Dissection of transcriptional and *cis*-regulatory control of differentiation in human pancreatic cancer

Giuseppe R Diaferia^{1,†}, Chiara Balestrieri^{1,†}, Elena Prosperini¹, Paola Nicoli¹, Paola Spaggiari², Alessandro Zerbi² & Gioacchino Natoli^{1,*}

Abstract

The histological *grade* of carcinomas describes the ability of tumor cells to organize in differentiated epithelial structures and has prognostic and therapeutic impact. Here, we show that differential usage of the genomic repertoire of transcriptional enhancers leads to grade-specific gene expression programs in human pancreatic ductal adenocarcinoma (PDAC). By integrating gene expression profiling, epigenomic footprinting, and loss-of-function experiments in PDAC cell lines of different grade, we identified the repertoires of enhancers specific to high- and low-grade PDACs and the cognate set of transcription factors acting to maintain their activity. Among the candidate regulators of PDAC differentiation, KLF5 was selectively expressed in pre-neoplastic lesions and low-grade primary PDACs and cell lines, where it maintained the acetylation of grade-specific enhancers, the expression of epithelial genes such as keratins and mucins, and the ability to organize glandular epithelia in xenografts. The identification of the transcription factors controlling differentiation in PDACs will help clarify the molecular bases of its heterogeneity and progression.

Keywords differentiation; enhancers; epigenetics; epithelial–mesenchymal transition; pancreatic ductal adenocarcinoma

Subject Categories Cancer; Chromatin, Epigenetics, Genomics & Functional Genomics; Transcription

DOI 10.15252/embj.201592404 | Received 30 June 2015 | Revised 3 December 2015 | Accepted 7 December 2015 | Published online 14 January 2016

The EMBO Journal (2016) 35: 595–617

See also: J Ferrer & FX Real (March 2016)

Introduction

Epithelial tumors (carcinomas) display various degrees of histological alteration compared to the tissue they derive from, a property that is commonly referred to as tumor grade (Devita *et al*, 2015).

Such diversity reflects the intrinsic ability of cancer cells to express genes characteristic of the tissue of origin and to interact among them and with the tumor stroma to organize differentiated epithelial structures. In addition to prognosis, grade may also impact therapeutic options, being cells of distinct grade differentially sensitive to various therapies. In many carcinomas, distinct areas of the same tumor display a similar degree of differentiation. However, in the specific case of pancreatic ductal adenocarcinoma (PDAC), the most frequent cancer of the exocrine pancreas (Hidalgo, 2010), a great intratumor variation in the degree of differentiation is the standard, with foci of poorly differentiated cell nests frequently coexisting with well- or moderately differentiated glandular structures (Kloppel *et al*, 1985). This histological variability indicates that mechanisms involved in the maintenance of PDAC differentiation commonly deteriorate, thus leading to the various levels of histological disorganization observed in G2–G3 areas of the tumors. Moreover, the distinct sensitivity to treatment of low- and high-grade PDAC cells coexisting in the same tumor lays the ground for resistance to chemotherapy and may critically contribute to the almost invariably deadly outcome of PDAC patients (Hidalgo, 2010).

Pancreatic ductal adenocarcinoma grade is also correlated with the gene expression signatures that distinguish three recently described PDAC subtypes with different prognosis and response to treatments (Collisson *et al*, 2011): The “classical” subtype is associated with lower grade tumors with better prognosis and is characterized by high expression of epithelial and adhesion genes; the “quasi-mesenchymal” subtype is associated with higher grade tumors and worse prognosis, and shows high expression of mesenchymal genes; and the less common “exocrine” subtype is characterized by the expression of genes encoding pancreatic acinar enzymes. PDAC cell lines include representatives of the classical and the quasi-mesenchymal, but not the exocrine subtype (Collisson *et al*, 2011). The transcription factors (TFs) and *cis*-regulatory circuits that enforce the maintenance of the gene expression programs characteristic of high- and low-grade PDAC cells are unknown.

Normal cell and tissue differentiation reflects the activity of gene regulatory networks supervised by TFs involved in lineage

1 Department of Experimental Oncology, European Institute of Oncology (IEO), Milan, Italy

2 Division of Pancreatic Surgery, Humanitas Clinical Institute, Milan, Italy

*Corresponding author. Tel: +39 02 57489953; E-mail: gioacchino.natoli@ieo.eu

†These authors contributed equally to this work (order determined by drawing)

determination and maintenance of cell identity (Huang *et al*, 2009). Combinatorial binding of sequence-specific TFs to *cis*-regulatory genomic elements (enhancers and gene promoters) (Spitz & Furlong, 2012) triggers the recruitment of enzymes (transcriptional coregulators) that control the deposition of histone modifications (Heintzman *et al*, 2007; Rada-Iglesias *et al*, 2010; Ostuni *et al*, 2013). Thus, cell type-specific differences in the repertoire of genomic regions marked by such histone modifications reflect the underlying activity of different combinations of TFs (Heintzman *et al*, 2009; Visel *et al*, 2009; Andersson *et al*, 2014; Lara-Astiaso *et al*, 2014). The differential usage of the genomic *cis*-regulatory information by sequence-specific TFs determines the transcriptional programs and eventually the functional properties characteristic of individual cell types (Neph *et al*, 2012).

Within this conceptual framework, new properties acquired by cancer cells reflect a change in the usage of the genomic *cis*-regulatory information driven by an altered expression and/or activity of sequence-specific TFs and their associated coregulators. For example, the invasiveness of high-grade gliomas was linked to the activity of two TFs, C/EBP β , and STAT3, that synergize to reprogram tumor cells toward a mesenchymal behavior (Carro *et al*, 2010). Therefore, the undifferentiated state of high-grade tumors may represent the consequence of cell reprogramming events driven by newly established TF networks. In fact, for every cancer type, tumor cells with different grade may have to be considered as truly distinct cell types, with their own stable gene regulatory networks programmed by specific combinations of active TFs.

In this study, we set out to identify the TFs responsible for the control of differentiation in human PDACs. To this aim, we first defined the transcriptomes and epigenomes of PDAC cells representative of well-differentiated (low-grade, Lo-G) and undifferentiated (high-grade, Hi-G) tumors. Then, we computationally deconvoluted the different types of next-generation sequencing data to identify candidate TFs controlling grade-specific transcriptional and epigenomic properties; and finally, we experimentally validated selected candidates.

We identified thousands of enhancers that were specific to Lo-G PDACs and a panel of cognate TFs acting on them, including known regulators of pancreas differentiation or endoderm development. Among the TFs identified, we selected for further investigation Krüppel-like factor 5 (KLF5) both because of its known involvement in the differentiation of endoderm-derived epithelia (Wan *et al*, 2008; Bell *et al*, 2011, 2013) and because of its implication in other tumors of the gastrointestinal tract (Nandan *et al*, 2004, 2008). In a large panel of human PDACs, KLF5 was selectively expressed in low-grade (G1) epithelia and absent in areas of poor differentiation (G3). *In vitro*, KLF5 was required for the expression of a subset of epithelial identity genes, such as keratins and mucins, for the maintenance of histone acetylation at a large fraction of enhancers selectively active in well-differentiated PDAC cells and, to a lesser extent, for the repression of genes associated with the mesenchymal program and with stemness. Consistently, KLF5-deleted PDAC cells were unable to form differentiated epithelia in xenografts. KLF5-dependent genes were largely distinct from those activated by ELF3, another regulator of epithelial identity (De Craene & Berx, 2013) also selectively expressed in low-grade PDACs, and from those suppressed by the transcriptional repressor ZEB1, a master inducer of mesenchymal properties that is instead selectively expressed in

high-grade PDACs. Therefore, maintenance of the epithelial identity in low-grade PDACs results from the complementary activities of multiple transcriptional regulators.

Results

Grade-specific expression of transcription factors in cell lines and tumors

We first used RNA sequencing (RNA-seq) (Appendix Fig S1A) to analyze the transcriptome of a panel of nine human PDAC cell lines that have been extensively characterized for their *in vitro* and *in vivo* properties and include representative of both low- and high-grade PDACs (Sipos *et al*, 2003; Collisson *et al*, 2011). For instance, CAPAN-2 and CFPAC-1 self-organized into differentiated epithelia similar to those characteristics of Lo-G PDACs *in vitro* and *in vivo* (Sipos *et al*, 2003) and were associated with the classical PDAC signature (Collisson *et al*, 2011). Conversely, MiaPaCa-2 and PANC-1 constitutively expressed mesenchymal genes (Collisson *et al*, 2011), showed fibroblastoid morphology, and were unable to generate differentiated epithelia *in vitro* and *in vivo*, thus representing Hi-G PDACs (Sipos *et al*, 2003). The Lo-G and the Hi-G PDACs clustered at the opposite ends of a correlation matrix (Fig 1A), while two cell lines (AsPC-1 and BxPc3) showed an intermediate transcriptional profile. A Gene Set Enrichment Analysis (GSEA) (Table EV1) revealed that the transcriptomes of the Lo-G cell lines were mainly enriched for gene sets associated with epithelial differentiation, while those of the Hi-G PDACs were associated with gene sets linked to the acquisition of mesenchymal properties. Lo-G PDACs also showed an enrichment of gene sets associated with the interferon (IFN) response and IFN signaling (Table EV1). Two representative GSEA plots are shown in Appendix Fig S1B.

Using a $\log_2(\text{fold change}) \geq \pm 1$ and an FDR ≤ 0.1 , we identified 1,164 differentially expressed genes, including 689 genes preferentially expressed in low-grade PDACs and 475 in high-grade PDACs (Appendix Fig S1A and Table EV2). From this list, we extracted a set of 60 sequence-specific TFs (Fig 1B and C) whose differential expression hinted at a possible role in controlling the epigenome and the transcriptional programs specific to the Lo-G and the Hi-G PDAC subtypes. The 38 TFs specific to Lo-G PDACs included known regulators of pancreatic development such as FOXA1 (Gao *et al*, 2008) and TFs implicated in maintenance of epithelial identity in other tumor types, such as GRHL2 (De Craene & Berx, 2013). This analysis also retrieved ELF3 and FOXQ1, which are both part of the gene signature that defines the classical PDAC subtype (Collisson *et al*, 2011) (Fig 1B). Lo-G PDACs selectively expressed KLF5 (Fig 1B and C), a TF involved in terminal differentiation of epithelia of endodermal origin, where it is also required to maintain the expression of ELF3 (Bell *et al*, 2011, 2013). The higher expression of some interferon regulatory factors (IRF1 and 9) in Lo-G PDACs may relate to the IFN response gene signatures detected in these cell lines (Table EV1 and Appendix Fig S1B) and is consistent with the observation that IRF1 expression in human PDACs correlates with tumor differentiation and better prognosis (Sakai *et al*, 2014).

Among the 22 TFs selectively expressed in the Hi-G cell lines, ZEB1 stands out as one of the best-characterized inducers of the epithelial-to-mesenchymal transition (EMT) (De Craene & Berx,

Figure 1. A set of TFs specific to distinct PDAC subtypes.

- A** Correlation matrix of Pearson correlation coefficients for RNA-seq data obtained from multiple PDAC cell lines (two biological replicates for each of the nine cell lines were used). The matrix was hierarchically clustered. The relationship with grade is shown. Source data were deposited in GEO.
- B** Heatmap showing TFs differentially expressed in Lo-G and Hi-G PDAC cell lines. Differentially expressed genes were identified based on a $\log_2(\text{fold change}) \geq \pm 1$ and an $\text{FDR} \leq 0.1$. Hierarchical clustering was based on Pearson correlation coefficients.
- C** Representative RNA-seq snapshots.
- D** Western blots for selected TFs. One experiment out of three independent biological replicates is shown.
- E** IHC showing TF staining in G1 and G3 areas from human PDACs. The areas in the insets are shown at 4× higher magnification in the panels below. Data from two representative patients out of nine tested are shown.

2013). Finally, it should be noted that some TFs involved in pancreas differentiation, such as GATA6 (Lango Allen *et al*, 2012; Martinelli *et al*, 2013) and PBX1 (Dutta *et al*, 2001), or expressed in early-to-intermediate stage pancreatic cancers, such as HNF4A (Kim *et al*, 2013), were also preferentially expressed in Lo-G PDACs, but fell just below our cutoff. A representative panel of Western blots is shown in Fig 1D.

We next determined whether we could confirm a grade-specific expression of the TFs identified in our RNA-seq analysis also in primary human PDACs. As discussed above, human PDACs contain histologically heterogeneous areas embedded in a very abundant stroma (accounting for up to 90–95% of the tumor mass). Therefore, we used immunohistochemistry (IHC) to determine TF expression in areas of different grade belonging to the same tumor.

Figure 1E shows representative examples referring to those TFs for which we obtained IHC-grade antibodies. Consistent with the data in cell lines (Fig 1B), ELF3, HNF1B, FOS, and JUNB showed high expression in low-grade and low-to-undetectable expression in high-grade areas; conversely, ZEB1, GATA2, ETV5, and TCF12 showed an opposite behavior, indicating that PDAC cell lines retain—at least in part—the essential transcriptional properties of primary human PDACs.

Chromatin profiling reveals a grade-specific usage of the *cis*-regulatory information

The data shown above suggest a possible contribution of differentially expressed TFs to the gene expression programs specific to low- and high-grade PDACs. Specifically, differential expression of TFs may lead to a grade-specific usage of the genomic repertoire of *cis*-regulatory elements. To address this possibility, we used ChIP-seq to profile selected chromatin marks (Kouzarides, 2007), namely histone H3 lysine 4 monomethylation (H3K4me1), a mark associated with both poised and active enhancers (Heintzman *et al*, 2007); H3K27Ac, which selectively marks active promoters and enhancers (Creyghton *et al*, 2010; Rada-Iglesias *et al*, 2010; Ostuni *et al*, 2013); H3K4me3, which is associated with active and poised transcription start sites (Bernstein *et al*, 2005; Kim *et al*, 2005); and H3K9me3, which marks inactive chromatin, including heterochromatin (Mikkelsen *et al*, 2007). By identifying epigenomic patterns that are common to cell lines of one subtype but different from those of the other, specificities of individual cell lines (such as cell line-specific mutations and copy number variations) are averaged, thus allowing the identification of commonalities in the usage of the *cis*-regulatory information that underlie transcriptional programs associated with low- and high-grade PDACs.

We centered our initial analysis on H3K27Ac, as this mark is informative of the activity state of both promoters and enhancers. Based on the H3K27Ac profiles, Lo-G and Hi-G PDAC cell lines clustered separately (Fig 2A), indicating that their distinct transcriptional profiles were associated with a different usage of the genomic *cis*-regulatory information. We next classified acetylated peaks based on their detection in the low- and/or the high-grade PDAC cell lines (Fig 2B and Table EV3). A total of 10,860 acetylated peaks were detected in all the cell lines (“common” peaks, shown in purple); 4,666 were specific to the Lo-G PDACs (gray) and 1,050 of the Hi-G group (red). A large number of acetylated peaks (indicated in yellow) were present in different subsets of the cell lines analyzed

and therefore were not included in this classification scheme. Importantly, almost 75% of the common peaks were associated with gene promoters. Conversely, the overwhelming majority of grade-specific H3K27Ac peaks was distal from the transcriptional start sites (TSSs) and therefore represented putative enhancers (Fig 2C). Grade-specific enhancers were generally closer to genes expressed in the corresponding PDAC grade (Fig 2D), suggesting that they may directly contribute to the activation of genes located within the same regulatory domain. We also analyzed enhancer clusters (or super-enhancers, SE), namely large genomic regions characterized by an unusually high density of regulatory elements (Whyte *et al*, 2013) and often dedicated to the regulation of only one adjacent gene (Downen *et al*, 2014). We defined as SE the genomic regions derived from the merge of multiple acetylated enhancers distant less than 12.5 kb from each other (see Discussion in Pott & Lieb, 2015). While only 12 SE were specific to Hi-G PDACs, 50 were selectively identified in Lo-G PDACs and were often associated with genes encoding TFs selectively expressed in these cells such as *KLF5* and *PBX1* (Appendix Fig S2 and Table EV4). Overall, most of the *cis*-regulatory variation between PDAC subtypes resided at enhancers.

The heatmaps in Fig 2E show the relationship between histone modifications at the differentially acetylated regions. TSS-distal acetylated regions specific to Lo-G PDACs were also associated with H3K4me1 levels higher than those observed in Hi-G PDACs (Fig 2E, top panel). The same regions were instead associated with higher levels of the repressive H3K9me3 mark in Hi-G PDACs. The reciprocal trend was observable at the TSS-distal acetylated regions specific to Hi-G PDACs (Fig 2E, bottom panel), although the differences in H3K4me1 and H3K9me3 were less evident.

When considering TSSs/promoters, peaks selectively acetylated in Lo-G PDAC cells were associated with levels of H3K4me3 higher than those measured in Hi-G cells (Fig 2F). Conversely, Hi-G-specific acetylation at TSS was associated with only a minor gain in H3K4me3. A representative snapshot is shown in Fig 2G.

To gain insight into the transcriptional responses controlled by enhancers selectively acetylated in low- vs. high-grade PDACs, we used the GREAT tool (McLean *et al*, 2010). GREAT links sets of enhancers to putative biological functions based on the functional annotations of the nearby genes, with a score that takes into consideration the enhancer-gene distance and therefore the likelihood of correct assignment. Enhancers specific to Lo-G PDACs were enriched for functional terms related to endoderm and epithelial development, as well as with some endodermal cancers, a result

Figure 2. PDAC grade-specific chromatin landscapes.

- A Pearson correlation matrix for the H3K27Ac genomic profiles in different PDAC cell lines. The matrix was hierarchically clustered. Source data were deposited in GEO.
- B Histogram showing common and grade-specific genomic regions based on their histone acetylation. The number of genomic regions in each category is indicated in the pie charts in (C). Detailed computational and statistical procedures for the identification of different groups of genomic regions are provided in the methods section.
- C Genomic distribution of common and grade-specific H3K27Ac peaks.
- D Distance of grade-specific enhancers identified by H3K27Ac from genes expressed in PDAC cells of different grade (* $P \leq 1E-50$; ** $P \leq 1E-100$, two-tailed Welch's *t*-test). Central values represent the median, the bars the 25th and 75th percentile and the dashed lines the lower and upper whiskers.
- E, F Heatmaps showing grade-specific, TSS-distal (E), or TSS-proximal (F) acetylated peaks and their relation to other histone marks. The box plots indicate the signal intensity for each of the modifications (normalized values were transformed into a 0–1 range; central values represent the median, the bars the 25th and 75th percentile and the dashed lines the lower and upper whiskers).
- G A representative snapshot of a genomic region with several enhancers selectively marked by H3K27Ac and H3K4me1 in Lo-G (CAPAN-1 and CFPAC-1) but not in Hi-G (MiaPaCa-2 and PANC1) PDAC cell lines. The *NFKB1Z* gene in this region is selectively expressed in Lo-G PDACs.

Figure 2.

consistent with the possibility that these differentially active elements contribute to grade-specific gene expression (Table EV5). Functional terms retrieved using high-grade-specific enhancers were more heterogeneous and less informative, which may reflect the less differentiated state of these cells but also, more trivially, the lower abundance of the enhancers in this set, which reduces the statistical power of the analysis.

Overall, a large set of enhancers associated with epithelial—and specifically endodermal—genes is selectively active in low-grade PDACs and may thus contribute to maintain their characteristic gene expression profile and their functional properties.

TF binding sites associated with grade-specific enhancers

To identify the TFs that control deposition and activity of grade-specific enhancers, we determined the TF consensus DNA binding sites (described by position weight matrixes, PWMs) that are statistically overrepresented in each set (Zambelli *et al*, 2009). Specifically, we compared the acetylated enhancers of Lo-G and Hi-G PDACs with each other and also with the FANTOM5 collection of human enhancers active in different tissues (Andersson *et al*, 2014). Only PWMs that were enriched in both comparisons were retained. Moreover, we tested whether the consensus DNA binding sites of the TFs overexpressed in low- or high-grade PDACs (Fig 1) were also overrepresented in the corresponding set of enhancers.

The upper part of Fig 3 shows the PWMs statistically overrepresented in this analysis (P -values $\leq 10E-5$ in both comparisons) and the cognate TF families. This analysis retrieved the PWMs for many of the TFs overexpressed in low-grade PDACs, including ELF3 and FOXQ1, two TFs also included in the classical PDAC gene signature (Collisson *et al*, 2011); FOXA and HNF1 TFs, which are involved in pancreatic development (Gittes, 2009); IRF family TFs, which may be responsible for the interferon signature observed in these tumor cell lines (Table EV1). The same set was enriched in PWMs for other TFs contributing to pancreatic development such as GATA6 and HNF4A (Gittes, 2009; Lango Allen *et al*, 2012; Martinelli *et al*, 2013), which—as discussed above—fell just below the threshold we applied to call TFs as expressed in a grade-specific manner.

Conversely, the PWMs of some TFs overexpressed in Lo-G PDACs, such as KLF5, FOS, and JUNB, were not statistically overrepresented and are shown in the lower part of Fig 3. This unexpected result can be explained by the high frequency of the KLF and FOS/AP-1 DNA binding sites in the sequence sets used as background in the overrepresentation analysis (Fig 3). For instance, KLF sites were present in 64.8% of Lo-G-specific enhancers but also in 68.1% of Hi-G-specific enhancers. Since different KLF family members recognize very similar consensus DNA binding sites, it is likely that the many KLF sites associated with Hi-G-specific enhancers are recognized by KLF family TFs other than KLF5.

Finally, the PWM of ZEB1, which is selectively expressed in Hi-G PDACs and is a master regulator of the epithelial-to-mesenchymal transition, was also overrepresented at enhancers specific to Lo-G PDACs, which may seem counterintuitive. However, this result is in line with the notion that ZEB1 acts mainly as a transcriptional repressor (De Craene & Berx, 2013), therefore, lack of

ZEB1-mediated repression in Lo-G PDACs would result in increased acetylation at its target sites.

The full list of PWMs used in this analysis and their associated statistical values and frequencies at Lo-G and Hi-G enhancers are available in Table EV6.

TF binding to enhancers of low-grade PDAC cells

We next determined in the enhancers specific to low-grade PDACs, the frequency of the consensus DNA binding sites for those TFs overexpressed in the same cells (Fig 4A). KLF and ELF consensus DNA binding sites were the most common ones, followed by consensus binding sites for TFs of the JUN/AP-1, FOXA, and IRF families; consensus binding sites for HNF1A/B and GATA family members were less numerous albeit detectable in a sizeable fraction of these enhancers.

We used ChIP-seq to profile in a Lo-G PDAC cell line (CFPAC-1) a panel of differentially expressed TFs for which we obtained ChIP-grade antibodies, namely KLF5, IRF1, HNF1B, ELF3, and FOXA1. Two representative snapshots (Fig 4B) show extensive cobinding of these TFs to the keratin gene cluster on chromosome 17 (left) and to the *CEACAM6* gene (right), which is part of the classical PDAC signature. The genomic distribution of these TFs (Appendix Fig S3A and Table EV7) indicated a strong preference for genomic regions that were selectively acetylated in Lo-G PDAC cell lines.

In keeping with the specificity of the antibodies used, *de novo* motif discovery analyses retrieved binding sites similar to those previously reported for each of the TFs under study, with the notable exception of IRF1 which was associated with a canonical AP-1 site instead (Appendix Fig S3B). This result was not unexpected, since in other cellular systems, IRF family members are recruited to chromatin in complexes with AP-1 proteins through either canonical AP-1 binding sites (Li *et al*, 2012) or composite API-1/IRF sites (Glasmacher *et al*, 2012; Mancino *et al*, 2015). The same analyses also retrieved binding sites for other TFs that were likely cobound at the same regulatory regions contacted by the immunoprecipitated TF (Appendix Fig S3B). For instance, the FOXA1-bound regions were enriched for ELF3, AP-1, and KLF5 DNA binding sites; the IRF1-bound regions, in addition to AP-1 sites, contained binding sites for ELF3 and KLF5; and the KLF5-bound regions were associated with binding sites for ELF3, AP-1, and FOX family TFs. Consistent with this *de novo* motif discovery analysis, a TF binding motif overrepresentation analysis corroborated the existence of a network of TFs acting at the enhancers of Lo-G PDACs (Appendix Fig S3C). For instance, the IRF1-bound regions, in addition to AP-1 sites, contained binding sites for ELF3 and KLF5, while AP-1, HNF1B, and FOXA1 sites were overrepresented in the ELF3 ChIP-seq. Binding sites for other TFs that are overexpressed in Lo-G PDACs were also frequently overrepresented.

The matrix representation in Fig 4C provides a synthetic view of the overlap between the TFs analyzed by ChIP at the enhancers specific to the Lo-G PDACs and indicates the high frequency of the combination of the five TFs analyzed. Figure 4D shows the distance of the summits of TF peaks from the center of the acetylated regions specific to Lo-G PDACs and indicates the tendency of these TFs to bind close to enhancer cores.

TF FAMILY	PWM	<i>p</i> vs. Hi-G	<i>p</i> vs. <i>hs</i> enh	<i>f</i> Lo-G	<i>f</i> Hi-G	TF	OTHER TFs
Ets domain		7.21E-34	9.58E-39	42.39	22.43	ELF3	ETV7 ELF1 ELF5
POU domain		4.28E-17	3.01E-22	17.00	11.29	HNF1B	HNF1A HMBOX1 POU1F1
HOX related TFs		6.61E-17	5.73E-08	14.78	9.55	HOXA3	HOXA1 GSX1 PDX1
Forkhead box		1.94E-15	5.57E-15	23.26	16.06	FOXA1	FOXC1 FOXQ1
SOX		4.75E-15	1.46E-14	13.79	8.39	SOX4	SOX12 SRY
Paired-related HD		8.76E-14	2.08E-07	12.29	6.95	DRGX	ALX1 HESX1 VSX1
HD-LIM		1.23E-13	1.12E-08	8.82	4.63	LHX6	LHX1 ISL1 LMX1A
NK		3.54E-13	4.50E-09	23.00	15.63	BARX2	BARX1 BARHL1 MSX2
HD-ZF		1.11E-11	4.09E-68	40.87	30.68	ZEB1	—
IRF		2.61E-11	1.93E-08	22.00	14.62	IRF9	IRF1 IRF2 IRF5
TALE-type HD		3.53E-11	1.62E-06	17.10	12.59	IRX3	IRX2 IRX3 IRX5 IRX6
TCF-7-related TFs		7.46E-11	9.06E-10	8.21	4.78	TCF1	LEF1
> 3 ZnF TFs		9.29E-11	1.95E-36	34.88	24.89	SNAI2	SCRT1 ZNF333 ZNF652
TBP-related TFs		3.64E-10	5.07E-07	19.81	16.06	TBP	—
ARID-related TFs		8.55E-10	6.12E-09	12.61	9.41	ARID5B	ARID5A
MEF2		2.51E-09	4.55E-07	19.69	14.76	MEF2A	MEF2B MEF2C MEF2D
GATA		2.96E-09	1.30E-08	15.80	11.29	GATA3	GATA6
Grainyhead		3.41E-09	3.49E-14	10.51	5.64	GRHL2	—
Paired plus HD		4.17E-09	1.27E-06	8.07	6.66	PAX7	PAX3 PAX4 PAX6
HD-CUT		7.59E-09	1.58E-08	12.51	9.70	ONECUT1	ONECUT2 ONECUT3 CUX1
HMGA		1.77E-08	8.75E-15	17.71	14.91	HMGA1	—
NR2 (RXR-like NR)		1.95E-08	4.85E-18	23.74	19.10	NR2F6	HNF4A RXRA RXRG
HD-SINE		2.19E-08	3.80E-11	12.78	9.70	SIX6	—
TEF		1.13E-07	1.15E-13	22.05	15.63	TEAD3	TEAD1 TEAD4
NFAT related TFs		2.82E-07	5.06E-08	33.82	26.77	NFATC4	—
CP2-related TFs		3.31E-07	2.50E-07	8.26	6.66	TFCP2	—
Nuclear receptor		4.72E-07	4.46E-21	25.27	19.39	RARG	VDR
TBrain-related TFs		2.10E-06	2.38E-07	22.39	16.64	TBR1	—
AIRE		2.24E-06	6.04E-09	17.97	14.33	AIRE	—
NR6 (GCN-rel. NR)		3.08E-06	1.29E-16	24.83	21.27	NR6A1	—
Brachyury-related TFs		4.86E-06	6.20E-09	15.65	11.14	T	—
NR1H2::RXRA		9.45E-06	1.43E-12	22.32	17.66	NR1H2::RXRA	—
3 ZnF Krüppel TFs		9.98E-01	9.79E-01	64.81	68.16	KLF5	—
B-ATF-related TFs		9.68E-01	6.05E-01	32.75	36.76	BATF	—
NGFI-B related TFs		1.42E-01	8.48E-08	29.83	29.38	NR4A1	—
Fos/AP-1 TFs		9.80E-01	1.86E-01	27.61	32.56	FOS	FOSB JUNB
STAT		4.35E-05	2.66E-04	23.86	19.10	STAT6	—
Hairy related TFs		2.27E-01	4.26E-01	23.45	20.12	BHLHE40	—
Multiple dispersed ZnF		1.33E-07	2.91E-03	19.32	16.79	MECOM	—
C/EBP-related TFs		2.76E-02	9.98E-01	16.64	17.37	CEBPD	—
PAS domain		1.00E+00	1.00E+00	6.40	8.68	AHR	—

Figure 3. TF binding sites representation at enhancers specific to low-grade PDACs.

The upper part of the table shows the PWMs overrepresented (*P*-values determined by a two-tailed Welch's *t*-test are shown in red) in Lo-G PDAC cell lines relative to both enhancers specific to Hi-G PDACs (*p* vs. Hi-G) and the FANTOM5 collection of human enhancers (*p* vs. *hs* enh) (Andersson *et al.*, 2014). The frequency (*f*) of these PWMs in the enhancers specific to Lo-G and Hi-G PDACs is also shown. The cognate TFs assigned to these overrepresented PWMs are indicated in bold if overexpressed in Lo-G PDACs. The bottom part of the table shows the result of the statistical overrepresentation analysis for those TFs expressed selectively in Lo-G PDACs but whose PWMs were not overrepresented in this analysis (*P* > 1E-5 determined by two-tailed Welch's *t*-test).

Figure 4.

Figure 4. Identifying TFs acting at subtype-specific enhancers.

- A Consensus DNA binding sites for the TFs overexpressed in Lo-G PDACs at enhancers specific to the same cell lines were detected using FIMO (Grant *et al*, 2011) (P -value $\leq 2E-4$). The orange dot indicates DNA binding sites overrepresented relative to Hi-G-specific enhancers (as determined by Pscan using as threshold $P \leq 1E-5$, see Materials and Methods).
- B ChIP-seq snapshots from the low-grade CFPAC-1 cell line showing histone modifications, RNA-seq data, and the recruitment of the five TFs analyzed to two representative genomic regions containing differentially expressed genes (*KRT23*, left and *CEACAM6*, right). Sequencing tracks of the high-grade PANC-1 cell line are shown at the bottom for comparison.
- C Cobinding of the five TFs analyzed by ChIP-seq at the enhancers selectively acetylated in Lo-G PDACs. The matrix indicates the number of regions with the indicated combinations of bound TFs.
- D TF distance from the center of acetylated regions specific to Lo-G PDACs. The box plots indicate the $\log_2$ (tag counts) of the five TFs at enhancers specific to Lo-G PDACs. Central values represent the median, the bars the 25th and 75th percentile and the dashed lines the lower and upper whiskers.
- E Cumulative distributions showing histone acetylation and H3K4 mono- and trimethylation levels of the genomic regions bound by KLF5. Data are shown for CFPAC-1 (continuous line) and PANC-1 cells at both TSS-distal (top panels) and TSS-proximal (bottom panels) regions.
- F KLF5 signal intensity at genomic regions corresponding to distinct classes of enhancers. Central values represent the median, the bars the 25th and 75th percentile and the dashed lines the lower and upper whiskers.

KLF5 as a candidate regulator of the epithelial program in low-grade PDACs

The motivation to further investigate KLF5 came from several considerations: First, it was one of the most highly and selectively overexpressed TFs in the low-grade PDAC cells (Fig 1); second, its binding site was pervasively detected in the enhancer set specific to this PDAC subtype (Fig 4A); third, KLF5 is required for terminal differentiation of several endoderm-derived epithelia (Wan *et al*, 2008; Bell *et al*, 2011, 2013; McConnell *et al*, 2011), suggesting that it may contribute to the transcriptional output of well-differentiated PDACs as well.

The cumulative distributions (Fig 4E, top) of KLF5 ChIP-seq peaks confirmed that KLF5 was preferentially associated with candidate enhancers selectively acetylated and monomethylated at H3K4 in the low-grade PDAC cell lines. Specifically, of the 4,140 enhancers specific to Lo-G PDACs, those bound by KLF5 were 2,691 (65%) using as threshold a MACS score of 100 and 3,298 (79.7%) using a score of 50. Conversely, the promoters bound by KLF5 had a similar profile of histone acetylation and methylation in the two PDAC subtypes (Fig 4E, bottom), although clearly this observation does not rule out a role of KLF5 at promoters. Moreover, KLF5 binding to low-grade-specific enhancers was significantly higher than that to enhancers shared by low- and high-grade PDAC cell lines (Fig 4F).

Many of the genes that specify epithelial identity were associated with prominent KLF5 binding, as exemplified by the junctional plakoglobin (*JUP*, encoding a desmosome component) and *ELF3* genes (Table EV7). Consistently, a GREAT analysis on the KLF5-bound regions retrieved categories related to cellular adhesion and epithelial differentiation, as well as to various cancers of endodermal origin (Table EV5).

These data motivated us to accurately determine whether KLF5 expression in human PDACs correlates with their degree of epithelial differentiation. We first used immunohistochemistry to analyze KLF5 expression in a small panel of patients for which we obtained histological sections containing areas representative of all three tumor grades. A strong KLF5 immunoreactivity was detected in tumor areas with evident duct-like structures (Fig 5A). Glandular epithelia containing mucin-producing neoplastic cells with pale cytoplasm and basally located round-to-ovoid nuclei—typical of low-grade tumors (G1)—displayed the strongest signal, while KLF5-expressing cells were less frequent and showed overall weaker signals in G2 areas (characterized by large duct-like structures

embedded in desmoplastic stroma and showing nuclear crowding and loss of polarity). Finally, KLF5 expression was almost completely absent in the small pleomorphic cell nests with little or no mucin production and marked nuclear polymorphism that represent G3 areas. As control, upon subcutaneous injection in nude mice the Lo-G PDAC cell line CFPAC-1 generated a moderately differentiated tumor with strong KLF5 expression; in contrast, the Hi-G PDAC cell line MiaPaCa-2, that generated poorly differentiated and densely packed cell sheets and nests, did not show any KLF5 immunoreactivity, thus confirming the specificity of the staining.

We next extended this analysis using tissue microarrays (TMA) containing 77 tumors areas from 45 patients and representing the three tumor grades. Consistent with the results above, data from automatically acquired and quantified images (Fig 5B) showed a strong and statistically significant correlation of KLF5 staining with low-grade tumor areas. Representative areas from the TMA are shown in Fig 5C.

Finally, we asked at what stage of pancreatic carcinogenesis KLF5 expression is upregulated. Therefore, we analyzed KLF5 expression in areas of acinar-to-ductal metaplasia (ADM) and pancreatic intraepithelial neoplasia (PanIN) (Reichert & Rustgi, 2011). ADM is the consequence of the transdifferentiation of acinar cells into ductal cells and is considered an early event in pancreatic carcinogenesis, while PanINs (which can arise from ADMs) are intraepithelial neoplastic lesions that can be divided into three different grades (I–III) of increasing histological severity (Takaori *et al*, 2004; Reichert & Rustgi, 2011). Both ADM and PanINs are induced by activating mutations in *KRAS*, which represent the earliest and most common genetic lesions in PDACs (Hezel *et al*, 2006). KLF5 expression was detected in early ADMs and throughout all PanIN grades, with a reduction in the transition from PanIN-2 to PanIN-3 (Fig 5D).

Overall, these data tentatively link KLF5 to maintenance of differentiation and epithelial identity in pancreatic pre-neoplastic lesions and PDACs and indirectly suggest that loss of KLF5 expression may contribute to histological grade progression of human PDACs. Therefore, we undertook a genetic approach to directly address its role in the maintenance of the epigenome and transcriptional program of low-grade PDAC cells.

KLF5 contributes to control the epithelial program in PDACs

To determine the role of KLF5 in maintaining the transcriptional output of low-grade PDACs, we generated knockout clones using

Figure 5. KLF5 expression in human PDACs.

- A KLF5 immunoreactivity in normal human pancreas, human PDACs of different histological grade, and mouse xenografts of CFPAC-1 and MiaPaCa-2 cells. One representative patient out of nine is shown. Xenografts: $n = 4$ for each cell line.
- B Violin plot showing the fraction of tumor cell nuclei that were positive for KLF5 in the tissue microarray. Images were automatically acquired and quantified using the ImmunoRatio software. P -values were calculated using a two-tailed Welch's t -test. The number of tissue core sections for each grade is indicated (n total = 77). White central dots represent the median.
- C Representative sections from the tissue microarray stained with the anti-KLF5 antibody are shown. The panels on the right were pseudocolored for automated quantification.
- D KLF5 expression in primary human samples of ADM and PanIN. One representative patient out of nine is shown.

CRISPR/Cas9-mediated genome editing in the CFPAC-1 cell line (Appendix Fig S4A). KLF5-KO cells grew slower and in a more dispersed fashion compared to their wild-type counterpart and generated smaller tumors *in vivo* (Appendix Fig S4B and C). To measure differential gene expression between KLF5-KO and control clones, we carried out a pairwise comparison of their transcriptomes. Using a $\log_2(\text{fold change}) \geq 1$ and an $\text{FDR} \leq 0.1$, we found 2,343 differentially expressed genes, with 1,221 up- and 1,122 downregulated genes in the KLF5-KO clones (Fig 6A; the complete list of genes is reported in Table EV8). Of the 689 genes specific to the Lo-G PDACs in our dataset (Appendix Fig S1 and Table EV2), 113 (16.4%) were downregulated in KLF5-KO cells. Finally, when considering the 24 genes that define the classical PDAC signature (Collisson *et al*, 2011), 7 of them (highlighted in green in Fig 6A) were downregulated in a statistically significant manner in KLF5-KO clones.

A gene ontology (GO) analysis of differentially expressed genes is shown in Fig 6B. Top-ranking categories were related to: (i) cell cycle, DNA replication and repair, which is consistent with the slower growth of these cells (Appendix Fig S4) and may be in part explained by the reduced expression of *WNT10A*, a direct KLF5 target (Appendix Fig S4D and Table EV8), (ii) cell adhesion, migration and organization of the extracellular matrix, and (iii) inflammatory response. The full list of enrichments is reported in Table EV9.

KLF5-dependent genes included many “epithelial identity genes”, such as: (i) most of the keratin genes expressed in PDACs and particularly all those in the large chromosome 12 cluster (Appendix Fig S4E), (ii) focal adhesion genes (such as *FERMT1*, encoding kindlin-1, *LCK*, and *FYN*, which encode Src family kinases), (iii) *CEACAM* genes, which are biomarkers of PDACs and other epithelial tumors (Beauchemin & Arabzadeh, 2013), (iv) genes encoding gap junction proteins (connexins) such as *GJB3*, *GJB4*, and *GJB5*, (v) mucins (*MUC4* and *MUC1*), (vi) genes encoding TFs specific to classical PDACs (and part of their characteristic signature) (Collisson *et al*, 2011) such as *ELF3* and *FOXQ1*.

Conversely, genes upregulated in KLF5-KO cells included: (i) genes associated with EMT such as those encoding fibronectin 1 (*FN1*), vimentin (*Vim*), N-cadherin (*Cdh2*) and lysyl oxidase (*LOX*), and (ii) genes preferentially expressed in PDAC cancer stem cells (Hermann *et al*, 2007; Li *et al*, 2007) such as *CD44*, *PROM1* (encoding CD133) and *SHH* (encoding sonic hedgehog). Some of the genes mentioned are highlighted in the volcano plot in Fig 6A.

Instead, other genes that impact on the adhesive properties of epithelial tumors and are typically downregulated in EMT (De Craene & Berx, 2013), notably *CGN* (encoding cingulin, a component of tight junctions) and *JUP* (encoding plakoglobin, a desmosome component), were less dramatically (albeit significantly) reduced in KLF5-KO cells. *CDH1* (encoding E-cadherin) was only mildly affected by KLF5 loss *in vitro*. The proliferation defect as well as the morphological and the gene expression changes was reversed by the re-expression of KLF5 in the KO clones (Appendix Fig S5).

We next analyzed the impact of *KLF5* loss on the CFPAC-1 epigenome. A total of 6,391 H3K27Ac peaks were affected by KLF5 loss (Table EV10), with 88.7% of acetylation changes occurring at TSS-distal regions that represent candidate enhancers (Fig 6C and D). Although hyperacetylated distal peaks ($n = 3,264$) in KLF5-KO cells were more frequent than the hypoacetylated ones ($n = 2,406$), they likely represented indirect effects of KLF5 loss, as indicated by their preferential occurrence at regions with no or low KLF5 binding (Fig 6E). Conversely, loss of acetylation occurred preferentially at regions strongly bound by KLF5 (Fig 6E).

When considering the 4,140 enhancers specific to Lo-G PDACs, KLF5 deletion resulted in reduced acetylation at 934 (22.5%) of them (MACS score ≥ 50). Enhancer hypoacetylation was associated with reduced transcription of the closest expressed gene in 57% of cases, thus indicating that reduced acetylation correlates with the transcriptional outcome (Table EV8). A GREAT analysis on the hypoacetylated regions (Table EV5) retrieved terms related to various endodermal cancers, inflammation, metabolism as well as epithelial differentiation.

We also analyzed changes in H3K4me1 due to KLF5 loss. A total of 4,122 H3K4me1 peaks were hypomethylated in KLF5 KO cells (Fig 6C), and 93.4% of them were putative enhancers. By crossing the H3K27Ac and the H3K4me1 data, we found that 39.3% of the hypoacetylated enhancers also showed reduced H3K4me1 levels.

In Fig 6F, we show two snapshots representative of our main findings. In the upper panel, loss of *MUC4* gene expression is associated with reduced H3K27Ac and, to a lesser extent, H3K4me1 across a large regulatory domain. In the bottom panel, it is shown the upregulation of the *VIM* gene, encoding the mesenchymal marker vimentin, which is associated with a gain in H3K27Ac at the KLF5-bound promoter. Since *VIM* is a direct KLF5 target, it may represent an unusual case of direct repression by KLF5.

Finally, we analyzed the impact of KLF5 loss on the maintenance of epithelial properties of low-grade PDACs *in vivo*. Specifically, we

Figure 6. KLF5 and maintenance of epithelial identity in PDACs.

- Volcano plot showing the transcriptional effects of KLF5 deletion in CFPAC-1 cells. Differentially expressed genes were identified by comparing two KLF5-deleted clones with two wt clones. The y-axis shows the $-\log_{10}$ of the q -values determined by Cuffdiff. The genes indicated in green are part of the signature of classical PDACs (Collisson *et al*, 2011).
- Gene ontology analysis on the set of differentially expressed genes in KLF5-KO cells. Bubble color indicates the FDR (color scale shown on the right), while size indicates the frequency of the GO term in the GOA database (bubbles of more general GO terms are larger).
- Scatter plots showing H3K27Ac (left) and H3K4me1 (right) levels at putative enhancers (TSS-distal genomic regions) in control vs. KLF5-KO cells. The box plots in the insets show KLF5 tag counts in the hypo- and hyperacetylated or methylated regions. Central values represent the median, the bars the 25th and 75th percentile and the dashed lines the lower and upper whiskers.
- Genomic distribution of common and differential acetylation in control and KLF5-KO cells.
- Histogram showing the percentage of hypo- and hyperacetylated peaks (divided in TSS-proximal and distal) that contain a KLF5 peak.
- Representative RNA-seq and ChIP-seq snapshots centered on a KLF5-dependent gene (*MUC4*) and a KLF5-repressed gene (*VIM*).
- Subcutaneous xenografts of wild-type or KLF5-deleted CFPAC-1 cells were stained with the indicated antibodies. MiaPaCa-2 xenografts are shown for comparison ($n = 4$ animals analyzed per group).

Figure 6.

compared the ability of wild-type and KLF5-deleted CFPAC-1 cells to give rise to organized and well-differentiated epithelia in xenografts. Wild-type CFPAC-1 cells generated differentiated tubular structures that were positive for mucins, CEACAMs, and E-cadherin (Fig 6G). Instead, KLF5-deleted cells lost the expression of those canonical epithelial markers and formed compact cell nests without any obvious lumen that were similar to those generated by the quasi-mesenchymal cell line MiaPaCa-2. Interestingly, while the effects of KLF5 loss on *CDH1* expression *in vitro* were of limited magnitude, they were apparently much stronger *in vivo*, indicating that in the context of the tumor tissue KLF5 greatly impacts E-cadherin expression.

Overall, these data indicate that KLF5 acts as a transcriptional activator contributing to control the epithelial gene expression program of low-grade PDACs.

KLF5 requirement for recruitment of other Lo-G-specific TFs to cis-regulatory elements

The effects of KLF5 loss on the CFPAC-1 epigenome (Fig 6C–E) prompted us to determine whether it is required for the genomic recruitment of other Lo-G-specific TFs. We analyzed the ELF3 and

FOXA1 ChIP-seq profiles in control and KLF5-KO clones. KLF5 loss selectively affected a subset of ELF3 and FOXA1 peaks. Specifically, 2,986/51,835 (5.7%) ELF3 peaks and 3,343/46,651 (7.2%) FOXA1 peaks were significantly reduced in KLF5-deficient cells (Fig 7A and B and Table EV11). A motif overrepresentation analysis on the unaffected and the reduced ELF3 and FOXA1 peaks indicated fundamental differences in the underlying DNA sequences (Appendix Fig S6 and Table EV12). At regions where FOXA1 and ELF3 binding was dependent on KLF5, the most enriched PWMs matched KLF motifs, while at regions where binding was KLF5-independent, ELF and FOXA PWMs were retrieved. Moreover, the genomic regions hypoacetylated in KLF5-KO cells (Table EV10) showed strongly reduced ELF3 and FOXA1 binding in KLF5-KO cells (Fig 7C). ELF3 and FOXA1 binding was also reduced at classical PDAC signature genes (Appendix Fig S7). These data suggest that KLF5 tethers ELF3 and FOXA1 to a subset of enhancers and that the combined loss of recruitment of these TFs contributes to reduction of acetylation. Importantly, although many genomic regions showed increases in ELF3 and FOXA1 recruitment in KLF5-KO cells (orange dots in Fig 7A and B), such regions were characterized by significantly lower levels of KLF5 binding than those with reduced ELF3 and FOXA1 recruitment

Figure 7. Effects of KLF5 deficiency on binding of other TFs.

- A, B Scatter plots showing ELF3 and FOXA1 levels (as determined by ChIP-seq) at bound genomic regions in wild-type and KLF5-KO cells. Light blue dots indicate regions where ELF3 or FOXA1 signal is reduced in KLF5-deficient cells.
- C Box plot showing ELF3 and FOXA1 signal intensity in WT and KLF5-KO clones at genomic regions whose histone acetylation is reduced in KLF5-deficient cells (** $P \leq 1E-100$, two-tailed Welch's t-test). Central values represent the median, the bars the 25th and 75th percentile and the dashed lines the lower and upper whiskers.
- D KLF5 signal intensity at genomic regions where ELF3 or FOXA1 signal is reduced in KLF5-KO clones relative to WT clones. Central values represent the median, the bars the 25th and 75th percentile and the dashed lines the lower and upper whiskers.
- E Representative snapshots showing genomic regions where a subset of ELF3 and/or FOXA1 signals (indicated by red boxes) are reduced in KLF5-KO cells.
- F The Venn diagram shows the overlap between KLF4, KLF5, and KLF6 ChIP-seq signals in CFPAC-1 cells. The number of ChIP-seq peaks is indicated.

(Fig 7D), hinting at indirect effects of KLF5 deficiency. Two representative snapshots are shown in Fig 7E.

KLF family TFs share highly similar DNA binding specificities and several of them are coexpressed in PDAC cell lines irrespective of their grade. Therefore, we asked whether the genomic distribution of KLF5 differs from that of non-grade-specific KLF family members (KLF4 and KLF6). In spite of the different efficiency of individual antibodies, ChIP-seq data indicated largely overlapping genomic distributions of KLF5, KLF4, and KLF6 (Fig 7F and Table EV13). Also when restricting this analysis to the regions that lost acetylation in KLF5-KO cells (Fig 6C–E), 81.8% of them were non-selectively bound by all three KLF TFs. Overall, the specific requirement for KLF5 in Lo-G PDAC cells did not depend on its unique DNA binding properties, but rather on its ability to interact with, and recruit to, selected genomic regions a group of partner TFs.

Integrated control of the PDAC epithelial program by multiple transcriptional regulators

The observation that some classical epithelial identity genes were not affected by KLF5 loss prompted us to compare the effects of KLF5 deletion with those generated by modulating the expression of validated positive or negative regulators of the epithelial program (De Craene & Berx, 2013).

ELF3 has been linked to epithelial differentiation and renewal in normal tissues (Ng *et al*, 2002; Oliver *et al*, 2011) and in various cancers (De Craene & Berx, 2013). It is a component of the gene signature characteristic of classical PDACs (Collisson *et al*, 2011) and is extensively associated with the enhancer repertoire characteristic of Lo-G PDACs (Fig 4). Nevertheless, the overall transcriptional impact of *ELF3* deletion in CFPAC-1 cells (generated by genome editing, Appendix Fig S4) was milder than that of *KLF5*, both in terms of number of affected genes and magnitude of the effects (Fig 8A and B and Table EV8). Deregulated genes were included in functional categories only partially overlapping those affected by KLF5 loss, such as motility and proliferation (Fig 8C). Moreover, we did not detect obvious morphological changes or proliferation defects in *ELF3* KO clones compared to control cells (data not shown).

Among the genes related to epithelial identity, *ELF3* shared with KLF5 the control of some mucins, kallikreins, and several keratin genes, although the effects of *ELF3* deletion on the mRNAs of these genes were in general more limited and of lower magnitude than those of *KLF5* loss (Table EV8). Moreover, *ELF3* loss did not have any obvious effects on the genes mentioned above that encode components of focal adhesions, gap junctions, and tight junctions (Table EV8). An example of a discordantly regulated gene (*LCK*, encoding a Src family protein kinase) bound by both TFs is provided in Appendix Fig S4F.

Consistent with transcriptomic data, also the effects of *ELF3* loss on the CFPAC1 epigenome were milder than those caused by the *KLF5* deletion. *ELF3*-KO and control cells shared more than 45,000 H3K27Ac peaks but, remarkably, they differed by only about 200 peaks (119 hypoacetylated peaks and 91 hyperacetylated peaks in the *ELF3*-KO cells) (Fig 8D, left). Similarly, only 138 H3K4me1 peaks were hypo- and 50 hypermethylated in *ELF3*-KO cells compared to their control clones (Fig 8D, right).

ZEB1 is expressed at higher levels in high-grade PDACs and is a well-characterized suppressor of epithelial differentiation, acting at least in part by repressing the expression of tight junction proteins such as *CDH1* (De Craene & Berx, 2013). To directly address how ZEB1 controls the expression of epithelial genes in PDACs, we overexpressed it in the low-grade CFPAC-1 cells (Appendix Fig S8A). ZEB1 expression was associated with morphological changes consistent with a transition toward a mesenchymal phenotype (Appendix Fig S8B). An RNA-seq analysis on these cells showed comparatively limited changes in mRNA profiles with 128 downregulated and 127 upregulated genes ($\log_2FC \geq \pm 1$, $FDR \leq 0.1$) showing only a partial overlap with those affected by KLF5 or *ELF3* loss (Fig 8A and E, and Table EV8). Some notable genes are indicated in the volcano plot in Fig 8E. Genes concordantly downregulated by ZEB1 overexpression and KLF5 loss included several keratin genes, mucins, and *GRHL2*, a TF that antagonizes the mesenchymal transition (Cieply *et al*, 2012), and is selectively expressed in low-grade PDACs (Table EV2). However, many others that were repressed by ZEB1 (notably *CLDN4* and *CLDN7*, encoding tight junction proteins) did not require KLF5 (Fig 8E and Table EV8), as shown in the representative snapshot in Fig 8F. Differentially expressed genes were enriched in functional categories consistent with the ability of ZEB1 to drive a mesenchymal transition (Fig 8G).

A large number of candidate enhancers active in low-grade PDACs are not acetylated or marked by H3K4me1 in high-grade PDACs (Fig 2). Therefore, one possibility is that ZEB1 inhibits the epithelial program by binding and repressing enhancers specifically active in low-grade PDACs. If this model is correct, ZEB1 peaks should be enriched at genomic regions that are not acetylated in high-grade PDACs and instead acetylated in low-grade PDACs. To address this possibility, we profiled ZEB1 genomic occupancy by ChIP-seq in the high-grade cell line PANC1. ZEB1 showed a comparatively limited number of genomic binding sites (5,063) that in 56% of cases were distant from mapped TSS, thus representing putative enhancers. In contrast with a direct enhancer repression model, however, TSS-distal ZEB1 peaks corresponded to genomic regions that were not highly acetylated in low-grade PDACs (Fig 9A). The specificity of the observed signals is indicated by a *de novo* motif analysis that retrieved three degenerate PWMs all containing the known core ZEB1 binding site (Fig 9B). Moreover, the ZEB1 PWM was also the most statistically overrepresented motif in this dataset. The snapshots in Fig 9C provide representative examples, including a prominent ZEB1 peak in the genomic region containing the MIR-200 cluster (left), a validated ZEB1 target (Park *et al*, 2008) and a peak just downstream of the *ADAM8* gene.

Discussion

In this study, we integrated different layers of analyses to determine the transcriptional bases that underlie the existence of PDACs retaining a well-differentiated epithelial phenotype and PDACs with an almost complete loss of the ability to generate organized glandular epithelia.

The main findings of this study can be summarized as follows. First, we identified a set of genomic *cis*-regulatory elements that were selectively acetylated in differentiated, low-grade PDACs. The

Figure 8. Comparison between the transcriptional effects of KLF5 loss and those of ELF3 loss and ZEB1 overexpression in CFPAC-1 cells.

- A Venn diagram showing the overlap between genes downregulated or upregulated in response to the indicated perturbations.
- B Volcano plot showing the genes differentially expressed in ELF3-deleted CFPAC-1 cells. Differentially expressed genes were identified by comparing three ELF3-deleted clones with two wt clones. The y-axis shows the $-\log_{10}$ of the q -values determined by Cuffdiff.
- C Gene ontology analysis of the set of differentially expressed genes in ELF3-KO cells.
- D Scatter plots showing H3K27Ac (left) and H3K4me1 (right) levels at putative enhancers (TSS-distal genomic regions) in control vs. ELF3-KO cells.
- E Volcano plot showing genes differentially expressed upon overexpression of ZEB1 in the low-grade PDAC cell line CFPAC-1. The y-axis shows the $-\log_{10}$ of the q -values determined by Cuffdiff.
- F A representative snapshot showing the differential effects of ZEB1 overexpression and KLF5 loss on the *CLDN7* gene.
- G Gene ontology analysis on the set of differentially expressed genes in ZEB1-overexpressing CFPAC1 cells.

partial loss of epithelial identity in high-grade PDAC cells was associated with the loss of acetylation at such enhancers; instead, the *de novo* acquisition of enhancers specific to poorly differentiated PDACs was a rather uncommon event. Second, we identified a cognate panel of TFs acting on such enhancers specific to low-grade PDACs and controlling their activity. Among them, the TF KLF5 was selectively expressed in low-grade human PDACs and was required for the maintenance of transcriptional and epigenomic features that

determine their epithelial identity and thus their ability to form differentiated glandular structures when xenografted. Third, we demonstrated that maintenance of the epithelial program in PDAC cells reflects the complementary activities of distinct TFs, with KLF5 controlling the expression of genes largely distinct from those activated by ELF3 as well as from those suppressed by the transcriptional repressor ZEB1, a prominent inducer of the mesenchymal fate.

Figure 9. Genomic distribution of ZEB1.

A Heatmaps and cumulative distributions showing ZEB1 genomic distribution in the high-grade PDAC cell line PANC1. For comparison, the acetylation and methylation levels at the same regions are also shown in the low-grade PDAC cell line CFPAC1.

B PWMs identified by *de novo* motif discovery at the top 2,500 ZEB1-bound genomic regions. The MEME E-values of individual PWMs are indicated.

C Two representative snapshots showing (left) a ZEB1 peak upstream of the MIR200A/B cluster and immediately downstream of the ADAM8 gene (right).

Regarding the first group of findings, it is important to consider that in addition to the impact of differentially expressed TFs, the biology of enhancers in PDACs may have additional layers of complications. Specifically, mutations in genes encoding transcriptional coregulators with a defined role in enhancer biology such as MLL3—which catalyzes the deposition of H3K4me1 at enhancers (Herz *et al*, 2012)—frequently occur in human PDACs (Biankin *et al*, 2012) as well as in other cancers (Smith & Shilatifard, 2014). Therefore, alterations of the enhancer landscape (and the ensuing transcriptional effects) may arise not only from differential TF expression or activity, but also because of alterations of the general machinery controlling the deposition of histone modifications.

KLF5 stood out among the TFs specific to differentiated PDACs for two main reasons: its high and selective expression in low-grade PDACs and its pervasive association with their specific enhancer repertoire. These observations are consistent with the demonstrated role of KLF5 in the terminal differentiation and homeostasis of epithelia of endodermal origin (including lung epithelium, urothelium, and intestinal epithelium) (Wan *et al*, 2008; Bell *et al*, 2011, 2013) (McConnell *et al*, 2011; Nandan *et al*, 2014). Moreover, a

recent systems-level study based on the transcriptomic analysis of highly purified cell populations from embryonic and postnatal mouse pancreas identified KLF5 and other TFs, retrieved from our analysis as specific to low-grade PDACs, as predicted regulators of pancreatic development (Benítez *et al*, 2014). Finally, a GWAS study retrieved a SNP associated with pancreatic cancer risk in the chromosome 13 intergenic region located between KLF5 and KLF12 (Petersen *et al*, 2010) and contained in the KLF5-associated super-enhancer we identified.

Reducing or abrogating KLF5 expression in low-grade PDACs resulted in decreased proliferation, downregulation of a broad panel of genes whose products impact epithelial identity (such as keratins, mucins, and E-cadherin) and instead upregulation of pancreatic cancer stem cell (*CD44*, *PROM1*, and *SHH*) (Li *et al*, 2007) and mesenchymal genes (*FN1*, *VIM*, and *CDH2*) which is consistent both with the reduced proliferative ability of cancer stem cells and with the known relationship between EMT and acquisition of cancer stem cell properties (Mani *et al*, 2008). Along the same line, quasi-mesenchymal, high-grade PDACs were shown to lose the addiction to K-Ras that instead characterizes epithelial

PDACs (Singh *et al*, 2009); some of the genes that were shown as critically required for survival of K-Ras-addicted epithelial PDACs, notably *SYK* and *MST1R* (encoding the RON kinase), were dependent on KLF5.

While data on its expression and role in pancreatic development are not available, KLF5 was not clearly detectable in the normal human adult pancreas. Therefore, a possible scenario is that its expression is reactivated in the pathologically proliferating pre-neoplastic and neoplastic exocrine epithelium where it may work to coordinate active proliferation with maintenance of the epithelial identity of proliferating cells. In this context, it is remarkable that the expression of KLF5 in a large panel of human PDACs displayed a strong and statistically significant correlation with tumor grade, being the highest in G1, still detectable in G2 and almost completely absent in G3 areas.

The interplay between KLF5 and other grade-specific TFs in the control of the transcriptional output of differentiated PDACs remains to be determined. In addition to ELF3, another TF reported to control epithelial identity in tumors (De Craene & Berx, 2013) and that is activated by KLF5 also in normal development (Bell *et al*, 2011, 2013), a few other TFs specific to low-grade PDACs (such as FOXQ1 and GRHL2) were downregulated upon KLF5 loss. These data suggest that KLF5 does not have an apical position in the transcriptional network that controls PDAC epithelial identity, but that conversely it may act as a direct activator of grade-specific enhancers. Reduced KLF5 expression during PDAC progression may reflect either a change in the upstream transcriptional regulatory circuits that maintain KLF5 expression in differentiated PDACs or the emergence of more aggressive subclones lacking KLF5 expression from the onset. Defining the identity of the TFs controlling KLF5 regulation in normal and tumoral pancreas will thus be crucial to deconstruct the regulatory network controlling epithelial identity of PDACs.

An important issue that is also highlighted by our findings is the complementarity and division of labor among the TFs that control the epithelial identity of PDACs. The maintenance of the properties characteristic of epithelia (in epithelial tumors just like in normal tissues) is in fact a highly integrated task that requires the expression of numerous sets of genes controlling distinct epithelial features (from keratin genes that impact mechanical properties of epithelial cells, to genes encoding junctional proteins and even metabolic regulators). On the one hand, our data indicate a broad role of KLF5 in maintaining the expression of many epithelial identity genes; on the other hand, they also imply that additional TFs, such as ELF3, which control largely distinct sets of epithelial genes, must complement its activities. For instance, when considering proteins of the epithelial cell surface, KLF5 appeared to be only marginally involved in the activation of various junctional protein genes such as claudins, but it was absolutely required for the regulation of several mucins and connexins. Along the same line, when comparing the effects of KLF5 and ELF3 loss with those of ZEB1 overexpression, an obvious complementarity of effects became apparent, as exemplified by the strong inhibition of claudin genes and junctional plakoglobin (*JUP*) by ZEB1.

In conclusion, this study provides a basic framework for the transcriptional basis of PDAC grading and also suggests a general approach to undermine transcriptional control of differentiation in other types of cancers.

Materials and Methods

Antibodies

The following antibodies were used for ChIP-sequencing experiments: H3K4me1 (ab8895, Abcam), H3K27Ac (ab4729, Abcam), H3K9me3 (ab8898, Abcam), H3K4me3 (#39159, Active Motif); KLF5 (#07-1580, Millipore), ELF3 (HPA003479, Sigma), HNF1B (HPA002083, Sigma), FOXA1 (ab23738, Abcam), ZEB1 (sc-25388, Santa Cruz Biotechnology), IRF1 (sc-640, Santa Cruz Biotechnology), KLF4 (AF3640, R&D), KLF6 (sc-7158, Santa Cruz Biotechnology). Immunohistochemistry was performed with anti-HNF1B (HPA002083, Sigma), JUNB (sc-8051, Santa Cruz Biotechnology), FOS (HPA018531, Sigma), ZEB1 (sc-25388, Santa Cruz Biotechnology), ETV5 (ab102010, Abcam), GATA2 (ab153820, Abcam), TCF12 (sc-357, Santa Cruz Biotechnology), MUC4 (sc-33654, Santa Cruz Biotechnology), CEACAM6 (ab137859, Abcam), and CDH1 (M3612, Dako). The affinity-purified rabbit polyclonal anti-KLF5 and anti-ELF3 antibodies used for Western blot and immunohistochemistry experiments were generated in-house.

Western blots

CFPAC-1 cells were lysed in RIPA buffer containing protease inhibitors, 1 mM PMSF, 1 mM EDTA, and 1 mM NaF, and 80 µg of clarified cell extracts was resolved on SDS-polyacrylamide gel, blotted onto nitrocellulose membranes, and probed with anti-KLF5 (1 µg/ml), anti-ELF3 (0.5 µg/ml), anti-IRF1 (2 µg/ml), and anti-ZEB1 (2 µg/ml).

Cells, reagents and transfection

The following human PDAC cell lines were used: CFPAC-1 (G1-G2, established from liver metastases, ATCC CRL-1918), CAPAN-1 (G1 from liver metastases, ATCC HTB-79), CAPAN-2 (G1 from primary PDAC, ATCC HTB-80), HPAF-II (G1-G2 from ascites, ATCC CRL-1997), MiaPaCa-2 (G3 from primary tumor, ATCC CRL-1420), PANC-1 (G3 from primary tumor, ATCC CRL-1469), PT45P1 (G3 from primary tumor, obtained from Paola Allavena, Humanitas Research Hospital, Milan), AsPC-1 (G2 from ascites, ATCC CRL-1682), BxPC-3 (G2 from primary tumor, ATCC CRL-1687).

Cells were maintained in IMDM/10% FBS (CFPAC-1), RPMI/10% FBS (HPAF-II, PT45P1, AsPC-1), RPMI/15% FBS (CAPAN-2), RPMI/20% FBS (CAPAN-1), DMEM/10% FBS (MiaPaCa-2, PANC-1, BxPC-3). Media were all supplemented with 2 mM L-glutamine. RPMI 1640 and DMEM were provided by Lonza, IMDM by Sigma, and fetal bovine serum (FBS) by Hyclone. ~0.5 × 10⁶ CFPAC-1 cells were electroporated with 2.5 µg of plasmid DNA using the Neon[®] Transfection system (Life Technologies) setting the microporator at 1,250 V for 2 pulses of 20 ms each. Cells were seeded in 6-well dishes and cultured for 48 h before further manipulations.

The ZEB1 expression plasmid was generated by subcloning into the BamHI sites of the pCDH-EF1-MCS-BGH-PGK-GFP-T2A-Puro vector, the human full-length cDNA sequence generated by PCR (using the following primers: ZEB1-BamHI_F: 5'-CGCGGATCCATG CCGGATGGCCCCAGGTGT-3', ZEB1-BamHI_R: 5'-CGCGGATCCTTA

GGCTTCATTTGTCT-3'). The construct was verified by Sanger sequencing. Cells electroporated with the ZEB1 construct or empty vector were grown for 5 days in culture medium supplemented with 1.5 µg/ml puromycin before harvesting for RNA and protein isolation.

ChIP-seq and RNA-seq

ChIP-seq and RNA-seq were carried out using previously described protocols (Austena et al, 2012) on an Illumina HiSeq2000 platform. 2–5 × 10E6 (histone marks) or 20–40 × 10E6 (TFs) fixed cells were lysed to prepare nuclear extracts. After chromatin shearing by sonication, lysates were incubated overnight at 4°C with protein G Dynabeads (Invitrogen) coupled with 3 µg of anti-histone antibody or 10 µg of anti-TF antibody. After immunoprecipitation, beads were recovered using a magnet and washed; chromatin was eluted and cross-links reverted overnight at 65°C. DNA was either purified with QiaQuick columns (QIAGEN) or solid-phase reversible immobilization (SPRI) beads (Agencourt AMPure XP, Beckman Coulter) and quantified with PicoGreen (Invitrogen). DNA libraries were prepared for HiSeq2000 sequencing as previously described (Ostuni et al, 2013). Total RNA was extracted from 1–4 × 10E6 cells using Maxwell[®] 16 LEV Simply RNA cells kit (Promega) and run on Agilent Bioanalyzer 2100 to assess sample integrity. mRNA-seq library preparation from 1 µg of total RNA (RIN > 8) was performed with TruSeq RNA Sample Prep Kit (Illumina) according to the manufacturer's instructions.

RT-qPCR

cDNA was prepared from 1 µg of total RNA with ImProm-II[®] Reverse Transcription System (Promega) following manual instructions. RT-qPCR was assembled with Fast SYBR[®] Green Master Mix and run on 7500HT ABI Prism machine (Applied Biosystems). Analysis (SDS v2.0.6 software, Applied Biosystems) and primer design were performed following MIQE guidelines using primer sets selected from the suggested validated databases: *MUC1* (ID 3226) and *FN1* (ID 1092) from RTprimerDB (<http://www.rtprimerdb.org>), *LCK* (ID 336285427c1), and *WNT10A* (ID 16936519c3) from PrimerBank (<http://pga.mgh.harvard.edu/primerbank>). *C1ORF43* (F: 5'-GGATGAAAGCTCTGGATGCC-3', R: 5'-GCTTTGCGTACACCC TTGAA-3') was used as reference gene based on the analysis of data from the Human BodyMap (HBM) 2.0 Project.

CRISPR/Cas9 genome editing

Single-guide sequences specific to KLF5 (exon 2) and ELF3 (exon 7, encoding the ETS domain) were designed using the CRISPR design tool (<http://tools.genome-engineering.org>) (Ran et al, 2013) and cloned into pSpCas9(BB)-2A-GFP (Ran et al, 2013). The sequences selected (based on the lowest number of predicted off-targets in exons and the highest predicted efficiency) were the following ones: KLF5: 5'-CACCGAAGAACTGGTCTACGACTG-3', ELF3: 5'-CACCGAT CCACCGGAGCTCAACGA-3'.

After electroporation, single cells were seeded in 96-well plates by dilution and expanded. Clones were screened using the Surveyor assay (Ran et al, 2013) using the following primers:

KLF5_Surveyor-F: 5'-ATTGGGGGAGCAGTTTTACC-3', *KLF5_Surveyor-R*: 5'-GGCATTGTGTATGTGCAAGG-3'; *ELF3_Surveyor-F*: 5'-CTCTGGGACACCCTCAATGT-3', *ELF3_Surveyor-R*: 5'-ACAAGCC TCACAGGCGATAC-3'.

Positive clones were subjected to Sanger sequencing.

Immunohistochemistry and tissue microarrays

Human PDAC specimens were provided by the Humanitas Clinical Institute (Milan, Italy) with written consent for tissue donation and under a protocol approved by the HCI ethical committee. Human pancreatic cancer tissue microarray slides (PA721 and PA1002a) were obtained from US Biomax, Inc. and contained cores for a total of 45 ductal adenocarcinomas of different pathological grade. Formalin-fixed paraffin-embedded (FFPE) sections were rehydrated and subjected to heat-induced antigen retrieval in EDTA buffer (1 mM EDTA, 0.05% Tween-20, pH 8.0). After blocking, samples were incubated with the following antibodies: anti-KLF5 (2 µg/ml), ELF3 (1 µg/ml), HNF1B (0.25 µg/ml), JUNB (2 µg/ml), FOS (0.25 µg/ml), ZEB1 (1 µg/ml), ETV5 (2 µg/ml), GATA2 (2 µg/ml), TCF12 (0.2 µg/ml), MUC4 (2 µg/ml), CEACAM6 (2 µg/ml), CDH1 (1 µg/ml), or corresponding amount of control IgG and detected with HRP-conjugated anti-rabbit or anti-mouse IgG antibody (Envision, DAKO). Signal was obtained by incubation with diaminobenzidine (DAB chromogen system, DAKO) and nuclei were counterstained with hematoxylin. Images were acquired using an Olympus upright BX51 microscope linked to a Nikon DS-5Mc Color camera. Quantitative image analysis of TMA slides was performed with the publicly available web application ImmunoRatio, <http://jvsmicroscope.uta.fi/immunoratio>.

Mouse xenografts

Mouse xenografts were generated and collected in accordance with the Italian laws (D.L.vo 116/92 and following additions), which enforce the EU 86/609 directive, and under the control of the institutional organism for the animal welfare (Cogentech OPBA). Nude mice ($n = 4$ for each group) were injected with 5 × 10E6 cells (CFPAC-1 control clones, CFPAC-1 KLF5-KO clones, and MiaPaCa-2 cells) resuspended in 100 µl of PBS under the skin of their hind flank. Subcutaneously injected mice were harvested 4 weeks after injection, fixed in 4% paraformaldehyde, and processed for paraffin embedding.

RNA-seq analyses

Read mapping, transcript assembly, and expression analysis

After quality filtering according to the Illumina pipeline, 51-bp paired-end reads were mapped to the *Homo sapiens* genome (assembly hg19), using TopHat (version 2.0.14) (Trapnell et al, 2009). TopHat was run with default parameters except for the mean distance between pairs (-r), which was set to 200 bp and Bowtie 0.12.7. Transcript abundance was reported as fragments per kilobase of transcript sequence per million mapped fragments (FPKM) using Cufflinks (version 2.2.1) (Trapnell et al, 2012). For subsequent analyses, we considered the information at the level of genes. In particular, Cufflinks was run setting the option upper-quartile-norm (which improves sensitivity without loss of specificity

for differential expression calls) (Bullard *et al*, 2010), -u (which allows a better weighting of the multimapping reads). Genes with an FPKM ≥ 4 in at least one sample were marked as expressed. Tracks for the UCSC genome browser (Fujita *et al*, 2011) were generated with samtools 0.1.18 and bedtools 2.16.1 (Quinlan & Hall, 2010) using the uniquely aligned reads. Tracks were linearly re-scaled to the same sequencing depth. For all comparisons, the differentially expressed genes (identified using two replicates or more per sample) were identified using Cuffdiff 2.2.1, setting as additional option the model of dispersion *per condition* (each replicated condition received its own model).

Determination of grade-specific transcripts

We identified two groups of genes (Table EV2): (i) a set of differentially expressed genes (DEGs) specific to either Lo-G or Hi-G PDAC cell lines by applying a threshold of FPKM value ≥ 4 in at least 6/8 Lo-G or 4/6 Hi-G cell line datasets, FC ≥ 2 , and FDR ≤ 0.1 , (ii) a set of commonly expressed genes that showed no statistically significant differences, FPKM values ≥ 4 in all samples, and a coefficient of variation ≤ 0.6 .

Determination of differentially expressed transcripts in KLF5-KO, ELF3-KO, and ZEB1-OE CFPAC1 cells

The DEGs for each comparison were identified by applying a threshold of FPKM ≥ 4 in all samples of a given group, FC ≥ 2 , and FDR ≤ 0.1 .

Heatmaps of differentially expressed genes

DEGs were clustered according to their FPKM values. The minimum FPKM value was forced to 0.1 and subjected to log₂-transformation. The distribution of values for each gene was then converted to z-scores. Genes were hierarchically clustered using complete linkage and Pearson correlation as measure of distance. R was used to perform hierarchical clustering and to draw the heatmaps.

Gene set enrichment analysis

Gene Set Enrichment Analysis (GSEA) (Subramanian *et al*, 2005) was used to investigate whether a gene set was significantly over-represented in the transcriptome of either Lo-G or Hi-G PDACs. Transcripts were ranked by the *difference of classes* (metric for gene ranking) and using the following settings: number of permutations = 1,000, permutation type = gene set, chip platform = GENE_SYMBOL.chip, enrichment statistic = weighted, gene list sorting mode = real, gene list ordering mode = descending, max gene set size = 500, min gene set size = 15. The *curated gene sets* collection (c2.all.v5.0.symbols.gmt) was downloaded from the GSEA website (<http://www.broadinstitute.org/gsea/index.jsp>). A gene set was identified as significantly enriched when associated with *q*-value scores ≤ 0.05 .

Functional enrichment analysis

Functional enrichment analyses were generated with the GOrilla tool (Eden *et al*, 2009). The GO enrichment analysis was carried out in the “two lists mode”, using the lists of DEGs and as background the corresponding list of expressed genes. We restricted the analysis to the biological process category and selected GO terms with enrichment (*q*-value $\leq 1E-3$). Data visualization was carried out using REVIGO (<http://revigo.irb.hr/index.jsp>) (Supek *et al*, 2011).

The analysis was run by selecting default parameters except for the resulting list that was setting as *small* size.

ChIP-seq analyses

For each ChIP-seq dataset, bad quality reads were filtered out according to the Illumina pipeline. The remaining reads were then aligned onto the hg19 release of the human genome using Bowtie 0.12.7 (Langmead *et al*, 2009). Options -m 1 -v 2 ensured that only reads showing a unique match to the genome and with two or fewer mismatches were retained for further analyses. Peak calling was performed using MACS 1.4 (Zhang *et al*, 2008), using default parameters except for the following ones: gsize = 2.7E9, nomodel, tsize = 51, and bw = 300 (except for H3K27Ac of PDACs for which a bw = 100 was used). A matched input was used as control. We used a *P*-value threshold of 1E-10 for peak calling (both ChIP vs. input DNA and ChIP vs. ChIP). Gene Interval Notator (GIN), a tool included in the CARPET suite (Cesaroni *et al*, 2008), was then used to annotate all regions over hg19 RefSeq genes extracted from the UCSC genome browser. GIN was run with priority set to “gene” and “-20000” as promoter definition. In order to visualize the raw profiles on the UCSC Genome Browser, wiggle files were generated with MACS and converted to bigwig file format. Tracks were linearly re-scaled to the same sequencing depth.

Classification of grade-specific acetylated regions

Peak calling was performed using MACS for each PDAC vs. input DNA and both ways (Lo-G vs. Hi-G PDACs and Hi-G vs. Lo-G PDACs) for each sample separately. These sets of acetylated regions were merged when located 500 bp within each other and subsequently filtered keeping only the peaks found to overlap at least one region enriched against the input. Peaks that remained unchanged in the comparison of Lo-G vs. Hi-G PDACs (and *vice versa*) and enriched against their respective inputs were considered as *common*; a set of regions enriched in more than 75% of the comparisons between Lo-G- vs. Hi-G PDACs (or *vice versa*) and filtered against the input were defined as *Lo-G-* and *Hi-G-grade-specific*, respectively; finally, all other regions were considered unclassified and not considered for further analyses.

Heatmap of grade-specific acetylated regions

The read counts for H3K27ac, H3K4me1, H3K4me3, and H3K9me3 for each one of the grade-specific acetylated regions were normalized by the sequencing depth. Besides, since these regions are not homogeneous in length, counts were normalized to the size of the region in kbp. To avoid any bias due to the outliers, a saturation procedure was performed considering each antibody independently: Counts exceeding the 95th percentile were set to this value and the distribution of values was transformed into a 0–1 range. R was used to perform hierarchical clustering and to draw the heatmaps.

Identification of super-enhancer-like acetylation clusters

SE-like clusters were identified by adapting a published approach (Whyte *et al*, 2013). First, H3K27ac peaks within ± 2.5 kbp from annotated TSSs of RefSeq genes were excluded. The remaining acetylated regions were merged using mergeBed (Quinlan & Hall, 2010) if located < 12.5 kbp from each other. CoverageBed (Quinlan & Hall, 2010) was then used to quantify the number of reads

contained in each one of the clusters. Finally, regions were ranked based on read count and the top 5% were retained. SE-like clusters were classified as specific to Lo-G or Hi-G if present in at least 3/4 or 2/3 Lo-G and Hi-G PDAC cell lines, respectively.

TFs peaks and associated chromatin signatures

ChIP-seq data analyses for each TF were performed as above. The resulting lists of peaks were then used to highlight the association with the chromatin modifications assayed in this study (H3K27ac, H3K4me1, H3K4me3, and H3K9me3), using cell line-specific ChIP-seq data. ChIP-seq signals were assayed in a window of 5 kbp centered on the summit of the TFs. These windows were then divided in bins of 10 bp each, and the number of reads spanning each one of them was quantified. Read counts from each chromatin modification were then linearly scaled, saturated to 95th percentile, and forced to 0–1. Finally, regions were sorted according to their intensity levels.

Functional enrichment analysis of ChIP-seq enriched regions using GREAT

For each list of ChIP-seq peaks of interest, GREAT 2.0.2 (McLean et al, 2010) was used with default parameters and selecting the whole hg19 genome as background.

Motif enrichment analysis

In order to identify statistically overrepresented motifs corresponding to known TF binding sites, position-specific weight matrices (PWMs) were collected from specific databases and the literature (Hallikas & Taipale, 2006; Berger et al, 2008; Badis et al, 2009; Jolma et al, 2010, 2013; Wei et al, 2010; Kulakovskiy et al, 2013; Mathelier et al, 2014), and used to build a custom set of 1,744 models. The complete list of PWMs used in the statistical overrepresentation analyses is in Table EV14 and the corresponding motifs are provided in MEME motif format (Table EV15). Significantly overrepresented PWMs between any two sets were identified using a modified version of Pscan, in which a *t*-test was implemented in place of the original *z*-test (Zambelli et al, 2009). Any PWM showing a *P*-value equal or lower than 1E-5 was considered as significantly overrepresented. The window considered for these analyses was set to ± 250 bp around the center of grade-specific acetylated regions or around the summit of the peaks in case of TF ChIP-seqs. For each class of interest, different comparisons were performed, using in turn each dataset as foreground or background depending on the type of enrichment to be determined. Specifically, for the set of grade-specific acetylated regions, we compared the acetylated enhancers of the Lo-G and Hi-G subtypes with each other and with the set of human enhancers identified by the FANTOM5 consortium on the basis of the presence of bidirectional, capped RNA within H3K27Ac+ve and H3K4me1+ve regions (Andersson et al, 2014). To identify PWMs enriched in TF ChIP-seq, the FANTOM5 enhancer set was used as background.

De novo motif discovery

We used MEME v4.6.1 (Bailey & Elkan, 1994) for *de novo* identification of TF binding sites associated with the top 5,000 genomic regions bound by each TF. We used 500-bp windows centered on the summit of the ChIP-seq peaks. The following parameters were

used: -dna -mod zoops -nmotifs 10 -minw 6 -maxw 12 -revcomp. We then used TOMTOM (Gupta et al, 2007) (default parameters except for -dist ed) in order to assess the similarity of the identified motifs to known TF consensus binding sites.

Heatmap of TF cobinding

FIMO (Grant et al, 2011) (version included in Meme 4.6.1) scans an input region of DNA for PWM occurrences. It computes a log-likelihood ratio score with respect to each sequence position and converts these scores into *P*-values. FIMO was run on the ChIP-seq peaks with MACS score $\geq 1,000$, using a 500-bp window centered on peak summit. In order to assign the presence of a high-scoring match for a given PWM, a *P*-value threshold of 2E-4 was considered. If below this value, a region was scored as positive for a certain PWM, otherwise it was scored as negative. This way, each region was annotated with the presence or absence of a list of possible TF binding sites and the resulting binary matrices showed as heatmaps.

Data deposition

Raw datasets are available for download at the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/gds/>) under the Super-Series accession number GSE64560.

Expanded View for this article is available online.

Acknowledgements

This project was supported by the Italian Association for Research on Cancer (A.I.R.C.). We thank Bruno Amati (IEO/IIT, Milan), Paco Real (CNIO, Madrid), Silvia Monticelli (IRB, Bellinzona), and Andrea Viale (MD Anderson, Houston) for comments on the manuscript. We also thank Luca Rotta, Thelma Capra, and Salvatore Bianchi (IEO and IIT@SEMM Genomic Unit) for the preparation and processing of the sequencing libraries; Iros Barozzi and Pasquale Laise for help on the computational analyses; Giuseppina Giardina, Marika Zanotti, and Cristina Spinelli for help with cell line culture. Sequence data in fastq format were prepared by Heiko Muller and the Computational Research Unit of the Center for Genomic Sciences of IIT@SEMM and hosted on the IIT@SEMM computational infrastructure.

Author contributions

GRD generated all data with contributions by EP and PN. CB analyzed all sequencing data. GRD and CB contributed equally and the order of their names was determined by drawing. PS and AZ provided clinical samples and their histological assessment. GN designed and supervised the project and wrote the paper with contributions from all authors.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, Ntini E, Arner E, Valen E, Li K, Schwarzfischer L, Glatz D, Raithel J, Lilje B, Rapin N, Bagger FO et al (2014) An atlas of active enhancers across human cell types and tissues. *Nature* 507: 455–461

- Austena L, Barozzi I, Chronowska A, Termanini A, Ostuni R, Prosperini E, Stewart AF, Testa G, Natoli G (2012) The histone methyltransferase Wbp7 controls macrophage function through GPI glycolipid anchor synthesis. *Immunity* 36: 572–585
- Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X, Kuznetsov H, Wang CF, Coburn D, Newburger DE, Morris Q, Hughes TR, Bulyk ML (2009) Diversity and complexity in DNA recognition by transcription factors. *Science* 324: 1720–1723
- Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2: 28–36
- Beauchemin N, Arabzadeh A (2013) Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis. *Cancer Metastasis Rev* 32: 643–671
- Bell SM, Zhang L, Mendell A, Xu Y, Haitchi HM, Lessard JL, Whitsett JA (2011) Kruppel-like factor 5 is required for formation and differentiation of the bladder urothelium. *Dev Biol* 358: 79–90
- Bell SM, Zhang L, Xu Y, Besnard V, Wert SE, Shroyer N, Whitsett JA (2013) Kruppel-like factor 5 controls villus formation and initiation of cytodifferentiation in the embryonic intestinal epithelium. *Dev Biol* 375: 128–139
- Benitez CM, Qu K, Sugiyama T, Pauerstein PT, Liu Y, Tsai J, Gu X, Ghodasara A, Arda HE, Zhang J, Dekker JD, Tucker HO, Chang HY, Kim SK (2014) An integrated cell purification and genomics strategy reveals multiple regulators of pancreas development. *PLoS Genet* 10: e1004645
- Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Pena-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, Khalid F, Zhang W, Newburger D, Jaeger SA, Morris QD, Bulyk ML, Hughes TR (2008) Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* 133: 1266–1276
- Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ 3rd, Gingeras TR, Schreiber SL, Lander ES (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 120: 169–181
- Biankin AV, Waddell N, Kassahn KS, Gingras MC, Muthuswamy LB, Johns AL, Miller DK, Wilson PJ, Patch AM, Wu J, Chang DK, Cowley MJ, Gardiner BB, Song S, Harliwong I, Idrisoglu S, Nourse C, Nourbakhsh E, Manning S, Wani S et al (2012) Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* 491: 399–405
- Bullard JH, Purdom E, Hansen KD, Dudoit S (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* 11: 94
- Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, Snyder EY, Sulman EP, Anne SL, Doetsch F, Colman H, Lasorella A, Aldape K, Califano A, Iavarone A (2010) The transcriptional network for mesenchymal transformation of brain tumours. *Nature* 463: 318–325
- Cesaroni M, Cittaro D, Brozzi A, Pelicci PG, Luzzi L (2008) CARPET: a web-based package for the analysis of ChIP-chip and expression tiling data. *Bioinformatics* 24: 2918–2920
- Cieply B, Pt R, Pifer PM, Widmeyer J, Addison JB, Ivanov AV, Denvir J, Frisch SM (2012) Suppression of the epithelial-mesenchymal transition by Grainyhead-like-2. *Cancer Res* 72: 2440–2453
- Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, Cooc J, Weinkle J, Kim GE, Jakkula L, Feiler HS, Ko AH, Olshen AB, Danenberg KL, Tempero MA, Spellman PT, Hanahan D, Gray JW (2011) Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat Med* 17: 500–503
- Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA* 107: 21931–21936
- De Craene B, Berx G (2013) Regulatory networks defining EMT during cancer initiation and progression. *Nat Rev Cancer* 13: 97–110
- Devita VT, Lawrence T, Rosenberg SA (2015) *Devita, Hellman, and Rosenberg's Cancer: Principles & Practice of Oncology*. Philadelphia: Wolters Kluwer
- Downen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, Weintraub AS, Schuijers J, Lee TI, Zhao K, Young RA (2014) Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* 159: 374–387
- Dutta S, Gannon M, Peers B, Wright C, Bonner-Weir S, Montminy M (2001) PDX:PBX complexes are required for normal proliferation of pancreatic cells during development. *Proc Natl Acad Sci USA* 98: 1065–1070
- Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10: 48
- Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, Diekhans M, Dreszer TR, Giardine BM, Harte RA, Hillman-Jackson J, Hsu F, Kirkup V, Kuhn RM, Learned K, Li CH et al (2011) The UCSC Genome Browser database: update 2011. *Nucleic Acids Res* 39: D876–D882
- Gao N, LeLay J, Vatamaniuk MZ, Rieck S, Friedman JR, Kaestner KH (2008) Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes Dev* 22: 3435–3448
- Gittes GK (2009) Developmental biology of the pancreas: a comprehensive review. *Dev Biol* 326: 4–35
- Glasmacher E, Agrawal S, Chang AB, Murphy TL, Zeng W, Vander Lugt B, Khan AA, Ciofani M, Spooner CJ, Rutz S, Hackney J, Nurieva R, Escalante CR, Ouyang W, Littman DR, Murphy KM, Singh H (2012) A genomic regulatory element that directs assembly and function of immune-specific AP-1-IRF complexes. *Science* 338: 975–980
- Grant CE, Bailey TL, Noble WS (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27: 1017–1018
- Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS (2007) Quantifying similarity between motifs. *Genome Biol* 8: R24
- Hallikas O, Taipale J (2006) High-throughput assay for determining specificity and affinity of protein-DNA binding interactions. *Nat Protoc* 1: 215–222
- Heintzman N, Stuart R, Hon G, Fu Y, Ching C, Hawkins R, Barrera L, Van Calcar S, Qu C, Ching K, Wang W, Weng Z, Green R, Crawford G, Ren B (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39: 311–318
- Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanov VV, Stewart R, Thomson JA, Crawford GE, Kellis M et al (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459: 108–112
- Hermann PC, Huber SL, Herrler T, Aicher A, Ellwart JW, Guba M, Bruns CJ, 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
- Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, Mickey K, Voets O, Verrijzer CP, Shilatifard A (2012) Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes Dev* 26: 2604–2620

- Hezel AF, Kimmelman AC, Stanger BZ, Bardeesy N, Depinho RA (2006) Genetics and biology of pancreatic ductal adenocarcinoma. *Genes Dev* 20: 1218–1249
- Hidalgo M (2010) Pancreatic cancer. *N Engl J Med* 362: 1605–1617
- Huang S, Ernberg I, Kauffman S (2009) Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Semin Cell Dev Biol* 20: 869–876
- Jolma A, Kivioja T, Toivonen J, Cheng L, Wei G, Enge M, Taipale M, Vaquerizas JM, Yan J, Sillanpää MJ, Bonke M, Palin K, Talukder S, Hughes TR, Luscombe NM, Ukkonen E, Taipale J (2010) Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Res* 20: 861–873
- Jolma A, Yan J, Whittington T, Toivonen J, Nitta KR, Rastas P, Morgunova E, Enge M, Taipale M, Wei G, Palin K, Vaquerizas JM, Vincetelli R, Luscombe NM, Hughes TR, Lemaire P, Ukkonen E, Kivioja T, Taipale J (2013) DNA-binding specificities of human transcription factors. *Cell* 152: 327–339
- Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B (2005) A high-resolution map of active promoters in the human genome. *Nature* 436: 876–880
- Kim J, Hoffman JP, Alpaugh RK, Rhim AD, Reichert M, Stanger BZ, Furth EE, Sepulveda AR, Yuan CX, Won KJ, Donahue G, Sands J, Gumbs AA, Zaret KS (2013) An iPSC line from human pancreatic ductal adenocarcinoma undergoes early to invasive stages of pancreatic cancer progression. *Cell Rep* 3: 2088–2099
- Kloppel G, Lingenthal G, von Bulow M, Kern HF (1985) Histological and fine structural features of pancreatic ductal adenocarcinomas in relation to growth and prognosis: studies in xenografted tumours and clinico-histopathological correlation in a series of 75 cases. *Histopathology* 9: 841–856
- Kouzarides T (2007) Chromatin modifications and their function. *Cell* 128: 693–705
- Kulakovskiy IV, Medvedeva YA, Schaefer U, Kasianov AS, Vorontsov IE, Bajic VB, Makeev VJ (2013) HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Res* 41: D195–D202
- Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: R25
- Lango Allen H, Flanagan SE, Shaw-Smith C, De Franco E, Akerman I, Caswell R, Ferrer J, Hattersley AT, Ellard S (2012) GATA6 haploinsufficiency causes pancreatic agenesis in humans. *Nat Genet* 44: 20–22
- Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, Zaretzky I, Jaitin DA, David E, Keren-Shaul H, Mildner A, Winter D, Jung S, Friedman N, Amit I (2014) Immunogenetics. Chromatin state dynamics during blood formation. *Science* 345: 943–949
- Li C, Heidt DG, Dalerba P, Burant CF, Zhang L, Adsay V, Wicha M, Clarke MF, Simeone DM (2007) Identification of pancreatic cancer stem cells. *Cancer Res* 67: 1030–1037
- Li P, Spolski R, Liao W, Wang L, Murphy TL, Murphy KM, Leonard WJ (2012) BATF-JUN is critical for IRF4-mediated transcription in T cells. *Nature* 490: 543–546
- Mancino A, Termanini A, Barozzi I, Ghisletti S, Ostuni R, Prosperini E, Ozato K, Natoli G (2015) A dual cis-regulatory code links IRF8 to constitutive and inducible gene expression in macrophages. *Genes Dev* 29: 394–408
- Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, Brooks M, Reinhard F, Zhang CC, Shipitsin M, Campbell LL, Polyak K, Briskin C, Yang J, Weinberg CA (2008) The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 133: 704–715
- Martinelli P, Canamero M, del Pozo N, Madriles F, Zapata A, Real FX (2013) Gata6 is required for complete acinar differentiation and maintenance of the exocrine pancreas in adult mice. *Gut* 62: 1481–1488
- Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen CY, Chou A, Ienasescu H, Lim J, Shyr C, Tan G, Zhou M, Lenhard B, Sandelin A, Wasserman WW (2014) JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res* 42: D142–D147
- McConnell BB, Kim SS, Yu K, Ghaleb AM, Takeda N, Manabe I, Nusrat A, Nagai R, Yang VW (2011) Kruppel-like factor 5 is important for maintenance of crypt architecture and barrier function in mouse intestine. *Gastroenterology* 141: 1302–1313.e6
- McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 28: 495–501
- Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C et al (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448: 553–560
- Nandan MO, Yoon HS, Zhao W, Ouko LA, Chanchevalap S, Yang VW (2004) Kruppel-like factor 5 mediates the transforming activity of oncogenic H-Ras. *Oncogene* 23: 3404–3413
- Nandan MO, McConnell BB, Ghaleb AM, Bialkowska AB, Sheng H, Shao J, Babbitt BA, Robine S, Yang VW (2008) Kruppel-like factor 5 mediates cellular transformation during oncogenic KRAS-induced intestinal tumorigenesis. *Gastroenterology* 134: 120–130
- Nandan MO, Ghaleb AM, Liu Y, Bialkowska AB, McConnell BB, Shroyer KR, Robine S, Yang VW (2014) Inducible intestine-specific deletion of Kruppel-like factor 5 is characterized by a regenerative response in adult mouse colon. *Dev Biol* 387: 191–202
- Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, John S, Sandstrom R, Johnson AK, Maurano MT, Humbert R, Rynes E, Wang H, Vong S, Lee K, Bates D, Diegel M, Roach V, Dunn D et al (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* 489: 83–90
- Ng AY, Waring P, Risteovski S, Wang C, Wilson T, Pritchard M, Hertzog P, Kola I (2002) Inactivation of the transcription factor Elf3 in mice results in dysmorphogenesis and altered differentiation of intestinal epithelium. *Gastroenterology* 122: 1455–1466
- Oliver JR, Kushwah R, Wu J, Pan J, Cutz E, Yeger H, Waddell TK, Hu J (2011) Elf3 plays a role in regulating bronchiolar epithelial repair kinetics following Clara cell-specific injury. *Lab Invest* 91: 1514–1529
- Ostuni R, Piccolo V, Barozzi I, Polletti S, Termanini A, Bonifacio S, Curina A, Prosperini E, Ghisletti S, Natoli G (2013) Latent enhancers activated by stimulation in differentiated cells. *Cell* 152: 157–171
- Park SM, Gaur AB, Lengyel E, Peter ME (2008) The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev* 22: 894–907
- Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, Jacobs KB, Arslan AA, Bueno-de-Mesquita HB, Gallinger S, Gross M, Helzlsouer K, Holly EA, Jacobs EJ, Klein AP, LaCroix A, Li D, Mandelsohn MT, Olson SH, Risch HA, Zheng W et al (2010) A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat Genet* 42: 224–228
- Pott S, Lieb JD (2015) What are super-enhancers? *Nat Genet* 47: 8–12
- Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842

- Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J (2010) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470: 279–283
- Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F (2013) Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8: 2281–2308
- Reichert M, Rustgi AK (2011) Pancreatic ductal cells in development, regeneration, and neoplasia. *J Clin Invest* 121: 4572–4578
- Sakai T, Mashima H, Yamada Y, Goto T, Sato W, Dohmen T, Kamada K, Yoshioka M, Uchinami H, Yamamoto Y, Ohnishi H (2014) The roles of interferon regulatory factors 1 and 2 in the progression of human pancreatic cancer. *Pancreas* 43: 909–916
- Singh A, Greninger P, Rhodes D, Koopman L, Violette S, Bardeesy N, Settleman J (2009) A gene expression signature associated with “K-Ras addiction” reveals regulators of EMT and tumor cell survival. *Cancer Cell* 15: 489–500
- Sipos B, Moser S, Kalthoff H, Torok V, Lohr M, Kloppel G (2003) A comprehensive characterization of pancreatic ductal carcinoma cell lines: towards the establishment of an in vitro research platform. *Virchows Arch* 442: 444–452
- Smith E, Shilatifard A (2014) Enhancer biology and enhanceropathies. *Nat Struct Mol Biol* 21: 210–219
- Spitz F, Furlong EE (2012) Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* 13: 613–626
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102: 15545–15550
- Supek F, Bosnjak M, Skunca N, Smuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* 6: e21800
- Takaori K, Hruban RH, Maitra A, Tanigawa N (2004) Pancreatic intraepithelial neoplasia. *Pancreas* 28: 257–262
- Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25: 1105–1111
- Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7: 562–578
- Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457: 854–858
- Wan H, Luo F, Wert SE, Zhang L, Xu Y, Ikegami M, Maeda Y, Bell SM, Whitsett JA (2008) Kruppel-like factor 5 is required for perinatal lung morphogenesis and function. *Development* 135: 2563–2572
- Wei GH, Badis G, Berger MF, Kivioja T, Palin K, Enge M, Bonke M, Jolma A, Varjosalo M, Gehrke AR, Yan J, Talukder S, Turunen M, Taipale M, Stunnenberg HG, Ukkonen E, Hughes TR, Bulyk ML, Taipale J (2010) Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *EMBO J* 29: 2147–2160
- Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153: 307–319
- Zambelli F, Pesole G, Pavesi G (2009) Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res* 37: W247–W252
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9: R137