggttgcccaatgccca
Id1_region5_sg3	caccgcacatagctgggttaccgag	caccgaccacaactgccaaccaggg
Id1_c-term	caccggatcgcatcttgtgtcgctg	aaaccagcgacacaagatgcgatcc

Table 2- 4. Primers

	Forward Primer	Reverse Primer	Purpose
hSMAD7	ttcctccgctgaaacaggg	cctcccagtatgccaccac	qPCR
hGAPDH	ctgggctacactgagcacc	aagtggtcgttgagggcaatg	qPCR
hID1	ctgctctacgacatgaacgg	gaaggtccctgatgtagtcgat	qPCR
mId1_1	cctagctgttcgctgaaggc	ctccgacagaccaagtaccac	qPCR
mId1_4	gcgaggtggtacttggtctg	aggatctccaccttgctcac	qPCR
mSmad7	ggccggatctcaggcattc	ttgggtatctggagtaaggagg	qPCR
mRxra	cacacccacattgggcttc	gaggccatatttcctgaggga	qPCR
mBrca1	cgaatctgagtcccctaaagagc	aagcaacttgaccttggggta	qPCR
mGapdh	aggtcggtgtgaacggatttg	tgtagaccatgtagttgaggtca	qPCR
Id1_HAL	atatatgaattcaataagtagagatcacagcc	atatatggtaccgcgacacaagatgcgatcgtc	cloning
Id1_HAR	atatatggatccggcggcgcactgagggaccag	atatattctagactttagaacccatctgtgcc	cloning
eGFP	atatatggtaccgtgagcaagggcgaggagctgttca	atatatggatccctacttgtacagctcgtccatgccgaga	cloning
Id1	gatagggatccatgaggtcgccagtggcagtg	aagctgaattctcagcgacacaagatgcgatcgtc	cloning

Table 2- 5. Locus screen sgRNAs

Name	Top_oligo	Bottom_oligo
1__17	caccgggggagactcggcaggtgt	aaacacacctgccgagtctccccc
1__11	caccgcggggaggtaagttgaccct	aaacagggtcaacttacctccccgc
1__22	caccgtgggtgcggttgctgaccaa	aaacttggtcagcaaccgcacccac
1__4	caccggtcagcaaccgcacccacg	aaaccgtgggtgcggttgctgacc
1__39	caccgaatgctttcctcgtgggtg	aaaccacccacgaggaaagcattc
1__32	caccgcccacgaggaaagcattccc	aaacgggaatgctttcctcgtgggc
1__12	caccgagacttagggcgcgaagcct	aaacaggcttcgcgccctaagtctc
1__19	caccgcttcgcgccctaagtctgc	aaacgcagacttagggcgcgaagc
1__27	caccgcgccctaagtctgcaggtga	aaactcacctgcagacttagggcgc
1__37	caccgtctgcaggtgacgggctca	aaactgagcccgtcacctgcagac
1__56	caccggtgacgggctcaggggcgg	aaacccgcccctgagcccgtcacc
1__47	caccgatggagacggcaaactgggc	aaacgcccagtttgccgtctccatc
1__23	caccgcgggcggtcgccatggaga	aaactctccatggcgaccgcccgc
1__8	caccgctccatggcgaccgcccgcg	aaaccgcgggcggtcgccatggagc
1__28	caccgtcaggctggcgccgcgcggg	aaaccccgcgcggcgccagcctgac
1__35	caccgcgccagcctgacagcccgtc	aaacgacgggctgtcaggctggcgc
1__6	caccgtaaaacccggacgggctgtc	aaacgacagcccgtccgggttttac
1__3	caccgcccgtccgggttttatgaa	aaacttcataaaacccggacgggc
1__25	caccgcgtcacccattcataaaacc	aaacggttttatgaatgggtgacgc
1__43	caccgtttatgaatgggtgacgtca	aaactgacgtcacccattcataaac
1__26	caccgaatgggtgacgtcacgggcc	aaacggcccgtgacgtcacccattc
1__10	caccgtcacgggcctggcgtctaa	aaacttagacgccaggcccgtgac
1__15	caccggctcagaccgttagacgcc	aaacggcgtctaacggtctgagcc
1__30	caccgtgtgtcagcgtctgaacaag	aaaccttgttcagacgctgacacac
1__44	caccgacgctgacacagaccagcc	aaacggctggtctgtgtcagcgtc
1__45	caccgacacagaccagcccgggaa	aaacttcccgggctggtctgtgtc
1__31	caccgtctccgcagctgctgcgccg	aaaccggcgcagcagctgcggagac
1__46	caccgttgctcggaggccttcaaga	aaactcttgaaggcctccgagcaac
2__8	caccgcataggtagagcagctagt	aaacactagctgctctacctatgc
2__40	caccgggctcaagaactgaaaggg	aaaccccttttcagttcttgagccc
2__29	caccgtcacctctccatcagcctaa	aaacttaggctgatggagaggtgac
2__16	caccgctagtacacccttaggctga	aaactcagcctaagggtgtactagc
2__2	caccggatctactagtacaccctt	aaacaagggtgtactagtagatcc
2__5	caccgggtgtactagtagatccag	aaacctggatctactagtacaccc
2__18	caccgcgtggctttgagcgcctct	aaacagaggcgctcaaagccacgc
2__19	caccgccaataggctgctagaggcg	aaaccgcctctagcagcctattggc
2__24	caccgaacgttcaaaagcaaccaat	aaacattggttgcttttgaacgttc
2__3	caccgaacgttctgaacccgccct	aaacagggcgggttcagaacgttc
2__20	caccgtataagggcagaccaagggc	aaacgcccttggtctgcccttatac

2__25	caccggtctgcccttataaaagac	aaacgtcttttataagggcagacc
2__32	caccgctggagccagtcttttata	aaactataaaagactggctccagc
2__42	caccgagaaaggttgtacaatgaga	aaactctcattgtacaacctttctc
2__35	caccgctcccctccgcctgttctc	aaacgagaacaggcggaggggagc
2__26	caccgatgatcctgagaacaggcgg	aaacccgcctgttctcaggatcatc
2__27	caccgaggatcatgaaggtcgccag	aaacctggcgaccttcatgatcctc
2__21	caccgcgaacagctagggcctgcag	aaacctgcaggccctagctgttcgc
2__4	caccgaggccctagctgttcgctga	aaactcagcgaacagctagggcctc
2__6	caccgctgttcgctgaaggcgggc	aaacgcccgccttcagcgaacagc
2__12	caccgacagcgggcgaggtggtact	aaacagtaccacctcgcccgctgtc
2__13	caccgcgaggtggtacttggtctgt	aaacacagaccaagtaccacctcgc
2__14	caccgtggtctgtcggagcaaagcg	aaaccgctttgctccgacagaccac
2__7	caccgtggccatctcgcgctgcgc	aaacgcgcagcgcgagatggccac
3__11	caccgcccaggtgtacggaggcgcc	aaacggcgcctccgtacacctgggc
3__8	caccgtacggaggcgcctggctgtc	aaacgacagccaggcgcctccgtac
3__12	caccgtcggcgggcccagacagcc	aaacggctgtctgggcccgccgac
3__10	caccgctgtctgggcccgccgactt	aaacaagtcggcgggcccagacagc
3__9	caccgtgggcccgccgactttggca	aaactgccaaagtcggcgggcccac
3__5	caccgcggccgtgccaaagtcggc	aaacgccgactttggcacggccgc
3__3	caccgcggccgctgtttgtctagcg	aaaccgctagacaaacagcggccgc
3__26	caccgcagctgcgggcggaggggca	aaactgcccctccgcccgcagctgc
3__23	caccgccgcagctgcagcccggcct	aaacaggccgggctgcagctgcggc
3__14	caccgcgcccgccccgttcccaggc	aaacgcctgggaacggggcgggcgc
3__17	caccgccgacgcccgccccgttccc	aaacgggaacggggcgggcgtcggc
3__4	caccgcgtcggcagcttgcataag	aaaccttatgcaagctgccgacgc
3__15	caccgagcttgcataagaggcacca	aaactggtgcctcttatgcaagctc
4__10	caccgaaccctaaaccaagggacag	aaacctgtcccttggtttagggttc
4__4	caccgtaggcatacacacacgttgg	aaacccaacgtgtgtgtatgcctac
4__16	caccgatttccaaaagaggagggtt	aaacaaccctcctcttttggaaatc
4__2	caccgtagcagcaccaagtctgtct	aaacagacagacttggtgctgctac
4__1	caccgcgcccaaggaacttgtcctg	aaaccaggacaagttccttgggcgc
4__3	caccgaggacaagttccttgggcgc	aaacgcgcccaaggaacttgtcctc
5__8	caccgaccacaactgccaaccaggg	aaacccctggttggcagttgtggtc
5__7	caccgggcattgggcaaccctccc	aaacgggagggttgcccaatgccc
5__4	caccggagggttgcccaatgccca	aaactgggcattgggcaaccctcc
5__2	caccgcacatagctgggttaccgag	aaacctcggtaacccagctatgtgc
5__5	caccgtagctgggttaccgagtggt	aaacaccactcggtaacccagctac
5__3	caccgctggctgccgtagtgaaagg	aaaccctttcactacggcagccagc
5__24	caccgcacagggccaggcaggagtt	aaacaactcctgcctggccctgtgc
5__1	caccgaggaccaggcgctactttcc	aaacggaaagtagcgcctggtcctc
5__6	caccgttttccaggaaagtagcgcc	aaacggcgctactttcctggaaaac

6__7	caccgtgggggcagctctacgaaga	aaactcttcgtagagctgcccccac
6__8	caccgcgtagagctgcccccaagag	aaacctcttgggggcagctctacgc
6__10	caccgcccccaagagaggtgcccc	aaacggggcacctctcttgggggc
6__11	caccgaggtgccccaggcccttaca	aaactgtaagggcctggggcacctc
6__12	caccgtgcaaccctgtaagggcctg	aaaccaggcccttacagggttgcac
6__4	caccgcattgtgcaaggagcggtt	aaacaaccgctccttgcacaatgc
6__30	caccgctgtgggccaagagggaagc	aaacgcttccctcttggcccacagc
6__6	caccgtccgatcagaagtggcactg	aaaccagtgccacttctgatcggac
7__4	caccgtgtccactcgcgccccctct	aaacagagggggcgcgagtggacac
7__8	caccgcgccccctcttggaagaac	aaacgttcttccaagagggggcgc
7__9	caccgaactggaactgcgcccgca	aaactgcgggcgcagttccagttc
7__2	caccgcagggtttcgcgcccatgc	aaacgcatgggcgcgaaaccctgc
7__3	caccgatgggcgcgaaaccctgcct	aaacaggcagggtttcgcgcccatc
7__16	caccgttcaggctgtctctaactgc	aaacgcagttagagacagcctgaac
7__10	caccgaaccctggtggcttctaat	aaacattagaagccaccagggttc
7__7	caccgacatcccattagaagccacc	aaacggtggcttctaatgggatgtc

Table 2- 6. Factor screen sgRNAs.

Name	Top_guide	Bottom_guide
NonTargeting_66406	caccgcgaggtattcggctccgcg	aaaccgcggagccgaatacctcgc
NonTargeting_66407	caccgctttcacggaggttcgacg	aaaccgtcgaacctccgtgaaagc
NonTargeting_66408	caccgatgttgcagttcggctcgat	aaacatcgagccgaactgcaacatc
Atf3_05592	caccgattctgagcccggacgatgc	aaacgcatcgtccgggctcagaatc
Atf3_05593	caccgtaccgtcaacaacagacccc	aaacggggtctgttgttgacggtac
Atf3_05594	caccgtgcatcgtccgggctcagaa	aaacttctgagcccggacgatgcac
Bhlhe40_06867	caccgggcaatgcactcgttaatc	aaacgattaacgagtgcattgccc
Bhlhe40_06868	caccgagtgttctcatgcttcgcc	aaacggcgaagcatgagaacactc
Bhlhe40_06869	caccgcagcttgtaaaccgctctgc	aaacgcagagcggtttacaagctgc
Cbx3_08244	caccgaacacagtgctgataatact	aaacagtattatcagcactgtgttc
Cbx3_08245	caccgtatacaggctgacaaacca	aaactggtttgtcagcctgtatac
Cbx3_08246	caccgcaacaggaaggactcggacg	aaaccgtccgagtccttcctgttgc
Ctcf_12271	caccgcatagcccgaaaaagtgatt	aaacaatcacttttttcgggctatgc
Ctcf_12272	caccgcaagtgcccagactgcgata	aaactatcgcagtctgggcacttgc
Ctcf_12273	caccgatgccgcaccaattctccac	aaacgtggagaattggtgcggcatc
Elf1_15931	caccgatgacgacatcacccttac	aaacgtaagggtgatgtcgtcatc
Elf1_15932	caccgtcgtcggggaatctggccg	aaaccggccagattccccgacgac
Elf1_15933	caccgttctttctcttgcgctgctc	aaacgagcagcgcaagagaaagaac
Ets1_16711	caccgcagaaacccacgtccgggac	aaacgtcccggacgtgggtttctgc
Ets1_16712	caccgcttactgatgaagtaatccg	aaaccggattacttcatcagtaagc
Ets1_16713	caccgagagtcggcttgagatcga	aaactcgatctcaagccgactctc
Hdac1_23860	caccgcttaccgacagagcctcccg	aaaccgggaggctctgtcggtaagc
Hdac1_23861	caccgttacgtcaatgacatcgtcc	aaacggacgatgtcattgacgtaac
Hdac1_23862	caccgctttcaagtcaatgttggtg	aaaccaccaacattgacttgaaagc
Hdac2_23863	caccgctcaccaactgaaccacccg	aaaccgggtggttcagttggtgagc
Hdac2_23864	caccgtaacgtcggagaagattgtc	aaacgacaatcttctccgacgttac
Hdac2_23865	caccgctggggctgtgaaattaaac	aaacgtttaatttcacagccccagc
Maff_30007	caccgcccggcgccgacttgacga	aaactcgtcaagtcggcgccgggc
Maff_30008	caccgcgagaacgccgccatgcgcc	aaacggcgcatggcggcgttctcgc
Maff_30009	caccgcttcagggctttgctagata	aaactatctagcaaagccctgaagc
Max_30508	caccgagcgataacgatgacatcg	aaaccgatgtcatcgttatcgctc
Max_30509	caccgatggtgccgcccttggcgt	aaacacgccaagggcggcaccatc
Max_30510	caccgcagcctctacaccaacgcca	aaactggcgttggtgtagaggctgc
Mbd4_30535	caccggtgggtgctgatcgcgagc	aaacgctcgcgatcagcacccacc
Mbd4_30536	caccgtcgcgactatatttctcaat	aaacattgagaaatatagtcgcgac
Mbd4_30537	caccgtgcgctcactgcttatcacc	aaacggtgataagcagtgagcgcac
Mxi1_32743	caccgctcgttccaagttctctagc	aaacgctagagaacttggaacgagc
Mxi1_32744	caccgtcccctccatgccgagcccc	aaacggggctcggcatggaggggac
Mxi1_32745	caccgcggatacgaatggacagcat	aaacatgctgtccattcgtatccgc

Mybl2_32764	caccggctatcagcgtccactgtt	aaacaacagtggacgctgatagcc
Mybl2_32765	caccgagatcgccaagatgctacc	aaacggtagcatcttggcgatctc
Mybl2_32766	caccgtctggatgagttacactacc	aaacggtagtgtaactcatccagac
Myc_32806	caccggacgtagcgaccgcaacat	aaacatgttgcggtcgctacgtcc
Myc_32807	caccgctgtacggagtcgtagtcg	aaaccgactacgactccgtacagc
Myc_32808	caccgtcgggctcatctccatcccg	aaaccgggatggagatgagcccgac
Nfic_33883	caccgaccgcgctcccgcacaaagt	aaacactttgtgcgggagcgcggtc
Nfic_33884	caccgtgtgtgcagccgcaccatat	aaacatatggtgcggctgcacacac
Nfic_33885	caccgagctgactcacccagcgtta	aaactaacgctgggtgagtcagctc
Rcor1_44958	caccgttgactgcgccttgccagtt	aaacaactggcaaggcgcagtcaac
Rcor1_44959	caccgcccgcttctgtttccgagca	aaactgctcggaaacagaagcgggc
Rcor1_44960	caccgagtcaagaacgagacaatct	aaacagattgtctcgttcttgactc
Rest_45135	caccgcatcatctgcacgtacacga	aaactcgtgtacgtgcagatgatgc
Rest_45136	caccgatcgcctgcgaaacctcccc	aaacggggaggtttcgcaggcgatc
Rest_45137	caccgaccaacgtgccaaagggcgg	aaacccgccctttggcacgttggtc
Rxra_46790	caccgccatggagcctcgacccgt	aaacacgggtcgaggctccatggc
Rxra_46791	caccgcttcgggactggtagccccc	aaacgggggctaccagtcccgaagc
Rxra_46792	caccgtccaccagcagtgccaacg	aaaccgttggcactgctggtggac
Sin3a_48515	caccgtgcagaatgaagcgacgtt	aaacaacgtcgcttcattctgcac
Sin3a_48516	caccgcgccctgtcctatcttgacc	aaacggtcaagataggacagggcgc
Sin3a_48517	caccgccggggcaaggacccggtgc	aaacgcaccgggtccttgccccggc
Smad4_49955	caccgccaagtaatcgcgcatcaa	aaacttgatgcgcgattacttggc
Smad4_49956	caccgtccgttgatgcgcgattact	aaacagtaatcgcgcatcaacggac
Smad4_49957	caccgacaacccgctcatagtgata	aaactatcactatgagcgggttgtc
Sp1_50666	caccgatctccgagcacacttgcgc	aaacgcgcaagtgtgctcggagatc
Sp1_50667	caccgcatatactttgccgcatcct	aaacaggatgcggcaaagtatatgc
Sp1_50668	caccggttggtgtccgaatgatga	aaactcatcattcggacaccaacc
Tbl1xr1_52979	caccgagacgtgacatccttgttac	aaacgtaacaaggatgtcacgtctc
Tbl1xr1_52980	caccgccgaagcacaaccgctttat	aaacataaagcggttgtgcttcggc
Tbl1xr1_52981	caccgtgtctaagcacgagctgcgt	aaacacgcagctcgtgcttagacac
Tcf7l2_53147	caccgaaacagctcctccgattccg	aaaccggaatcggaggagctgtttc
Tcf7l2_53148	caccgcgagagcgatccgttgggg	aaaccccccaacggatcgctctcgc
Tcf7l2_53149	caccgacggcgaacgagcatccttg	aaaccaaggatgctcgttcgccgtc
Tgfbr2_53621	caccgaccgcaccgccattgtcgc	aaacgcgacaatggcggtgcggtc
Tgfbr2_53622	caccgccttgtagacctcggcgaag	aaaccttcgccgaggtctacaaggc
Tgfbr2_53623	caccgccacgcgaagggcaacctgc	aaacgcaggttgcccttcgcgtggc
Trim28_55757	caccggacctgctaagactcgaga	aaactctcgagtcttagcaggtcc
Trim28_55758	caccgataattctccacgatgtctt	aaacaagacatcgtggagaattatc
Trim28_55759	caccgttacgagcttcgaacctcct	aaacaggaggttcgaagctcgtaac
Yy1_59977	caccgcttcttattacaaccgtcga	aaactcgacggttgtaataagaagc
Yy1_59978	caccgccctatgtgtgcccttcga	aaactcgaaggggcacacatagggc

Yy1_59979	caccgtcaatgccaggtatccctcc	aaacggagggatacctggcattgac
Zbtb33_60064	caccgaccgaaaattccgggcccat	aaacatgggcccggaattttcggtc
Zbtb33_60065	caccgttggtgcggtattctgcaag	aaaccttgcagaataccgcaccaac
Zbtb33_60066	caccggcgagttattgctagcgca	aaactgcgctagcaataactcgcc
Zbtb7a_60118	caccgatcggcccccagccggcga	aaactcgccggctgggggccgatc
Zbtb7a_60119	caccgcaggtcgtagttgtgcgcga	aaactcgcgcacaactacgacctgc
Zbtb7a_60120	caccgcgccctctgcagcgcctgac	aaacgtcaggcgctgcagagggcgc
Smad2_49949	caccgttagaatctccgtgtgccg	aaaccggcacacggagattctaac
Smad2_49950	caccgtctaacagaactgccgcccc	aaacggggcggcagttctgttagac
Smad2_49951	caccgcttcaggtttcacaccggaa	aaacttccggtgtgaaacctgaagc
Smad3_49952	caccgctccatggcccgtaattca	aaactgaattacgggccatggagc
Smad3_49953	caccgacctacctggaatattgctc	aaacgagcaatattccaggtaggtc
Smad3_49954	caccgcaggtcgggccatcgccac	aaacgtggcgatggcccgacctgc

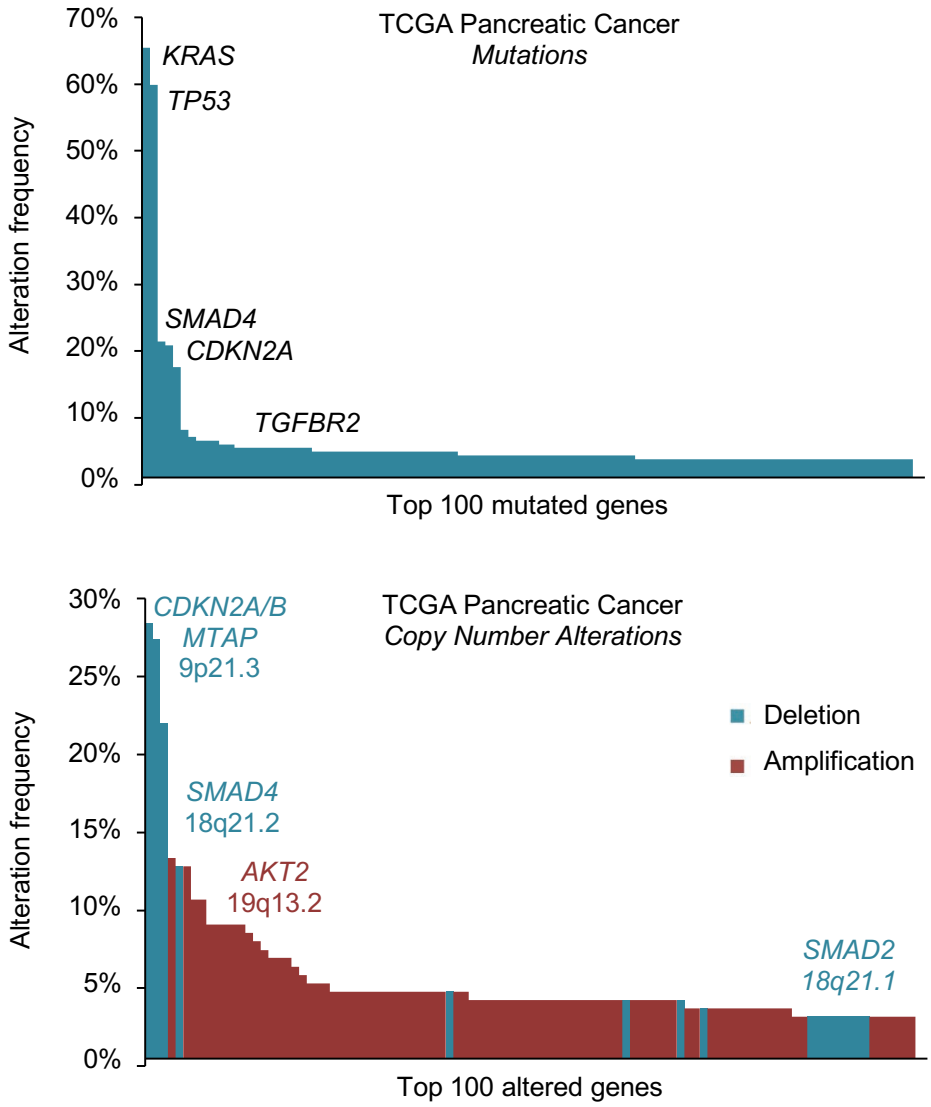


Figure 1-1. Mutational spectrum of pancreatic cancer.
The top 100 mutated or copy number altered genes in pancreatic cancer (184 cases) were curated from the TCGA PanCancer database via cBioPortal.

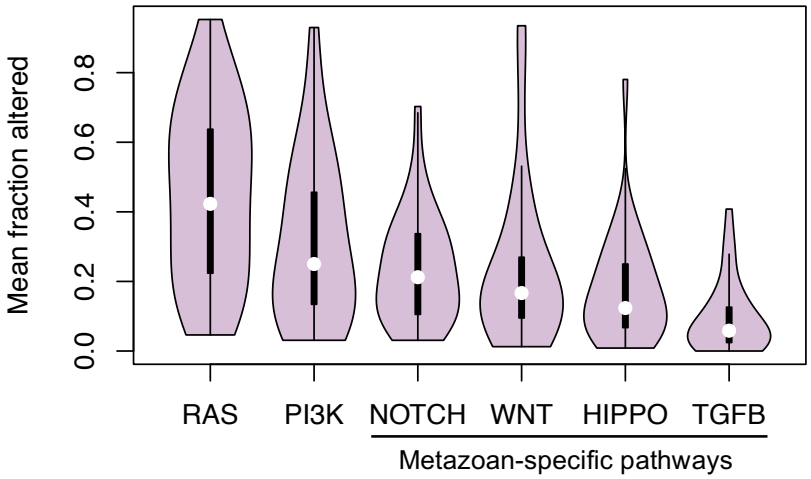


Figure 1-2. Signaling pathways altered in 33 tumor types. Data from the TCGA PanCancer Atlas (Sanchez-Vega et al, Cell, 2018) was analyzed for the distribution of pathway alterations.

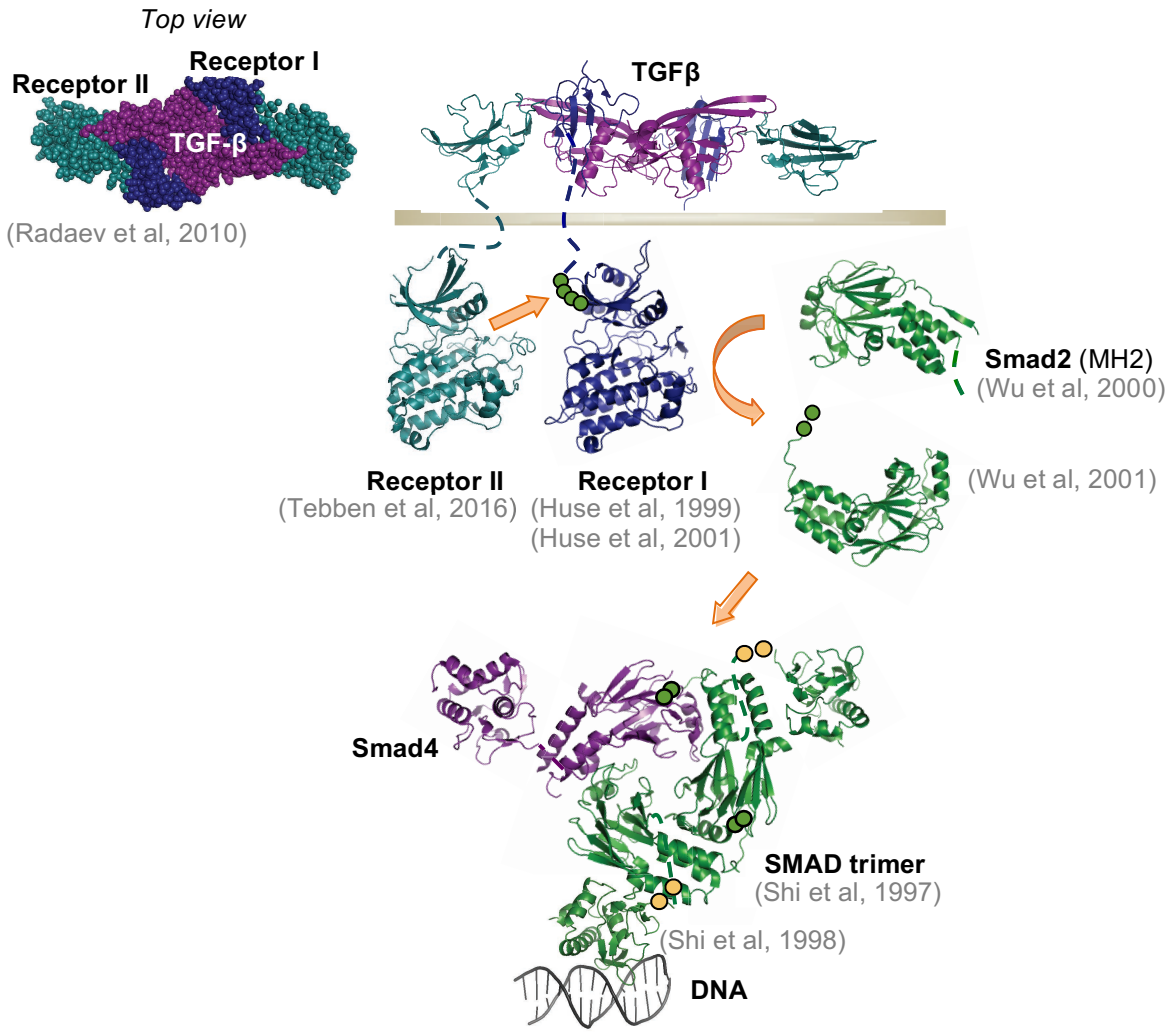


Figure 1-3. TGFβ signaling via the SMAD proteins.
TGFβ dimers bind to type II receptors, which recruit and phosphorylate type I receptors. The heterotetrameric receptor complex phosphorylates the C-terminal tail of receptor SMADs, which heterotrimerize with SMAD4 and bind to DNA to modulate transcription.

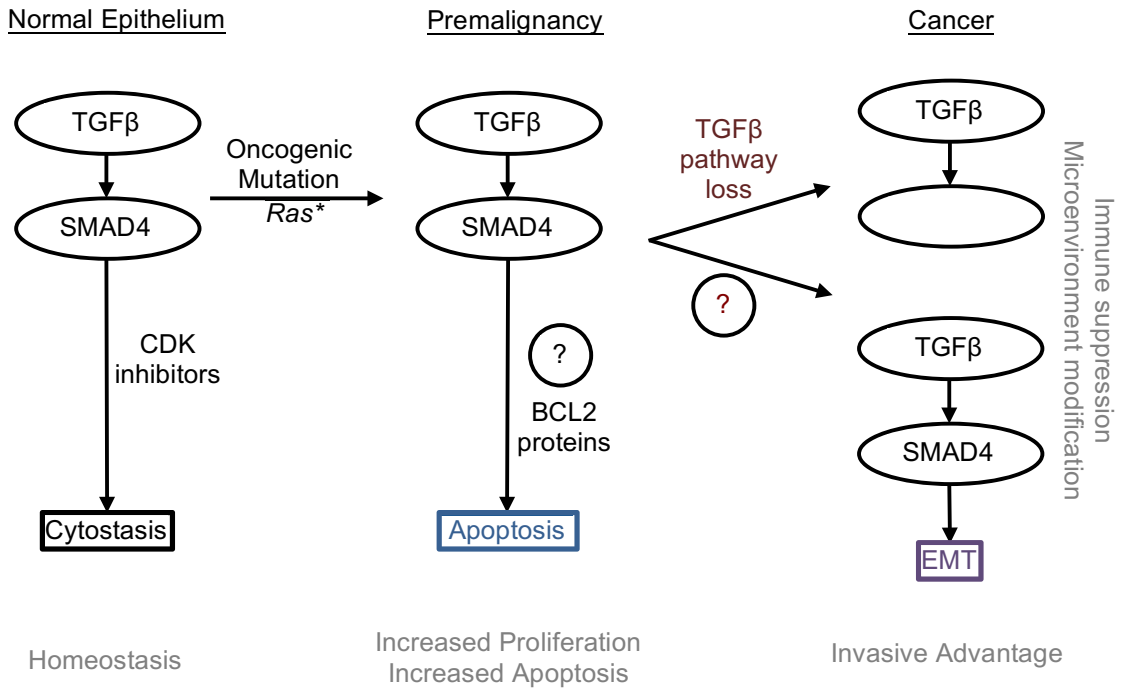


Figure 1-4. The changing roles of TGFβ throughout tumor development.
The tumor suppressive effects of TGFβ include cytostasis, apoptosis, and suppression of growth factors from the microenvironment. In malignant tumors, TGFβ can promote tumor growth, invasion, and metastasis via both cell-autonomous and non-cell-autonomous means. The mechanisms of TGFβ-mediated apoptosis and non-genetic loss of TGFβ tumor suppression are less well understood.

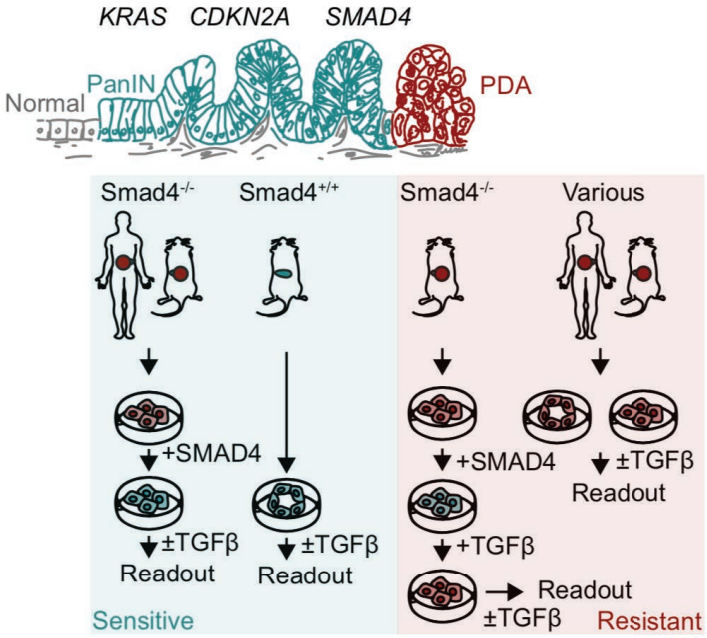


Figure 2-1. Models used to simulate the TGFβ-sensitive and –resistant states.
Sensitive models: SMAD4 was re-expressed in human or mouse SMAD4-null PDA cells or cells were isolated from premalignant mouse pancreata.
Resistant models: SMAD4-restored cells were selected in the presence of TGFβ for 3 weeks or isolated from human or mouse PDAs.

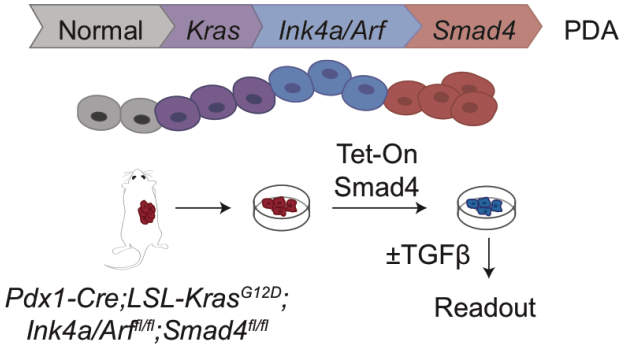


Figure 3-1. Experimental model.
Schematic of order of genetic alterations in PDA and experimental model to study TGFβ-mediated tumor suppression by expression of SMAD4 in SMAD4-null PDA.

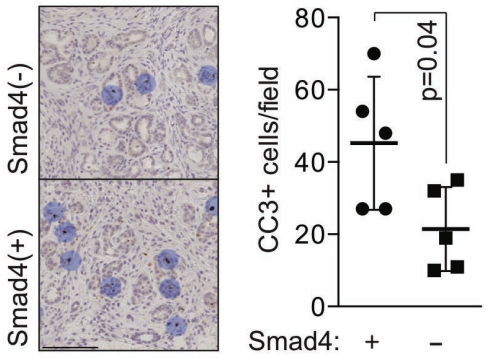


Figure 3-2. Effects of SMAD4 loss in the inflamed pancreas.
Immunohistochemistry (IHC) for cleaved caspase 3 (CC3) in KC or KSC pancreata collected 2 days after caerulein treatment and quantification. Scale bars, 100 μm. Mean ± SD.

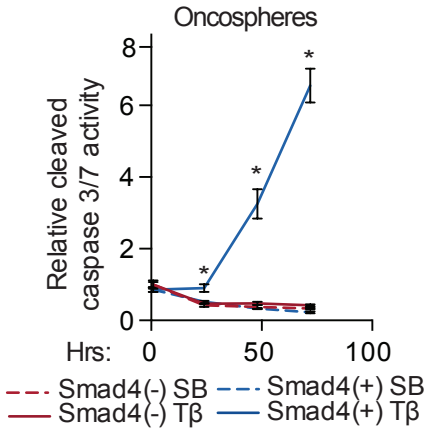


Figure 3- 3. TGFβ in SMAD4-restored PDA oncospheres.
Caspase 3/7 activity in the treated SMAD4-null and SMAD4-restored oncospheres (p values ≤ 0.01).

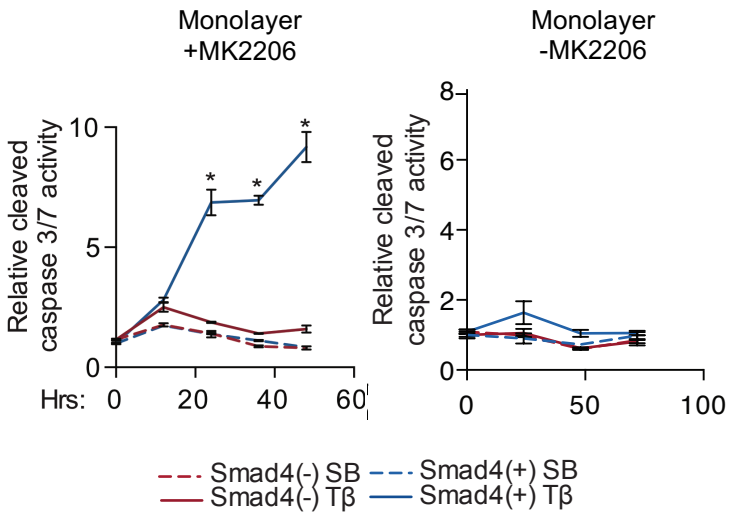


Figure 3- 4. TGFβ response of SMAD4-restored PDA monolayers.
Cells grown in monolayer in 10% serum medium were pre-treated with 2.5 μM MK2206 or not and then treated with TGFβ (100 pM) or SB505124 (SB; 2.5 μM) and assayed for cleaved caspase 3/7 activity. * $p \leq 0.01$. Data courtesy of CD.

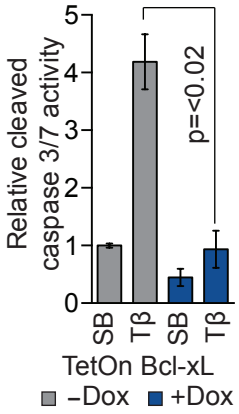


Figure 3- 5. Rescue of TGFβ apoptosis by BCL-XL.
Caspase 3/7 activity in SMAD4-restored cells bearing Tet-On *Bcl-xL* construct treated as indicated.

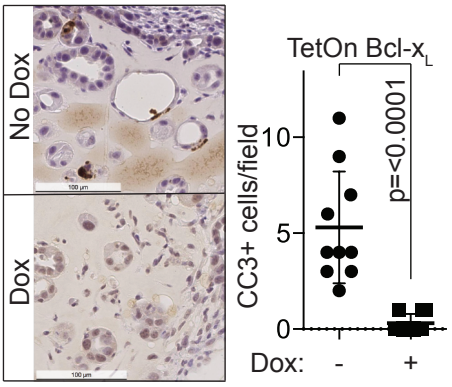


Figure 3- 6. *In vivo* rescue of apoptosis by BCL-XL.
Orthotopic implantation of SMAD4-restored KSIC cells with a Tet-On *Bcl-xL* construct. Acute pancreatitis was induced with 16 hourly injections of 50 mg/kg caerulein over two days. Pancreata were collected after 2 days followed by CC3 IHC staining. Error bars represent ± SD, two-tailed t test.

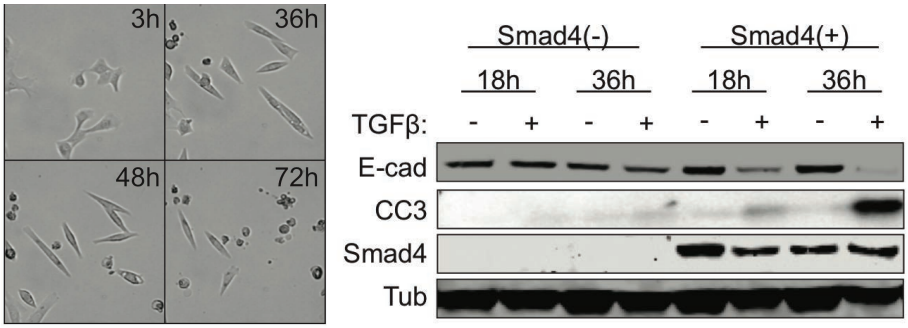


Figure 3- 7. Morphologic response of SMAD4-restored PDA cells to TGFβ.
SMAD4-restored KSIC cells treated with 100 pM TGFβ. Data courtesy of CD.

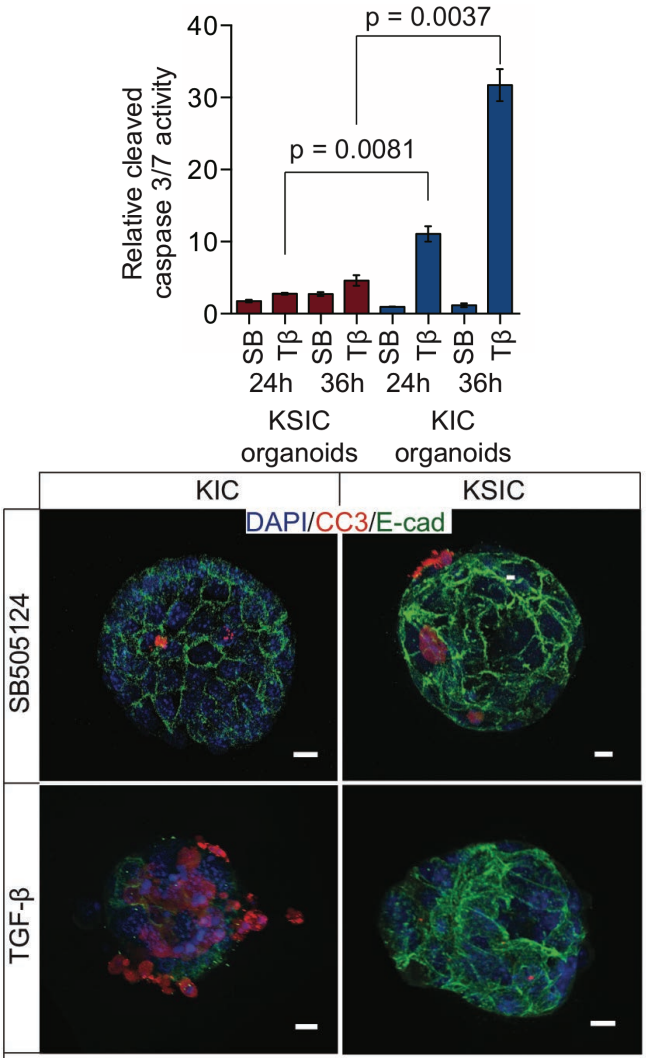


Figure 3- 8. Response of pancreatic organoids to TGFβ.
Organoids derived from PDA-prone mouse models prior to development of invasive PDA were treated with 2.5 μM SB505124 or 100 pM TGFβ. Data courtesy of JS.

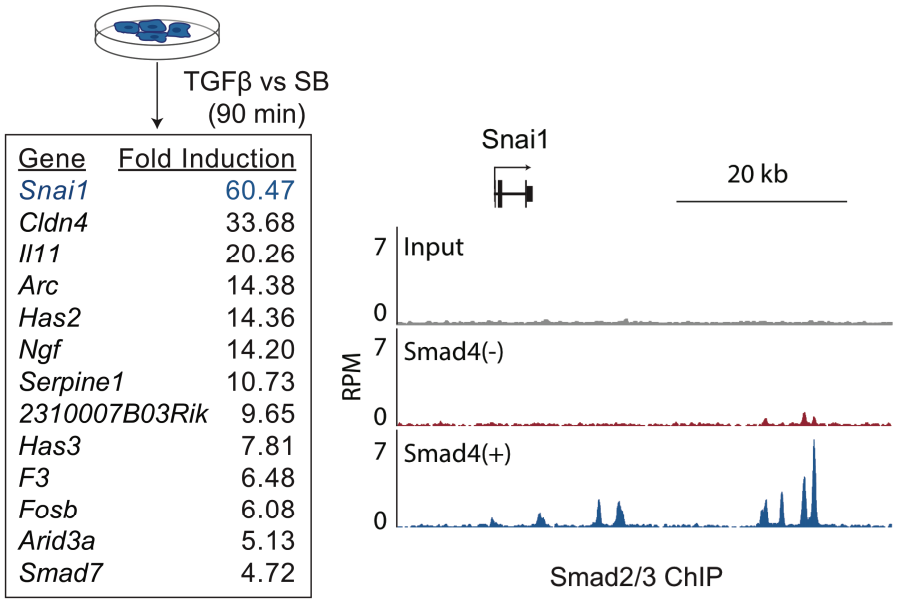


Figure 3- 9. Gene expression changes of SMAD4-restored PDA cells.
Left panel: List of most strongly induced genes after 90 min of TGFβ treatment in SMAD4+ cells. *Right panel*: SMAD2/3 binding to the *Snai1* locus after 1h of TGFβ treatment in SMAD4 mutant or SMAD4+ cells. Data courtesy of CD, MC, and YZ.

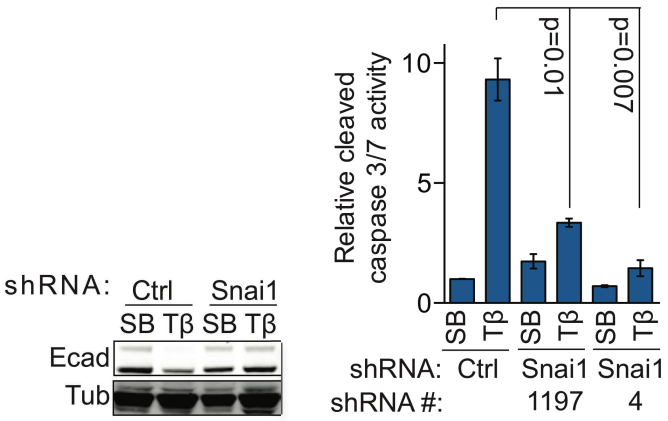


Figure 3- 10. SNAIL functionality in lethal EMT.
Left panel: Immunoblot for E-cadherin in SMAD4+ cells expressing the indicated shRNAs 24h post-TGFβ/SB treatment. *Right panel*: SMAD4+ cells expressing *Snai1*-targeting shRNAs were treated with TGFβ or SB for 36h then assayed for caspase 3/7 activity. Data courtesy of CD.

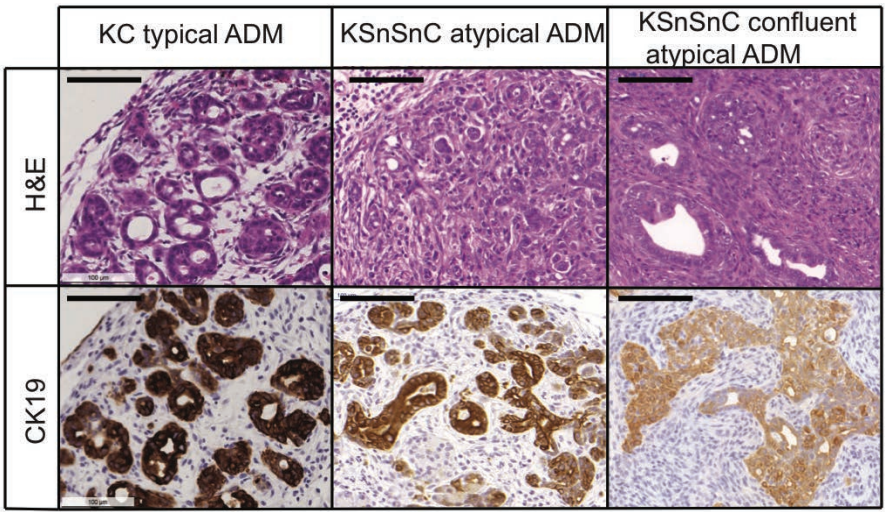


Figure 3- 11. SNAIL in a mouse model of PDA.
H&E and CK19 IHC for representative ADM or atypical ADM lesions in KC or KSnSnC mice. Scale bars, 100 μm. Data courtesy of CD.

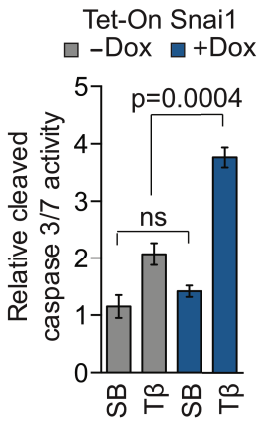


Figure 3- 12. Insufficiency of SNAIL in lethal EMT.
SMAD4+ cells transduced with Tet-On *Snai1* were treated with or without doxycycline for 12h followed by TGFβ or SB, then assayed for caspase activity after 18h.

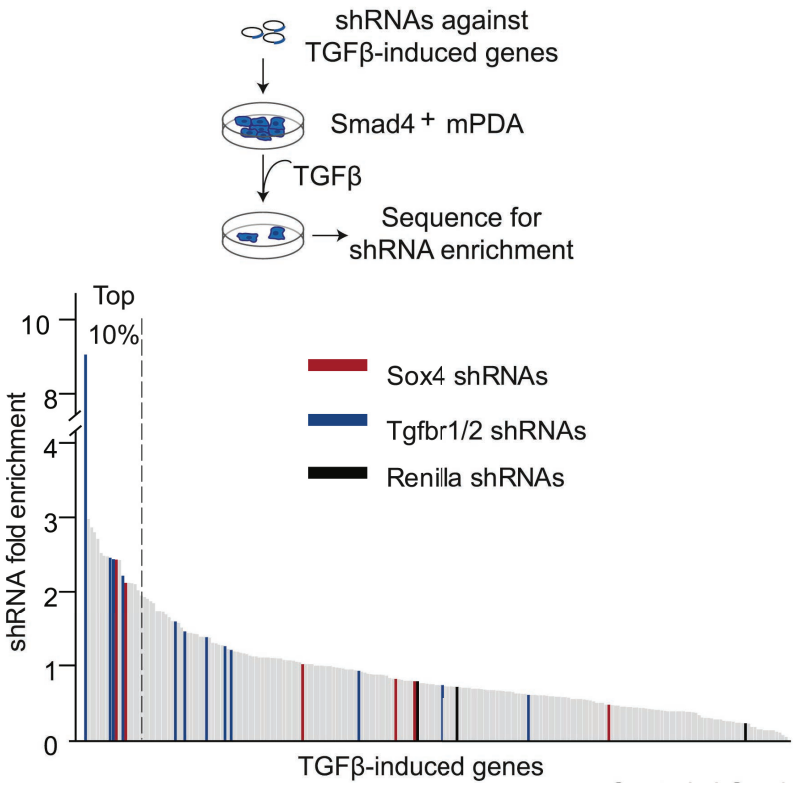


Figure 3- 13. Screen to identify factors promoting lethal EMT.
SMAD4-restored PDA cells were transduced with shRNAs against genes induced by TGFβ after 90 min. Cells containing the shRNA libraries were treated with 100 pM TGFβ or 2.5 μM SB505124 for 5 days and then collected for targeted sequencing of the shRNAs. Positive control shRNAs targeting *Tgfbr1* and *Tgfbr2* and negative control shRNAs targeting Renilla luciferase are indicated. Data courtesy of CD.

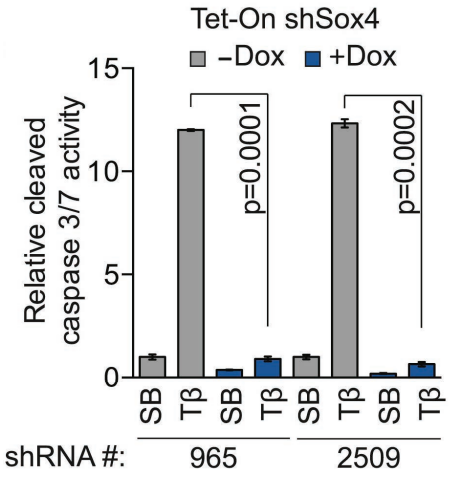


Figure 3- 14. Necessity of SOX4 in lethal EMT.
Indicated Tet-On *Sox4* shRNAs were transduced into SMAD4+ cells; cells were pretreated with or without doxycycline for 24h, then treated as indicated. Caspase activity was assayed after 36h.

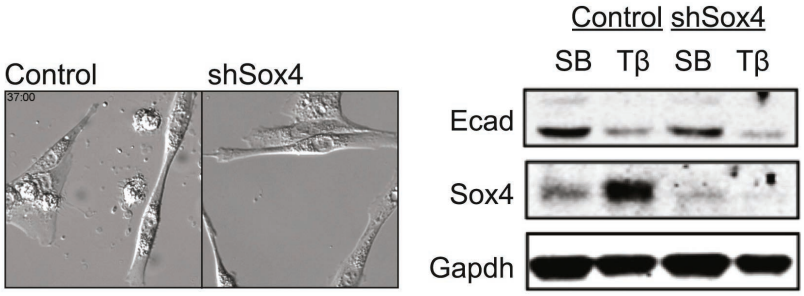


Figure 3- 15. Effect of SOX4 depletion on EMT.
Left panel: Images of control or *Sox4* KD cells treated with TGFβ for 37h. *Right panel*: Immunoblot for the indicated factors after SOX4 depletion and 24h of SB/TGFβ treatment.

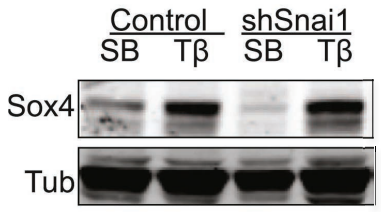


Figure 3- 16. Independent regulation of SNAIL and SOX4
SOX4 immunoblot after 24h of TGFβ treatment in cells bearing control shRNAs or shRNAs targeting *Snai1*. Data courtesy of CD.

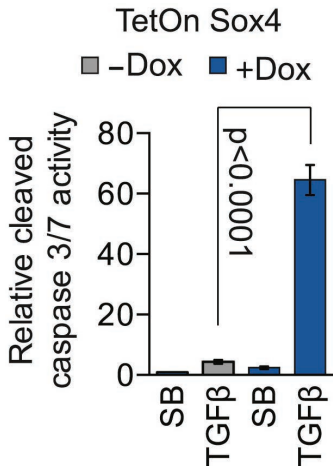


Figure 3- 17. Synergy of SOX4 and TGFβ in lethal EMT.
Cells transduced with doxycycline-inducible *Sox4* cDNA were treated for 36 hours with SB505124 or TGFβ. Apoptosis and cell viability were measured using CaspaseGlo 3/7 and CellTiterGlo. Data courtesy of CD.

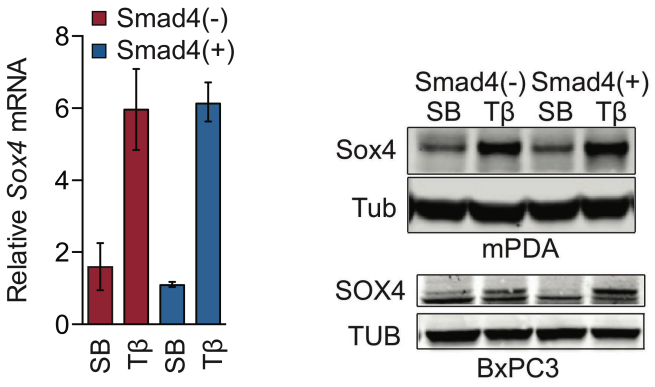


Figure 3- 18. SMAD4-independence of SOX4.
Left panel: *Sox4* mRNA levels in the indicated cells after 24h of the indicated treatment. Values are the mean and range of four technical replicates and representative of at least three separate experiments. *Right panel*: Immunblot for SOX4 after 24h of the indicated treatment in mouse PDA (top) and human PDA cells (bottom). Data by YH and CD.

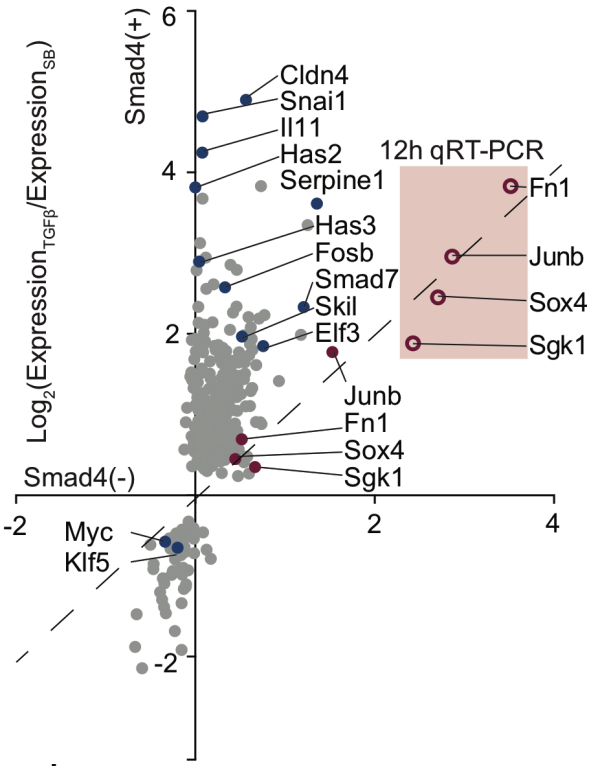


Figure 3- 19. SMAD4-dependence of immediate TGFβ gene responses.
SMAD4 dependence of TGFβ-induced genes after 90 min treatment of KSIC cells. Fold induction in SMAD4 mutant and SMAD4+ cells is plotted on the x and y axes, respectively. Select SMAD4-independent genes were assayed by qRT-PCR after 12h induction and plotted (red-shaded region). RNAseq by CD, analysis by YH.

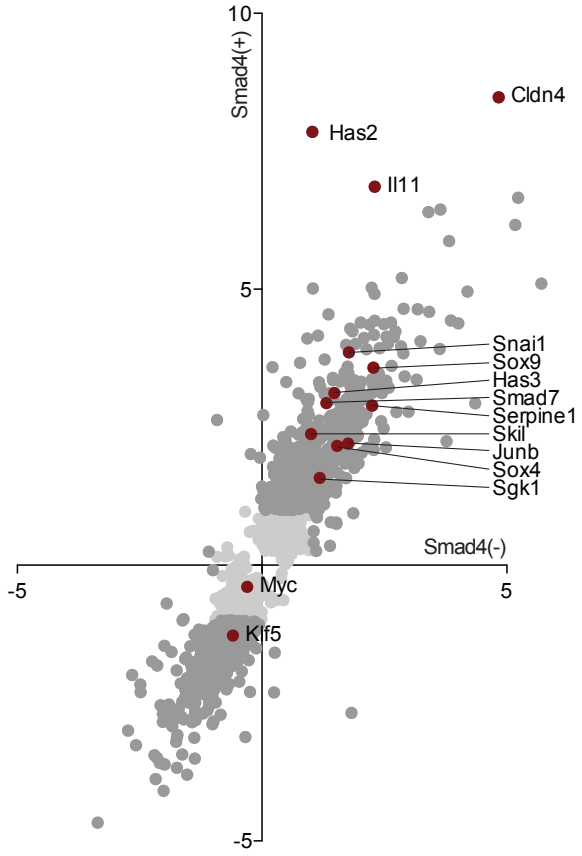


Figure 3- 20. Long-term SMAD4-dependence of TGFβ gene responses.
SMAD4 dependence of TGFβ-induced genes after 12h treatment of KSIC cells. Fold induction in SMAD4 mutant and SMAD4+ cells is plotted on the x and y axes, respectively. Plotted genes were detected at >100 readcounts, $p_{adj}<10^{-10}$, and had a SMAD2/3 peak present at 90 min.

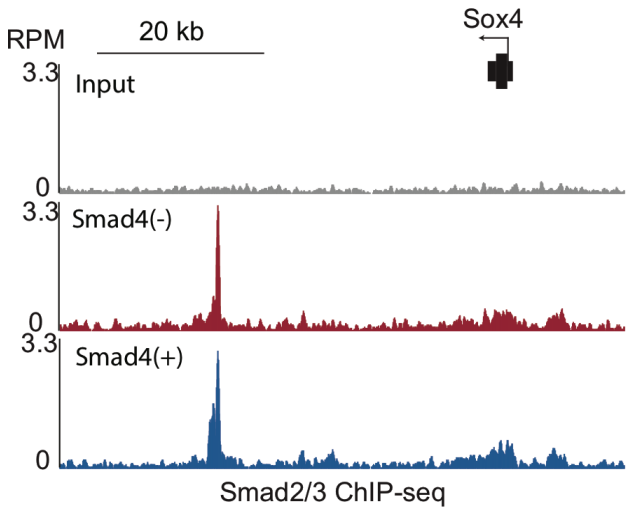


Figure 3- 21. SMAD2/3 binding to the *Sox4* locus.
SMAD2/3 binding to the *Sox4* locus as determined by ChIP-seq performed in SMAD4 mutant and SMAD4+ cells. Data courtesy of MC and YZ.

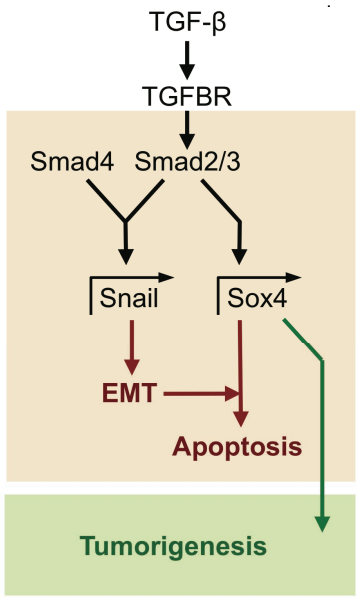


Figure 3- 22. SMAD4-dependent and –independent branches in lethal EMT

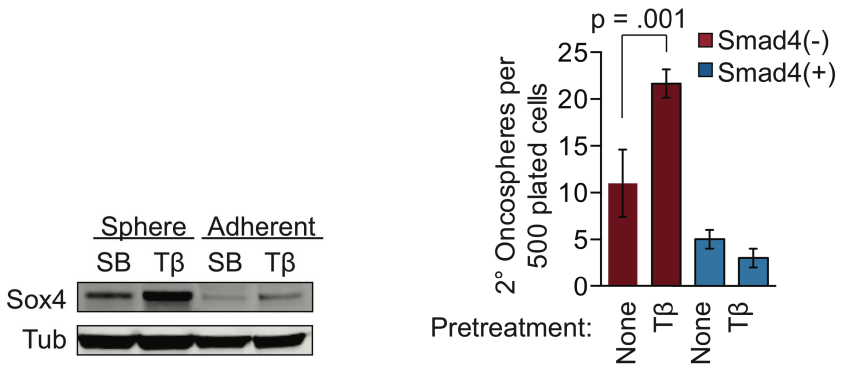


Figure 3- 23. Correlation of *Sox4* and oncosphere potential
Left panel: SOX4 protein expression after SB/TGFβ treatment of cells grown in adherent conditions or after 3 days in oncosphere conditions. *Right panel*: Secondary oncosphere formation of SMAD4 mutant and SMAD4+ PDA cells. Primary oncospheres were cultured for 2 weeks in the presence or absence of 20 pM TGFβ. Spheres were dissociated with trypsin and replated for 1 week prior to counting. Data courtesy of CD.

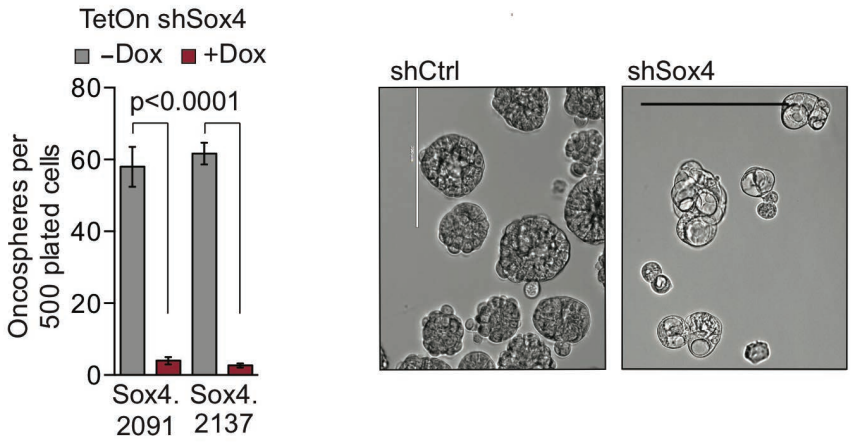


Figure 3- 24. SOX4 in PDA oncospheres.
PDA cells expressing the indicated shRNAs were plated as oncospheres and assayed after 1 week. Scale bar = 200 μm

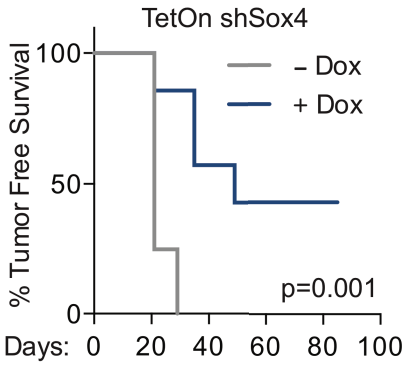


Figure 3- 25. SOX4 in subcutaneous tumor formation.
Tumor-free survival after KSIC mPDA cells tranduced with Tet-On *Sox4* shRNAs were injected subcutaneously into mice maintained on or off doxycycline diet. Plot reflects results of pooled experiments using Sox4.965 and Sox4.2509 shRNAs. For each condition, n = 10 mice; survival compared by log-rank test.

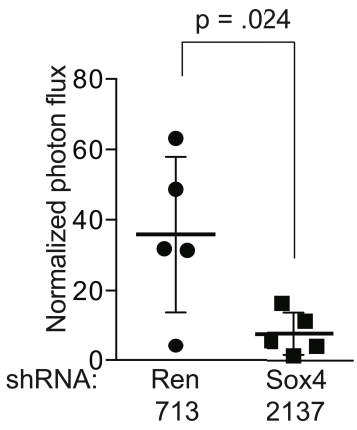


Figure 3- 26. Role of SOX4 in orthotopic PDA.
Luciferase-labeled KSIC cells transduced with the indicated shRNAs were injected orthotopically in mice and imaged after 4 weeks. Luminescence was normalized to day 0 values. Statistical significance was calculated by a two-tailed unpaired t-test. Error bars represent mean ± SD.

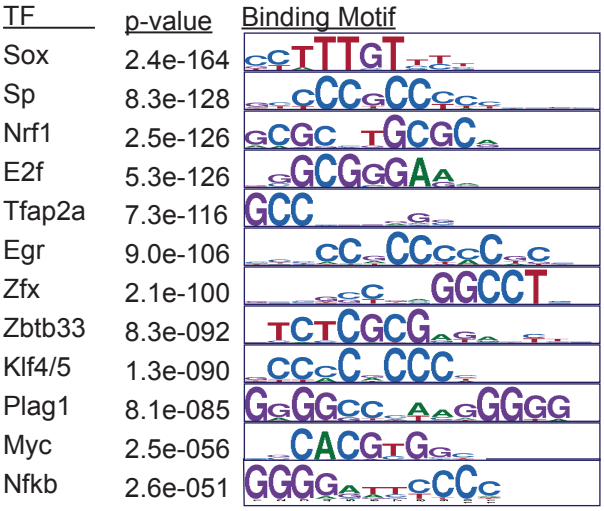


Figure 3- 27. TF motifs within SOX4 ChIP-seq peaks
Enriched TF motifs within the SOX4 ChIP-seq peaks, determined using PSCAN ChIP with the Jaspar motif database. Data courtesy of CD, MC & YZ.

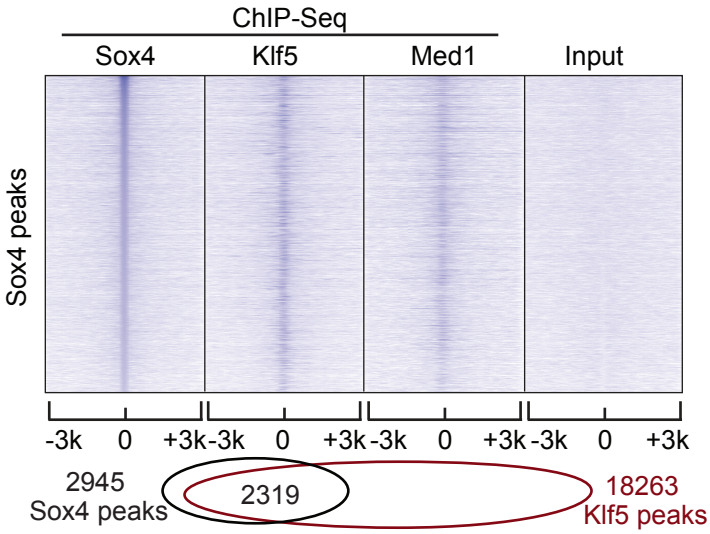


Figure 3- 28. Correlation of SOX4 and KLF5 peaks.
Correlation plots of SOX4, KLF5, and MED1 ChIPseq. Data courtesy of MC & YZ.

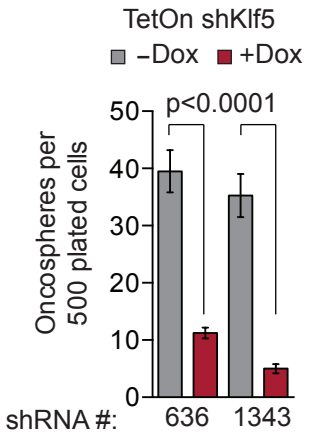


Figure 3- 29. KLF5 in oncosphere formation.
KSIC mPDA cells transduced with Tet-On *Klf5* shRNAs were grown as oncospheres for 2 weeks as indicated then quantified. Data courtesy of CD.

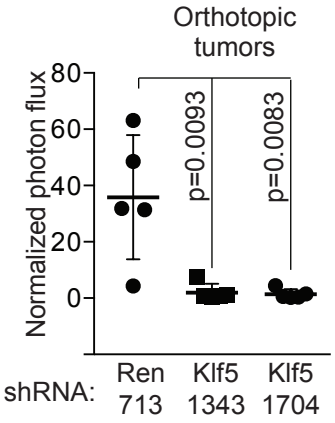


Figure 3- 30. KLF5 in orthotopic PDA tumors.
KSIC cells were transduced with the indicated shRNAs and implanted orthotopically in mice maintained on doxycycline feed. Mice were imaged after 4 weeks. Control group same as in Figure 3-26. Error bars represent mean ± SD. Data courtesy of CD.

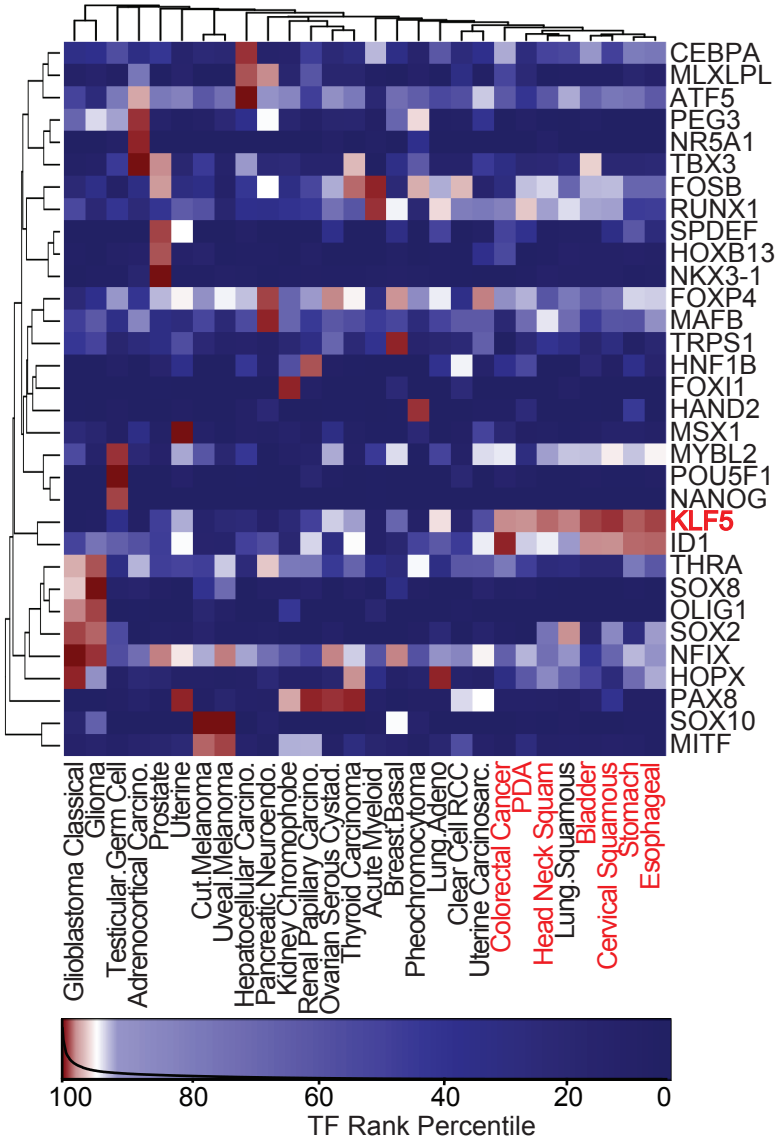


Figure 3- 31. Highly expressed TFs across cancer types.
Heatmap showing highly expressed TFs in cancer exhibiting lineage specificity. TCGA RNA-seq data were used to rank TFs by average expression levels in the indicated tumor types. To identify lineage-specific TFs, any TF with expression in the top 0.5% in at least one tumor type was included, excluding TFs that exhibited expression in the top 10% in more than half of the tumor types analyzed. The latter step excluded 22 TFs. Tumor types exhibiting >10% of loss of TGFβ signaling are indicated in red. Raw data sourced from TCGA.

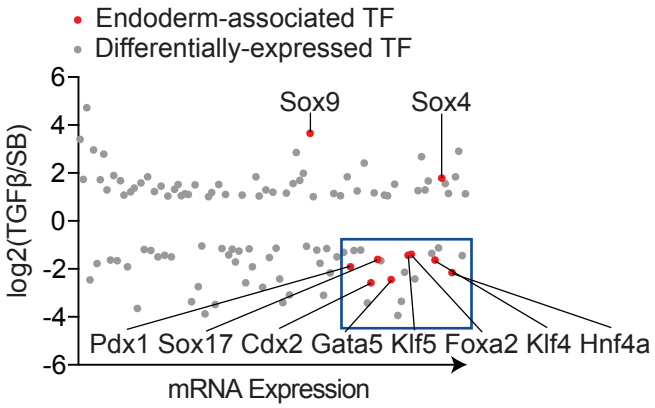


Figure 3- 32. TGFβ regulation of TFs.
Scatterplot showing all TFs differentially regulated during TGFβ-driven EMT. Overall expression levels (x axis) are plotted versus $\log_2$(fold change) in TGFβ versus SB treated SMAD4+ cells (y axis). Only TFs with > 2x fold change shown (gray) with endoderm-associated TFs highlighted (blue). Box indicates highly expressed TFs repressed by TGFβ.

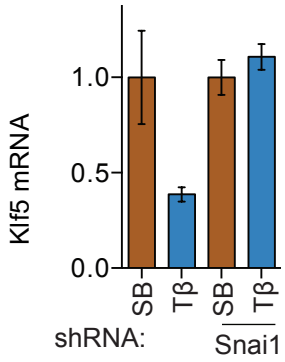


Figure 3- 33. SNAIL regulation of *Klf5*.
mRNA levels of *Klf5* in SMAD4+ cells expressing control, or *Snail* shRNA. Error bars represent the mean ± SD of 4 technical replicates, representative of three separate experiments. Data courtesy of CD.

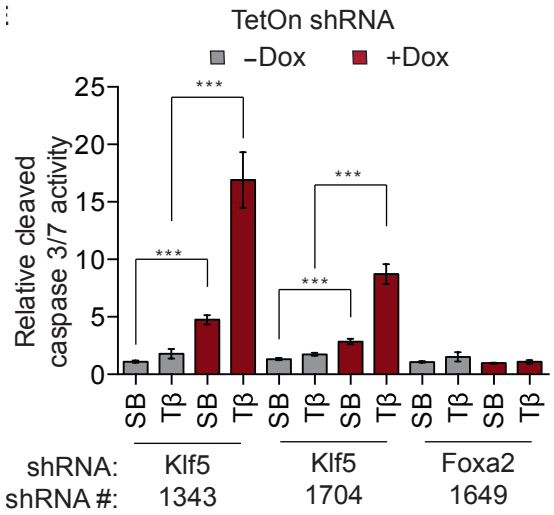


Figure 3- 34. Synergy of KLF5 and lethal EMT.
SMAD4 mutant PDA cells transduced with the indicated Tet-On shRNAs were treated with doxycycline for 48h to deplete the target protein, followed by treatment with SB/TGFβ. Cells were assayed for caspase 3/7 activity after 36h SB/TGFβ treatment. Data courtesy of CD.

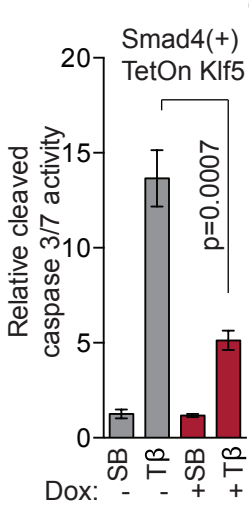


Figure 3- 35. Rescue of lethal EMT by KLF5.
A Tet-On *Klf5* construct was introduced into constitutive SMAD4+ cells. *Klf5* was induced with doxycycline for 12h prior to the addition of MK2206. Cells were then treated with SB/TGFβ and assayed for caspase 3/7 activity at 36h. Data courtesy of CD.

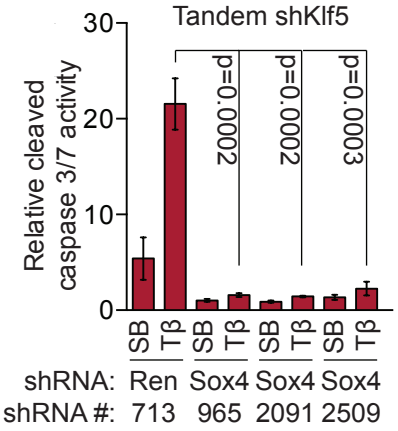


Figure 3- 36. Synergy of *Sox4* and *Klf5* depletion in lethal EMT.
Klf5.1343 shRNA was cloned into a tandem Tet-On shRNA vector together with shRNAs targeting *Renilla* luciferase or *Sox4*. Cells were incubated with doxycycline for 48h, then treated as indicated and assayed for caspase 3/7 activity after 36h. Data courtesy of CD.

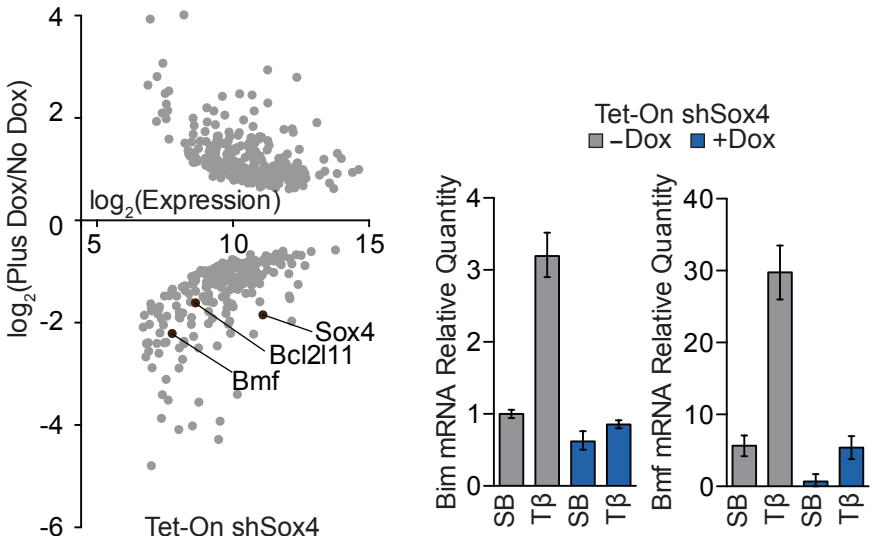


Figure 3- 37. SOX4-dependent TGFβ responses.
Left panel: SOX4-dependent gene expression in TGFβ-treated SMAD4+ cells. Cells bearing Tet-On *Sox4* shRNAs were cultured ± doxycycline for 24h prior to treatment with TGFβ. After another 24h, RNA was extracted and RNA-seq was performed. SOX4-dependent differentially expressed genes were then plotted. Right panel: qRT-PCR validation of *Bim* and *Bmf* as SOX4-dependent gene responses in TGFβ-treated SMAD4+ cells. Values are the result of four technical replicates.

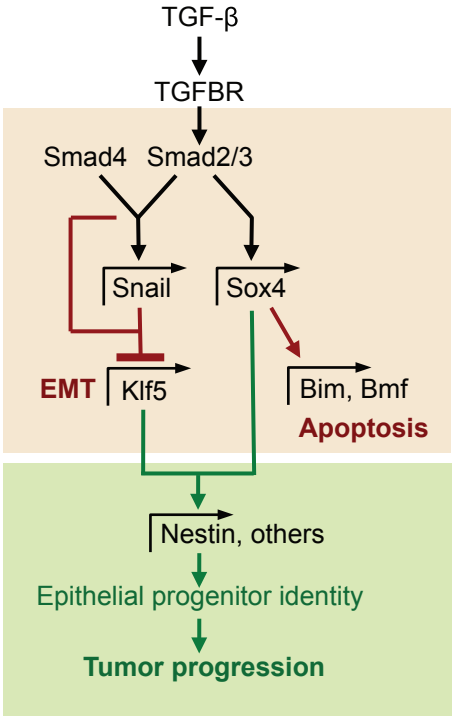


Figure 3- 38. Summary of effects of TGFβ signaling in PDA cells.

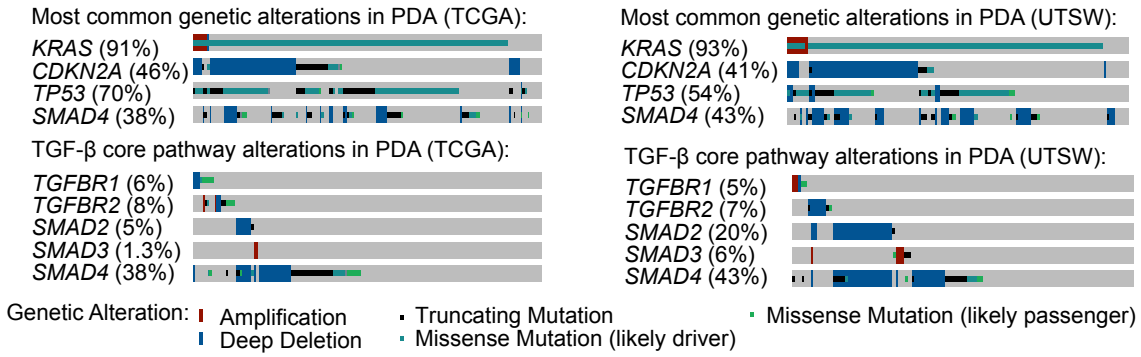


Figure 4- 1. cBioportal oncoprints of common genetic alterations in PDA.
Each column represents one case of PDA.

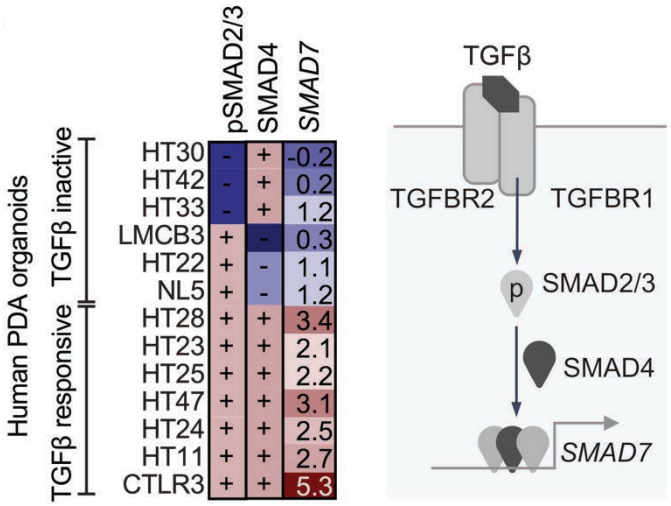


Figure 4- 2. TGFβ response of human PDA organoids.
Human PDA organoids were treated ±100 pM TGFβ for 2 hours. SMAD4 and pSMAD2 were detected by Western immunoblotting (WB) and *SMAD7* transcript by qRT-PCR. Values reported for *SMAD7* represent $\log_2$ of the fold change induction by TGFβ. (+) represents a strong band by WB and (-) represents a diminished or absent band detected by WB. A schematic representation of the core TGFβ pathway components and *SMAD7* as a target gene is included. In collaboration with NL & CD.

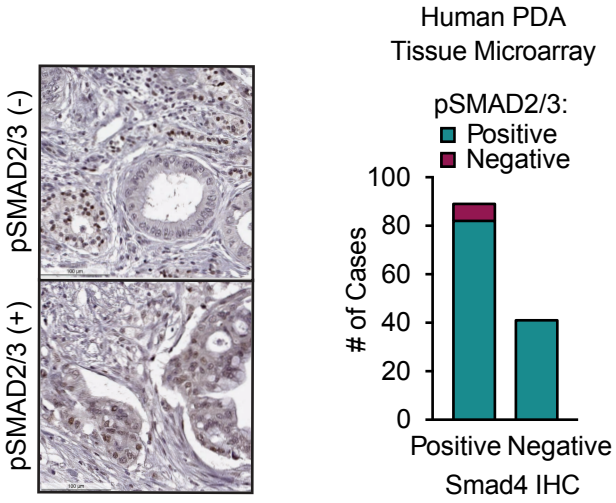
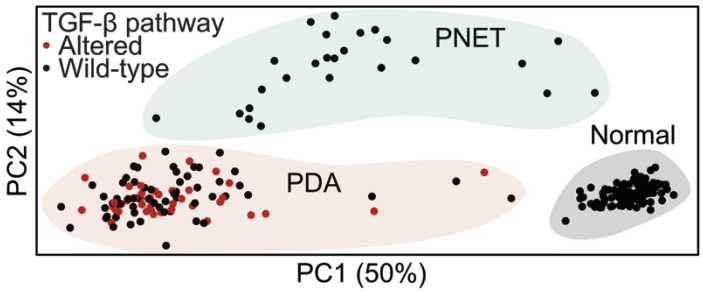


Figure 4- 3. pSMAD2 in human PDA samples.
A formaldehyde-fixed, paraffin-embedded tissue microarray was constructed of human PDA samples collected at surgical resection and stained for pSMAD2. Samples were scored positive if ≥50% of spots contained pSMAD2 in the tumor cells. TMA by PA.



AHR	*FOS*	*JUND*	*NKX2-2*	*STAT1*
ATF4	*FOSB*	*KLF13*	*NKX6-3*	*STAT2*
BCL6	*HIF1A*	*KLF5*	*NPAS2*	*STAT3*
BHLHA15	*HMGA1*	*KLF6*	*NR4A1*	*STAT6*
BHLHE40	*HMGB1*	*MAFB*	*NR5A2*	*TFDP1*
BPTF	*HMGB2*	*MAZ*	*PAX6*	*TSC22D1*
CREB3L1	*HMGB3*	*MEF2A*	*PEG3*	*TSC22D3*
DMTF1	*HSF4*	*MEIS1*	*PPARG*	*TSHZ3*
DRAP1	*ID1*	*MEIS2*	*PRRX1*	*XBP1*
EGR1	*ID2*	*MLXIP*	*RBPJ*	*ZBTB16*
EHF	*ID3*	*MLXIPL*	*RBPJL*	*ZBTB38*
ELF1	*ISL1*	*MYC*	*RUNX1*	
ELF3	*JUN*	*NFE2L1*	*SREBF2*	
EPAS1	*JUNB*	*NFIC*	*ST18*	

Figure 4- 4. Highly expressed TFs in pancreatic tissues.
Top panel: RNAseq datasets of normal pancreas, PDA, and PNET from GTEx and ICGC were curated for transcription factors. Principal component analysis (PCA) was performed of the ranked factors represented within the top 5 of at least one case. *Bottom panel*: complete list of factors included in the PCA.

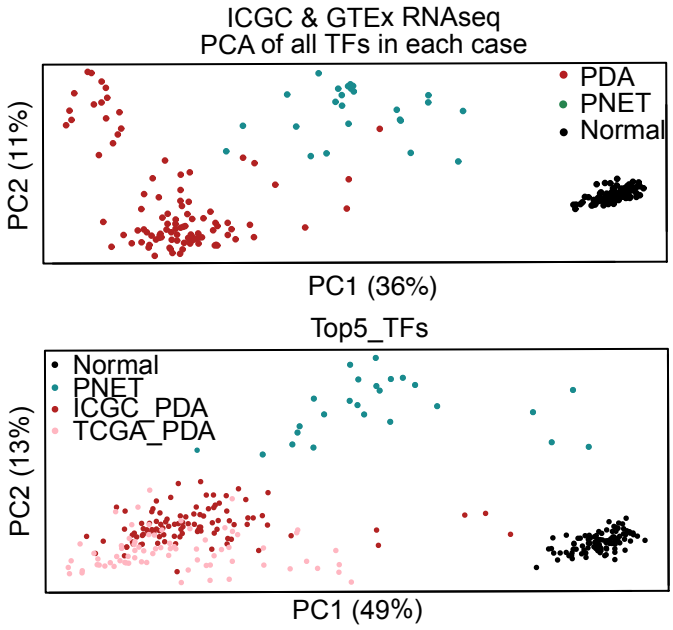


Figure 4- 5. TFs expressed in pancreatic tissues.
Top panel: PCA of RNAseq datasets from GTEx and ICGC using a complete list of TFs. *Bottom panel*: PCA of RNAseq datasets from GTEx, TCGA, and ICGC using factors listed in Figure 4-4.

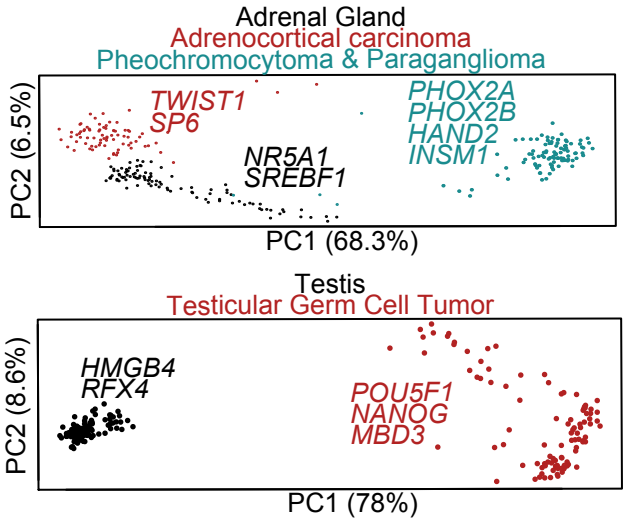


Figure 4- 6. Highly expressed TFs in other tissues.
RNAseq datasets from GTEx and TCGA were curated for transcription factors. Principal component analysis (PCA) was performed of the ranked factors represented within the top 5 of at least one case.

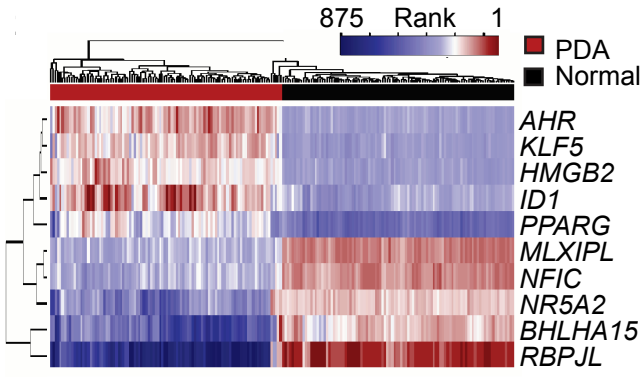


Figure 4- 7. Clustering of PDAs and normal pancreas.
Using the top 10 PC1 components from Figure 4-4, unsupervised clustering was performed of the rank-based analysis of RNAseq datasets from GTEx and ICGC.

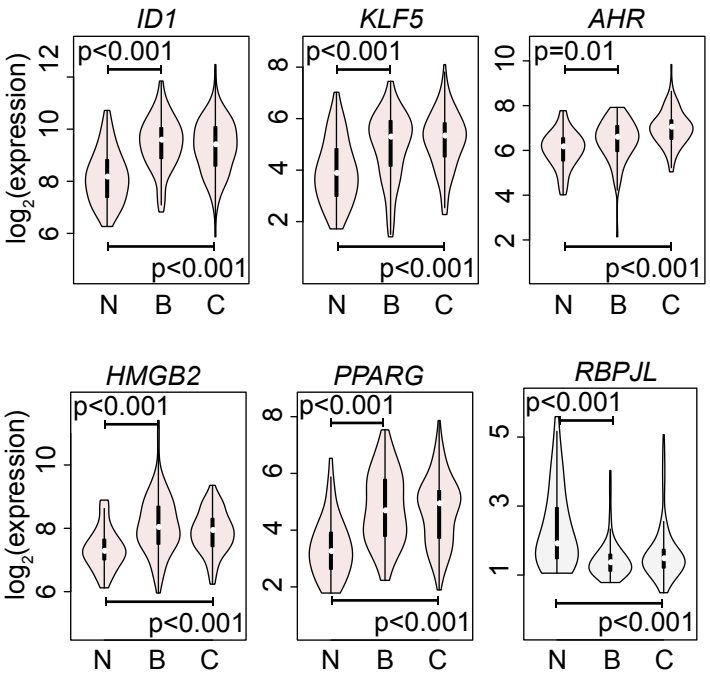


Figure 4- 8. Similarity of PDA-high TFs in PDA subtypes.
Analysis via GEO2R of PC1 genes from Figure 4-7 in the pancreatic microarray dataset GSE71729. N=Normal, B=Basal, C=Classical. P-values from two-sided, unpaired t-tests of B vs N and C vs. N.

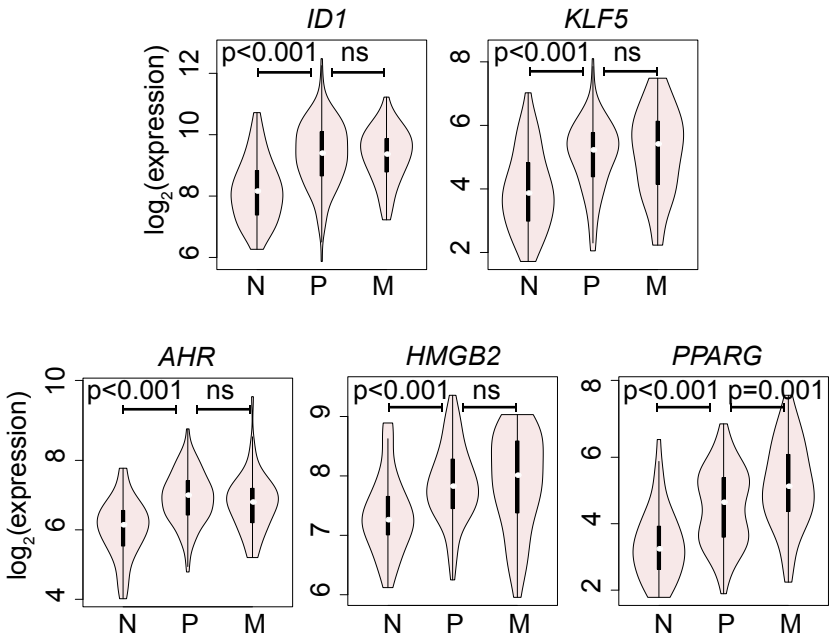


Figure 4- 9. Similarity of PDA-high TFs in primary and metastatic PDA.
Analysis via GEO2R of PDA-enriched PC1 genes (Figure 4-7) in pancreatic microarray data GSE71729. N=Normal, P=Primary, M=Metastasis. P-values from two-sided, unpaired t-tests of P vs N and M vs. P.

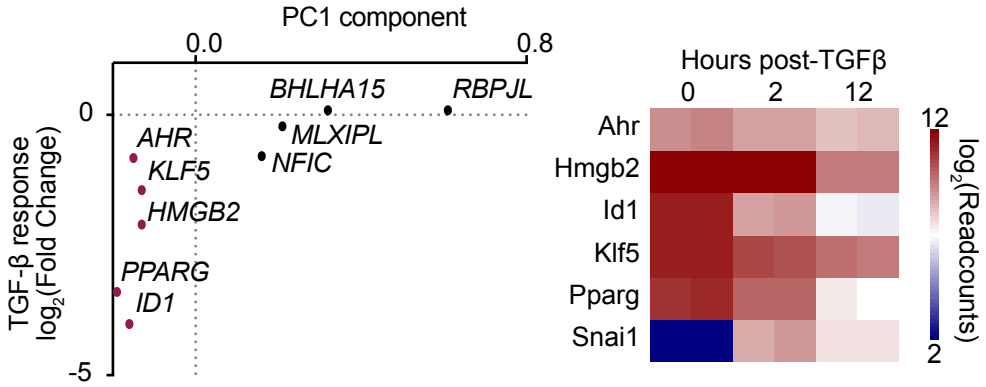


Figure 4- 10. Regulation of PDA-high TFs by TGFβ.
RNAseq of SMAD4-restored mouse PDA cells +100 pM TGFβ or 2.5 μM SB505124 (TGFBRI kinase inhibitor) 2 replicates per sample. *Left panel:* Analysis of PC1 genes from Figure 4-7 at 12h. *Right panel*: Analysis of PC1 genes and *Snai1* as an indicator of EMT.

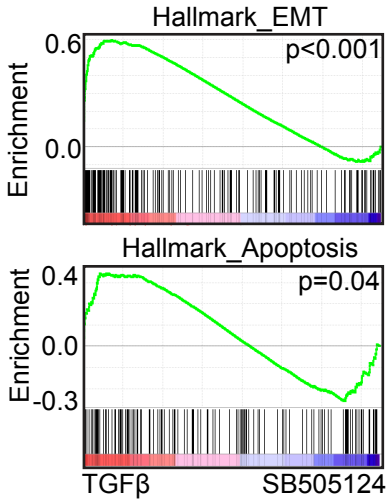


Figure 4- 11. Lethal EMT with repression of PDA-high TFs.
Gene set enrichment analysis of RNAseq of SMAD4-restored mouse PDA cells +100 pM TGFβ or 2.5 μM SB505124 (TGFBRI kinase inhibitor).

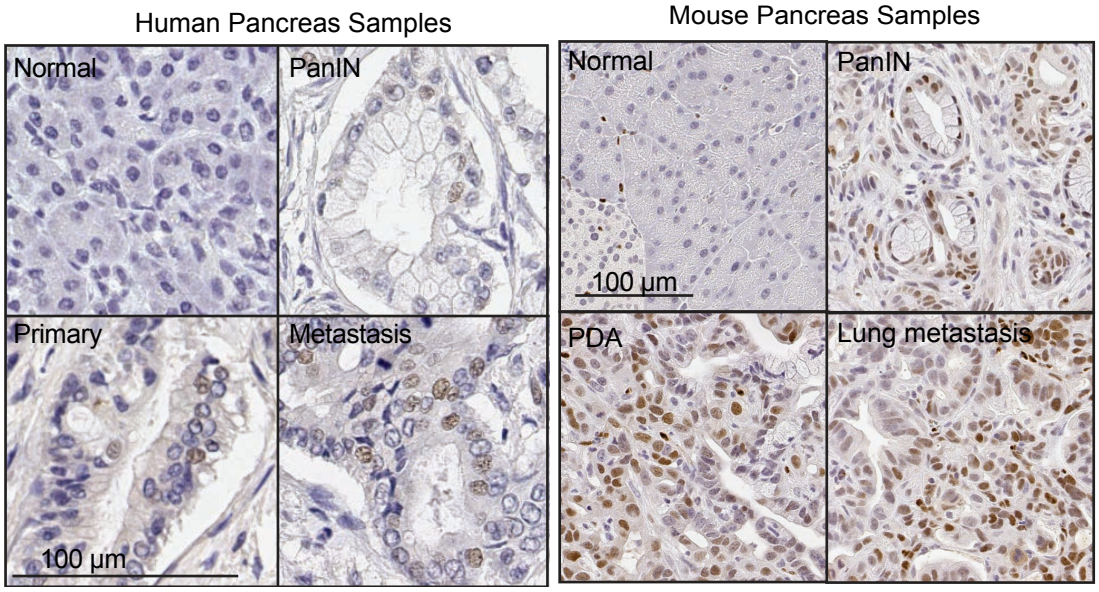


Figure 4- 12. ID1 expression in pancreatic tissue.
IHC for ID1 of normal pancreas and PDAs.

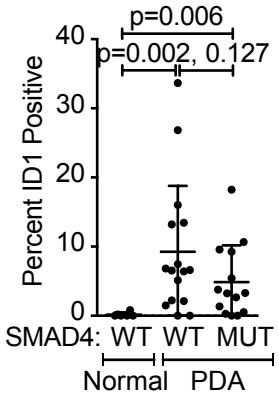


Figure 4- 13. ID1 staining in human pancreas.
Quantification of ID1 staining in human pancreatic samples. Two fields of view (average of 129 epithelial cells/field) per sample were quantified with blinding towards SMAD4 status. n=6 normal, n=16 SMAD4-WT PDAs, n=14 SMAD4-MUT PDAs. Two-sided unpaired t-test with Welch's correction for unequal variances.

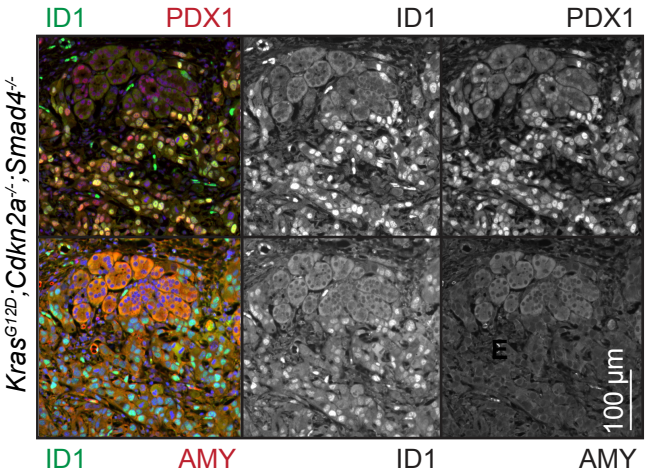


Figure 4- 14. Correlation of ID1 with markers of pancreatic differentiation. Autochthonous mouse PDA tumors were co-stained for ID1 and PDX1, a marker of pancreatic progenitors, or amylase (AMY), a marker of differentiated acinar cells.

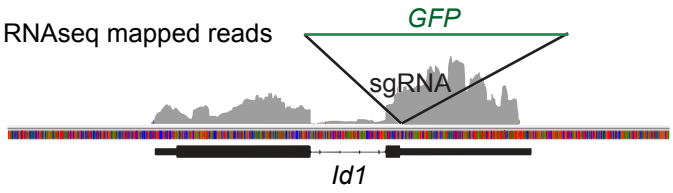


Figure 4- 15. Generation of endogenous ID1-GFP reporter.
ID1-GFP reporter was constructed using CRISPR/Cas9 to insert GFP fused to the 3' end of *Id1* at the endogenous locus in *Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$* mouse PDA cells.

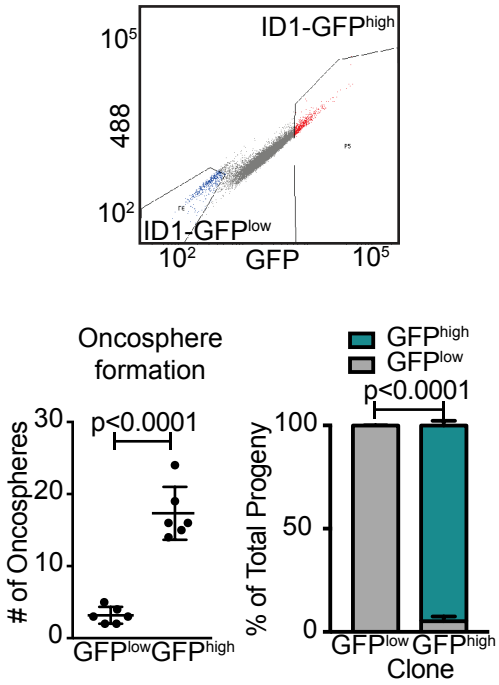


Figure 4- 16. Properties of ID1-GFP$^{high/low}$ PDA cells
Top panel: ID1-GFP *Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$* mouse PDA cells were sorted. *Bottom left panel:* Sorted cells were plated at 500 cells/well in low-adhesion 96-well plates, grown for 1 week, and counted. n=6 per group. *Bottom right panel:* Cells were plated at 1 cell/well in 96-well plates. Clones were expanded for 3 weeks and 10,000 cells were analyzed by flow cytometry. n=3 per group, 10,000 cells analyzed per clone.

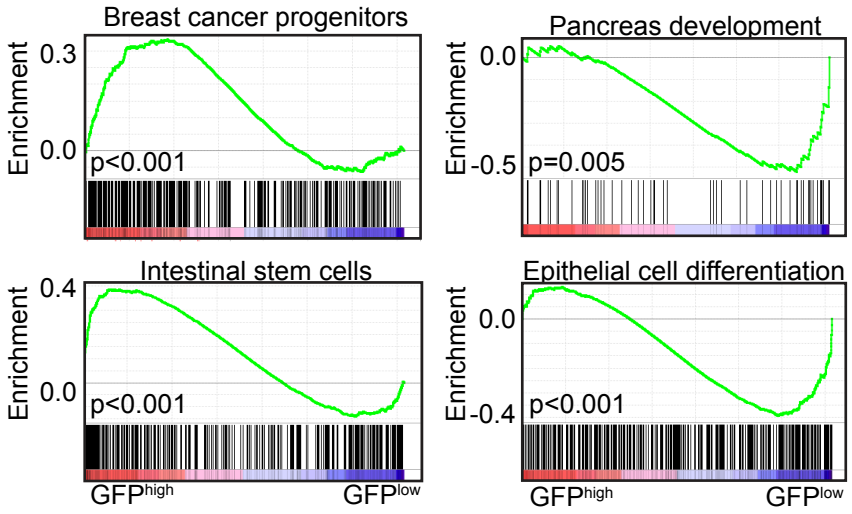


Figure 4- 17. Gene expression profiles of ID1-GFP$^{high/low}$ PDA cells.
SMAD4-restored ID1-GFP^{high} and ID1-GFP^{low} mouse PDA cells were collected for RNAseq. Differentially expressed genes were determined by DESeq2 and ranked by fold change in expression for gene set enrichment analysis. n=2 per group.

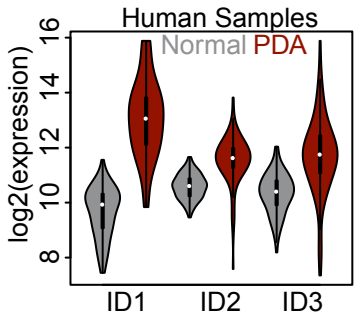


Figure 4- 18. ID1-3 expression in human tissues.
GTEx and ICGC RNAseq datasets were analyzed for ID expression levels.

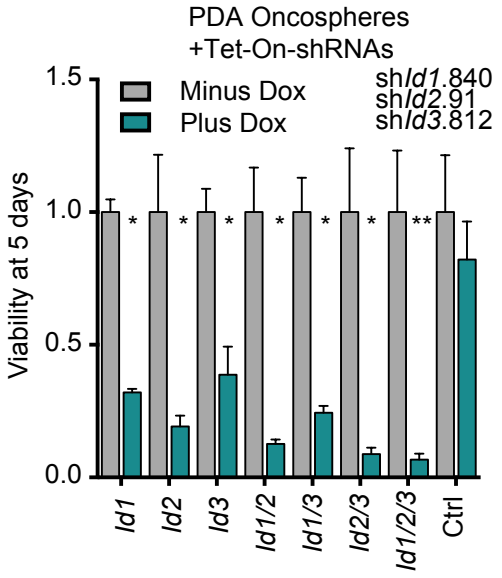


Figure 4- 19. Requirement for IDs in PDA oncospheres.
Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$ mouse PDA cells with Tet-On shRNAs were plated at 500 cells/well in low-adhesion 96-well plates ±doxycycline. Viability was measured by CellTiter-Glo 5 days after induction of indicated shRNAs. n=3 per group, mean±SD, all comparisons against Ctrl+Dox, * $p<0.01$, ** $p<0.001$.

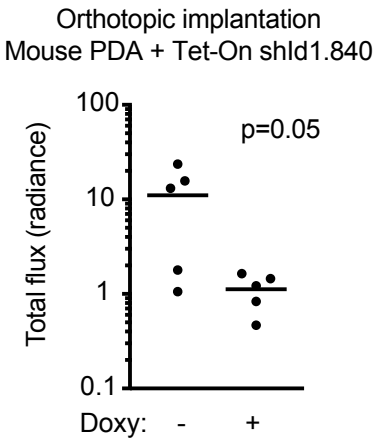


Figure 4- 20. ID1 in orthotopic PDAs.
Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$ mouse PDA cells with Tet-On shRNAs were orthotopically implanted at 500 cells per pancreas. Mice were started on doxycycline diet three days after implantation. n=5 per group.

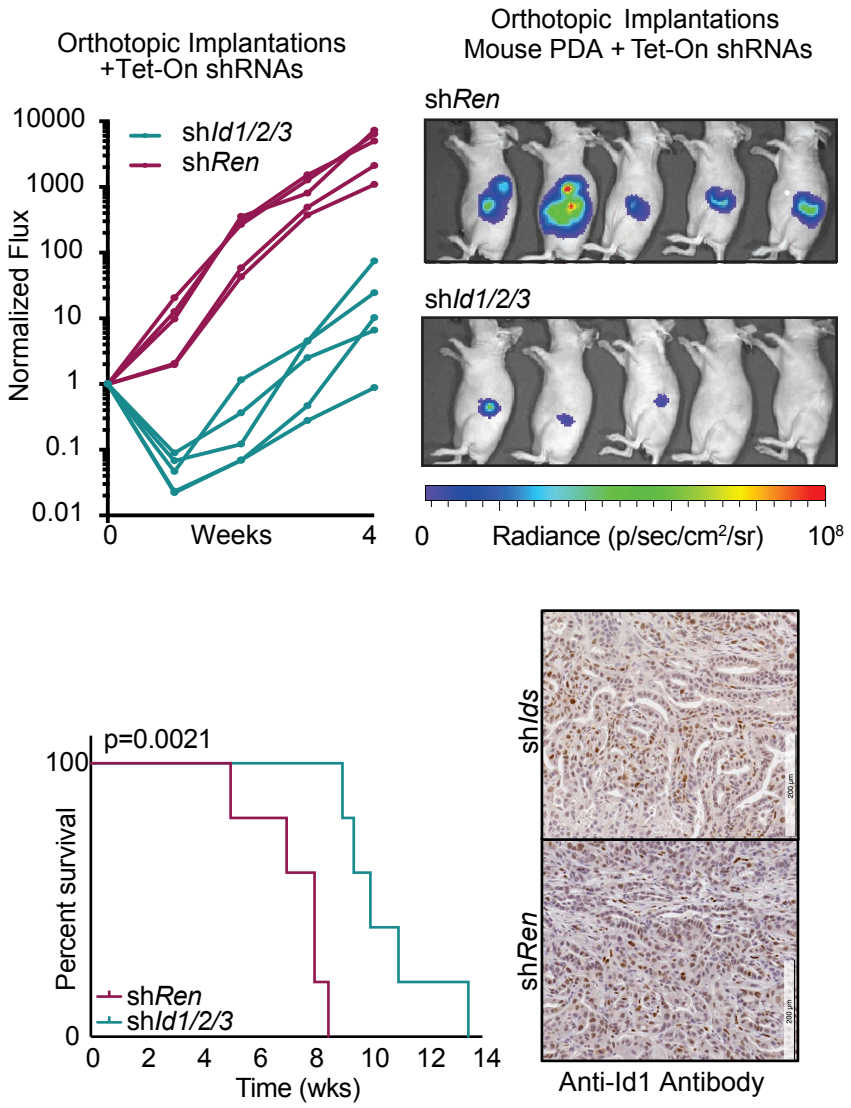


Figure 4- 21. ID proteins in orthotopic PDAs.
$Kras^{G12D}$; $Cdkn2a^{-/-}$; $Smad4^{-/-}$ mouse PDA cells with Tet-On shRNAs were orthotopically implanted at 500 cells per pancreas. Mice were started on doxycycline diet three days after implantation. n=5 per group *Top panel:* Bioluminescence was measured weekly. *Bottom left panel:* Survival analysis of mice with orthotopically implanted mouse PDA cells. Log-rank test. *Bottom right panel:* ID1 IHC in tumors collected at endpoint.

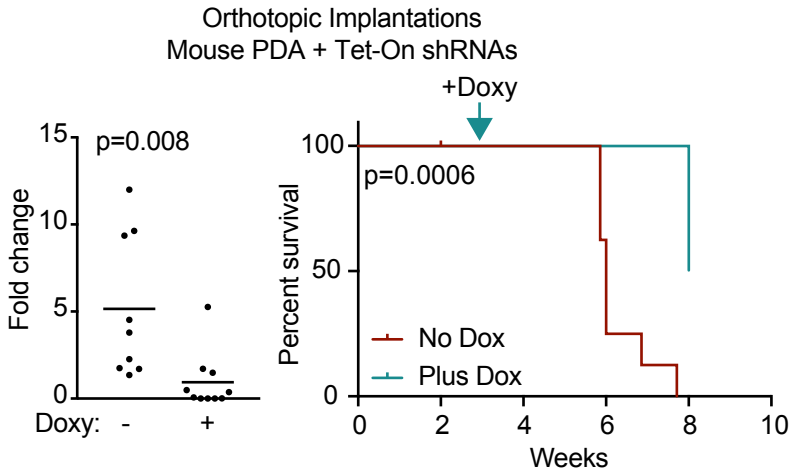


Figure 4- 22. ID proteins in PDA tumor maintenance.
Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$ mouse PDA cells with Tet-On shRNAs were orthotopically implanted at 500 cells per pancreas. Tumors were allowed to grow and bioluminescence was measured weekly. Mice were randomized to matched bioluminescent groups and started on doxycycline diet three weeks after implantation. n=5 per group. *Left panel:* bioluminescence fold change after 1 week of doxycycline treatment. *Right panel*: Kaplan-Meier, log-rank p-value.

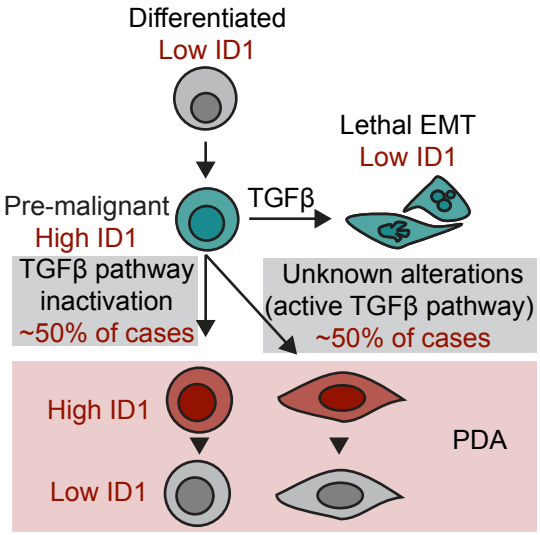


Figure 4- 23. Patterns of ID1 in normal pancreas and PDA.
ID1 is lowly expressed in the normal pancreas, up-regulated in the premalignant state, and repressed by TGFβ. PDAs escape TGFβ tumor suppression with high ID1, either through loss of TGFBR1/2 or SMAD2/3/4, or through other mechanisms downstream of the SMADs.

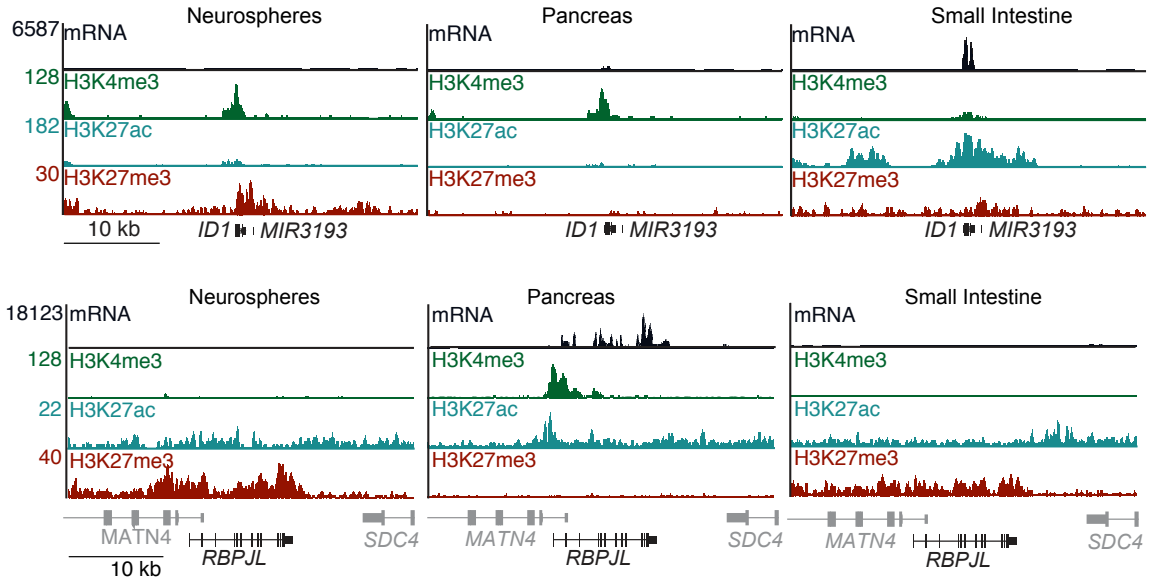


Figure 4- 24. ChIP-seq tracks from the Roadmap Epigenomics Project.

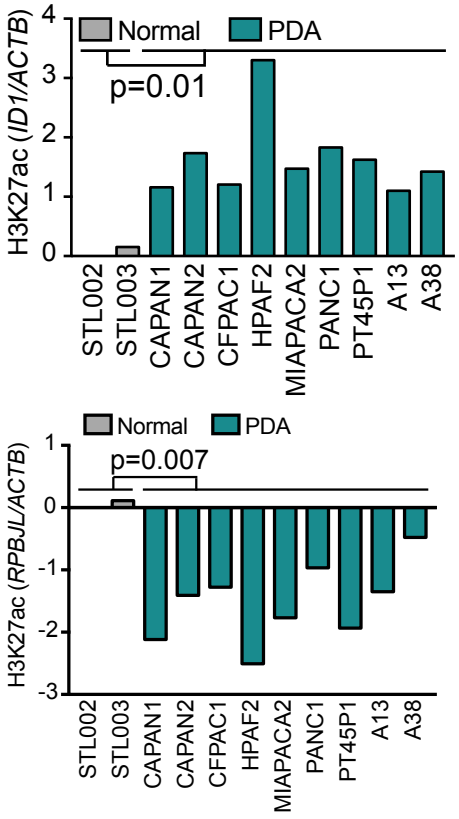


Figure 4- 25. *ID1* enhancer correlation with expression.
ChIP-seq signal at the *ID1* (*top panel*) or *RPBJL* (*bottom panel*) enhancer based on normal tissue samples from the Roadmap Epigenomics Project and PDA cells from GSE64557, normalized to signal at *ACTB*.

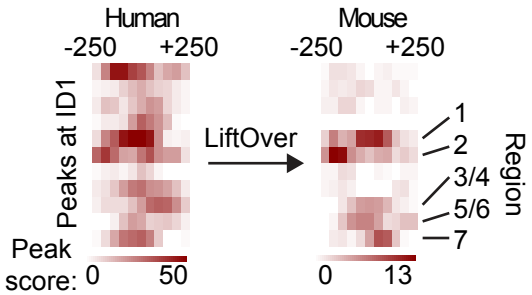


Figure 4- 26. Conservation of *ID1* accessible chromatin regions.
Heatmap representation of accessible chromatin detected by ATAC-seq at *ID1* in human Panc1 PDA cells (left) and at the corresponding regions in mouse *Kras^{G12D};Cdkn2a$^{-/-}$;Smad4$^{-/-}$* PDA cells. Each row represents 500 bp centered at a peak detected in Panc1 cells. Coordinates were converted from hg19 to mm10.

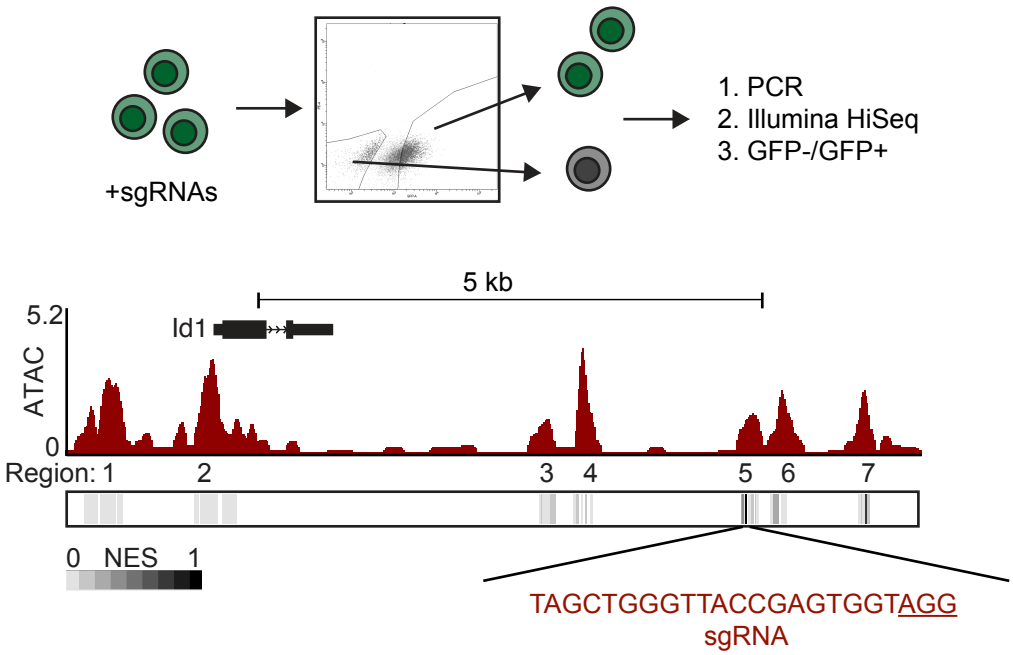


Figure 4- 27. Screen for loci contributing to high Id1 expression.
Top panel: CRISPR/CAS9 screen. Mouse PDA cells with the ID1-GFP reporter were transduced with 96 sgRNAs tiled across conserved regions of accessible chromatin using GuideScan to select guides with low off-target predictions (Perez et al., 2017). One week later, cells were sorted GFP^{high} vs GFP^{low} and replated. Sorting was repeated once and cells were collected for HiSeq. >1000x representation was maintained in 3 replicates. *Bottom panel:* ATAC-seq of mouse PDA (*Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$)* cells. Fold enrichment of sgRNAs in ID1-GFP^{low} vs ID1-GFP^{high} cells was determined and rank-based normalized enrichment score (NES) was calculated using RIGER (Luo et al., 2008). Sequence of the most highly enriched sgRNA is indicated.

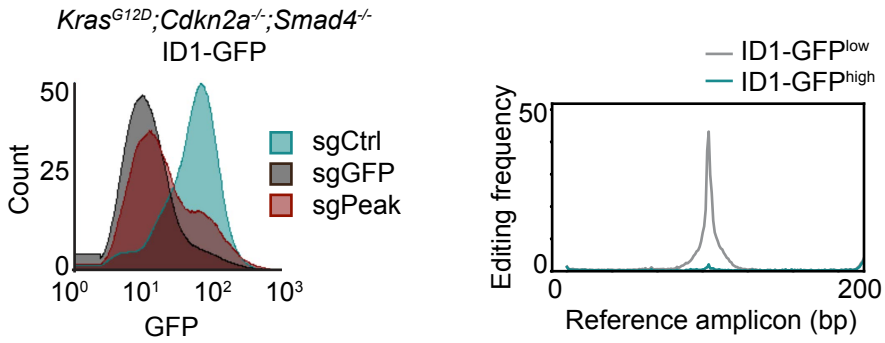


Figure 4- 28. Validation of locus necessary for high ID1 expression.
Left panel: Six sgRNAs were designed to flank Region 5 and transfected into ID1-GFP reporter cells for CRISPR/Cas9-mediated knockout of Region 5. One week after transfection and puromycin selection, cells were treated with SB505124 or TGFβ for 36h and then 10,000 cells per sample were analyzed by flow cytometry. *Right panel:* ID1-GFP reporter cells were transfected with the sgRNA indicated in Figure 4-27, sorted into ID1-GFP^{high} or ID1-GFP^{low} populations and amplicon-sequenced at the sgRNA cut site.

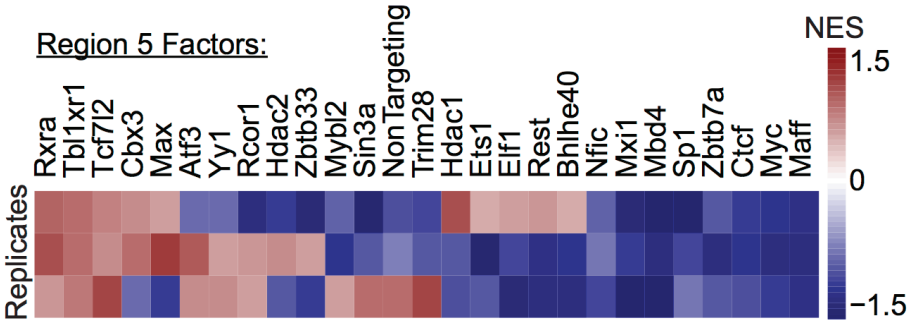


Figure 4- 29. Screen for factors contributing to high ID1 expression.
Factors reported to bind to Region 5 were identified using the ENCODE transcription factor ChIP track in UCSC genome browser. Three sgRNAs per gene were selected from the Gecko_v2 library. Screen was performed as described in Figure 4-27.

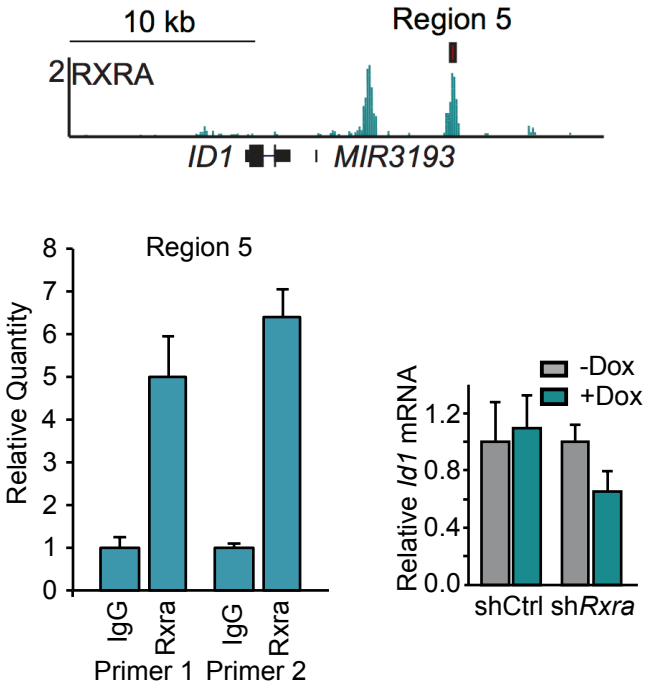


Figure 4- 30. Validation of RXRA in ID1 expression.
Top panel: ENCODE RXRA ChIP-seq track of HepG2 cells. *Bottom left panel:* RXRA ChIP was performed from *Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$* cells and quantification of Region 5 DNA was measured by qPCR. *Bottom right panel: Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$* mouse PDA cells with Tet-On shRNAs were plated ±doxycycline for 3 days and collected for qRT-PCR. Mean±range of 3 replicates, representative of 2 independent experiments.

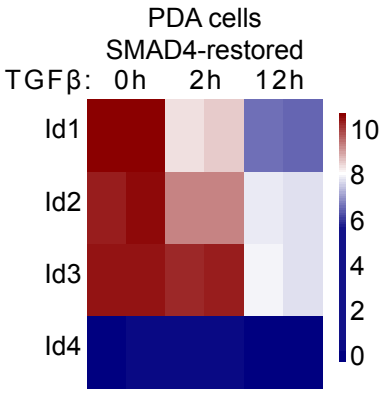


Figure 4- 31. *Id1-3* repression accompanies lethal EMT.
RNAseq of SMAD4-restored PDA cells treated with TGFβ for the indicated times.

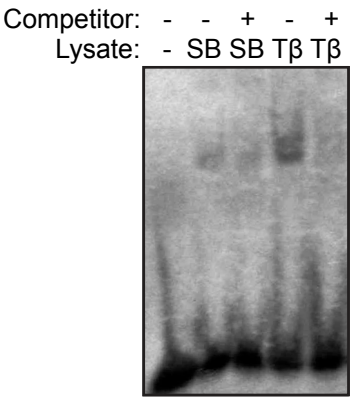


Figure 4- 32. E-protein binding in response to TGFβ
Electrophoretic mobility shift assay performed on SMAD4-restored PDA cells treated with 2.5 μM SB505124 or 100 pM TGFβ.

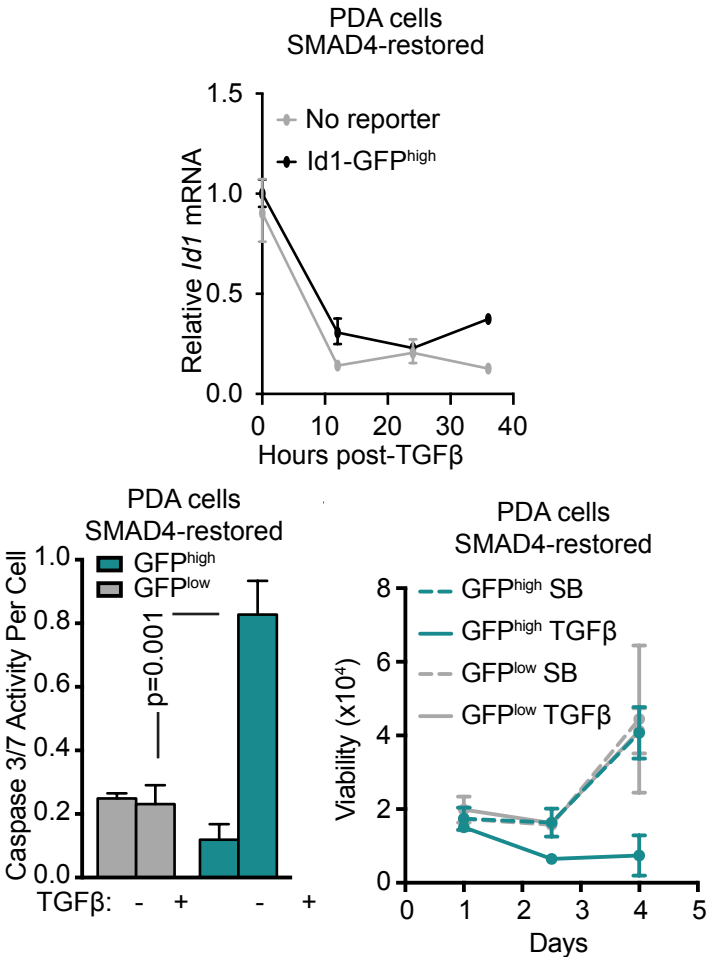


Figure 4- 33. TGFβ sensitivity of ID1-GFP^{high} PDA cells.
Top panel: qRT-PCR for *Id1* in SMAD4-restored PDA cells. Mean±range of 3 replicates. *Bottom panel:* SMAD4-restored mouse PDA cells were sorted and plated at 500 cells in 96-well format. Cells were treated for 36h with 100 pM TGFβ or 2.5 μM SB505124 (TGFBR1 kinase inhibitor). Caspase-Glo 3/7 and CellTiter-Glo were used to measure apoptosis and cell viability. n=3 per group, mean±SD.

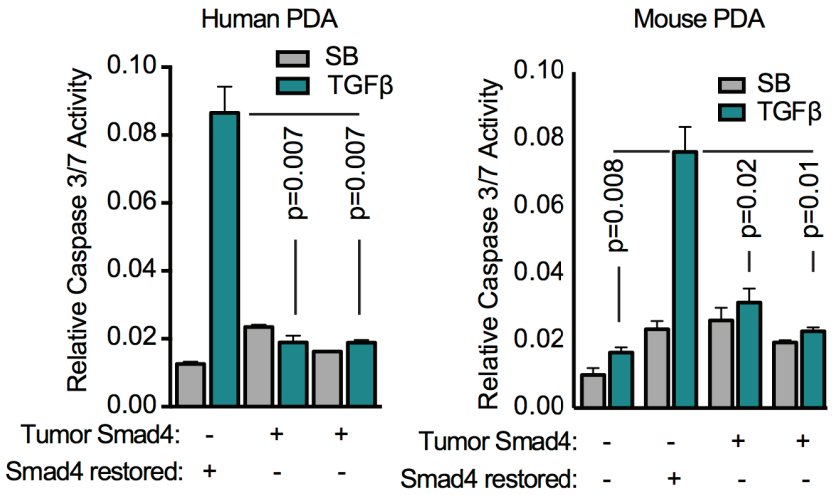


Figure 4- 34. TGFβ resistance of PDA cell lines.
Human and mouse PDA cells were treated with 2.5 μM SB505124 or 100 pM TGFβ for 36h and assayed using CaspaseGlo 3/7 and CellTiter-Glo. *Left panel:* SMAD4-restored BxPC3, MiaPaca2, and Panc1 human PDA cell lines. *Right panel: $Kras^{G12D}$; $Cdkn2a^{-/-}$; $Smad4^{-/-}$* mouse PDA cells (806), SMAD4-restored *$Kras^{G12D}$; $Cdkn2a^{-/-}$; $Smad4^{-/-}$* mouse PDA cells (806+S4), and *$Kras^{G12D}$; $Cdkn2a^{-/-}$* mouse PDA cells (NB44 and 4279).

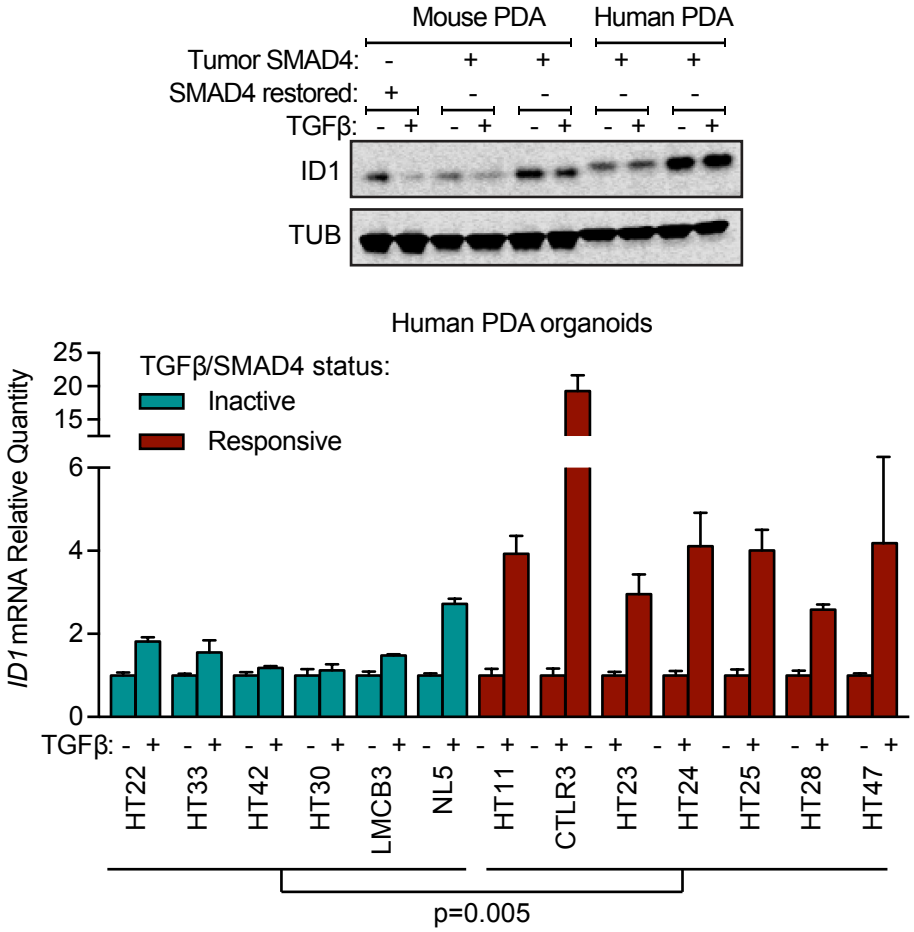


Figure 4- 35. ID1 regulation in PDA cell lines and organoids.
Top panel: 806+S4, NB44, 4279, MiaPaca2, and Panc1 cells were treated with 2.5 μM SB505124 or 100 pM TGFβ for 24h and collected for Western blot. *Bottom panel:* qRT-PCR for *ID1* in human PDA organoids 2 hours post TGFβ treatment. Mean±range of 3 replicates. Samples collected by NL & CD.

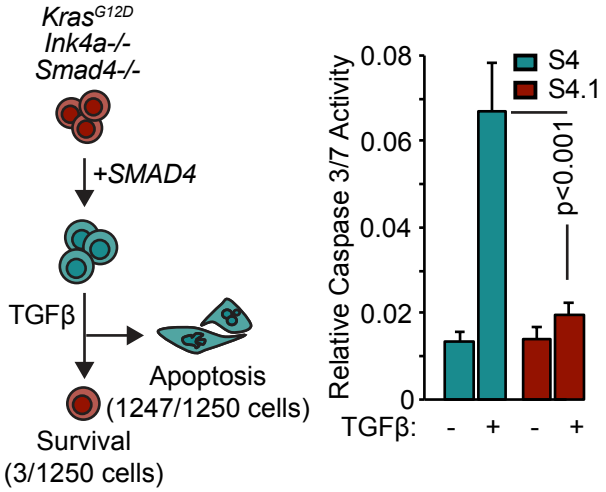


Figure 4- 36. Selection for syngeneic cells with differential TGFβ response.
Left panel: SMAD4-restored *Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$* mouse PDA cells were cultured in 100 pM TGFβ for 3 weeks and surviving clones were selected. *Right panel*: SMAD4-restored *Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$* mouse PDA cells (S4) and a resistant population were treated with 2.5 μM SB505124 or 100 pM TGFβ for 36h. Apoptosis was measured using CaspaseGlo 3/7 normalized to CellTiter-Glo. n=6 per group, mean±SD.

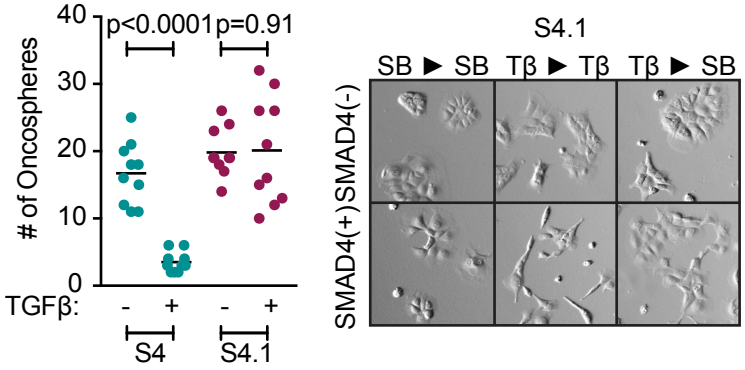


Figure 4- 37. ID1-inducing cells survive TGFβ with reversible EMT.
Left panel: S4 and S4.1 cells were seeded at 500 cells per well in low attachment plates with minimal growth factors and treated with 2.5 μM SB505124 or 100 pM TGFβ for 1 week. *Right panel*: S4.1 cells were seeded with or without doxycycline and then treated with 2.5 μM SB505124 for 48h, 100 pM TGFβ for 48h, or 100 pM TGFβ for 24h followed by 2.5 μM SB505124 for 24h.

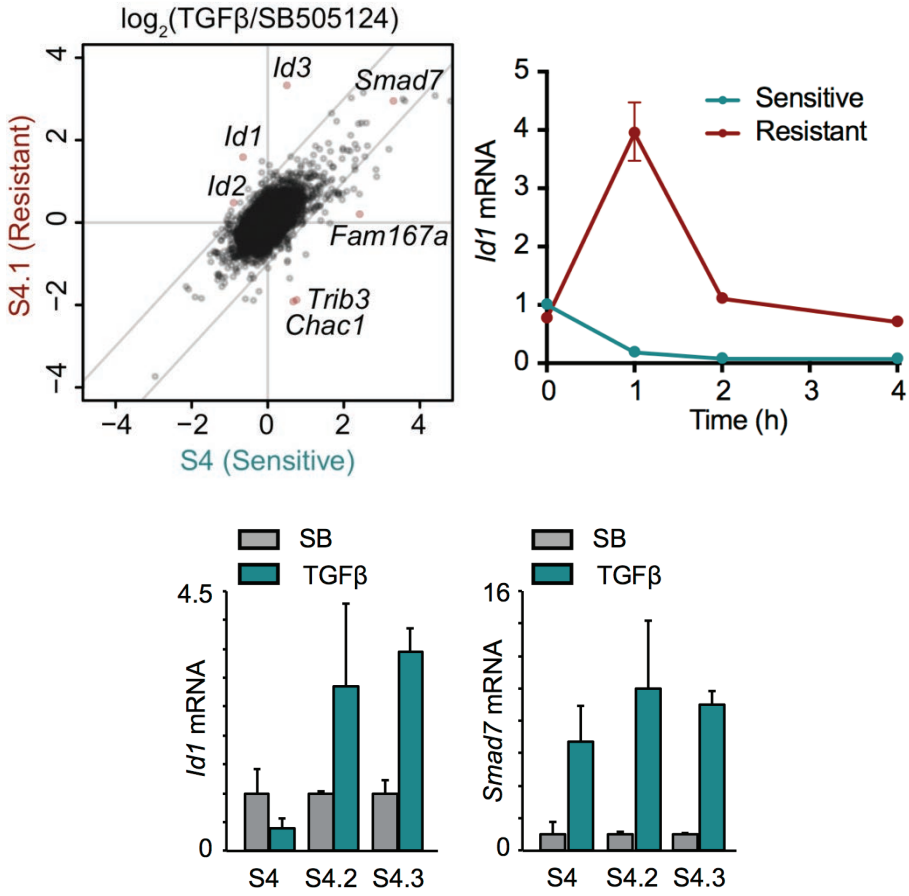


Figure 4- 38. ID1 dysregulation by TGFβ correlates with survival.
Top left panel: S4 and S4.1 cells were treated with 2.5 μM SB505124 or 100 pM TGFβ for 1.5h and collected for RNAseq, 2 replicates per sample. *Top right panel*: S4 and S4.1 cells were treated with 100 pM TGFβ for the indicated times and collected for qRT-PCR. Mean±range of 3 replicates, representative of 2 independent experiments. *Bottom panel:* S4, S4.2, and S4.3 cells were treated with 100 pM TGFβ for 1.5h and collected for qRT-PCR. Mean±range of 3 replicates, representative of 2 independent experiments.

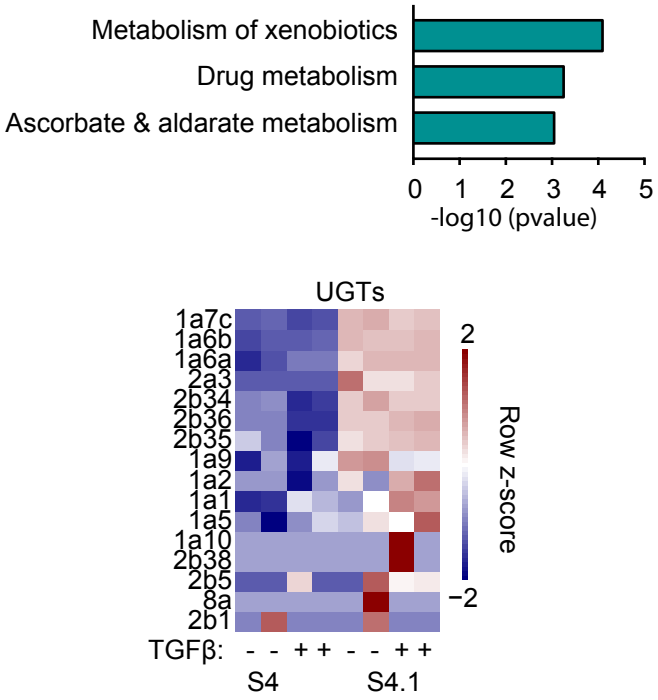


Figure 4- 39. Increased UGTs in selected TGFβ-resistant cells
S4 and S4.1 cells were treated with 2.5 μM SB505124 or 100 pM TGFβ for 1.5h and collected for RNAseq, 2 replicates per sample. *Top panel*: DAVID functional annotation of genes differentially expressed between S4 and S4.1 cells (average expression >10 reads, fold change >2, padj<0.05). *Bottom panel:* Expression of UGTs in RNAseq data.

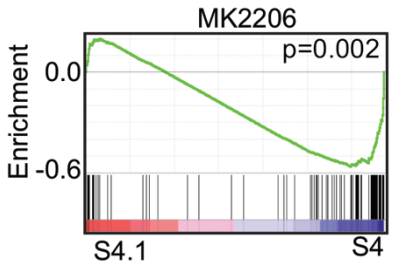


Figure 4- 40. Decreased efficacy of AKT inhibition in TGFβ-resistant cells
Gene set enrichment analysis of expression profiles of S4 and S4.1 cells. MK2206 signature created based on RNAseq of S4 cells treated with 2.5 μM MK2206 for 16h (average expression >10 reads, fold change >2, p_{adj}<0.05).

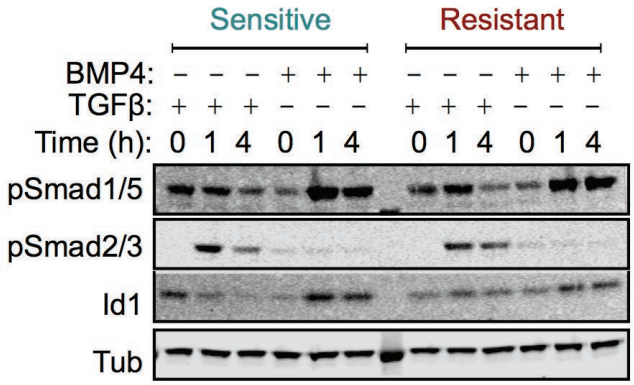


Figure 4- 41. Comparison of TGFβ and BMP responses of SMAD4-restored cells.
S4 and S4.1 cells were treated with 100 pM TGFβ or 20 ng/mL BMP4 for the indicated times and collected for Western blot.

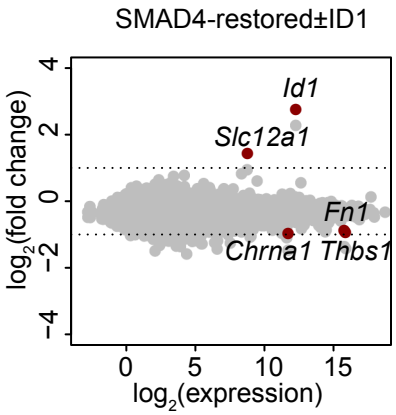


Figure 4- 42. Gene expression profile of ID1-enforced cells
SMAD4-restored *Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$* mouse PDA cells transduced with a Tet-On *Id1* construct were plated ± doxycycline for 12 hours and then treated with 2.5 μM SB505124 for 24h and collected for RNAseq.

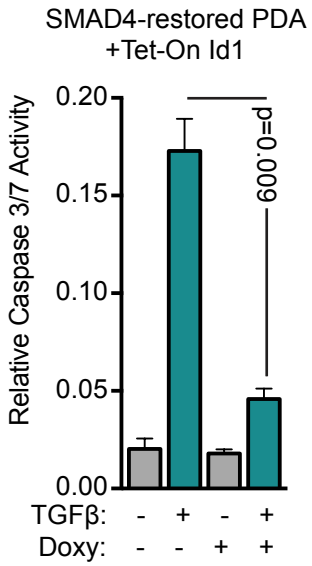


Figure 4- 43. Protection of SMAD4-restored cells from TGFβ by ID1.
SMAD4-restored *Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$* mouse PDA cells transduced with a Tet-On Id1 construct were plated ± doxycycline for 12 hours and then treated with 2.5 μM SB505124 or 100 pM TGFβ for 36h. Apoptosis was assayed using CaspaseGlo 3/7 and CellTiter-Glo. n=2, mean±SD.

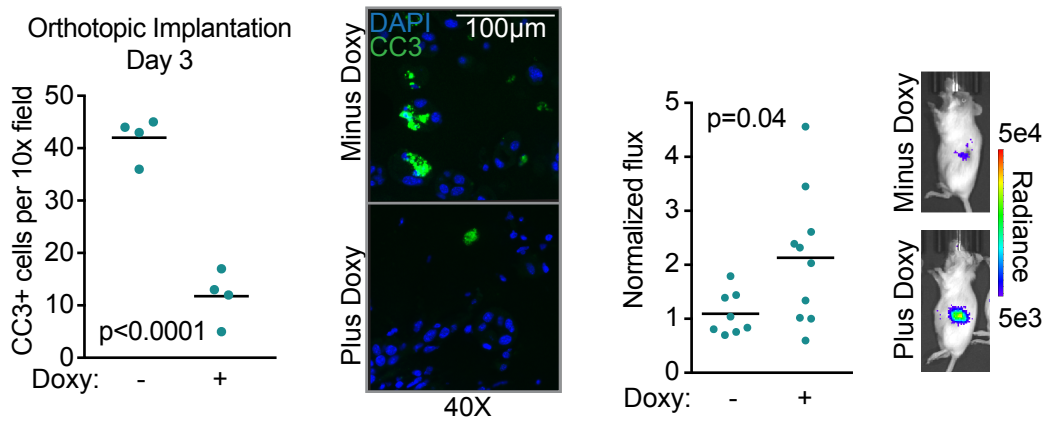


Figure 4- 44. ID1 promotes tumor formation.
SMAD4-restored *Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$* mouse PDA cells transduced with a Tet-On *Id1* construct were plated ± doxycycline for 12 hours and then orthotopically implanted into mice with caerulein-induced acute pancreatitis. Pancreata were collected at Day 3 for IF and bioluminescence was measured 1 week after implantation.

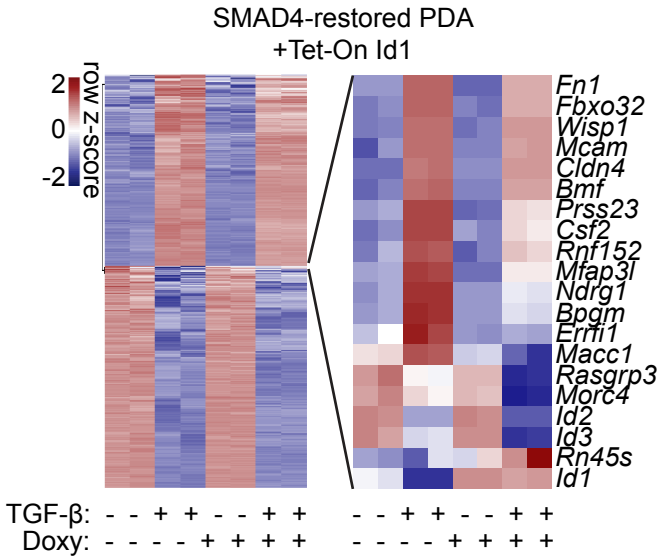


Figure 4- 45. TGFβ-mediated expression changes with ID1 enforcement.
SMAD4-restored Tet-On *Id1* mouse PDA cells were plated ±doxycycline for 12 hours and then treated with 2.5 μM SB505124 or 100 pM TGFβ for 24h and collected for RNAseq. 2 replicates per sample.

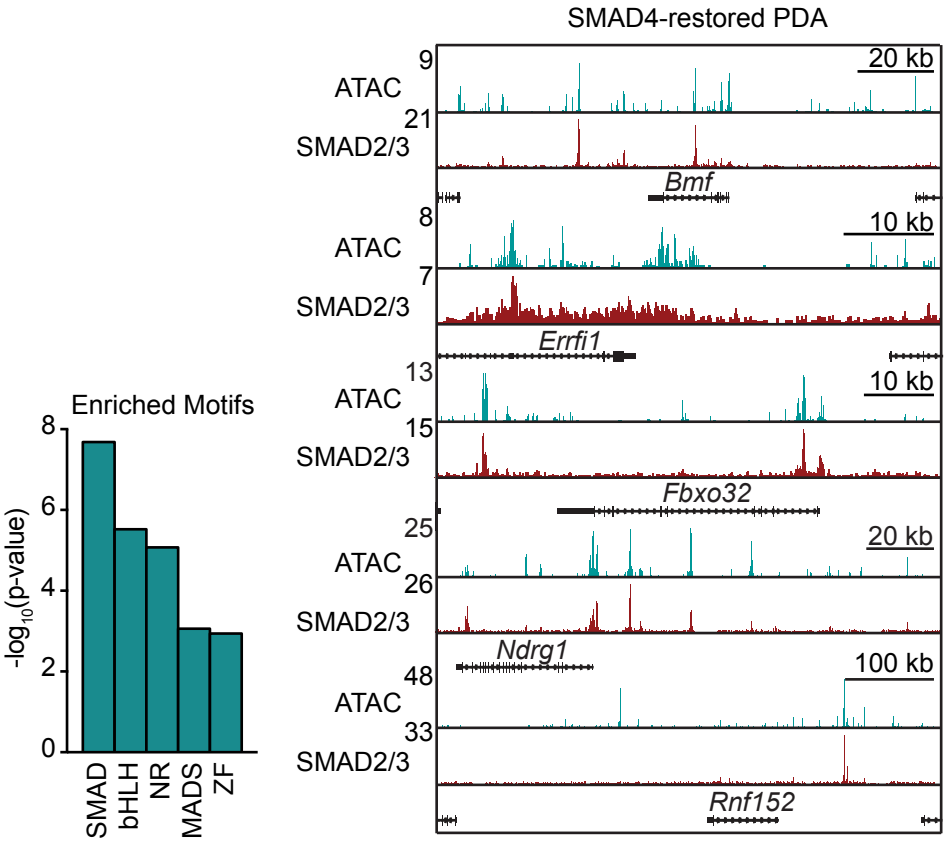


Figure 4- 46. Regulators of genes differentially regulated by TGFβ with Id1
Left panel: Accessible chromatin regions detected by ATAC-seq were assigned to the nearest transcription start site, and peaks associated with ID1-regulated, pro-apoptotic genes were selected. De novo motif enrichment versus a background list of ATAC-seq accessible regions was performed via Homer and compared with known motifs. *Right panel*: ATAC-seq and SMAD2/3 ChIP-seq of SMAD4-restored PDA cells treated with 100 pM TGFβ for 1.5h.

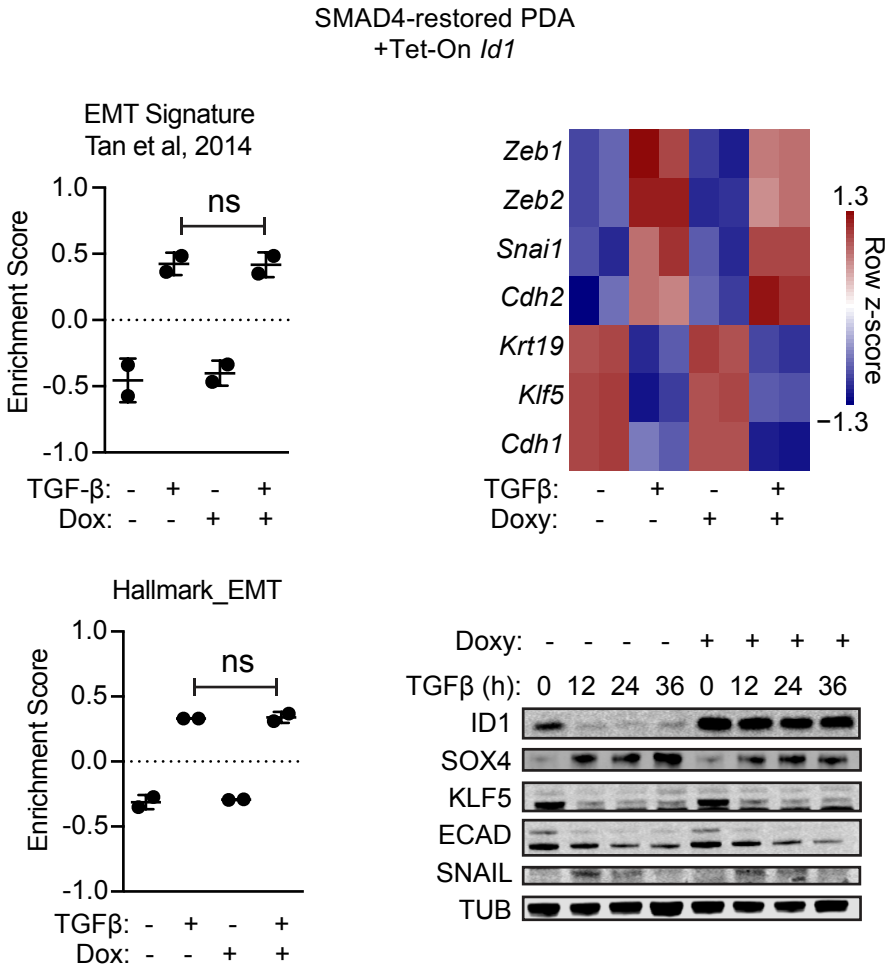


Figure 4- 47. Characterization of TGFβ –mediated EMT regulation by ID1.
Left panel: Gene set variation analysis of RNAseq data. n=2, mean±SD. *Top right panel:* EMT genes expressed in RNAseq data (Figure 4-45). *Bottom right panel:* SMAD4-restored Tet-On *Id1* mouse PDA cells were plated ±doxycycline for 12 hours and then treated with 2.5 μM SB505124 or 100 pM TGFβ and collected for Western blot.

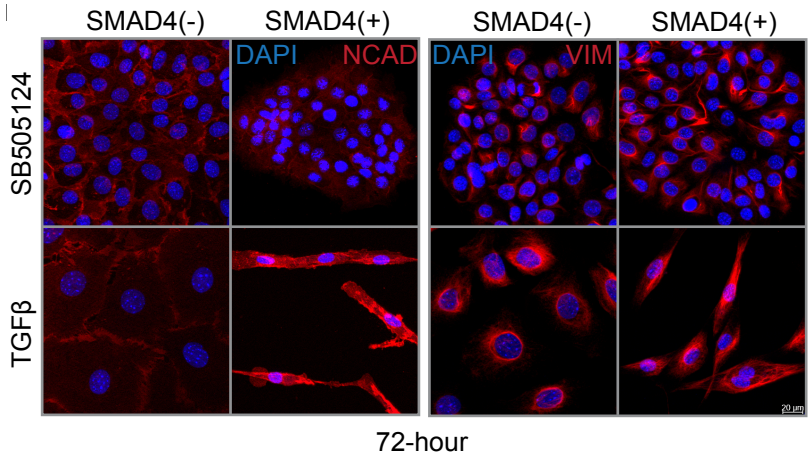


Figure 4- 48. Morphology of SMAD4(+) vs SMAD4(-) survivors of lethal EMT.
Kras^{G12D}; Cdkn2a$^{-/-}$; Smad4$^{-/-}$ mouse PDA cells and SMAD4-restored Tet-On *Id1* mouse PDA cells were treated with TGFβ for 72 hours and collected for immunofluorescence.

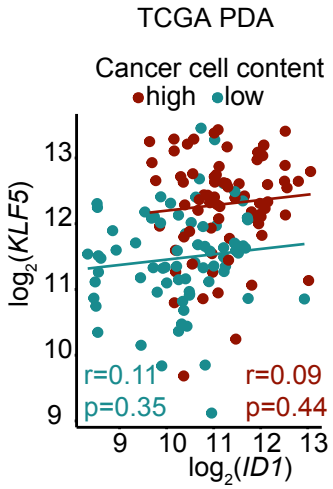


Figure 4- 49. Correlation of *KLF5* and *ID1* in human PDAs.
TCGA PDA RNAseq data were queried for *KLF5* and *ID1* expression levels and Pearson correlation coefficients were calculated.

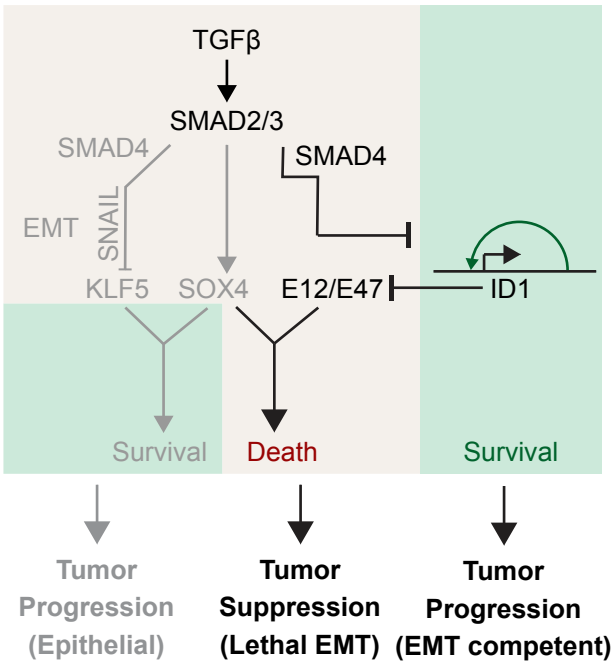


Figure 4- 50. ID1 in the context of lethal EMT.
ID1 favors cell survival and uncouples TGFβ-mediated lethality and EMT.

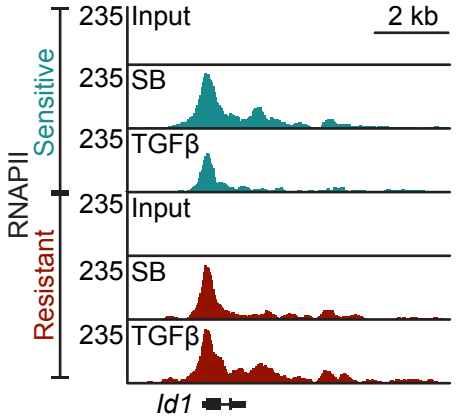


Figure 4- 51. RNAPII occupancy of the *Id1* locus.
S4 and S4.1 cells were treated with 2.5 μM SB505124 or 100 pM TGFβ for 1.5h and collected for RNAPII ChIPseq.

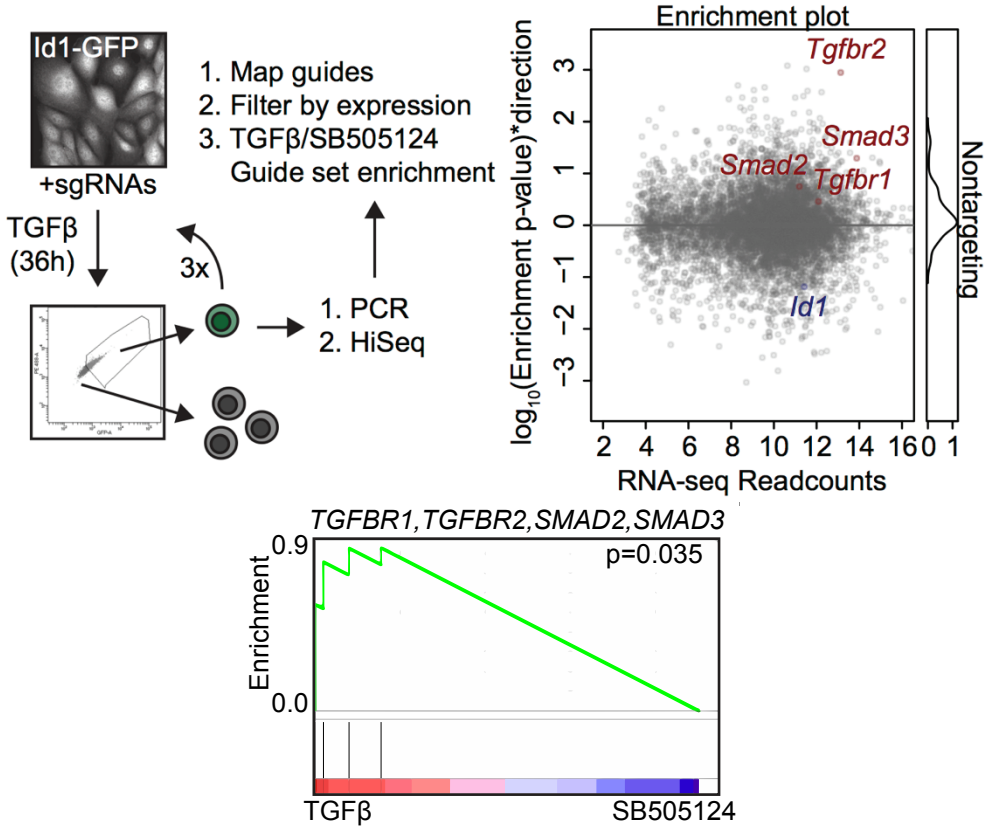


Figure 4- 52. Genome-wide screen for regulators of ID1 repression by TGFβ.
Top left panel: ID1-GFP^{high} SMAD4-restored PDA cells with CAS9 were transduced with the Gecko_v2 genome-wide sgRNA library, selected in puromycin for 1 week, and treated for 36h with 2.5 μM SB505124 or 100 pM TGFβ. ID1-GFP^{high} cells were sorted. After 3 rounds of selection, cells were collected for HiSeq. 200x representation was maintained throughout the screen. *Top right panel:* sgRNAs were mapped and sgRNAs targeting genes expressed in the PDA cells were selected. Gene enrichment in the TGFβ vs SB505124 samples was determined using the camera competitive gene set test. Direction * p-value was calculated as p-value * (±1), where +1=enriched and -1=depleted. n=2 per group. *Bottom panel:* Genes enriched in the genome-wide screen were ranked by direction * p-value and gene set enrichment analysis was performed for TGFβ pathway members.

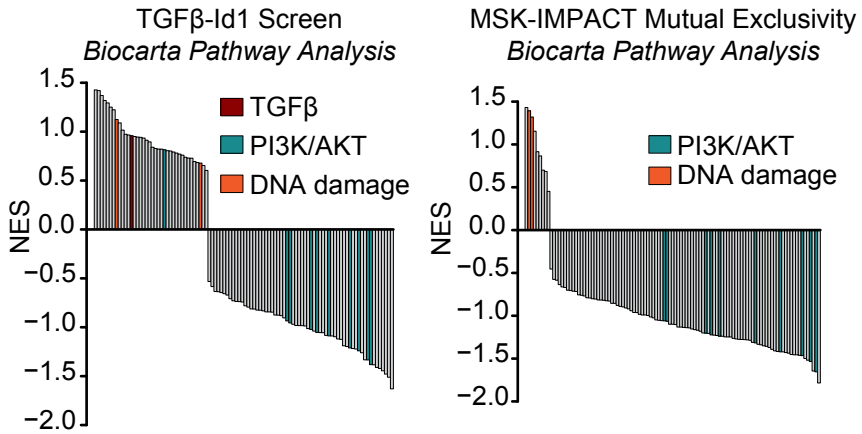


Figure 4- 53. Pathway analysis of candidates for regulators of ID1 repression
Left panel: Genes enriched in the genome-wide screen were ranked by (direction * p-value) and gene set enrichment analysis of Biocarta pathways was performed. *Right panel:* MSK-IMPACT mutations mutually exclusive to TGFβ pathway alterations were analyzed via cBioPortal and ranked by $-\log_{10}$(p-value) for mutual exclusivity. Gene set enrichment analysis of Biocarta pathways was performed.

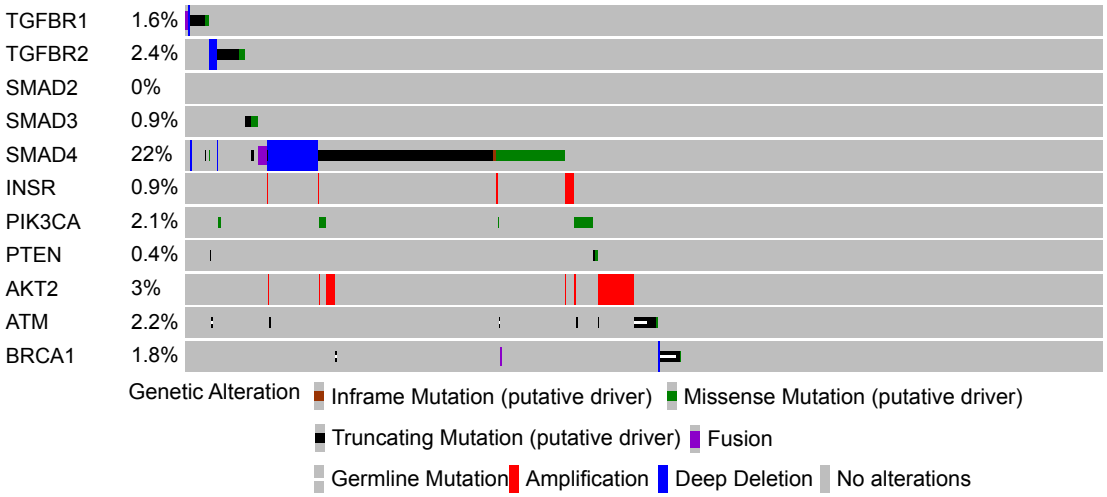


Figure 4- 54. Oncoprint of alterations in candidate modulators of ID1.
cBioportal oncoprint of MSK-IMPACT alterations in the TGFβ pathway and frequently altered PI3K/AKT signaling or DNA damage proteins in 1013 cases of PDA. Each column represents one case of PDA, unaltered and hypermutated cases are not included in the analysis.

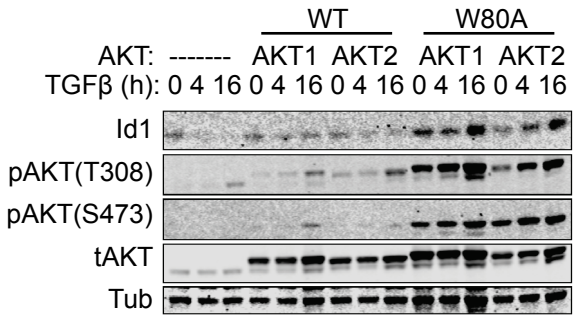


Figure 4- 55. AKT activation and TGFβ repression of ID1
SMAD4-restored mouse PDA cells transduced with the indicated constructs were treated with 2.5 μM MK2206 and 100 pM TGFβ for the indicated times and collected for Western blot.

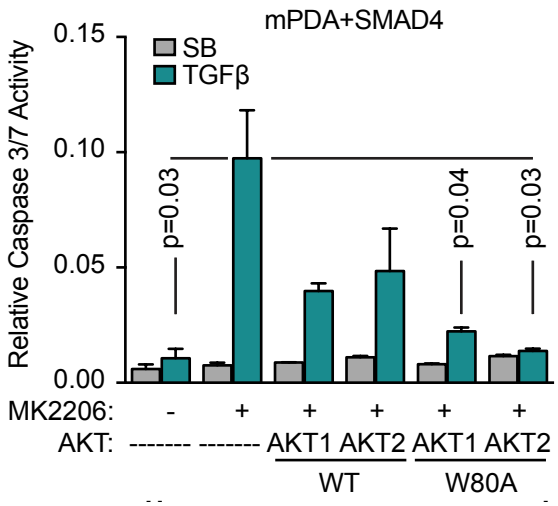


Figure 4- 56. AKT protection from TGFβ-mediated apoptosis.
SMAD4-restored mouse PDA cells transduced with the indicated constructs were treated ± 2.5 μM MK2206 and 2.5 μM SB505124 or 100 pM TGFβ for 36h. Apoptosis was determined by CaspaseGlo 3/7 and CellTiter-Glo.

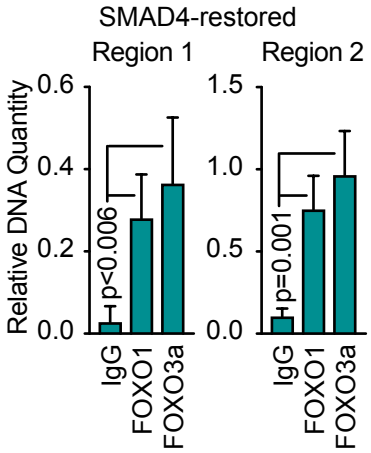


Figure 4- 57. FOXO1/3 at the *Id1* promoter and 5' enhancer
SMAD4-restored PDA cells were collected for ChIP. qPCR of the 5' accessible chromatin regions of the *Id1* locus was performed.

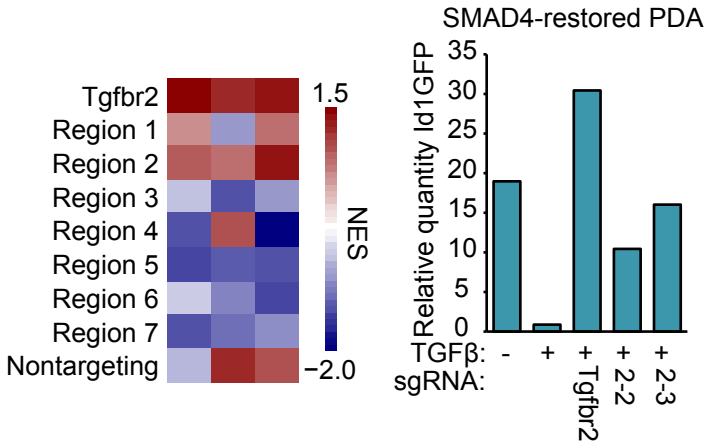


Figure 4- 58. Screen for loci contributing to TGFβ repression of *Id1*.
Left panel: CRISPR/CAS9 screen using the library and method described in 4-27, except that cells were treated with 2.5 μM SB505124 or 100 pM TGFβ for 36h prior to sorting for GFP^{high} populations. Analysis was performed with sgRNAs grouped by Region identified by ATACseq (see Figure 4-27). *Right panel:* Validation of the top two scoring guides from the screen, both from Region 2. Guides were introduced individually into SMAD4-restored reporter cells, treated with TGFβ for 36h, and analyzed by FACS.

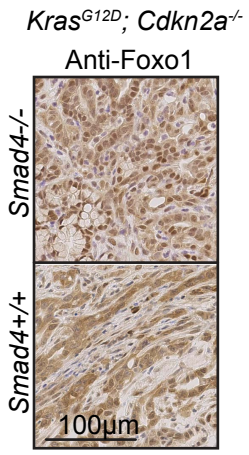


Figure 4- 59. FOXO1 localization in the mouse pancreas.
Paraffin-embedded sections of mouse PDAs were stained by IHC for FOXO1 localization. Nuclear localization = AKT inactive, cytoplasmic = AKT active.

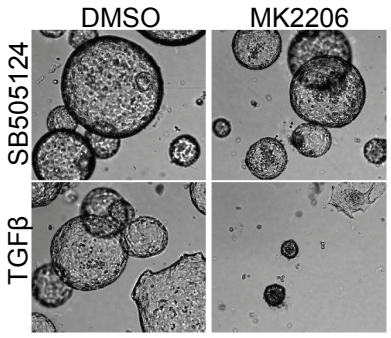


Figure 4- 60. Synergism of AKT inhibition and TGFβ.
Kras^{G12D}; Cdkn2a$^{-/-}$ mouse pancreatic organoids were treated ± 2.5 μM MK2206 (AKT inhibitor) and 2.5 μM SB505124 or 100 pM TGFβ for 5 days.

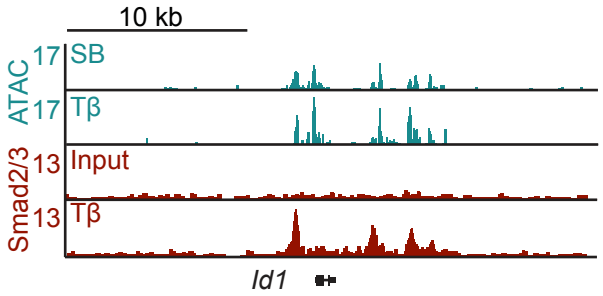


Figure 4- 61. ATACseq and SMAD2/3 ChIPseq of *Id1* locus.
SMAD4-restored mouse PDA cells were treated with 2.5 μM SB505124 or 100 pM TGFβ for 1.5h and collected for ATACseq and SMAD2/3 ChIPseq.

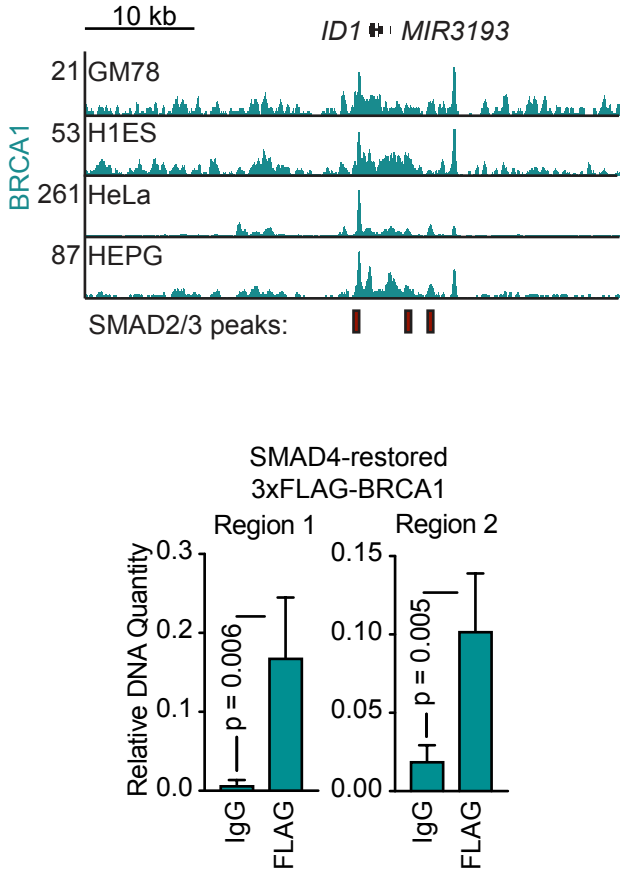


Figure 4- 62. BRCA1 at the ID1 locus.
Top panel: ENCODE BRCA1 ChIPseq tracks in human cell lines, visualized in UCSC genome browser. SMAD2/3 peaks are based on Figure 4-63 and converted from mouse to human coordinates using LiftOver.
Bottom panel: ChIP-PCR of FLAG-tagged BRCA1 in SMAD4-restored mouse PDA cells.

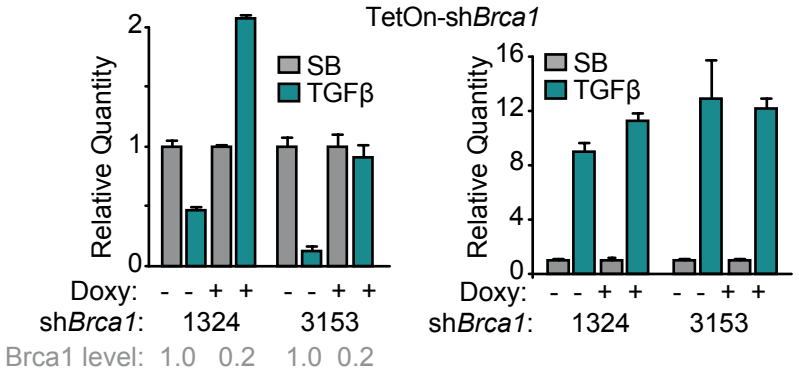


Figure 4- 63. BRCA1 and TGFβ regulation of *Id1*.
SMAD4-restored mouse PDA cells with Tet-On shRNAs were treated with 2.5 μM SB505124 or 100 pM TGFβ for 1.5h and collected for qRT-PCR. Mean±range of 3 replicates, representative of 2 independent experiments.

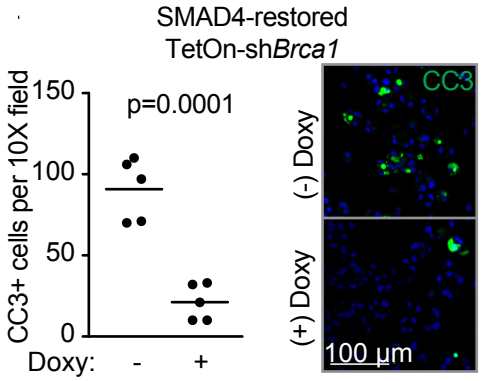


Figure 4- 64. BRCA1 and survival in a TGFβ-rich microenvironment.
SMAD4-restored mouse PDA cells with Tet-On shRNAs were implanted in mice with caerulein-induced acute pancreatitis. Pancreata were collected after 72h for IF.

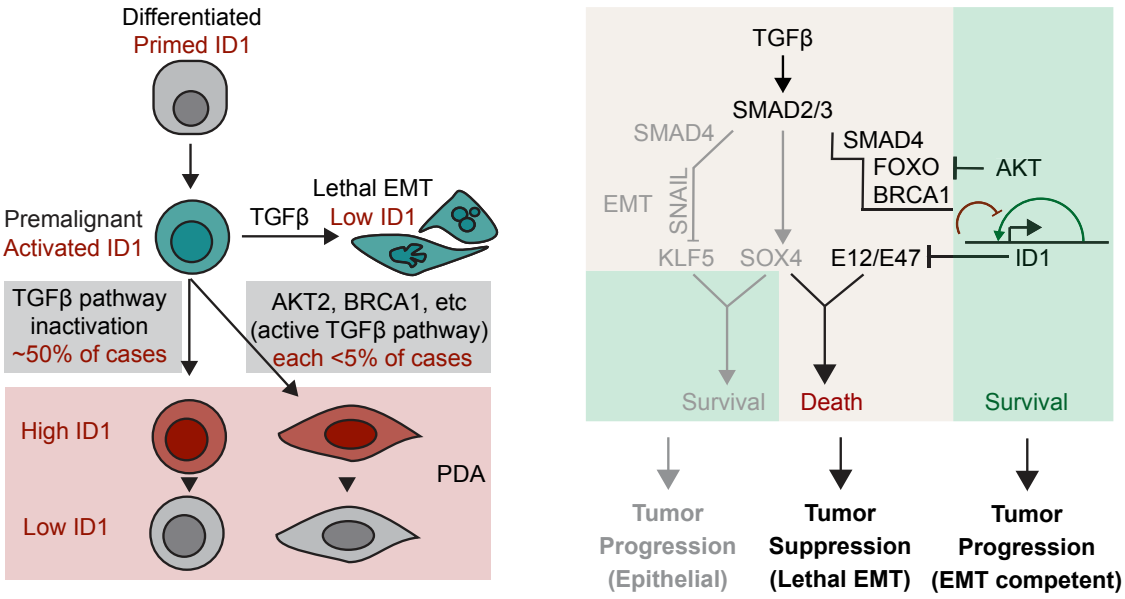


Figure 4- 65. Rewiring of ID1 regulation during PDA progression.

Bibliography

Adams, J. M., Harris, A. W., Pinkert, C. A., Corcoran, L. M., Alexander, W. S., Cory, S., Palmiter, R. D., and Brinster, R. L. (1985). The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. Nature *318*, 533-538.

Aguirre, A. J., Bardeesy, N., Sinha, M., Lopez, L., Tuveson, D. A., Horner, J., Redston, M. S., and DePinho, R. A. (2003). Activated Kras and Ink4a/Arf deficiency cooperate to produce metastatic pancreatic ductal adenocarcinoma. Genes & development *17*, 3112-3126.

Ahn, D. H., Li, J., Wei, L., Doyle, A., Marshall, J. L., Schaaf, L. J., Phelps, M. A., Villalona-Calero, M. A., and Bekaii-Saab, T. (2015). Results of an abbreviated phase-II study with the Akt Inhibitor MK-2206 in Patients with Advanced Biliary Cancer. Sci Rep *5*, 12122.

Almoguera, C., Shibata, D., Forrester, K., Martin, J., Arnheim, N., and Perucho, M. (1988). Most human carcinomas of the exocrine pancreas contain mutant c-K-ras genes. Cell *53*, 549-554.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome biology *11*, R106.

Andrews, S. (2010). FastQC. In Babraham Bioinformatics.

Anido, J., Saez-Borderias, A., Gonzalez-Junca, A., Rodon, L., Folch, G., Carmona, M. A., Prieto-Sanchez, R. M., Barba, I., Martinez-Saez, E., Prudkin, L.*, et al.* (2010). TGF-beta Receptor Inhibitors Target the CD44(high)/Id1(high) Glioma-Initiating Cell Population in Human Glioblastoma. Cancer cell *18*, 655-668.

Anzano, M. A., Roberts, A. B., Meyers, C. A., Komoriya, A., Lamb, L. C., Smith, J. M., and Sporn, M. B. (1982). Synergistic interaction of two classes of transforming growth factors from murine sarcoma cells. Cancer research *42*, 4776-4778.

Ardito, C. M., Gruner, B. M., Takeuchi, K. K., Lubeseder-Martellato, C., Teichmann, N., Mazur, P. K., Delgiorno, K. E., Carpenter, E. S., Halbrook, C. J., Hall, J. C.*, et al.* (2012). EGF receptor is required for KRAS-induced pancreatic tumorigenesis. Cancer cell *22*, 304-317.

Artner, I., Blanchi, B., Raum, J. C., Guo, M., Kaneko, T., Cordes, S., Sieweke, M., and Stein, R. (2007). MafB is required for islet beta cell maturation. Proceedings of the National Academy of Sciences of the United States of America *104*, 3853-3858.

Atchley, D. P., Albarracin, C. T., Lopez, A., Valero, V., Amos, C. I., Gonzalez-Angulo, A. M., Hortobagyi, G. N., and Arun, B. K. (2008). Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *26*, 4282-4288.

Bailey, M. H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M. C., Kim, J., Reardon, B.*, et al.* (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell *173*, 371-385 e318.

Bailey, P., Chang, D. K., Nones, K., Johns, A. L., Patch, A. M., Gingras, M. C., Miller, D. K., Christ, A. N., Bruxner, T. J., Quinn, M. C.*, et al.* (2016). Genomic analyses identify molecular subtypes of pancreatic cancer. Nature *531*, 47-52.

Bardeesy, N., Aguirre, A. J., Chu, G. C., Cheng, K. H., Lopez, L. V., Hezel, A. F., Feng, B., Brennan, C., Weissleder, R., Mahmood, U.*, et al.* (2006a). Both p16(Ink4a) and the p19(Arf)-p53 pathway constrain progression of pancreatic adenocarcinoma in the mouse. Proceedings of the National Academy of Sciences of the United States of America *103*, 5947-5952.

Bardeesy, N., Cheng, K. H., Berger, J. H., Chu, G. C., Pahler, J., Olson, P., Hezel, A. F., Horner, J., Lauwers, G. Y., Hanahan, D., and DePinho, R. A. (2006b). Smad4 is dispensable for normal pancreas development yet critical in progression and tumor biology of pancreas cancer. Genes & development *20*, 3130-3146.

Barrett, L. E., Granot, Z., Coker, C., Iavarone, A., Hambardzumyan, D., Holland, E. C., Nam, H. S., and Benezra, R. (2012). Self-renewal does not predict tumor growth potential in mouse models of high-grade glioma. Cancer cell *21*, 11-24.

Bartsch, D. K., Krysewski, K., Sina-Frey, M., Fendrich, V., Rieder, H., Langer, P., Kress, R., Schneider, M., Hahn, S. A., and Slater, E. P. (2006). Low frequency of CHEK2 mutations in familial pancreatic cancer. Familial cancer *5*, 305-308.

Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L., and Weintraub, H. (1990). The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell *61*, 49-59.

Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M. A., Beaudet, A. L., Ecker, J. R.*, et al.* (2010). The NIH Roadmap Epigenomics Mapping Consortium. Nature biotechnology *28*, 1045-1048.

Bhowmick, N. A., Chytil, A., Plieth, D., Gorska, A. E., Dumont, N., Shappell, S., Washington, M. K., Neilson, E. G., and Moses, H. L. (2004). TGF-beta signaling in fibroblasts modulates the oncogenic potential of adjacent epithelia. Science *303*, 848-851.

Bialkowska, A. B., Liu, Y., Nandan, M. O., and Yang, V. W. (2014). A colon cancer-derived mutant of Kruppel-like factor 5 (KLF5) is resistant to degradation by glycogen synthase kinase 3beta (GSK3beta) and the E3 ubiquitin ligase F-box and WD repeat domain-containing 7alpha (FBW7alpha). The Journal of biological chemistry *289*, 5997-6005.

Biankin, A. V., Morey, A. L., Lee, C. S., Kench, J. G., Biankin, S. A., Hook, H. C., Head, D. R., Hugh, T. B., Sutherland, R. L., and Henshall, S. M. (2002). DPC4/Smad4 expression and outcome in pancreatic ductal adenocarcinoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *20*, 4531-4542.

Birkenkamp, K. U., Essafi, A., van der Vos, K. E., da Costa, M., Hui, R. C., Holstege, F., Koenderman, L., Lam, E. W., and Coffer, P. J. (2007). FOXO3a induces differentiation of Bcr-Abl-transformed cells through transcriptional down-regulation of Id1. The Journal of biological chemistry *282*, 2211-2220.

Boj, S. F., Hwang, C. I., Baker, L. A., Chio, II, Engle, D. D., Corbo, V., Jager, M., Ponz-Sarvise, M., Tiriac, H., Spector, M. S.*, et al.* (2015). Organoid models of human and mouse ductal pancreatic cancer. Cell *160*, 324-338.

Bozic, I., Antal, T., Ohtsuki, H., Carter, H., Kim, D., Chen, S., Karchin, R., Kinzler, K. W., Vogelstein, B., and Nowak, M. A. (2010). Accumulation of driver and passenger mutations during tumor progression. Proceedings of the National Academy of Sciences of the United States of America *107*, 18545-18550.

Brennan, J., Lu, C. C., Norris, D. P., Rodriguez, T. A., Beddington, R. S., and Robertson, E. J. (2001). Nodal signalling in the epiblast patterns the early mouse embryo. Nature *411*, 965-969.

Brinster, R. L., Chen, H. Y., Messing, A., van Dyke, T., Levine, A. J., and Palmiter, R. D. (1984). Transgenic mice harboring SV40 T-antigen genes develop characteristic brain tumors. Cell *37*, 367-379.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., and Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature methods *10*, 1213-1218.

Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell *144*, 327-339.

Calon, A., Espinet, E., Palomo-Ponce, S., Tauriello, D. V., Iglesias, M., Cespedes, M. V., Sevillano, M., Nadal, C., Jung, P., Zhang, X. H.*, et al.* (2012). Dependency of colorectal cancer on a TGF-beta-driven program in stromal cells for metastasis initiation. Cancer cell *22*, 571-584.

Chapman, P. B., Hauschild, A., Robert, C., Haanen, J. B., Ascierto, P., Larkin, J., Dummer, R., Garbe, C., Testori, A., Maio, M.*, et al.* (2011). Improved survival with vemurafenib in melanoma with BRAF V600E mutation. The New England journal of medicine *364*, 2507-2516.

Chaudhary, J., Garland, W., and Salvador, R. (2009). A novel small molecule inhibitor of Id proteins (AGX-51) blocks cell survival in vitro and diminishes angiogenesis and tumor growth in vivo. The FASEB Journal *23*.

Chia, N. Y., Deng, N., Das, K., Huang, D., Hu, L., Zhu, Y., Lim, K. H., Lee, M. H., Wu, J., Sam, X. X.*, et al.* (2015). Regulatory crosstalk between lineage-survival oncogenes KLF5, GATA4 and GATA6 cooperatively promotes gastric cancer development. Gut *64*, 707-719.

Cho, H., Mu, J., Kim, J. K., Thorvaldsen, J. L., Chu, Q., Crenshaw, E. B., 3rd, Kaestner, K. H., Bartolomei, M. S., Shulman, G. I., and Birnbaum, M. J. (2001a). Insulin resistance and a diabetes mellitus-like syndrome in mice lacking the protein kinase Akt2 (PKB beta). Science *292*, 1728-1731.

Cho, H., Thorvaldsen, J. L., Chu, Q., Feng, F., and Birnbaum, M. J. (2001b). Akt1/PKBalpha is required for normal growth but dispensable for maintenance of glucose homeostasis in mice. The Journal of biological chemistry *276*, 38349-38352.

Christophorou, M. A., Ringshausen, I., Finch, A. J., Swigart, L. B., and Evan, G. I. (2006). The pathological response to DNA damage does not contribute to p53-mediated tumour suppression. Nature *443*, 214-217.

Cicchini, C., Filippini, D., Coen, S., Marchetti, A., Cavallari, C., Laudadio, I., Spagnoli, F. M., Alonzi, T., and Tripodi, M. (2006). Snail controls differentiation of hepatocytes by repressing HNF4alpha expression. Journal of cellular physiology *209*, 230-238.

Clem, R. J., Cheng, E. H., Karp, C. L., Kirsch, D. G., Ueno, K., Takahashi, A., Kastan, M. B., Griffin, D. E., Earnshaw, W. C., Veliuona, M. A., and Hardwick, J. M. (1998). Modulation of cell death by Bcl-XL through caspase interaction. Proceedings of the National Academy of Sciences of the United States of America *95*, 554-559.

Collado, M., Gil, J., Efeyan, A., Guerra, C., Schuhmacher, A. J., Barradas, M., Benguria, A., Zaballos, A., Flores, J. M., Barbacid, M.*, et al.* (2005). Tumour biology: senescence in premalignant tumours. Nature *436*, 642.

Collisson, E. A., Sadanandam, A., Olson, P., Gibb, W. J., Truitt, M., Gu, S., Cooc, J., Weinkle, J., Kim, G. E., Jakkula, L.*, et al.* (2011). Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nature medicine *17*, 500-503.

Connolly, E. C., Freimuth, J., and Akhurst, R. J. (2012). Complexities of TGF-beta targeted cancer therapy. International journal of biological sciences *8*, 964-978.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A.*, et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences of the United States of America *107*, 21931-21936.

Cui, W., Fowlis, D. J., Bryson, S., Duffie, E., Ireland, H., Balmain, A., and Akhurst, R. J. (1996). TGFbeta1 inhibits the formation of benign skin tumors, but enhances progression to invasive spindle carcinomas in transgenic mice. Cell *86*, 531-542.

Cybulski, C., Carrot-Zhang, J., Kluzniak, W., Rivera, B., Kashyap, A., Wokolorczyk, D., Giroux, S., Nadaf, J., Hamel, N., Zhang, S.*, et al.* (2015). Germline RECQL mutations are associated with breast cancer susceptibility. Nature genetics *47*, 643-646.

DaCosta Byfield, S., Major, C., Laping, N. J., and Roberts, A. B. (2004). SB-505124 is a selective inhibitor of transforming growth factor-beta type I receptors ALK4, ALK5, and ALK7. Molecular pharmacology *65*, 744-752.

David, C. J., Huang, Y. H., Chen, M., Su, J., Zou, Y., Bardeesy, N., Iacobuzio-Donahue, C. A., and Massague, J. (2016). TGF-beta Tumor Suppression through a Lethal EMT. Cell *164*, 1015-1030.

David, C. J., and Massague, J. (2018a). Contextual determinants of TGFbeta action in development, immunity and cancer. Nature reviews Molecular cell biology *19*, 419-435.

David, C. J., and Massague, J. (2018b). Contextual determinants of TGFβ action in development, immunity and cancer. Nature Reviews Molecular Cell Biology.

Davis, R. L., Weintraub, H., and Lassar, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987-1000.

DiGiuseppe, J. A., Hruban, R. H., Goodman, S. N., Polak, M., van den Berg, F. M., Allison, D. C., Cameron, J. L., and Offerhaus, G. J. (1994). Overexpression of p53 protein in adenocarcinoma of the pancreas. American journal of clinical pathology *101*, 684-688.

Ding, L., Bailey, M. H., Porta-Pardo, E., Thorsson, V., Colaprico, A., Bertrand, D., Gibbs, D. L., Weerasinghe, A., Huang, K. L., Tokheim, C*., et al.* (2018). Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics. Cell *173*, 305-320 e310.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Dong, X., Zhao, B., Iacob, R. E., Zhu, J., Koksal, A. C., Lu, C., Engen, J. R., and Springer, T. A. (2017). Force interacts with macromolecular structure in activation of TGF-beta. Nature *542*, 55-59.

Doulatov, S., Vo, L. T., Chou, S. S., Kim, P. G., Arora, N., Li, H., Hadland, B. K., Bernstein, I. D., Collins, J. J., Zon, L. I., and Daley, G. Q. (2013). Induction of multipotential hematopoietic progenitors from human pluripotent stem cells via respecification of lineage-restricted precursors. Cell stem cell *13*, 459-470.

Dow, L. E., O'Rourke, K. P., Simon, J., Tschaharganeh, D. F., van Es, J. H., Clevers, H., and Lowe, S. W. (2015). Apc Restoration Promotes Cellular Differentiation and Reestablishes Crypt Homeostasis in Colorectal Cancer. Cell *161*, 1539-1552.

Druker, B. J., Sawyers, C. L., Kantarjian, H., Resta, D. J., Reese, S. F., Ford, J. M., Capdeville, R., and Talpaz, M. (2001). Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the

blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. The New England journal of medicine *344*, 1038-1042.

Duda, D. G., Sunamura, M., Lefter, L. P., Furukawa, T., Yokoyama, T., Yatsuoka, T., Abe, T., Inoue, H., Motoi, F., Egawa, S.*, et al.* (2003). Restoration of SMAD4 by gene therapy reverses the invasive phenotype in pancreatic adenocarcinoma cells. Oncogene *22*, 6857-6864.

Evans, R. M., and Mangelsdorf, D. J. (2014). Nuclear Receptors, RXR, and the Big Bang. Cell *157*, 255-266.

Fellmann, C., Hoffmann, T., Sridhar, V., Hopfgartner, B., Muhar, M., Roth, M., Lai, D. Y., Barbosa, I. A., Kwon, J. S., Guan, Y.*, et al.* (2013). An optimized microRNA backbone for effective single-copy RNAi. Cell reports *5*, 1704-1713.

Friess, H., Yamanaka, Y., Buchler, M., Berger, H. G., Kobrin, M. S., Baldwin, R. L., and Korc, M. (1993). Enhanced expression of the type II transforming growth factor beta receptor in human pancreatic cancer cells without alteration of type III receptor expression. Cancer research *53*, 2704-2707.

Fusco, A., and Fedele, M. (2007). Roles of HMGA proteins in cancer. Nature reviews Cancer *7*, 899-910.

Gal, A., Sjoblom, T., Fedorova, L., Imreh, S., Beug, H., and Moustakas, A. (2008). Sustained TGF beta exposure suppresses Smad and non-Smad signalling in mammary epithelial cells, leading to EMT and inhibition of growth arrest and apoptosis. Oncogene *27*, 1218-1230.

Garraway, L. A., and Sellers, W. R. (2006). Lineage dependency and lineage-survival oncogenes in human cancer. Nature reviews Cancer *6*, 593-602.

Ghaleb, A. M., Nandan, M. O., Chanchevalap, S., Dalton, W. B., Hisamuddin, I. M., and Yang, V. W. (2005). Kruppel-like factors 4 and 5: the yin and yang regulators of cellular proliferation. Cell research *15*, 92-96.

Goggins, M., Shekher, M., Turnacioglu, K., Yeo, C. J., Hruban, R. H., and Kern, S. E. (1998). Genetic alterations of the transforming growth factor beta receptor genes in pancreatic and biliary adenocarcinomas. Cancer research *58*, 5329-5332.

Goto, Y., Nomura, M., Tanaka, K., Kondo, A., Morinaga, H., Okabe, T., Yanase, T., Nawata, H., Takayanagi, R., and Li, E. (2007). Genetic interactions between activin type IIB receptor and Smad2 genes in asymmetrical patterning of the thoracic organs and the development of pancreas islets. Developmental dynamics : an official publication of the American Association of Anatomists *236*, 2865-2874.

Greaves, M. (2015). When one mutation is all it takes. Cancer cell *27*, 433-434.

Grunseich, C., Wang, I. X., Watts, J. A., Burdick, J. T., Guber, R. D., Zhu, Z., Bruzel, A., Lanman, T., Chen, K., Schindler, A. B.*, et al.* (2018). Senataxin Mutation Reveals How R-Loops Promote Transcription by Blocking DNA Methylation at Gene Promoters. Molecular cell *69*, 426-437 e427.

GTEx (2013). The Genotype-Tissue Expression (GTEx) project. Nature genetics *45*, 580-585.

Guasch, G., Schober, M., Pasolli, H. A., Conn, E. B., Polak, L., and Fuchs, E. (2007). Loss of TGFbeta signaling destabilizes homeostasis and promotes squamous cell carcinomas in stratified epithelia. Cancer cell *12*, 313-327.

Guerra, C., Schuhmacher, A. J., Canamero, M., Grippo, P. J., Verdaguer, L., Perez-Gallego, L., Dubus, P., Sandgren, E. P., and Barbacid, M. (2007). Chronic pancreatitis is essential for induction of pancreatic ductal adenocarcinoma by K-Ras oncogenes in adult mice. Cancer cell *11*, 291-302.

Gupta, G. P., Perk, J., Acharyya, S., de Candia, P., Mittal, V., Todorova-Manova, K., Gerald, W. L., Brogi, E., Benezra, R., and Massague, J. (2007). ID genes mediate tumor reinitiation during breast cancer lung metastasis. Proceedings of the National Academy of Sciences of the United States of America *104*, 19506-19511.

Habbe, N., Shi, G., Meguid, R. A., Fendrich, V., Esni, F., Chen, H., Feldmann, G., Stoffers, D. A., Konieczny, S. F., Leach, S. D., and Maitra, A. (2008). Spontaneous induction of murine pancreatic intraepithelial neoplasia (mPanIN) by acinar cell targeting of oncogenic Kras in adult mice. Proceedings of the National Academy of Sciences of the United States of America *105*, 18913-18918.

Hahn, S. A., Schutte, M., Hoque, A. T., Moskaluk, C. A., da Costa, L. T., Rozenblum, E., Weinstein, C. L., Fischer, A., Yeo, C. J., Hruban, R. H., and Kern, S. E. (1996). DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. Science *271*, 350-353.

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

Hanzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC bioinformatics *14*, 7.

Hata, A., Lo, R. S., Wotton, D., Lagna, G., and Massague, J. (1997). Mutations increasing autoinhibition inactivate tumour suppressors Smad2 and Smad4. Nature *388*, 82-87.

Hatchi, E., Skourti-Stathaki, K., Ventz, S., Pinello, L., Yen, A., Kamieniarz-Gdula, K., Dimitrov, S., Pathania, S., McKinney, K. M., Eaton, M. L.*, et al.* (2015). BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. Molecular cell *57*, 636-647.

Heid, I., Steiger, K., Trajkovic-Arsic, M., Settles, M., Esswein, M. R., Erkan, M., Kleeff, J., Jager, C., Friess, H., Haller, B.*, et al.* (2017). Co-clinical Assessment of Tumor Cellularity in Pancreatic Cancer. Clinical cancer research : an official journal of the American Association for Cancer Research *23*, 1461-1470.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., and Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell *38*, 576-589.

Heldin, C. H., Vanlandewijck, M., and Moustakas, A. (2012). Regulation of EMT by TGFbeta in cancer. FEBS letters *586*, 1959-1970.

Henry, C., Close, A. F., and Buteau, J. (2014). A critical role for the neural zinc factor ST18 in pancreatic beta-cell apoptosis. The Journal of biological chemistry *289*, 8413-8419.

Hermann, P. C., Huber, S. L., Herrler, T., Aicher, A., Ellwart, J. W., Guba, M., Bruns, C. J., and Heeschen, C. (2007). Distinct populations of cancer stem cells determine tumor growth and metastatic activity in human pancreatic cancer. Cell Stem Cell *1*, 313-323.

Hill, R., Calvopina, J. H., Kim, C., Wang, Y., Dawson, D. W., Donahue, T. R., Dry, S., and Wu, H. (2010). PTEN loss accelerates KrasG12D-induced pancreatic cancer development. Cancer research *70*, 7114-7124.

Hingorani, S. R., Petricoin, E. F., Maitra, A., Rajapakse, V., King, C., Jacobetz, M. A., Ross, S., Conrads, T. P., Veenstra, T. D., Hitt, B. A.*, et al.* (2003). Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. Cancer cell *4*, 437-450.

Hingorani, S. R., Wang, L., Multani, A. S., Combs, C., Deramaudt, T. B., Hruban, R. H., Rustgi, A. K., Chang, S., and Tuveson, D. A. (2005). Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. Cancer cell *7*, 469-483.

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-Andre, V., Sigova, A. A., Hoke, H. A., and Young, R. A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934-947.

Hoadley, K. A., Yau, C., Hinoue, T., Wolf, D. M., Lazar, A. J., Drill, E., Shen, R., Taylor, A. M., Cherniack, A. D., Thorsson, V.*, et al.* (2018). Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell *173*, 291-304 e296.

Hodis, E., Watson, I. R., Kryukov, G. V., Arold, S. T., Imielinski, M., Theurillat, J. P., Nickerson, E., Auclair, D., Li, L., Place, C.*, et al.* (2012). A landscape of driver mutations in melanoma. Cell *150*, 251-263.

Holley, R. W. (1975). Control of growth of mammalian cells in cell culture. Nature *258*, 487-490.

Hopkins, S., Linderoth, E., Hantschel, O., Suarez-Henriques, P., Pilia, G., Kendrick, H., Smalley, M. J., Superti-Furga, G., and Ferby, I. (2012). Mig6 is a sensor of EGF receptor inactivation that directly activates c-Abl to induce apoptosis during epithelial homeostasis. Developmental cell *23*, 547-559.

Horiguchi, K., Shirakihara, T., Nakano, A., Imamura, T., Miyazono, K., and Saitoh, M. (2009). Role of Ras signaling in the induction of snail by transforming growth factor-beta. The Journal of biological chemistry *284*, 245-253.

Howe, J. R., Roth, S., Ringold, J. C., Summers, R. W., Jarvinen, H. J., Sistonen, P., Tomlinson, I. P., Houlston, R. S., Bevan, S., Mitros, F. A.*, et al.* (1998). Mutations in the SMAD4/DPC4 gene in juvenile polyposis. Science *280*, 1086-1088.

Hunter, K. E., Quick, M. L., Sadanandam, A., Hanahan, D., and Joyce, J. A. (2013). Identification and characterization of poorly differentiated invasive carcinomas in a mouse model of pancreatic neuroendocrine tumorigenesis. PloS one *8*, e64472.

Iacobuzio-Donahue, C. A., Fu, B., Yachida, S., Luo, M., Abe, H., Henderson, C. M., Vilardell, F., Wang, Z., Keller, J. W., Banerjee, P.*, et al.* (2009). DPC4 gene status of the primary carcinoma correlates with patterns of failure in patients with pancreatic cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *27*, 1806-1813.

Ijichi, H., Chytil, A., Gorska, A. E., Aakre, M. E., Fujitani, Y., Fujitani, S., Wright, C. V., and Moses, H. L. (2006). Aggressive pancreatic ductal adenocarcinoma in mice caused by pancreas-specific blockade of transforming growth factor-beta signaling in cooperation with active Kras expression. Genes & development *20*, 3147-3160.

Ikushima, H., Todo, T., Ino, Y., Takahashi, M., Miyazawa, K., and Miyazono, K. (2009). Autocrine TGF-beta signaling maintains tumorigenicity of glioma-initiating cells through Sry-related HMG-box factors. Cell stem cell *5*, 504-514.

Ischenko, I., Petrenko, O., and Hayman, M. J. (2014). Analysis of the tumor-initiating and metastatic capacity of PDX1-positive cells from the adult pancreas. Proceedings of the National Academy of Sciences of the United States of America *111*, 3466-3471.

Jen, Y., Manova, K., and Benezra, R. (1996). Expression patterns of Id1, Id2, and Id3 are highly related but distinct from that of Id4 during mouse embryogenesis. Developmental dynamics : an official publication of the American Association of Anatomists *207*, 235-252.

Jones, S., Zhang, X., Parsons, D. W., Lin, J. C., Leary, R. J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A.*, et al.* (2008). Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science *321*, 1801-1806.

Jonsson, J., Carlsson, L., Edlund, T., and Edlund, H. (1994). Insulin-promoter-factor 1 is required for pancreas development in mice. Nature *371*, 606-609.

Joung, J., Konermann, S., Gootenberg, J. S., Abudayyeh, O. O., Platt, R. J., Brigham, M. D., Sanjana, N. E., and Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. Nature protocols *12*, 828-863.

Kajno, E., McGraw, T. E., and Gonzalez, E. (2015). Development of a new model system to dissect isoform specific Akt signalling in adipocytes. The Biochemical journal *468*, 425-434.

Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J. F., Wyczalkowski, M. A*., et al.* (2013). Mutational landscape and significance across 12 major cancer types. Nature *502*, 333-339.

Kane, J. R., Laskin, W. B., Matkowskyj, K. A., Villa, C., and Yeldandi, A. V. (2014). Sarcomatoid (spindle cell) carcinoma of the pancreas: A case report and review of the literature. Oncology letters *7*, 245-249.

Kang, Y., Chen, C. R., and Massague, J. (2003). A self-enabling TGFbeta response coupled to stress signaling: Smad engages stress response factor ATF3 for Id1 repression in epithelial cells. Molecular cell *11*, 915-926.

Kang, Y., He, W., Tulley, S., Gupta, G. P., Serganova, I., Chen, C. R., Manova-Todorova, K., Blasberg, R., Gerald, W. L., and Massague, J. (2005). Breast cancer bone metastasis mediated by the Smad tumor suppressor pathway. Proceedings of the National Academy of Sciences of the United States of America *102*, 13909-13914.

Karlsson, O., Thor, S., Norberg, T., Ohlsson, H., and Edlund, T. (1990). Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain. Nature *344*, 879-882.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., and Haussler, D. (2002). The human genome browser at UCSC. Genome research *12*, 996-1006.

Kim, S. K., Hebrok, M., Li, E., Oh, S. P., Schrewe, H., Harmon, E. B., Lee, J. S., and Melton, D. A. (2000). Activin receptor patterning of foregut organogenesis. Genes & development *14*, 1866-1871.

Konturek, P. C., Dembinski, A., Warzecha, Z., Ihlm, A., Ceranowicz, P., Konturek, S. J., Stachura, J., and Hahn, E. G. (1998). Comparison of epidermal growth factor and transforming growth factor-beta1 expression in hormone-induced acute pancreatitis in rats. Digestion *59*, 110-119.

Kopp, J. L., von Figura, G., Mayes, E., Liu, F. F., Dubois, C. L., Morris, J. P. t., Pan, F. C., Akiyama, H., Wright, C. V., Jensen, K*., et al.* (2012). Identification of Sox9-dependent acinar-to-ductal

reprogramming as the principal mechanism for initiation of pancreatic ductal adenocarcinoma. Cancer cell *22*, 737-750.

Korchynskyi, O., and ten Dijke, P. (2002). Identification and functional characterization of distinct critically important bone morphogenetic protein-specific response elements in the Id1 promoter. The Journal of biological chemistry *277*, 4883-4891.

Krah, N. M., De La, O. J., Swift, G. H., Hoang, C. Q., Willet, S. G., Chen Pan, F., Cash, G. M., Bronner, M. P., Wright, C. V., MacDonald, R. J., and Murtaugh, L. C. (2015). The acinar differentiation determinant PTF1A inhibits initiation of pancreatic ductal adenocarcinoma. eLife *4*.

Kreider, B. L., Benezra, R., Rovera, G., and Kadesch, T. (1992). Inhibition of myeloid differentiation by the helix-loop-helix protein Id. Science *255*, 1700-1702.

Lagna, G., Hata, A., Hemmati-Brivanlou, A., and Massague, J. (1996). Partnership between DPC4 and SMAD proteins in TGF-beta signalling pathways. Nature *383*, 832-836.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods *9*, 357-359.

Lasorella, A., Benezra, R., and Iavarone, A. (2014). The ID proteins: master regulators of cancer stem cells and tumour aggressiveness. Nature reviews Cancer *14*, 77-91.

Lee, T. I., and Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. Cell *152*, 1237-1251.

Li, D., Frazier, M., Evans, D. B., Hess, K. R., Crane, C. H., Jiao, L., and Abbruzzese, J. L. (2006). Single nucleotide polymorphisms of RecQ1, RAD54L, and ATM genes are associated with reduced survival of pancreatic cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *24*, 1720-1728.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Liu, F., Pouponnot, C., and Massague, J. (1997). Dual role of the Smad4/DPC4 tumor suppressor in TGFbeta-inducible transcriptional complexes. Genes & development *11*, 3157-3167.

Loh, K. M., Ang, L. T., Zhang, J., Kumar, V., Ang, J., Auyeong, J. Q., Lee, K. L., Choo, S. H., Lim, C. Y., Nichane, M.*, et al.* (2014). Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations. Cell Stem Cell *14*, 237-252.

Lonardo, E., Hermann, P. C., Mueller, M. T., Huber, S., Balic, A., Miranda-Lorenzo, I., Zagorac, S., Alcala, S., Rodriguez-Arabaolaza, I., Ramirez, J. C.*, et al.* (2011). Nodal/Activin signaling drives self-renewal and tumorigenicity of pancreatic cancer stem cells and provides a target for combined drug therapy. Cell Stem Cell *9*, 433-446.

Lopes, R., Korkmaz, G., and Agami, R. (2016). Applying CRISPR-Cas9 tools to identify and characterize transcriptional enhancers. Nature reviews Molecular cell biology *17*, 597-604.

Lopez-Rovira, T., Chalaux, E., Massague, J., Rosa, J. L., and Ventura, F. (2002). Direct binding of Smad1 and Smad4 to two distinct motifs mediates bone morphogenetic protein-specific transcriptional activation of Id1 gene. The Journal of biological chemistry *277*, 3176-3185.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology *15*, 550.

Lyden, D., Young, A. Z., Zagzag, D., Yan, W., Gerald, W., O'Reilly, R., Bader, B. L., Hynes, R. O., Zhuang, Y., Manova, K., and Benezra, R. (1999). Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts. Nature *401*, 670-677.

Macias, M. J., Martin-Malpartida, P., and Massague, J. (2015). Structural determinants of Smad function in TGF-beta signaling. Trends in biochemical sciences *40*, 296-308.

Mariathasan, S., Turley, S. J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., Kadel, E. E., III, Koeppen, H., Astarita, J. L., Cubas, R.*, et al.* (2018). TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature *554*, 544-548.

Martin-Malpartida, P., Batet, M., Kaczmarska, Z., Freier, R., Gomes, T., Aragon, E., Zou, Y., Wang, Q., Xi, Q., Ruiz, L.*, et al.* (2017). Structural basis for genome wide recognition of 5-bp GC motifs by SMAD transcription factors. Nature communications *8*, 2070.

Massague, J. (1998). TGF-beta signal transduction. Annual review of biochemistry *67*, 753-791.

Massague, J. (2008). TGFbeta in Cancer. Cell *134*, 215-230.

Massague, J. (2012). TGFbeta signalling in context. Nature reviews Molecular cell biology *13*, 616-630.

Masui, T., Swift, G. H., Deering, T., Shen, C., Coats, W. S., Long, Q., Elsasser, H. P., Magnuson, M. A., and MacDonald, R. J. (2010). Replacement of Rbpj with Rbpjl in the PTF1 complex controls the final maturation of pancreatic acinar cells. Gastroenterology *139*, 270-280.

McWilliams, R. R., Bamlet, W. R., Cunningham, J. M., Goode, E. L., de Andrade, M., Boardman, L. A., and Petersen, G. M. (2008). Polymorphisms in DNA repair genes, smoking, and pancreatic adenocarcinoma risk. Cancer Res *68*, 4928-4935.

Moffitt, R. A., Marayati, R., Flate, E. L., Volmar, K. E., Loeza, S. G., Hoadley, K. A., Rashid, N. U., Williams, L. A., Eaton, S. C., Chung, A. H.*, et al.* (2015). Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nature genetics *47*, 1168-1178.

Morris, J. C., Tan, A. R., Olencki, T. E., Shapiro, G. I., Dezube, B. J., Reiss, M., Hsu, F. J., Berzofsky, J. A., and Lawrence, D. P. (2014). Phase I study of GC1008 (fresolimumab): a human anti-transforming growth factor-beta (TGFbeta) monoclonal antibody in patients with advanced malignant melanoma or renal cell carcinoma. PloS one *9*, e90353.

Morris, J. P. t., Cano, D. A., Sekine, S., Wang, S. C., and Hebrok, M. (2010). Beta-catenin blocks Kras-dependent reprogramming of acini into pancreatic cancer precursor lesions in mice. J Clin Invest *120*, 508-520.

Mullen, A. C., Orlando, D. A., Newman, J. J., Loven, J., Kumar, R. M., Bilodeau, S., Reddy, J., Guenther, M. G., DeKoter, R. P., and Young, R. A. (2011). Master transcription factors determine cell-type-specific responses to TGF-beta signaling. Cell *147*, 565-576.

Muller, W. J., Sinn, E., Pattengale, P. K., Wallace, R., and Leder, P. (1988). Single-step induction of mammary adenocarcinoma in transgenic mice bearing the activated c-neu oncogene. Cell *54*, 105-115.

Murray, I. A., Patterson, A. D., and Perdew, G. H. (2014). Aryl hydrocarbon receptor ligands in cancer: friend and foe. Nature reviews Cancer *14*, 801-814.

Nakaya, T., Ogawa, S., Manabe, I., Tanaka, M., Sanada, M., Sato, T., Taketo, M. M., Nakao, K., Clevers, H., Fukayama, M.*, et al.* (2014). KLF5 regulates the integrity and oncogenicity of intestinal stem cells. Cancer research *74*, 2882-2891.

Navas, C., Hernandez-Porras, I., Schuhmacher, A. J., Sibilia, M., Guerra, C., and Barbacid, M. (2012). EGF receptor signaling is essential for k-ras oncogene-driven pancreatic ductal adenocarcinoma. Cancer cell *22*, 318-330.

Niola, F., Zhao, X., Singh, D., Castano, A., Sullivan, R., Lauria, M., Nam, H. S., Zhuang, Y., Benezra, R., Di Bernardo, D.*, et al.* (2012). Id proteins synchronize stemness and anchorage to the niche of neural stem cells. Nature cell biology *14*, 477-487.

Nomura, M., and Li, E. (1998). Smad2 role in mesoderm formation, left-right patterning and craniofacial development. Nature *393*, 786-790.

Nordling, C. O. (1953). A new theory on cancer-inducing mechanism. British journal of cancer *7*, 68-72.

Norrander, J., Kempe, T., and Messing, J. (1983). Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene *26*, 101-106.

O'Toole, P. J., Inoue, T., Emerson, L., Morrison, I. E., Mackie, A. R., Cherry, R. J., and Norton, J. D. (2003). Id proteins negatively regulate basic helix-loop-helix transcription factor function by disrupting subnuclear compartmentalization. The Journal of biological chemistry *278*, 45770-45776.

Ohnishi, S., Laub, F., Matsumoto, N., Asaka, M., Ramirez, F., Yoshida, T., and Terada, M. (2000). Developmental expression of the mouse gene coding for the Kruppel-like transcription factor KLF5. Developmental dynamics : an official publication of the American Association of Anatomists *217*, 421-429.

Oshimori, N., Oristian, D., and Fuchs, E. (2015). TGF-beta promotes heterogeneity and drug resistance in squamous cell carcinoma. Cell *160*, 963-976.

Padua, D., Zhang, X. H., Wang, Q., Nadal, C., Gerald, W. L., Gomis, R. R., and Massague, J. (2008). TGFbeta primes breast tumors for lung metastasis seeding through angiopoietin-like 4. Cell *133*, 66-77.

Passegue, E., Jamieson, C. H., Ailles, L. E., and Weissman, I. L. (2003). Normal and leukemic hematopoiesis: are leukemias a stem cell disorder or a reacquisition of stem cell characteristics? Proceedings of the National Academy of Sciences of the United States of America *100 Suppl 1*, 11842-11849.

Peinado, H., Quintanilla, M., and Cano, A. (2003). Transforming growth factor beta-1 induces snail transcription factor in epithelial cell lines: mechanisms for epithelial mesenchymal transitions. The Journal of biological chemistry *278*, 21113-21123.

Perez, A. R., Pritykin, Y., Vidigal, J. A., Chhangawala, S., Zamparo, L., Leslie, C. S., and Ventura, A. (2017). GuideScan software for improved single and paired CRISPR guide RNA design. Nature biotechnology *35*, 347-349.

Perk, J., Iavarone, A., and Benezra, R. (2005). Id family of helix-loop-helix proteins in cancer. Nature reviews Cancer *5*, 603-614.

Pin, C. L., Rukstalis, J. M., Johnson, C., and Konieczny, S. F. (2001). The bHLH transcription factor Mist1 is required to maintain exocrine pancreas cell organization and acinar cell identity. The Journal of cell biology *155*, 519-530.

Pinon, J. D., Labi, V., Egle, A., and Villunger, A. (2008). Bim and Bmf in tissue homeostasis and malignant disease. Oncogene *27 Suppl 1*, S41-52.

Puthalakath, H., Villunger, A., O'Reilly, L. A., Beaumont, J. G., Coultas, L., Cheney, R. E., Huang, D. C., and Strasser, A. (2001). Bmf: a proapoptotic BH3-only protein regulated by interaction with the myosin V actin motor complex, activated by anoikis. Science *293*, 1829-1832.

Qin, Y., Garrison, B. S., Ma, W., Wang, R., Jiang, A., Li, J., Mistry, M., Bronson, R. T., Santoro, D., Franco, C.*, et al.* (2018). A Milieu Molecule for TGF-beta Required for Microglia Function in the Nervous System. Cell *174*, 156-171 e116.

Ramachandran, A., Vizan, P., Das, D., Chakravarty, P., Vogt, J., Rogers, K. W., Muller, P., Hinck, A. P., Sapkota, G. P., and Hill, C. S. (2018). TGF-beta uses a novel mode of receptor activation to phosphorylate SMAD1/5 and induce epithelial-to-mesenchymal transition. eLife *7*.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nature protocols *8*, 2281-2308.

Rhim, A. D., Mirek, E. T., Aiello, N. M., Maitra, A., Bailey, J. M., McAllister, F., Reichert, M., Beatty, G. L., Rustgi, A. K., Vonderheide, R. H.*, et al.* (2012). EMT and dissemination precede pancreatic tumor formation. Cell *148*, 349-361.

Richards, G. S., and Degnan, B. M. (2009). The dawn of developmental signaling in the metazoa. Cold Spring Harbor symposia on quantitative biology *74*, 81-90.

Roberts, A. B., Anzano, M. A., Wakefield, L. M., Roche, N. S., Stern, D. F., and Sporn, M. B. (1985). Type beta transforming growth factor: a bifunctional regulator of cellular growth. Proceedings of the National Academy of Sciences of the United States of America *82*, 119-123.

Roberts, N. J., Jiao, Y., Yu, J., Kopelovich, L., Petersen, G. M., Bondy, M. L., Gallinger, S., Schwartz, A. G., Syngal, S., Cote, M. L.*, et al.* (2012). ATM mutations in patients with hereditary pancreatic cancer. Cancer discovery *2*, 41-46.

Romero-Lanman, E. E., Pavlovic, S., Amlani, B., Chin, Y., and Benezra, R. (2012). Id1 maintains embryonic stem cell self-renewal by up-regulation of Nanog and repression of Brachyury expression. Stem cells and development *21*, 384-393.

Rosenbloom, K. R., Sloan, C. A., Malladi, V. S., Dreszer, T. R., Learned, K., Kirkup, V. M., Wong, M. C., Maddren, M., Fang, R., Heitner, S. G.*, et al.* (2013). ENCODE data in the UCSC Genome Browser: year 5 update. Nucleic acids research *41*, D56-63.

Ruzinova, M. B., and Benezra, R. (2003). Id proteins in development, cell cycle and cancer. Trends Cell Biol *13*, 410-418.

Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W. K., Luna, A., La, K. C., Dimitriadoy, S., Liu, D. L., Kantheti, H. S., Saghafinia, S.*, et al.* (2018). Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell *173*, 321-337 e310.

Sander, M., Neubuser, A., Kalamaras, J., Ee, H. C., Martin, G. R., and German, M. S. (1997). Genetic analysis reveals that PAX6 is required for normal transcription of pancreatic hormone genes and islet development. Genes & development *11*, 1662-1673.

Sanjana, N. E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nature methods *11*, 783-784.

SEER, N. (2000). SEER, Surveillance, Epidemiology, and End Results Program, (Bethesda, Md.: National Institutes of Health, National Cancer Institute).

Sekelsky, J. J., Newfeld, S. J., Raftery, L. A., Chartoff, E. H., and Gelbart, W. M. (1995). Genetic characterization and cloning of mothers against dpp, a gene required for decapentaplegic function in Drosophila melanogaster. Genetics *139*, 1347-1358.

Sengupta, S., and George, R. E. (2017). Super-Enhancer-Driven Transcriptional Dependencies in Cancer. Trends in cancer *3*, 269-281.

Seoane, J., Le, H. V., Shen, L., Anderson, S. A., and Massague, J. (2004). Integration of Smad and forkhead pathways in the control of neuroepithelial and glioblastoma cell proliferation. Cell *117*, 211-223.

Shi, G., Zhu, L., Sun, Y., Bettencourt, R., Damsz, B., Hruban, R. H., and Konieczny, S. F. (2009). Loss of the acinar-restricted transcription factor Mist1 accelerates Kras-induced pancreatic intraepithelial neoplasia. Gastroenterology *136*, 1368-1378.

Shi, Y., Hata, A., Lo, R. S., Massague, J., and Pavletich, N. P. (1997). A structural basis for mutational inactivation of the tumour suppressor Smad4. Nature *388*, 87-93.

Shi, Y., Wang, Y. F., Jayaraman, L., Yang, H., Massague, J., and Pavletich, N. P. (1998). Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-beta signaling. Cell *94*, 585-594.

Shivji, M. K. K., Renaudin, X., Williams, C. H., and Venkitaraman, A. R. (2018). BRCA2 Regulates Transcription Elongation by RNA Polymerase II to Prevent R-Loop Accumulation. Cell reports *22*, 1031-1039.

Siegel, P. M., Shu, W., Cardiff, R. D., Muller, W. J., and Massague, J. (2003). Transforming growth factor beta signaling impairs Neu-induced mammary tumorigenesis while promoting pulmonary metastasis. Proceedings of the National Academy of Sciences of the United States of America *100*, 8430-8435.

Smith, A. L., Robin, T. P., and Ford, H. L. (2012). Molecular pathways: targeting the TGF-beta pathway for cancer therapy. Clinical cancer research : an official journal of the American Association for Cancer Research *18*, 4514-4521.

Stankic, M., Pavlovic, S., Chin, Y., Brogi, E., Padua, D., Norton, L., Massague, J., and Benezra, R. (2013). TGF-beta-Id1 signaling opposes Twist1 and promotes metastatic colonization via a mesenchymal-to-epithelial transition. Cell reports *5*, 1228-1242.

Stein, S., Thomas, E. K., Herzog, B., Westfall, M. D., Rocheleau, J. V., Jackson, R. S., 2nd, Wang, M., and Liang, P. (2004). NDRG1 is necessary for p53-dependent apoptosis. The Journal of biological chemistry *279*, 48930-48940.

Stepanenko, A. A., Vassetzky, Y. S., and Kavsan, V. M. (2013). Antagonistic functional duality of cancer genes. Gene *529*, 199-207.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America *102*, 15545-15550.

Sur, I., and Taipale, J. (2016). The role of enhancers in cancer. Nature reviews Cancer *16*, 483-493.

Sussel, L., Kalamaras, J., Hartigan-O'Connor, D. J., Meneses, J. J., Pedersen, R. A., Rubenstein, J. L., and German, M. S. (1998). Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes due to arrested differentiation of pancreatic beta cells. Development *125*, 2213-2221.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Takeda, N., Jain, R., LeBoeuf, M. R., Wang, Q., Lu, M. M., and Epstein, J. A. (2011). Interconversion between intestinal stem cell populations in distinct niches. Science *334*, 1420-1424.

Tan, J., Yang, X., Zhuang, L., Jiang, X., Chen, W., Lee, P. L., Karuturi, R. K., Tan, P. B., Liu, E. T., and Yu, Q. (2007). Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells. Genes & development *21*, 1050-1063.

Tange, O. (2011). GNU Parallel - The Command-Line Power Tool. In, (The USENIX Magazine).

Tauriello, D. V. F., Palomo-Ponce, S., Stork, D., Berenguer-Llergo, A., Badia-Ramentol, J., Iglesias, M., Sevillano, M., Ibiza, S., Canellas, A., Hernando-Momblona, X.*, et al.* (2018). TGFbeta drives immune evasion in genetically reconstituted colon cancer metastasis. Nature *554*, 538-543.

TCGA (2017). Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. Cancer cell *32*, 185-203 e113.

Tetreault, M. P., Yang, Y., and Katz, J. P. (2013). Kruppel-like factors in cancer. Nature reviews Cancer *13*, 701-713.

Thiery, J. P., Acloque, H., Huang, R. Y., and Nieto, M. A. (2009). Epithelial-mesenchymal transitions in development and disease. Cell *139*, 871-890.

Tokheim, C. J., Papadopoulos, N., Kinzler, K. W., Vogelstein, B., and Karchin, R. (2016). Evaluating the evaluation of cancer driver genes. Proceedings of the National Academy of Sciences of the United States of America *113*, 14330-14335.

Tomasetti, C., Marchionni, L., Nowak, M. A., Parmigiani, G., and Vogelstein, B. (2015). Only three driver gene mutations are required for the development of lung and colorectal cancers. Proceedings of the National Academy of Sciences of the United States of America *112*, 118-123.

Tomasetti, C., and Vogelstein, B. (2015). Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science *347*, 78-81.

Tresini, M., Warmerdam, D. O., Kolovos, P., Snijder, L., Vrouwe, M. G., Demmers, J. A., van, I. W. F., Grosveld, F. G., Medema, R. H., Hoeijmakers, J. H.*, et al.* (2015). The core spliceosome as target and effector of non-canonical ATM signalling. Nature *523*, 53-58.

Tucker, R. F., Shipley, G. D., Moses, H. L., and Holley, R. W. (1984). Growth inhibitor from BSC-1 cells closely related to platelet type beta transforming growth factor. Science *226*, 705-707.

Valastyan, S., and Weinberg, R. A. (2011). Tumor metastasis: molecular insights and evolving paradigms. Cell *147*, 275-292.

Ventura, A., Kirsch, D. G., McLaughlin, M. E., Tuveson, D. A., Grimm, J., Lintault, L., Newman, J., Reczek, E. E., Weissleder, R., and Jacks, T. (2007). Restoration of p53 function leads to tumour regression in vivo. Nature *445*, 661-665.

Vervoort, S. J., Lourenco, A. R., van Boxtel, R., and Coffer, P. J. (2013a). SOX4 mediates TGF-beta-induced expression of mesenchymal markers during mammary cell epithelial to mesenchymal transition. PloS one *8*, e53238.

Vervoort, S. J., van Boxtel, R., and Coffer, P. J. (2013b). The role of SRY-related HMG box transcription factor 4 (SOX4) in tumorigenesis and metastasis: friend or foe? Oncogene *32*, 3397-3409.

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr., and Kinzler, K. W. (2013). Cancer genome landscapes. Science *339*, 1546-1558.

von Figura, G., Morris, J. P. t., Wright, C. V., and Hebrok, M. (2014). Nr5a2 maintains acinar cell differentiation and constrains oncogenic Kras-mediated pancreatic neoplastic initiation. Gut *63*, 656-664.

Waddell, N., Pajic, M., Patch, A. M., Chang, D. K., Kassahn, K. S., Bailey, P., Johns, A. L., Miller, D., Nones, K., Quek, K.*, et al.* (2015). Whole genomes redefine the mutational landscape of pancreatic cancer. Nature *518*, 495-501.

Wang, F., Herrington, M., Larsson, J., and Permert, J. (2003). The relationship between diabetes and pancreatic cancer. Molecular cancer *2*, 4.

Wiater, E., and Vale, W. (2012). Roles of activin family in pancreatic development and homeostasis. Molecular and cellular endocrinology *359*, 23-29.

Wilentz, R. E., Chung, C. H., Sturm, P. D., Musler, A., Sohn, T. A., Offerhaus, G. J., Yeo, C. J., Hruban, R. H., and Slebos, R. J. (1998). K-ras mutations in the duodenal fluid of patients with pancreatic carcinoma. Cancer *82*, 96-103.

Williams, J. A., and Goldfine, I. D. (1985). The insulin-pancreatic acinar axis. Diabetes *34*, 980-986.

Wilson, M. E., Yang, K. Y., Kalousova, A., Lau, J., Kosaka, Y., Lynn, F. C., Wang, J., Mrejen, C., Episkopou, V., Clevers, H. C., and German, M. S. (2005). The HMG box transcription factor Sox4 contributes to the development of the endocrine pancreas. Diabetes *54*, 3402-3409.

Winter, J. M., Tang, L. H., Klimstra, D. S., Brennan, M. F., Brody, J. R., Rocha, F. G., Jia, X., Qin, L. X., D'Angelica, M. I., DeMatteo, R. P.*, et al.* (2012). A novel survival-based tissue microarray of pancreatic cancer validates MUC1 and mesothelin as biomarkers. PloS one *7*, e40157.

Witkiewicz, A. K., McMillan, E. A., Balaji, U., Baek, G., Lin, W. C., Mansour, J., Mollaee, M., Wagner, K. U., Koduru, P., Yopp, A.*, et al.* (2015). Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. Nature communications *6*, 6744.

Wrana, J. L., Attisano, L., Wieser, R., Ventura, F., and Massague, J. (1994). Mechanism of activation of the TGF-beta receptor. Nature *370*, 341-347.

Wu, M. Y., and Hill, C. S. (2009). Tgf-beta superfamily signaling in embryonic development and homeostasis. Developmental cell *16*, 329-343.

Yachida, S., White, C. M., Naito, Y., Zhong, Y., Brosnan, J. A., Macgregor-Das, A. M., Morgan, R. A., Saunders, T., Laheru, D. A., Herman, J. M.*, et al.* (2012). Clinical significance of the genetic landscape of pancreatic cancer and implications for identification of potential long-term survivors. Clinical cancer research : an official journal of the American Association for Cancer Research *18*, 6339-6347.

Yamaguchi, J., Yokoyama, Y., Kokuryo, T., Ebata, T., and Nagino, M. (2018). Cells of origin of pancreatic neoplasms. Surgery today *48*, 9-17.

Yamamizu, K., Piao, Y., Sharov, A. A., Zsiros, V., Yu, H., Nakazawa, K., Schlessinger, D., and Ko, M. S. (2013). Identification of transcription factors for lineage-specific ESC differentiation. Stem cell reports *1*, 545-559.

Yan, W., Young, A. Z., Soares, V. C., Kelley, R., Benezra, R., and Zhuang, Y. (1997). High incidence of T-cell tumors in E2A-null mice and E2A/Id1 double-knockout mice. Molecular and cellular biology *17*, 7317-7327.

Ying, H., Kimmelman, A. C., Lyssiotis, C. A., Hua, S., Chu, G. C., Fletcher-Sananikone, E., Locasale, J. W., Son, J., Zhang, H., Coloff, J. L.*, et al.* (2012). Oncogenic Kras maintains pancreatic tumors through regulation of anabolic glucose metabolism. Cell *149*, 656-670.

Ying, Q. L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell *115*, 281-292.

Zehir, A., Benayed, R., Shah, R. H., Syed, A., Middha, S., Kim, H. R., Srinivasan, P., Gao, J., Chakravarty, D., Devlin, S. M.*, et al.* (2017). Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nature medicine *23*, 703-713.

Zhang, B., Zhang, Z., Xia, S., Xing, C., Ci, X., Li, X., Zhao, R., Tian, S., Ma, G., Zhu, Z.*, et al.* (2013). KLF5 activates microRNA 200 transcription to maintain epithelial characteristics and prevent induced epithelial-mesenchymal transition in epithelial cells. Molecular and cellular biology *33*, 4919-4935.

Zhang, M., Behbod, F., Atkinson, R. L., Landis, M. D., Kittrell, F., Edwards, D., Medina, D., Tsimelzon, A., Hilsenbeck, S., Green, J. E.*, et al.* (2008). Identification of tumor-initiating cells in a p53-null mouse model of breast cancer. Cancer Res *68*, 4674-4682.

Zhang, N., Yantiss, R. K., Nam, H. S., Chin, Y., Zhou, X. K., Scherl, E. J., Bosworth, B. P., Subbaramaiah, K., Dannenberg, A. J., and Benezra, R. (2014). ID1 is a functional marker for intestinal stem and progenitor cells required for normal response to injury. Stem cell reports *3*, 716-724.

Zhang, S., Wu, W., Wu, Y., Zheng, J., Suo, T., Tang, H., and Tang, J. (2010). RNF152, a novel lysosome localized E3 ligase with pro-apoptotic activities. Protein & cell *1*, 656-663.

Zhang, X., Chiang, H. C., Wang, Y., Zhang, C., Smith, S., Zhao, X., Nair, S. J., Michalek, J., Jatoi, I., Lautner, M.*, et al.* (2017). Attenuation of RNA polymerase II pausing mitigates BRCA1-associated R-loop accumulation and tumorigenesis. Nature communications *8*, 15908.

Zheng, X., Carstens, J. L., Kim, J., Scheible, M., Kaye, J., Sugimoto, H., Wu, C. C., LeBleu, V. S., and Kalluri, R. (2015). Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. Nature *527*, 525-530.

Zhou, Q., Law, A. C., Rajagopal, J., Anderson, W. J., Gray, P. A., and Melton, D. A. (2007). A multipotent progenitor domain guides pancreatic organogenesis. Developmental cell *13*, 103-114.

Zhu, Q., Pao, G. M., Huynh, A. M., Suh, H., Tonnu, N., Nederlof, P. M., Gage, F. H., and Verma, I. M. (2011). BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. Nature *477*, 179-184.

Zhu, Y., Richardson, J. A., Parada, L. F., and Graff, J. M. (1998). Smad3 mutant mice develop metastatic colorectal cancer. Cell *94*, 703-714.

Zorn, A. M., and Wells, J. M. (2007). Molecular basis of vertebrate endoderm development. International review of cytology *259*, 49-111.

Zuber, J., McJunkin, K., Fellmann, C., Dow, L. E., Taylor, M. J., Hannon, G. J., and Lowe, S. W. (2011). Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. Nature biotechnology *29*, 79-83.