PhD degree in Systems Medicine (curriculum in Molecular Oncology)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples "Federico II"

Settore disciplinare: Med/04

Involvement of the transcription factor MYRF in signaling from the endoplasmic reticulum to the nucleus in pancreatic cancer

Marta Milan

IEO, Milan

Matricola n. R11122

Supervisor: Prof. Gioacchino Natoli

IEO, Milan

Anno accademico 2017-2018

Table of contents

List of abbreviations

4sU	4-Thiouridine
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation followed by sequencing
CRISPR	Cluster regularly interspaced short palindromic repeats
DBD	DNA binding domain
DEG	Differentially expressed gene
EMSA	Electrophoretic mobility shift assay
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum associated protein degradation
FDR	False discovery rate
GO	Gene ontology
GSEA	Gene set enrichment analysis
ICA	Intramolecular chaperone auto-processing
KO	Knock out
PanIN	Pancreatic Intraepithelial Neoplesia
PDAC	Pancreatic ductal adenocarcinoma
RNA-seq	RNA-sequencing
RT-qPCR	Real time quantitative polymerase chain reaction
TF	Transcription factor
TMA	Tissue microarray
TSS	Transcription start site
UPR	Unfolded protein response
WT	Wild type

Figures index

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is one of the leading causes of cancer death worldwide. One reason for the poor prognosis is the high intra-tumor heterogeneity, with the coexistence of well- and poorly-differentiated cells in virtually all tumor cases. Thus, a better characterization of the circuitries regulating PDAC cell differentiation is required.

We found that MYRF, a poorly characterized transcription factor, is selectively expressed in well-differentiated PDAC cell lines. MYRF is synthetized as an endoplasmic reticulum (ER) membrane protein and self-cleaves after trimerization, releasing the N-terminal trimer that translocates into the nucleus and regulates transcription. We generated MYRF-KO PDAC cells and combined transcriptomic profiles and analyses of MYRF genomic occupancy to study its function. We retrieved the MYRF DNA binding motif from our ChIP-sequencing data and demonstrated that MYRF capability to bind this sequence and activate transcription is strictly dependent on its trimerization.

MYRF deletion resulted in the downregulation of cell replication-related genes and upregulation of ER stress-related genes. Consistently, MYRF loss resulted in an altered ER morphology and function, probably as a result of the overexpression of membrane and secreted proteins with complex folding, such as cysteine rich and highly glycosylated proteins.

Additionally, we found that MYRF creates a feed-forward loop with the transcription factors FOS and FOSB. In doing so, MYRF directly regulates the expression of these two transcription factors that in turn cooperate to generate the MYRF transcriptional outcome.

In conclusion, this work points to a role for MYRF as a key player in the communication between ER and nucleus, working as a sensor of proper ER function. MYRF appears to license cells for DNA replication and concomitantly to serve as a guardian against ER overload in highly secretory cancer cells.

Introduction

Pancreatic cancer

Pancreatic ductal adenocarcinoma (PDAC) is a tumor that originates from the exocrine pancreas. PDAC accounts for more than 90% of all pancreatic malignancies. The statistics have put a spotlight on this tumor which is characterized by a five year survival rate of only 6%. As such, it is projected to become the second cause of cancer death in developed countries by 2030[5].

This high lethality is attributable to different concomitant factors: the latent clinical manifestation, that in the majority of the cases leads to the appearance of symptoms only when the disease has already spread[6]; the lack of early detection markers and effective treatments; finally, the complex biology of the tumor.

The biological complexity of the tumor, in particular, is reflected by the high intra-tumor cell diversity, which urges for an increased molecular understanding and improved classification schemes.

PDAC stratification

Different efforts have been carried out to classify PDAC tumors based on both genomic and transcriptomic data[7,8]. Collisson and colleagues[7] used tissue expression microarray data to retrieve signatures that allowed for the classification of PDACs. With a 62-gene signature they clustered pancreatic tumors in subtypes: classical, *quasi*-mesenchymal and exocrine-like. Interestingly, these subtypes stratified patients in terms of survival rate, with the classical subtype having a better prognosis than the other types of tumor. Moreover, the authors were able to retrieve two of the three subpopulations in a cohort of mouse and human PDAC cell lines,

in which the classification correlated also with a differential sensitivity to drugs currently used in therapy[7].

Another study based on the combination of genomic and transcriptomic data led to the identification of four subclasses of PDACs, three of which were found to overlap with the previously published classes: squamous (quasi-mesenchymal), pancreatic progenitors (classical), aberrantly differentiated endocrine exocrine (exocrine-like) and the newly introduced immunological subtype[8], which is highly similar to the progenitor subtype but is characterized by a significant immune infiltrate.

Despite the different nomenclature, the overlap in classification of these studies demonstrates the existence of clear subtypes of PDAC, highlighting the necessity for further investigations to unveil the mechanisms that underlie their diversity.

PDAC intratumor heterogeneity

The existence of different subtypes of PDAC could partially explain why current therapies, which do not take into consideration such heterogeneity, are not effective in the treatment of this disease. However, another layer of complexity is that PDAC is characterized by a high tumor cellular heterogeneity, with the coexistence of a broad range of cells in distinct differentiation stages within the same tumor[9]. Furthermore, PDACs are associated with a desmoplastic reaction that might contribute to the creation of a protected and isolated niche that makes this tumor even harder to target[10].

Figure 1 shows a representative PDAC. Surrounded by the stroma, well differentiated areas (Grade 1) are characterized by ductal-like structures, discretely organized with cell nuclei lining the basal part of the cells, thus resembling non-neoplastic ducts normally found in the pancreas. Adjacent to these structures, a variety of additional differentiation grades can be found. These range from moderately differentiated areas (G2), where cells start to lose polarization,

cytoplasm is enlarged and ducts collapse, to nests of completely undifferentiated cells representative of G3 areas.

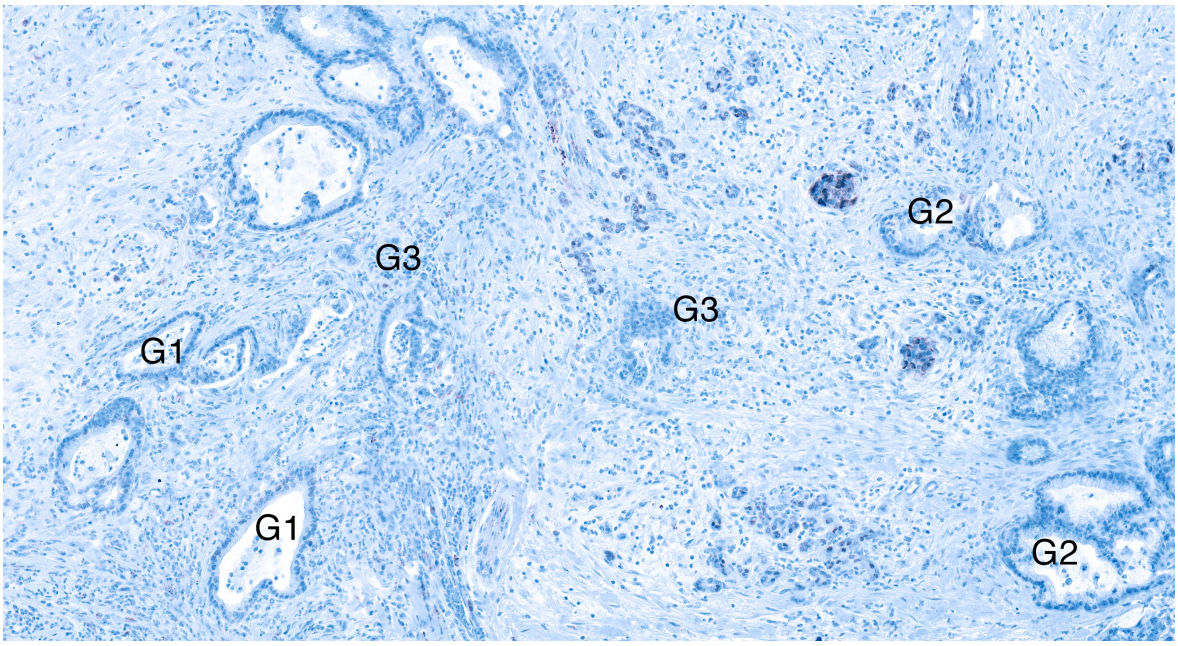


Figure 1: PDAC heterogeneity. Hematoxylin staining of a human PDAC tumor highlight all the typical features of PDAC: surrounded by an important stromal reaction, tumor cells span all the differentiation grades, from well-differentiated G1 ducts to poorly-differentiated G3 areas represented by unorganized clusters of cells.

It is therefore clear that this extreme morphological heterogeneity is linked to phenotypical differences that can explain why generalized therapies may only target specific subpopulations and thus are not effective in eradicating the entire tumor.

PDAC development and its contribution to intra-tumor heterogeneity

Historically, PDAC has been thought to develop from preneoplastic lesions through the gradual accumulation of mutations in key genes involved in the control of cell proliferation and migration. Among these pre-neoplastic lesions, PanINs (Pancreatic Intraepithelial Neoplasias) are the most commonly studied. PanIN initiation and development is thought to occur with the serial accumulation of mutations in KRAS, CDKN2A, TP53 and eventually SMAD4[11].

A recent study based on whole-genome sequencing of more than one hundred pancreatic tumors instead proposed a new and somewhat more dramatic

hypothesis of PDAC development[12]. Deconvolution of mutational patterns in tumors led to the inference that PDAC might develop following the early catastrophic acquisition of massive genetic changes due to mitotic errors (chromothripsis)[12]. Chromothripsis is a catastrophic event of acute and diffuse chromosome disruption that leads to the acquisition of up to hundreds of chromosomal rearrangements in a unique cell cycle event. If stabilized, this crisis can lead to the simultaneous accumulation of several mutations that give selective advantages to cancer cells[13]. This means that while in most other epithelial tumors the slow and progressive accumulation of mutations leads to the gradual increase in tumor heterogeneity, human PDACs show a high level of heterogeneity from very early stages of tumor progression. Such premature high variability could therefore be a key factor in the poor response to therapies and poor prognosis that characterize this tumor.

Molecular characterization of PDAC heterogeneity

This finding suggests that a better understanding of the circuitries underling PDAC heterogeneity could help clarify PDAC biology and, in the long term, lead to the development of more effective therapies. To this aim, our lab has recently undertaken a systematic study to unveil the basis of PDAC grading. In an initial study, a series of PDAC cell lines were used that represent punctual stages of differentiation, thus enabling a systematic analysis of PDAC grades. To this end, we used transcriptomic and epigenomic analysis to unveil key players in the regulation and maintenance of PDAC differentiation[1]. Given that transcription factors are key hubs in the control of transcriptional programs in cells, the research focused on the identification of transcription factors differentially expressed or active between Low-Grade and High-Grade cell lines, as reported in the heat map shown in Figure 2.

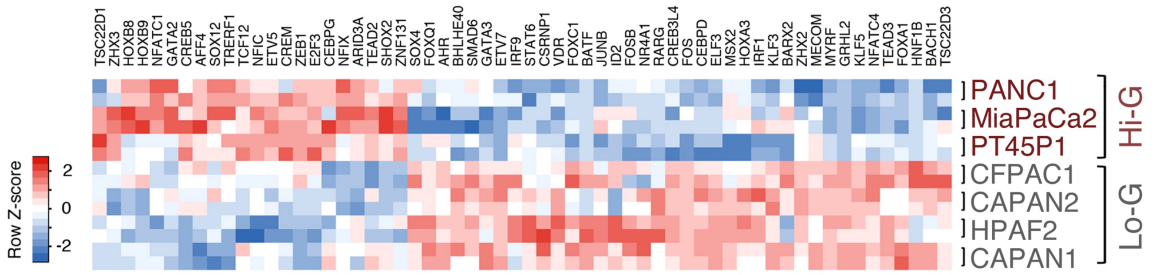


Figure 2: Transcription factors differentially expressed in PDAC grading. Heatmap showing transcription factors (TFs) differentially expressed between Low-Grade and High-Grade PDAC cell lines. Figure taken from Diaferia et al[1].

The next section will be dedicated to the general description of transcriptional regulation and why this process is highly important in cancer biology.

Transcription and epigenetic in cancer

General overview

Transcriptional regulation is a highly controlled process that in eukaryotes involves thousands of proteins, acting on different layers: transcriptional activation and repression, chromatin remodeling, and chromatin structure[14].

The primary players involved in this process are transcription factors (TFs), proteins that recognize specific DNA sequences and in doing so are able to control gene expression. Transcription factors bind to promoters and enhancers and their combinatorial expression is responsible for the establishment of the transcriptional network dictating cellular identity[15].

Another layer of complexity in transcription regulation is that in eukaryotes, DNA is organized in a specific chromatin structure with the basic unit represented by nucleosomes. Nucleosomes are formed by sequences of 147 base pairs of DNA surrounding histone octamers whose tails, in particular, are heavily post-translationally modified. These modifications imply the presence of proteins able to read these marks (i.e. readers), insert them (i.e. writers) and remove them (i.e. erasers). All together these proteins create combinatorial histone marks that

correlate with different transcriptional outputs. As such, it is clear that histone readers, writers and erasers are integral contributors in the maintenance of the transcriptional network of a specific cell state[16].

Moreover, an increasing amount of data is emerging on the role of 3D chromosome organization in transcription regulation. Enhancers and their regulated genes are in fact often inserted in unique isolated parts of the chromatin landscape, separated from the surrounding genomic regions by DNA sites bound by CTCF and cohesin. These insulated neighborhoods define the so called topologically associated domains (TADs) and create physical and functional constraints that are conserved in development and provide specific enhancer-gene interactions and are thus part of the control of the transcriptional network in a cell[17].

Transcription alterations in cancer

Since transcriptional programs are an integral part of cellular identity, they are intrinsically embedded in the global events characterizing tumor initiation and transformation. Therefore, two important aspects need to be taken in consideration. First, oncogenic transformation happens within a specific transcription network, thus the transforming cell needs to hijack the normal transcriptional landscape and exploit it. As a result, well-differentiated tumors often reflect the gene expression patterns of their normal counterparts, while more aggressive cancers express gene signatures that are associated with 'stemness'[18].

Secondly, there is increasing evidence to show that tumors depend on their specific transcriptional program, a phenomenon coined transcriptional addiction[14]. Their dependence thus unveils transcriptional susceptibilities and implies the possibility of targeted therapy. This is well exemplified by the use of anti-estrogen therapy in the treatment of Low-Grade breast cancer. Although not directly mutated, the estrogen receptor (ESR1) is part of the cellular circuitry on which breast cancer cells

rely and thus ESR1 antagonists have been demonstrated to have a great impact in patient treatment[14].

Given the importance of transcriptional networks in cancer development and progression, it is therefore not surprising that either mutation or hyperactivation of various components of the transcriptional machinery have been investigated for their role in cancer biology.

In the next section the impact of different components of the transcriptional and epigenetic machinery will be analyzed, with a particular focus on the field of pancreatic cancer.

Transcription factors in cancer

The transcription factors with a recognized role in cancer can essentially be divided in three groups: TFs controlling cell identity and tissue specificity; TFs involved in proliferation regulation; TFs that deconvolute external signals into cellular responses[14].

Regarding the first group of transcription factors, they fall into the aforementioned concept of transcription hijacking by oncogenic transformation. Given the role of developmental TFs in the control of cell identity and differentiation programs it is indeed not surprising that several of them have been implicated in cancer, including PDACs[11].

In particular, a genome-wide association study (GWAS) that analyzed a cohort of thousands of pancreatic cancer patients and matched controls, identified pancreatic development as the pathway most significantly associated to pancreatic cancer. Among the top-ranking genes in this pathway there were known developmental TFs including NR5A2, HNF1A, HNF4G and PDX1[19]. PDX1 is a master transcription factor in pancreas development whose involvement in PDAC has been clearly demonstrated. While knock-out of PDX1 in early stages of tumor progression in

mouse models inhibits the transition from PanIN preneoplastic lesions to PDAC, its deletion in established PDAC inhibits tumor cell growth[20], indicating a role of this TF in the control of differentiation at various stages of tumor progression.

The other class of TFs influencing cancer transcriptional networks are those that control cell proliferation. A pivotal example is given by P53. P53 is a TF that, in response to various stimuli like DNA damage, oncogene activation, hypoxia and stress, is activated and prevent potentially harmful situations by regulation of G1/S checkpoint, DNA repair and apoptosis[21].

TP53, the gene encoding for P53, is frequently mutated in cancer, in the majority of cases with missense mutations that cause the production of a mutated form of the protein. Mutant p53 not only loses its tumor suppressor capabilities, but also acquires additional oncogenic properties given both by the dominant negative effect on the eventual wild type residual allele, and by the acquisition of new capabilities that all together are referred as gain of function properties[22].

In pancreatic cancer, TP53 represents one of the hotspot mutations. TP53 is somatically mutated in up to 85% of tumor cases and the importance of the gain of function properties is highlighted by the fact that 66% of these mutations are missense and fall in the DNA binding domain of the protein[23].

Epigenetic regulators in cancer

As mentioned before, transcriptional responses cannot be achieved without the contribution of chromatin regulators that enforce gene activation or repression. Both DNA and histone modifications are part of the so called epigenome and epigenetic writers, readers and erasers serve as integral components of the machinery that controls transcriptional states together with transcription factors[14]. For the most part, epigenetic regulators work globally, thus targeting of these protein could have a stronger and broad effect than targeting single transcription factors[24].

Another significant aspect of chromatin regulators is that these proteins are enzymes, therefore the design of drugs to target them is easier than drugs aimed at transcription factors, against which only few cases currently exist. These drugs have different applications. Firstly, they can be used to treat tumors that directly depend on mutations or amplifications of chromatin regulators. Secondly, they can be exploited to target an undruggable transcription factor when the cooperation between the two proteins is known to have a crucial role in tumor biology[14].

This is clearly exemplified by the development of drugs targeting bromodomain containing protein 4 (BRD4). BRD4 recognizes acetylated histones through its two bromodomains and mediates transcriptional activation via interaction with the pTEFb elongation factor in the PolII transcriptional machinery. BRD4 fusions with NUT are found in an aggressive form of squamous carcinoma called NUT midline carcinoma (NMC). Treatment of these tumors with BRD4 inhibitor, in a pilot study, led to tumor regression[25]. In parallel, inhibition of BRD4 has also been exploited in the treatment of MYC driven tumors. Given the lack of drugs targeting this oncogenic TF, the concept of targeting its dependence on BRD4 for the activation of its transcriptional responses led to good results in treatment of various MYC dependent cancers such as multiple myeloma[26].

In the context of pancreatic cancer, the relevance of epigenetic regulators is highlighted by the high incidence of somatic mutations targeting this type of proteins. In particular, members of the SWI/SNF complex and of the histone-lysine N-methyltransferase 2 (KMT2) family have a high somatic mutation rate[27,28]. Interestingly, while single members of the SWI/SNF complex have mutation rates below 5% in a mutually exclusive manner. However, if summed and counted as contributions by the complex as a whole, these mutations may reach up to 30% of tumor cases in pancreatic cancers[29].

Further studies are therefore needed to better elucidate their role in tumor progression and their possible exploitation in terms of cancer therapy.

3D chromosome organization in cancer

The expanding knowledge and available technologies to study 3D chromatin organization is opening a highly dynamic field aimed to unveil the contribution of this aspect of transcriptional control in cancer.

As described before, 3D chromosome organization is achieved through the formation of insulated genomic neighborhoods flanked by CTCF binding sites that create isolated units in which promoter-enhancer interactions occur. The disruption of these insulated structures can bring enhancers to aberrantly loop with different genes. This could in turn either lead to the insertion of proto-oncogenes in a transcriptionally favorable environment or switch off the expression of active genes[14].

Cancer cells use different means to achieve insulation disruption. One such mechanism, that occurs frequently in glioma, is DNA methylation, that blocks CTCF binding and prevents the formation of CTCF-cohesin complex at TADs boundaries[30].

Interestingly, it has been shown that this abnormal chromatin organization can create targetable vulnerabilities in cancer. In glioma cells, mutations in IDH1 leads to inhibition of the demethylase TET and consequent aberrant DNA methylation. The aberrant methylation at CTCF anchors is associated with the hyperactivation of the receptor tyrosine kinase gene PDGFA due to the establishment of an interaction between this gene and a constitutive enhancer. Treatment of glioma cells with a small molecule that blocks PDGFA inhibits tumor growth[31], confirming that further characterization of the alterations in insulated neighborhoods could lead to the

discovery of additional mechanisms of cancer addiction and vulnerabilities exploitable in cancer therapy.

Insulated neighborhoods can also be destroyed through direct mutation of CTCF binding sites. As in the case of DNA methylation, this can serve to create novel enhancer-gene interactions, thus altering the transcriptional network of cancer cells. In pancreatic and other gastrointestinal cancers, the CTCF motif appears to be among the most mutated DNA binding recognition motifs when compared to normal tissues. These mutations often fall in the proximity of genes that are part of pathways dysregulated in cancers, supporting the notion of the functional significance of these motif-breaking mutations in tumors[32].

Therefore, transcriptional and epigenetic regulation are integral part of tumorigenesis and a better understanding of transcriptional networks in cancer could have a tremendous impact on the understanding and treatment of this disease.

The next chapter will be dedicated to the transcription factor representing the subject of this thesis, MYRF, with a focus on what is currently known on its structure and functions.

MYRF

In the aforementioned work of our laboratory, a series of transcription factors have been identified as differentially expressed in PDAC grading[1]. Among them, MYRF was revealed to be differentially expressed in Low-Grade PDAC cell lines and hence we decided to further investigate its role in this context.

MYRF, myelin regulatory factor, derives its name from its first characterization as a transcription factor expressed in the central nervous system where it plays a fundamental role in the maturation of oligodendrocytes. Its knock-out in these cells

in mice leads to lack of myelin production, consequent apoptosis of differentiating oligodendrocyte and ultimately death of the mice[33].

Before describing the known functions of this transcription factor, MYRF protein structure needs to be elucidated, given its peculiarity and possible key involvement in MYRF functions.

MYRF protein structure

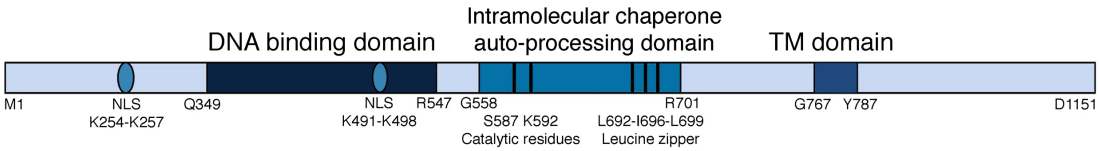


Figure 3: MYRF protein structure. Graphical representation of MYRF protein structure. Principal functional domains and their aminoacidic positions are indicated. NLS: nuclear localization signal; TM: trans-membrane.

As shown in Figure 3, MYRF possesses a transmembrane domain, via which it is inserted into the endoplasmic reticulum membrane, a DNA binding domain (DBD) and an intramolecular chaperone auto-processing domain (ICA domain)[34]. The ICA domain is the ortholog of the same domain found in the tailspike proteins of bacteriophages. When infecting bacteria, these proteins trimerize, attach to the surface of the host and the self-cleavage of the ICA-domain activates the endosialidase that subsequently degrades the bacteria wall[35]. In the MYRF protein, the ICA domain maintains its function and induces trimerization and self-cleavage and in doing so releases the N-terminal trimer that translocates into the nucleus and induces the transcription of the target genes[34,36].

MYRF DNA binding domain belongs to the family of s-type immunoglobulin (Ig) like domains, a family that comprises a number of important transcription factors including NFAT, NF-κB, p53 and STAT proteins. In this structure, there are three loops that are responsible for the DNA binding specificity of the various proteins: the a-b loop, which binds to the major grove of the DNA and makes sequence specific

contacts; the e-f loop, that binds the DNA backbone and finally, the C-terminus that is the most variable region and contributes to sequence specificity and recognition[37]. MYRF DBD is homolog to the DNA binding domain of the yeast protein Ndt80. Ndt80 is a transcription factor that recognizes a DNA sequence named middle sporulation element (MSE) and is responsible for meiosis completion in yeast[38]. Nevertheless, Ndt80 does not possess the ICA domain found in MYRF and its orthologs. It has indeed been observed that only its orthologs in higher eukaryotes, like MYRF, carry the Ndt80-like DBD fused with the ICA domain, possibly a result of lateral gene transfer from bacteriophages[34].

Interestingly, the crystal structure of human MYRF DBD has recently been resolved. Although the authors did not manage to co-crystalize MYRF in complex with DNA, it has been shown that it forms trimers when crystalized[39].

In addition to the DNA binding domain, MYRF most N-terminal portion contains a transactivation domain sufficient to induce transcription and whose sumoylation represent a further layer of control of MYRF transcriptional activity[40].

MYRF functions

As mentioned before, MYRF was first identified as a fundamental TF controlling myelin expression in the transition from pre-oligodendrocytes to mature oligodendrocytes in the central nervous system. The importance of MYRF function is highlighted by the observation that its knock-out in this cell type in mice leads to severe tremors and ataxia, followed by seizures and ultimately death during the third postnatal week[33].

Additionally, MYRF is important in the adult brain, where it plays a fundamental role in the myelination and plasticity needed for the acquisition of complex sensomotor skills[41] and for the maintenance of myelin production in adult oligodendrocytes[42].

MYRF is highly conserved throughout evolution: its first ortholog can be found in *Dictyostelium*, where it is involved in the regulation of extracellular matrix proteins and differentiation[43]. In *C. elegans*, MYRF ortholog was found in a mutant screening aimed to the identification of genes involved in molting, the process through which the external collagenous cuticle is lost and substituted by a new slightly bigger cuticle in the maturation toward the adulthood[44]. MYRF null mutants fail to resolve the larval cycles and to exit from the first molting stage, possibly due to a defect in the proteolytic release of the external cuticle[44].

Similarly, in *C. elegans*, MYRF is required for the rewiring of a specific population of neurons during larval development. This indicates that although invertebrates lack the myelination process, MYRF could have additional functions in development beyond the role in myelination acquired only in vertebrates[45].

MYRF protein processing and its possible functional role

Given the involvement of MYRF in all of these processes that are associated with secretion (i.e. myelination, molting), it has been postulated that the peculiar processing of MYRF, after which the protein is cleaved in an ER membrane-trimer and a transcription factor complex, could be essential to coordinate the functions of these two organelles. This hypothesis, however, remains to be demonstrated.

This is supported by the observation that MYRF in highly express in highly secretory tissue aside from the central nervous system. Among the most MYRF-expressing tissue are indeed the stomach, the lung and the pancreas, as shown in Figure 4, where RNA expression data from GTex database are reported.

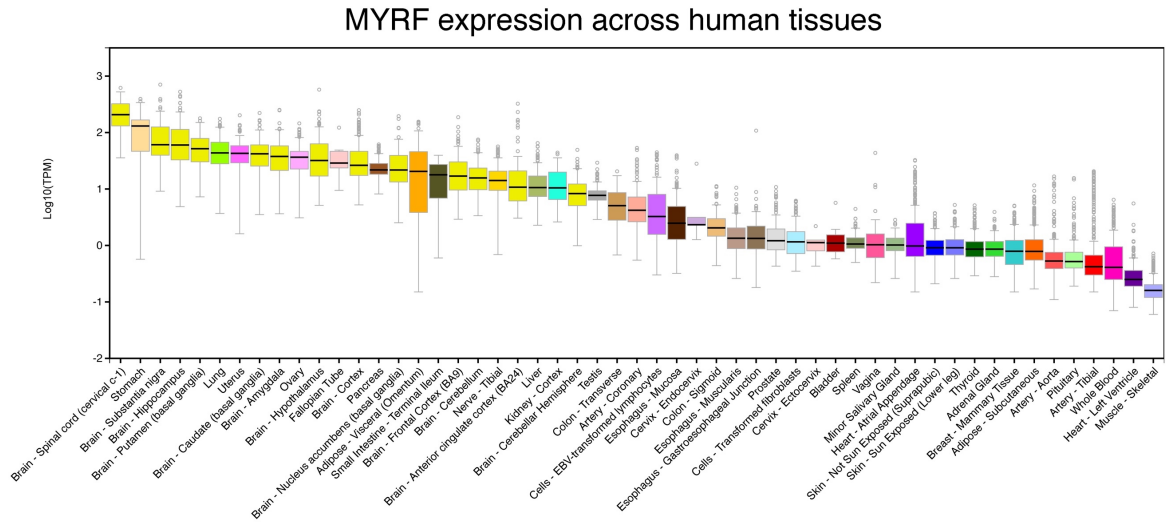


Figure 4: MYRF is expressed in highly secretory tissues. Box plot showing RNA expression levels (expressed as TPM) from different human tissues retrieved from GTex database. Pancreas is the brown box in the first quarter of tissues highly expressing MYRF.

The pancreas is an organ composed of two different compartments: the exocrine pancreas which is mainly constituted of acinar cells that produce digestive enzymes and ductal cells that transport these enzymes to the duodenum. The endocrine pancreas is instead composed by hormone-secreting cells that are aggregated in the islets of Langerhans and produce, among others, insulin, glucagon and somatostatin[46].

Both of these compartments are obviously highly secreting tissues. This observation, coupled with the fact that in the previous work of the laboratory MYRF was found as differentially expressed in Low-Grade and highly secretory PDAC cell lines, prompted us to better characterize the role of MYRF in PDAC function and biology.

Endoplasmic reticulum stress and unfolded protein response

The endoplasmic reticulum (ER) serves several functions in cells, one of which is the processing and folding of one third of the entire proteome of a cell. This requires a highly precise control of its functions and physiology, with the presence of

pathways aimed to sense and to resolve situations that could be dangerous for cell homeostasis. ER stress is characterized by the accumulation of misfolded proteins in the ER and activates a series of responses including unfolded protein response (UPR) and ER-associated degradation (ERAD)[47]. These responses are aimed to: reduce global protein synthesis, thus reducing the protein load in the ER; increase the capability of the ER to handle misfolded proteins; in case of failure in normal function restoration, trigger cell death through initiation of apoptosis[48].

ERAD

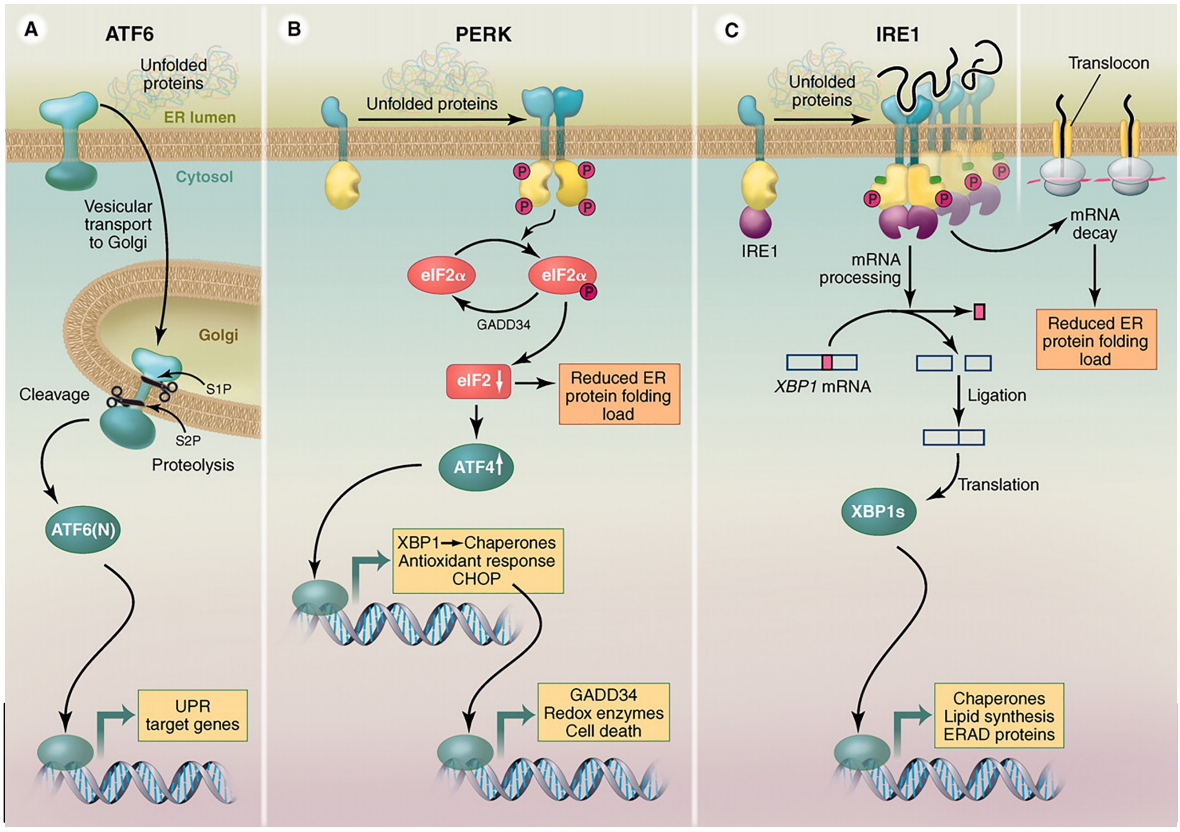
The ER lumen is characterized by the presence of a high concentration of molecular chaperones, including calnexin, calreticulin and BiP. These chaperones are primarily involved in folding and preventing aggregation of newly synthetized proteins where hydrophobic residues may be transiently exposed. In physiological conditions, these chaperones also bind and inhibit the transducers of the UPR response, rendering them in an inactive state[49].

When misfolded proteins accumulate in the ER, two major events occur: chaperone proteins are recruited to misfolded proteins, releasing the UPR sensors that activate their signaling cascade; ER-associated degradation (ERAD) is activated. The main function of ERAD is to mediate the degradation of misfolded proteins and alleviate ER stress. It is composed of E3 ligase machinery that connects the system of recognition of misfolded proteins in the ER lumen and the proteasome in the cytosol[47].

Two key components of ERAD are Synoviolin (SYVN1), an E3 ubiquitin ligase that targets a subset of ERAD targets to the proteasome, and SEL1L which is responsible for the nucleation of the ERAD complex[47].

UPR

As mentioned before, the principal function of ERAD is in relieving ER stress by retrotranslocation of misfolded proteins from the ER lumen to the cytosol. Ultimately, the sequestering of chaperones by misfolded proteins releases the three UPR sensors eventually activating the UPR pathway. These three sensors are IRE1α, PERK and ATF6. All three sensors are ER transmembrane proteins that, once unbound by the ER chaperones, get activated and induce a downstream cascade of events, namely unfolded protein response (UPR)[48].



Although the three branches of UPR involve different effectors, all of them contribute to the final goal of releasing ER stress or if necessary, induce apoptosis. A general scheme of the pathway is depicted in Figure 5. The next sections will give a description of the different effectors and their function.

IRE1α-XBP1 pathway

IRE1α (Inositol-Requiring protein 1α) is a type I ER-membrane protein. Once activated, it oligomerizes and trans-phosphorylates, leading to a conformational change and the activation of the endonuclease domain[47]. A principal target of this RNase domain is XBP1, in which the cleavage leads to the removal of a 26 bp intron and the production of a stable transcription factor. Once in the nucleus, XBP1 binds to stress element promoters and activates the transcription of genes involved in protein folding, ERAD, protein trafficking and lipid biosynthesis[50].

In addition to the endonuclease activity, IRE1α activates stress-induced Jun N-terminal kinase (JNK) and interacts with components of the apoptosis machinery, suggesting a contribution to the cell death program independent from its endonuclease activity[48].

Finally, it has been demonstrated that IRE1α possesses RNase activity on additional mRNA targets, a process called Regulated IRE1-dependent Decay (RIDD). This leads to the degradation of a variety of mRNA in proximity to the ER membrane, reducing the ER protein load[51].

PERK pathway

As IRE1α, PERK (protein kinase RNA (PKR)-like ER kinase) is an ER-membrane protein that oligomerizes and trans-phosphorylates in response to ER stress. Once activated, PERK phosphorylates the α-subunit of eukaryotic translation initiation factor 2 (eIF2α), leading to the inhibition of the interaction between eIF2α and eIF2B and consequent reduction in translational initiation[48]. This leads to a global attenuation of protein translation, blocking further accumulation of RNA that is translated on the ER membrane, with the parallel increase in translation of specific

UPR targets, including the transcription factor ATF4[52]. ATF4 subsequently activates the transcription of genes involved in protection from oxidative stress, amino acid import and synthesis, of XBP1 and the transcription factor CHOP[49]. CHOP, in turn, activates its target genes including GADD43, that dephosphorylates eIF2α restoring protein translation, and ERO1, an ER oxidase required for disulfide bond formation in protein folding[48].

ATF6 pathway

ATF6 is a transcription factor synthetized as an inactive precursor in the ER-membrane. After ER stress induction, it translocates to the Golgi where two subsequent cleavages operated by the Golgi-resident proteases S1P and S2P release the active TF that then translocates into the nucleus and activates the transcription of UPR-related genes[48].

UPR and cancer

Giving the pleiotropic responses activated in the UPR pathway, it cannot be considered either completely oncogenic or tumor suppressive. While some prosurvival aspects can promote tumor growth, the components responsible for apoptosis could be detrimental for cancer cells. Thus, the role of UPR components in tumor initiation and progression is controversial[49]. Here some examples are reported.

Hypoxia is a situation that often occurs in solid tumors and is characterized by the production of reactive oxygen species (ROS). While on one hand PERK can alleviate oxidative stress through the induction of intracellular antioxidants, thus promoting cell proliferation and tumor growth[53], on the other hand it can induce apoptosis both through induction of its proapoptotic target CHOP and through tethering of ER-mitochondria contacts that propel ROS-mediated apoptosis[54].

Another important aspect in tumor biology is the interplay between cancer cells and inflammation and inflammatory cells. In some contexts ER stress (possibly through direct transcriptional control by XBP1) induces the production of proinflammatory cytokines with known implication in tumor survival[49]. Paradoxically, ER stress has also been linked to the cell surface expression of damage-associated molecular patterns (DAMPs), like ectopic exposure of calreticulin, leading to immunogenic killing of cancer cells[55].

UPR in cancer therapy

Despite this bifunctional role of UPR in cancer, there are cases in which UPR inhibition represents a promising approach in cancer therapy.

For example, multiple myeloma, been a malignant proliferation of plasma cells, is characterized by an abundant production of secretory proteins, and consequently activation of UPR. This characteristic has been exploited as a possible vulnerability. More specifically, inhibition of IRE1α inhibits tumor growth *in vitro* and *in vivo*, suggesting the potential antitumoral effect of UPR inhibition in this context[56].

Another example of tumor susceptibility to UPR inhibition has emerged in the context of oncolytic viral therapy. An RNAi-based screen unveiled UPR as a modulator of oncolytic virus toxicity. Importantly, preconditioning of cancer cells with IRE1α inhibitors dramatically increases oncolytic efficacy, suggesting a combined therapy as a tool to target initially resistant cells[57].

UPR in pancreas development

Given the importance of ER homeostasis in tissues with elevated rates of secretion, it is therefore not surprising that genes involved in ER stress control and signaling have a central role in pancreatic development and physiology.

A truncating mutation in SEL1L, the gene involved in the formation of the ERAD complex, leads to systemic ER stress and embryonic lethality by day E15.5. In particular, SEL1L is essential for pancreatic epithelial induction and is required in both exocrine and endocrine compartment morphogenesis[58].

In line with the impact of ERAD deficiency in pancreas function, deletion of single branches of the UPR are associated with pancreatic disfunction. For example, ablation of PERK in the adult pancreas leads to massive ER expansion and death of insulin-producing β-cells, demonstrating the essential role of ER homeostasis in secretory β-cells[59].

Similarly, studies on the IRE1α/XBP1 pathway show that XBP1 deficiency is associated with abnormalities in secretory tissue and apoptosis of exocrine pancreas during embryogenesis[60]. Moreover, IRE1α deficient mice show defects in pancreatic organ architecture, impairment in β-cell proliferation and hyperglycemia associated with β-cell disfunction[61].

Taken together, these data suggest that a tight control of ER homeostasis is fundamental in pancreas development and functionality.

UPR and pancreatic cancer

Increasing evidence suggests a role for the ER stress response pathway in pancreatic cancer. In particular, studies aimed at identifying pathways active in mesenchymal aggressive sub-populations of PDAC revealed that these cells rely on UPR activation. Additionally, UPR inhibition, alone or in combination with classical therapeutic agents, caused a robust apoptotic response and prolonged survival both *in vitro* and *in vivo*[62].

Another elegant study recently identified a population of single disseminating cancer cells characterized by a mesenchymal state that are able to escape

immunosurveillance and eventually give rise to macrometastasis. This provides a possible explanation for why patients with local tumors undergoing surgery develop metastasis in 75% of the cases. Interestingly, these disseminating cells are undergoing ER stress and relief of these phenotype leads to a reduction of this aggressive sub-population[63].

In both these works ER stress is thus emerging as a condition associated with an aggressive phenotype in PDAC. Moreover, these observations suggest a possibility to target ER stress as a potential therapeutic strategy to prevent relapse and metastasis.

Aim of the project

In this study we characterized the role of the transcription factor MYRF in PDAC grading. Given the observation that MYRF is expressed in our system and that, although with just a few hints, it has been connected to ER functionality, we decided to better characterize the biology and function of this TF in PDAC biology.

First, we decided to define its molecular function given that some aspects of it are still controversial. Second, we generated transcriptomic and epigenetic data to unveil MYRF role in Low-Grade PDAC cells. Finally, we integrated the data in our hands to generate a working model that propose MYRF as a central TF controlling ER functionality in Low-Grade, highly secretory cancer cells. All these data better define the function of this poorly characterized transcription factor and add a piece to the comprehension of the basis of PDAC heterogeneity.

Material and Methods

Cell culture

Cell lines and media

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), 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).

Cells were maintained in IMDM / 10% FBS (CFPAC-1), RPMI / 10% FBS (PT45P1), RPMI / 15% FBS (CAPAN-2), RPMI / 20% FBS (CAPAN-1), DMEM / 10% FBS (MiaPaCa-2, PANC-1, BxPC-3). Media were all supplemented with 2mM L-Glutamine. RPMI 1640 and DMEM were provided by Lonza, IMDM by Sigma and Fetal Bovine Serum (FBS) by Hyclone.

All cell lines were authenticated by the Tissue Culture Facility of IEO using the GenePrint10 System (Promega) for the amplification of 10 short tandem repeat-containing loci, followed by Sanger sequencing.

3D spheroids culture

Three WT and three MYRF deleted clones were counted and seeded in Low attachment 24-well plates at a concentration of 25000 cells per well. Each well contained 500 µL of medium (DMEM, 10% FBS, 2mM L-Glutamine, Pen/Strep) with 0.24% Methylcellulose. After 8 days, bright-flied picture were taken with EVOS Cell imaging system (Thermo Fisher) and sphere size was measured using ImageJ software.

Proliferation curves

On day 0, three WT and three MYRF KO clones were seeded in order to have 20000 cells per well in 6-well plates. Every two days, two wells per clone were detached and counted until day 10. Alive cells were considered and the mean of the three clones was calculated.

UPR induction

200000 cells of either WT or MYRF deleted clones were seeded in 6-well plates. Two days after, cells were treated with different concentrations of thapsigargin (5, 10, 20 nM) for 6 hours: cells were then directly lysed on plate with Lysis buffer for subsequent RNA extraction.

Depletion and deletion techniques

siRNA mediated depletion

To generate MYRF and FOS depleted cells, 200000 CFPAC1 cells were seeded in 6-well plates 24 hours prior transfection. Cells were then transfected using Lipofectamine RNAiMax (ThermoFisher, 13778075) following manual instructions: cells were transfected with either Control siRNA-A (Santa Cruz, sc-37007), MYRF siRNA (SANTA CRUZ, sc-96447), FOS siRNA (SANTA CRUZ, sc-29221), or a combination of MYRF and FOS siRNAs. Final siRNA concentration was 40nM. Transfection was repeated in the same conditions the day after and cells were then collected 48 hours after the second round of transfection.

CRISPR/Cas9 genome editing

For the generation of MYRF deleted cells, single guide sequences targeting either MYRF second or third exon were designed using the CRISPR design tool (http://tools.genome-engineering.org) and cloned into lentiCRISPv2 plasmid (Addgene #52961). The following guide sequence were selected and cloned following manufacterer's instructions:

5'-GCCACGACATCAACGGTGCCC-3'

5'-^{64}GCCAGGGCACCGTTGATGTCG-3'

After infection and puromycin selection (1.5 µg/ml), single cells were seeded in 96-well plates by dilution and expanded. Clones were screened using Western Blot.

MYRF constructs generation

Full length MYRF sequence was reconstructed from RNA-seq data and cloned in pcDNA3.1 vector carrying Flag tag at the N-terminal of the cloned protein. For fragments and mutants design, data from previous publications were used: constructs corresponding to trimerizing or monomeric MYRF were amplified from full length vector while mutants were obtained through mutagenesis strategy. Briefly, primers were designed to target the desired mutation; full length MYRF containing vector was amplified using Turbo DNA Polymerase to incorporate the mutated primers in a nicked circular dsDNA. After digestion with DpnI, aimed to degrade the parental methylated DNA, the reaction was transformed in bacteria and colonies were screened. All constructs were verified by Sanger sequencing.

EMSA

Nuclear extracts

HEK-293 cells were transfected with different MYRF overexpressing constructs using calcium phosphate technique. Two days after transfection nuclear extracts

were isolated as follows: after cytoplasmic lysis with NP40 containing lysis buffer (10 mM Hepes pH 7.6, 1mM EDTA, 60mM KCl, 0,5% NP40) for 5 minutes on ice, pellets were washed to remove the detergent with washing buffer (10 mM Hepes pH 7.6, 1mM EDTA, 60mM KCl) and purified nuclei were lysed through three freeze/thawing cycles in nuclear resuspension buffer (250 mM Tris-HCl pH 7.8, 60 mM KCl, 1 mM DTT).

Nuclear extracts were then quantified and stored at -80°C.

Probe design

IRDye 700 labeled synthetic oligonucleotides (Metabion) harboring MYRF specific consensus sequence were designed from a MYRF-bound region containing a perfect match with the motif retrieved from de novo motif discovery analysis (chr15:66,751,381-66,752,145; the MYRF motif is underlined):

5'TCTGTGCCTGGCACCATG3';

The synthetic probe harboring 3 consensus sites was designed with the motif of the previously described probe repeated three times and spaced by a constant, not interfering sequence:

5'TCTGGTGCCTGGCACCATGATCTGGTGCCTGGCACCATGATCTGGTGCCTGGCACCATG3'.

Gel retardation assay

Binding reactions were assembled in 10 mM Tris-HCl (pH 7.6), 50 mM NaCl, 0.1mM EDTA, 2 mM DTT, 4 mM MgCl2, 1 µg of Salmon Sperm DNA, and the desired amount of nuclear extract (ranging from 0.25 to 32 µg). The mixtures were incubated with 0.1 pmol of labeled probe for 20 min at room temperature and complexes were resolved on 4% polyacrylamide Tris/Borate/EDTA (TBE) native gel in the dark using

0.5X TBE buffer (pH 8) and running for ~ 150 min at 0.01A at 4°C. The gel was scanned with the Li-Cor Odissey Infrared Imaging System.

In the competition EMSA assay, increasing amount of non-labeled probes were added to the binding reaction (ranging from 6,4 to 25.6 pmol) and assay was then performed as described above.

Luciferase assay

Construct design

MYRF bound regions were cloned in different NanoLuc expressing vectors: promoter regions were cloned in a promoter-less vector (pNL1.2, Promega #1011), while enhancer regions in a minimal promoter-containing vector (pNL3.2, Promega #N1041). Synthetic MYRF bound sequence corresponding to the three motifs containing sequence used for the EMSA assay was cloned in the pNL3.2 vector.

	Chromosomal location (hg38)	*Cloning primer - Fw*	*Cloning primer - Rev*	*Recipient vector*
TM4SF4 promoter	chr3:149,473,332-149,473,991	GGCATGACACCTTCTTCAC	CCTGATGAGTTGACTGGGGA	pNL1.2
MYRF motif sequence		GTGCCTGGCACCATGATCTGGTGCCTGGCA CCATGATCTGGTGCCTGGCACAAGCTTCCG	GTGCCAGGCACCAGATCATGGTGCCAGGCA CCAGATCATGGTGCCAGGCACCTCGAGCCCG	pNL3.2
FOS enhancer	chr14:75,293,931-75,294,320	AAAAGAAGGGACAGGAGGGC	CAAATATGGGCAGAGCGGAA	pNL3.2
CEACAM6 enhancer	chr19:41,774,356-41,774,903	TACTTGCACAGCCATTCCC	CAGAGAGAATGCAAGCCACA	pNL3.2
LCN2 promoter	chr9:128,149,019-128,149,430	ACCCTTCCCTGACCCTTAA	ATGGGATCTAGGGTGGGTTG	pNL1.2

Experiment procedure

For transcriptional activation experiments, 50000 HEK-293 cells were seeded into 24-well dish. After 24 hours, cells were co-transfected with 25ng NanoLuc vector, 2.5ng PGK-Firefly luciferase vector (pGL4.53) and 100 ng MYRF expressing pCDNA3 vector using Lipofectamine2000 (ThermoFisher, 11668019) following manual instructions.

In case of repressive luciferase, the assay was performed in CFPAC1 cells in order to obtain a basal activation of the regions selected on which MYRF putative repressive effect could be evaluated. 50000 cells per well were seeded into 24-well plates the day prior transfection. The cells were then transfected with 200 ng

NanoLuc vector, 10 ng PGK-Firefly luciferase vector (pGL4.53) and 400 ng MYRF expressing pCDNA3 vector using Lipofectamine2000.

In both cases, 24 hours post transfection luminescent activity was assessed with the Nano-Glo Dual Luciferase reporter assay kit and measured at GloMax Detection system (Promega). Data are presented as Relative Luminescence Unit (RLU) normalizing NanoLuc signal by Firefly luciferase activity.

RNA manipulation techniques

Total RNA extraction

Total RNA was extracted from 10^6 cells using Zymo Quick-RNA kit (Zymo Research, R1055) and used for either retrotranscription and RT-qPCR analysis, or RNA-seq.

4sU-labelled RNA purification

Three WT and three MYRF deleted clones were plated in order to have 10^6 cells in each 10 cm dish. Two days after plating, cells were treated with 300μM 4sU (4-Thiouridine, Sigma) for 30 minutes. After harvesting by scraping in cold PBS, total RNA was extracted using Maxwell® 16 miRNA Tissue kit (Promega). 4sU-labeled RNA was biotinylated using EZ-Link Biotin-HPDP (Pierce) previously dissolved in dimethylformamide (DMF) at a concentration of 1 mg/m. Biotinylation was done in labeling buffer (10 mM Tris pH 7.4, 1 mM EDTA) and 0.2 mg/ml Biotin-HPDP for 2 h at 25 °C. Unbound Biotin-HPDP was removed by chloroform/isoamylalcohol (24:1) extraction using MaXtract (high density) tubes (Qiagen). RNA was precipitated at 20,000g for 20 min with a 1:10 volume of 5 M NaCl and an equal volume of isopropanol. The pellet was washed with an equal volume of 75% ethanol and precipitated again at 20,000g for 10 min. The pellet was resuspended in 100 μl RNase-free water. Biotinylated RNA was captured using Dynabeads MyOne

Streptavidin T1 beads (Invitrogen) in rotation for 15 min at 25 °C. Beads were magnetically fixed and washed with 1× Dynabeads washing buffer. RNA-4sU was eluted with 100 μl of 10mM EDTA in 95% formamide through incubation for 10 minutes at 65°C. RNA was eventually recovered with RNeasy MinElute Spin columns (Qiagen).

In parallel, one untreated dish per clone was extracted using Maxwell® 16 miRNA Tissue kit (Promega) as control of total RNA expression.

RT-PCR

For RT-qPCR experiments, cDNA was prepared from 1 μg of total RNA or 500 ng of nascent (4sU-labelled) RNA with ImProm-II® Reverse Transcription System (Promega) following manual instructions. RT-qPCR was assembled with Fast SYBR® Green Master Mix and run on QuantStudio 5 Real-Time PCR machine (Applied Biosystems). Analysis (Thermo Fisher Cloud platform) and primer design were performed following MIQE guidelines using primer sets either selected from the suggested validated database (PrimerBank - https://pga.mgh.harvard.edu/primerbank/) or designed with Primer3.

For nascent RNA, primers were design on introns to minimize possible contributions from fully processed RNA contaminants.

Target		*Primer sequence*	*Source*
CEACAM6	Fw	TCAATGGGACGTTCCAGCAAT	209862773c1; PrimerBank
	Rev	CACTCCAATCGTGATGCCGA	209862773c1; PrimerBank
CHAC1	Fw	GAACCCTGGTTACCTGGGC	PrimerBank
	Rev	CGCAGCAAGTATTCAAGGTTGT	PrimerBank
CHOP (DDIT3)	Fw	GGAAACAGAGTGGTCATTCCC	Primer bank
	Rev	CTGCTTGAGCCGTTCATTCTC	Primer bank
DEFB1	Fw	ATGAGAACTTCCTACCTTCTGCT	50659096c1; Primer bank
	Rev	TCTGTAACAGGTGCCTTGAATTT	50659096c1; Primer bank
GAPDH	Fw	GTGGAAGGGCTCATGACCA	Primer3
	Rev	GGATGCAGGGATGATGTTCT	Primer3
HERPUD1	Fw	ATGGAGTCCGAGACCGAAC	Primer bank
	Rev	TTGGTGATCCAACAACAGCTT	Primer bank
IRE1a (ERN1)	Fw	CACAGTGACGCTTCCTGAAAC	PrimerBank
	Rev	GCCATCATTAGGATCTGGGAGA	PrimerBank
LCN2	Fw	GACAACCAATTCCAGGGGAAG	PrimerBank
	Rev	GCATACATCTTTTGCGGGTCT	PrimerBank
NECTIN4 (PRVL4)	Fw	AGGACGCAAAACTGCCCTG	222136610c1; Primer bank
	Rev	TGAAGCCCGTATTTGGAGTGC	222136610c1; Primer bank
S100A14	Fw	GAGACGCTGACCCCTTCTG	359385701c1; Primer bank
	Rev	CTTGGCCGCTTCTCCAATCA	359385701c1; Primer bank
S100P	Fw	AAGGATGCCGTGGATAAATTGC	PrimerBank
	Rev	ACACGATGAACTCACTGAAGTC	PrimerBank
SERPINA3	Fw	GCTCATCAACGACTACGTGAA	73858562c2; Primer bank
	Rev	CACCATTACCCACTTTTTCTTGC	73858562c2; Primer bank
TNFRSF6B	Fw	GTACGCGGAGTGGCAGAAA	323639466c1; Primer bank
	Rev	CAGAGGACGTTGCAGTAGC	323639466c1; Primer bank
XBP1 spliced	Fw	TGCTGAGTCCGCAGCAGGTG	van Schadewijk et al, 2012
	Rev	GCTGGCAGGCTCTGGGGAAG	van Schadewijk et al, 2012
CEACAM6 intron	Fw	TGTTGACAAGAGCCCTGGAA	Primer3
	Rev	ACATCAGGGCATTCTTGGGA	Primer3
DEFB1 intron	Fw	TCATTGGGGTGTGCTCTGAT	Primer3
	Rev	CTGGGTCCTCACATGGTCTT	Primer3
S100A14 intron	Fw	CCCACCTCTGCTTGAAACAC	Primer3
	Rev	TGTCCAGAGGGCACACTTAG	Primer3
S100P intron	Fw	GCTTTGCTATCACATGGGCA	Primer3
	Rev	GTCCCAAGCTCTCAGGTCTT	Primer3
SERPINA3 intron	Fw	GCTGCCAACCATTATCAGGG	Primer3
	Rev	AGTGACTCTCCCAGCCTCTA	Primer3
TNFRSF6B intron	Fw	GTCAGGTTGCTGGTCCCA	Primer3
	Rev	GAAGAACAGGGTCAGGGGAA	Primer3

Protein manipulation techniques

Co-Immunoprecipitation experiment

HEK-293 cells were transfected by calcium phosphate with either empty vector or the combination of FOS and MYRF overexpressing pCDNA3 vectors. Two days

after transfection, cells were lysed in RIPA buffer (50 mM Tris-HCl pH8, 150 mM NaCl, 1 mM EDTA, 1% NP-40, 0.5% Na-DOC). After clarification, lysetes were incubated with 3 μg of MYRF antibody or corresponding amount of control rabbit IgG. IPs were set as follows: every IP was performed with 1.8 mL of lysate, corresponding to the proteins collected from one 10 cm dish, while 50 μL were saved as Inputs. After over-night incubation, Dynabeads Protein G addition for two hours and washes, IPs were eluted directly in Laemmli Sample Buffer and loaded on gel.

Western blot

Cells were lysed in lysis buffer (50 mM TrisHCl pH 7.5, 300 mM NaCl, 1% NP-40, 1mM EDTA) containing protease inhibitors, 1 mM PMSF and 1 mM NaF, sonicated and clarified through centrifugation. 50 μg of cell extracts were resolved on SDS–polyacrylamide gel, blotted onto nitrocellulose membranes, and probed with the following antibodies:

Target	*Antibody*	*Antibody concentration*
MYRF	HomeMade	2 ug/mL
FOS	HPA018531 (Sigma)	0,4 ug/mL
Vinculin	V9131 (Sigma)	1:10000

Immunofluorescence analysis.

Two-color immunofluorescence and confocal analysis were performed on siRNA treated CFPAC1 cells grown on glass coverslips coated with collagen. Briefly, PFA fixed cells where permeabilized with 0.1% Triton X-100, blocked and incubated with the primary antibodies. Alexa488 labeled anti-rabbit IgGs secondary antibody (Jackson ImmunoResearch) was used to detect the primary antibody. Nuclei were counterstained with DAPI and samples mounted with Mowiol aqueous mounting medium supplemented with DABCO anti-fading agent (Sigma). Confocal

microscopy was performed on a Leica SP8 laser confocal microscope and four random images for each experimental point were acquired.

Primary antibodies were used as follows:

Target	*Antibody*	*Antibody concentration*
MYRF	LS-C109911 (LSBio)	20 ug/mL
FOS	HPA018531 (Sigma)	2 ug/mL

Next-generation sequencing

RNA-seq

RNA-seq was carried out using the SMART-seq2 protocol[65] with minor modifications. Briefly, 10ng of total RNA were copied into first strand cDNA by reverse transcription and template-switching using oligo(dT) primers and an LNA-containing template-switching oligo (TSO). The resulting cDNA was pre-amplified, purified and tagmented with Tn5 transposase produced in-house using a described protocol[64]. cDNA fragments generated after tagmentation were gap-repaired, enriched by PCR and purified to create the final cDNA library for Illumina HiSeq2000 platform.

ChIP-seq

50-150 x 10^6 CFPAC1 cells were either fixed for 10 minutes with 1% of formaldehyde (for MYRF ChIP-seq) or fixed with a double crosslinking protocol (for FOS ChIP-seq). In the second case, cells were scraped in PBS, incubated for 45 minutes with 2mM DSG in PBS, washed twice and then incubated with 1% formaldehyde in PBS for 10 minutes followed by quenching with 125mM glycine. 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 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 purified with solid-phase reversible immobilization (SPRI) beads (Agencourt AMPure XP, Beckman Coulter), and then quantified with QuantiFluor (Promega). DNA libraries were prepared for HiSeq2000 sequencing using a standard protocol.

The following antibodies were used: MYRF (Homemade generated antibody targeting the N-terminal portion of the protein), FOS (Sigma, HPA018531).

Immunohistochemistry and tissue microarrays

IHC

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 (TMA) slides (PA961c) were obtained from US Biomax, Inc while sections of PDAC cell lines derived xenografts were generated as previously described (Diaferia et al., 2016). Formalin fixed paraffin embedded (FFPE) sections were rehydrated and subjected to heat-induced antigen retrieval in either NaC buffer (10 mM sodium citrate, 0.05% Tween20, pH6.0) or EDTA buffer (1mM EDTA, 0.05% Tween20, pH8.0. After blocking, samples were incubated with the following antibodies or corresponding amount of control IgG.

Target	*Antibody*	*Antigen retrieval*	*Antibody concentration*
MYRF	LS-C109911 (LSBio)	EDTA	20 ug/mL
Ki67	BD-550609	NaC	5 ug/mL
CEACAM6	Ab137859 (Abcam)	NaC	1 ug/mL
FOS	HPA018531 (Sigma)	EDTA	0,25 ug/mL

Slides were incubated with HRP-conjugated anti-rabbit or anti-mouse IgG antibody (Envision, DAKO). Signal was revealed by incubation with diamonobenzidine (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 open source software QuPath.

Alcian Blue staining

Formalin fixed paraffin embedded (FFPE) sections were rehydrated and incubated with alcian blue solution for 5 minutes at room temperature. Nuclei were counterstained with Nuclear Fast Red solution (Sigma) and slides were mounted for image acquisition at Olympus upright BX51 microscope linked to a Nikon DS-5Mc Color camera.

Transmission electron microscopy (TEM)

Spheroids obtained as described in previous sections from WT and MYRF deleted clones were fixed after 7 days of culture at 4°C in 0.1 M sodium cacodylate buffer (pH 7.4), 2% paraformaldehyde, 2.5% glutaraldehyde (Electron Microscopy Sciences), and 3 μM CaCl2. Samples were then postfixed with osmium tetraoxide (1% wt/vol in H2O) and counterstained with uranyl acetate (2% wt/vol in H2O). After embedment in Durcupan resin (MilliporeSigma), ultrathin sections (70 nm) were prepared, mounted on 300 mesh gold grids, and counterstaining with uranyl acetate (1% wt/vol in H2O) and Sato lead (1% wt/vol in H2O). Ultrathin sections were imaged at 80 keV using an electron microscope (JEOL JEM-1230) equipped with an AMT XR80 CCD camera.

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). Two pools of three WT or three MYRF-KO clones were prepared and nude mice (n=5 for each group) were injected with 10^7 cells resuspended in 100ul of PBS under the skin of their hind flank. Subcutaneously injected tumors were harvested 4 weeks after injection, fixed in 4% paraformaldehyde and processed for paraffin embedding.

Computational methods

ChIP-seq data analysis

Short reads obtained from Illumina HiSeq 2000 were quality filtered according to the Illumina pipeline. Reads were then mapped to the human hg38 reference genome using Bowtie2 v2.2.6 [66] with the "–very-sensitive" parameter. Reads that did not align to the nuclear genome or aligned to the mitochondrial genome were removed. Moreover, duplicate reads were marked and removed using SAMtools [67]. Peak calling vs. the input genomic DNA was performed using MACS2 (version 2.1.0.20150731)[68] using the "--nomodel", "--extsize 200" and "--qvalue 0.01" flags and arguments. Peaks with a fold enrichment (FE) relative to input <5 (as determined by MACS2) and those blacklisted by the ENCODE consortium analysis of artifactual signals in human cells (https://sites.google.com/site/anshulkundaje/projects/blacklists) were removed using bedtools [69].

Annotation and classification of ChIP-seq peaks

To classify ChIP-seq peaks based on their genomic location and assign them to the nearest TSS, the RefSeq annotation of the hg38 version of the human genome was given as input to the annotatePeaks script from HOMER package [70]. We

classified each peak as either TSS-proximal or TSS-distal, depending on its distance (< or > 2.5 kb, respectively) from annotated transcription start sites (TSS).

Heatmap of MYRF ChIP-seq enrichment in CFPAC1 cell line

Reads Per Million (RPM) were measured in a window of 5 kb (500 bins of 10 bp) centered on the summits of MYRF peaks. To avoid any bias due to outliers, a saturation procedure was performed and values were then scaled to the range 0-1. Regions were sorted according to their intensity levels and visualized using heatmap.2 in R.

De novo motif discovery

Motif discovery was performed using MEME v4.10.1 [71] with the options "-dna -mod zoops -evt 1e-5 -nmotifs 10 -minw 6 -maxw 12 -revcomp -maxsize 10+7" using a window of +/-100 bp centered on the summits of the 1000 highest-scoring MYRF peaks. We next used TomTom [72], with default parameters except for "-dist ed", in order to assess the similarity of the identified motifs to the consensus binding sites collected in JASPAR database (http://jaspar.genereg.net).

Motif Enrichment Analysis

In order to identify statistically over-represented motifs corresponding to known TF binding sites, position-specific weight matrices (PWMs) were collected from specific databases and the literature and used to build a custom set of 1,744 models. Significantly over-represented 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 over-represented. The window considered for these

analyses was set to ± 150 bp around MYRF peak summit. The FANTOM5 enhancer set was used as background.

Smart-seq2 analysis

After quality filtering according to the Illumina pipeline, 50 bp single-end reads were aligned to the hg38 human reference genome and to the Homo sapiens transcriptome (NCBI build 37.2) using TopHat (version 2.1.0)[73] with the option "--b2-very-sensitive". Only uniquely mapped reads were retained. At the gene level, expression counts were estimated using featureCounts (Rsubread version 1.5.1)[74], summarized across all exons as annotated in NCBI GRCh38/hg38, with option "--largestOverlap". Both coding and long noncoding genes were retained for downstream analyses. Differentially expressed genes in biological triplicates of wild type and MYRF-KO CFPAC1 clones were identified using EdgeR R-package (version 3.2.2)[75]. Prior to normalization using the Trimmed Mean of M (TMM) method, only genes with at least 3 CPM (Count Per Million) in at least half of the samples were retained. A common dispersion was estimated for all genes to measure the global biological variation (with option robust = "TRUE"). A negative binomial generalized log-linear model was fitted to each gene, and likelihood ratio tests were performed to assess differential expression [76]. Genes were identified as differentially expressed when the following criteria were met: fold-changes (FC) ≥ |1.5|, false discovery rate (FDR) ≤ 0.01 and 1 TPM in all samples in one or both conditions. Then, Transcript Per Million (TPM) values were used as expression unit.[77]

Gene ontology analysis

Functional enrichment analyses were performed using the GOrilla tool. 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 either Biological Process or Cellular Component categories and selected GO terms with enrichment (p-value ≤ 1E-3). Data visualization was carried out using REVIGO (http://revigo.irb.hr/index.jsp)[78]. The analysis was run by selecting default parameters except for the resulting list that was setting as small size.

Gene Set Enrichment Analysis (GSEA)

Gene Set Enrichment Analysis (GSEA)[79] was used to investigate whether a gene set was significantly over-represented in the transcriptome of either WT or MYRF KO cells. Transcripts were ranked by the Difference of Classes (metric for gene ranking) and using the following settings: number of permutations = 1000, 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 $\underline{\leq}$ 0.01.

Results

MYRF is differentially expressed in PDAC grading

We previously identified a series of differentially expressed transcription factors[1] using a model of PDAC grading based on cell lines. In this study, MYRF was higher expressed in Low-grade cells which retain the high secretory capacity of normal pancreatic ductal cells. Therefore, given its possible involvement in the connection between ER and nuclear functions, we set out to study the impact of MYRF in PDAC biology and function.

To verify if MYRF differential expression was maintained at the protein level, Western Blot on PDAC cell lines was performed. MYRF protein resulted differentially expressed in Low-grade PDAC cell lines compared to High-grade cells (Figure 6A). These same cell lines, when xenotransplanted into nude mice, give rise to tumors that highly resemble the grade of origin, thus being a good model of differentiation and grading *in vivo*. Therefore, we evaluated MYRF levels in tumors generated by subcutaneous injection of low-grade (Capan1 and CFPAC1) and high-grade (MiaPaca2 and PANC1) cells. As shown in Figure 6B, MYRF staining in immunohistochemistry shows its differential expression in Low-Grade cell derived tumors, with high levels of MYRF expression in ductal-like structures typical of Low-grade G1 tumors.

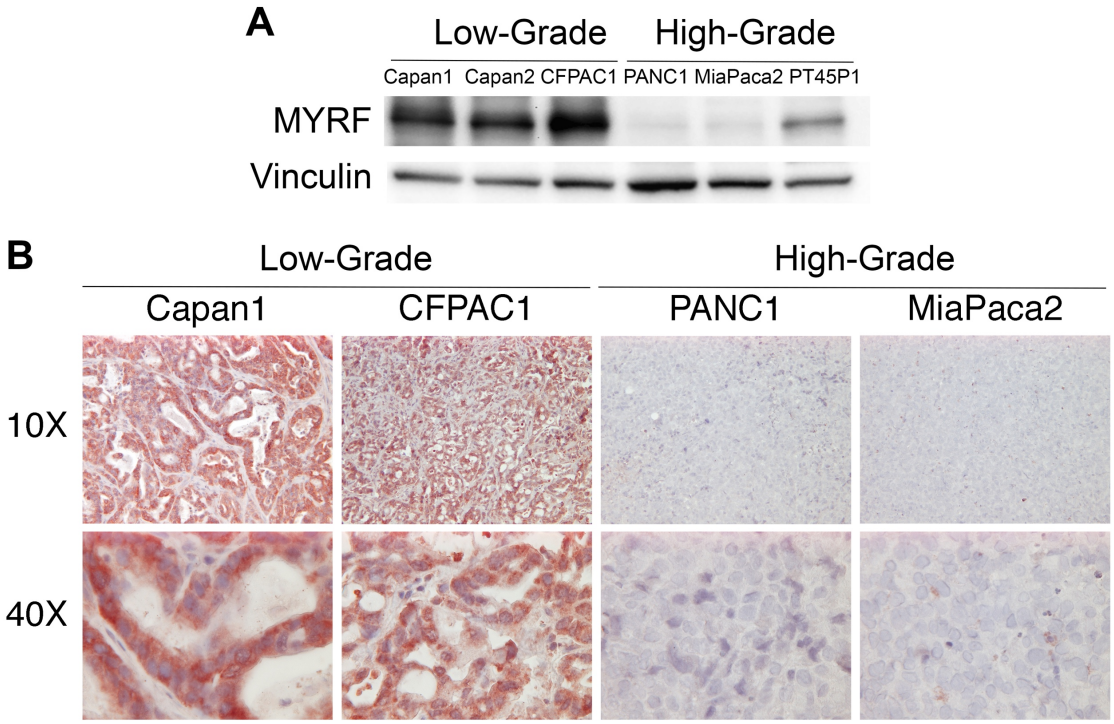


Figure 6: MYRF differential expression in PDAC cell lines. A) Western Blot showing differential expression of MYRF protein in Low-Grade PDAC cell lines. Vinculin is shown as loading control. B) Immunohistochemistry analysis of MYRF expression on tumors generated by subcutaneous transplantation of PDAC cell lines in nude mice.

We then expanded our analysis of MYRF levels to a cohort of human tumors using tissue microarrays (TMA). MYRF expression inversely correlates with grading, being highly expressed in G1 areas and less expressed or completely absent in G3 tumor areas (Figures 7A-7B).

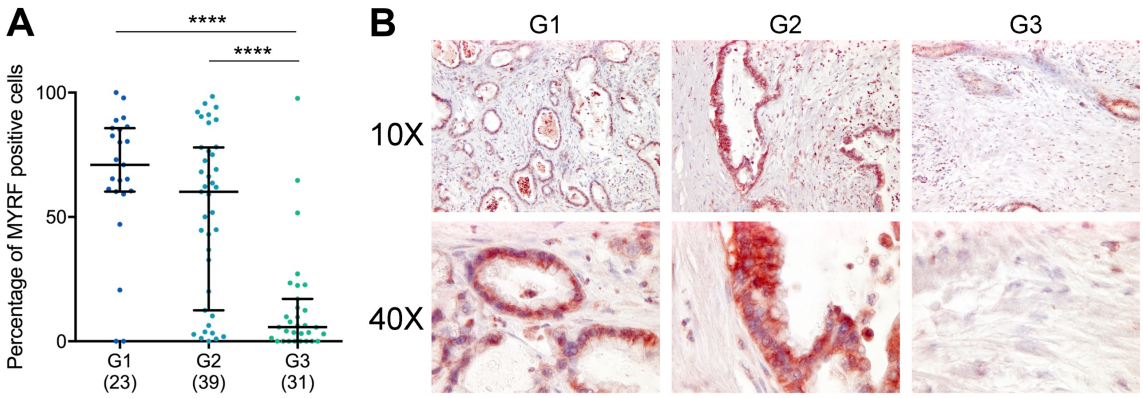


Figure 7: MYRF differential expression in PDAC tumors. A) Scatter plot reporting the fraction of MYRF positive cells in a tumor tissue microarray. Images were automatically acquired and quantified using QuPath. P-values were calculated using Mann-Whitney test. The number of quantified tumor areas is indicated (total number=93). B) Representative IHC of MYRF staining in PDAC tumor samples

Taken together, these results indicate that MYRF is differentially expressed in Low-grade PDAC cell lines and tumors.

MYRF genomic distribution in Low-Grade PDAC cells

Given that the grade-specific expression of MYRF might underlie a functional role in this PDAC subpopulation, we decided to better characterize MYRF function and mechanism of action in these cells. We first characterized MYRF genomic occupancy in the Low-grade cell line, CFPAC1, by performing MYRF chromatin immunoprecipitation followed by sequencing (ChIP-seq).

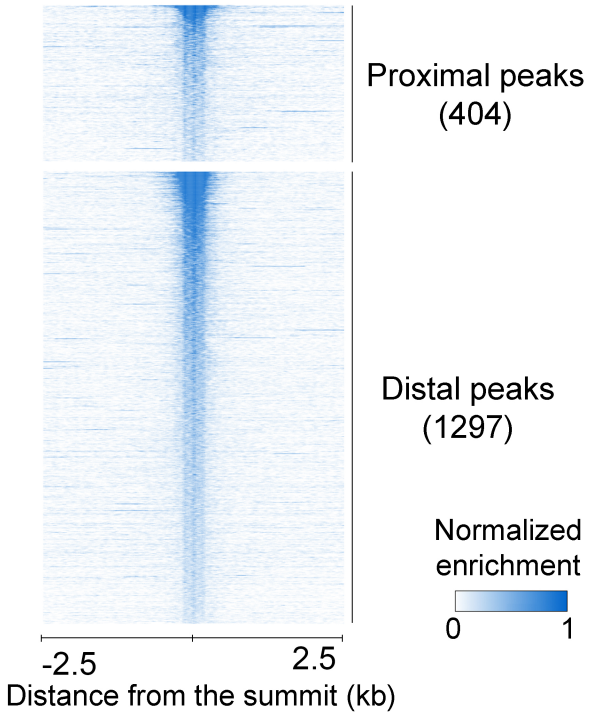


Figure 8: MYRF genomic occupancy in PDAC cells. Heatmap showing MYRF genomic distribution in CFPAC1 cells. Peaks located within 2.5 kb from the transcription start site of a gene are define as proximal peaks. Binding intensity is normalized as shown in the legend and peaks are sorted according to their intensity.

MYRF bound to 1701 genomic regions, 76% of which are transcription start site-distal regions representing candidate enhancers (Figure 8). Despite the relatively

small number of regions retrieved, the quality of the peaks was good, as shown in Figure 9 with representative snapshots, demonstrating the overall good quality of the ChIP-seq experiment.

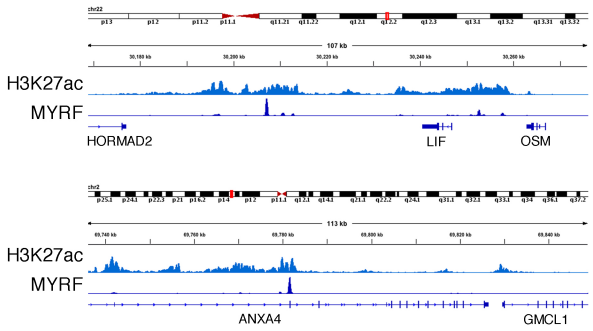


Figure 9: MYRF genomic distribution in CFPAC1 cells. Representative MYRF ChIP-seq snapshot in Low-Grade PDAC cells. H3K27ac ChIP-seq is shown as a reference for genomic active regions.

To date, only mouse MYRF protein has ever been chromatin immunoprecipitated.[34] Thus, we performed a *de novo* motif discovery analysis to identify the human MYRF DNA binding motif. The most enriched motif in the top 1000 MYRF-bound regions (Figure 10) is a palindromic sequence with a central nucleotide spacer and it resembles the published motif for MYRF mouse ortholog[2].

Figure 10: Identification of MYRF DNA binding motif. Top match retrieved in MEME *de novo* motif discovery. The 200 bp around the summit of MYRF 1000 top peaks were used in the analysis. Number of matches in the sequences submitted and E-value are reported.

To confirm MYRF binding to this motif, we analyzed its spatial distribution in relation to the summit of MYRF ChIP-seq peaks. Firstly, in over 50% of cases, this motif was

located within 50 bp from the summit of MYRF peaks (Figure 11). Secondly, MYRF binding strength correlated with the number of motifs under the peak (Figure 12).

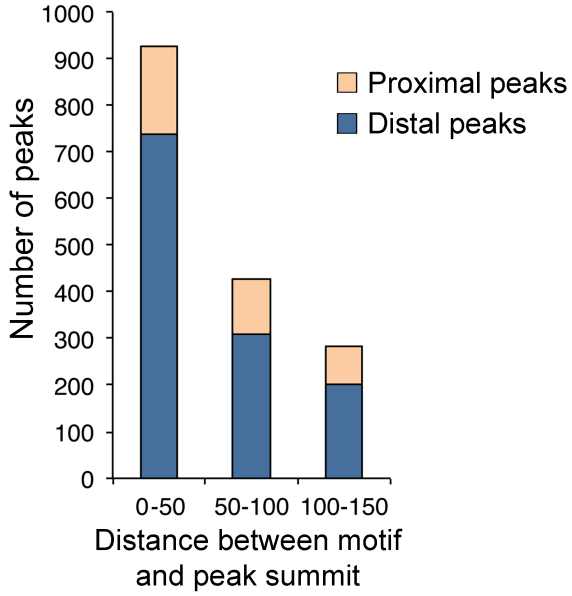


Figure 11: Validation of MYRF motif binding. Distribution of MYRF motif in relation to peak summit shows that, in more than half of the peaks, the motif falls within 50 bp from the summit. Peaks within 2.5 kb from the TSS of the nearest gene are defined as proximal peaks.

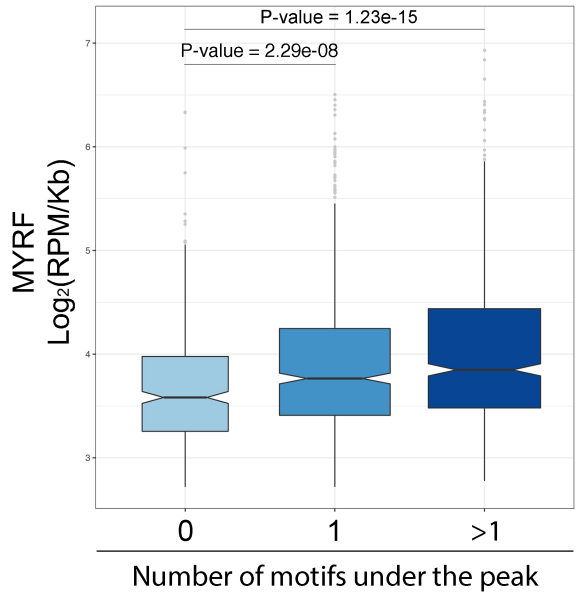


Figure 12: MYRF binding intensity correlates with the number of motifs under the peaks. MYRF peaks retrieved in ChIP-seq experiment were clustered based on the presence of none, one or more motifs underneath them and the box plot reports the correlation between MYRF binding intensity and number of MYRF motifs underneath each peak.

In conclusion, MYRF binds a specific set of genomic regions in Low-Grade PDAC cell lines thorough the recognition of a motif highly conserved from mouse.

Characterization of MYRF-DNA binding interaction

To confirm that MYRF physically interacts with the DNA motif that we retrieved from our ChIP-seq data, we performed an electrophoresis mobility shift assay (EMSA) by incubating a probe containing the MYRF motif with nuclear extract from MYRF-overexpressing HEK-293 cells. As shown in Figure 13, MYRF strongly bound to the probe and a mutation in the DNA binding domain responsible for MYRF-DNA interaction[36] completely abolished binding.

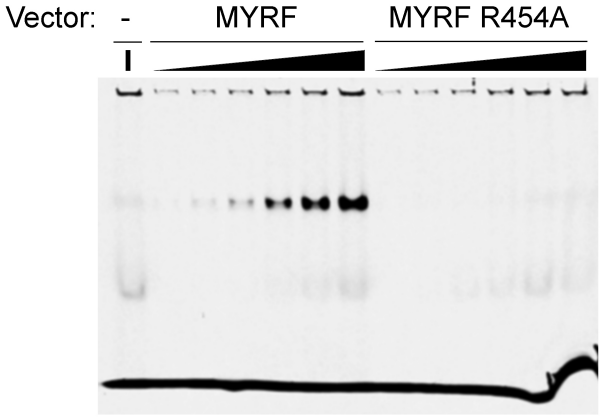


Figure 13: MYRF specifically recognize its DNA motif. EMSA experiment performed incubating a labelled probe containing MYRF motif with increasing amounts of nuclear lysates overexpressing MYRF. MYRF: wild type protein. MYRF R454A: Point mutant MYRF unable to bind DNA.

MYRF is synthetized as an ER-membrane protein and only after proper folding and trimerization it self-cleaves and translocates into the nucleus. While the trimerization is fundamental for self-cleavage, its role in enabling binding to the DNA and transcription is controversial[34,36]. To shed light on this aspect we performed an EMSA assay with different MYRF expression constructs. We used the wild type protein, a fragment that lacks the transmembrane domain but maintains the capability to trimerize and self-cleave (MYRF 1-765) and a protein that corresponds

to the product of cleavage but, as it lacks the catalytic domain, is not able to trimerize (MYRF 1-586).

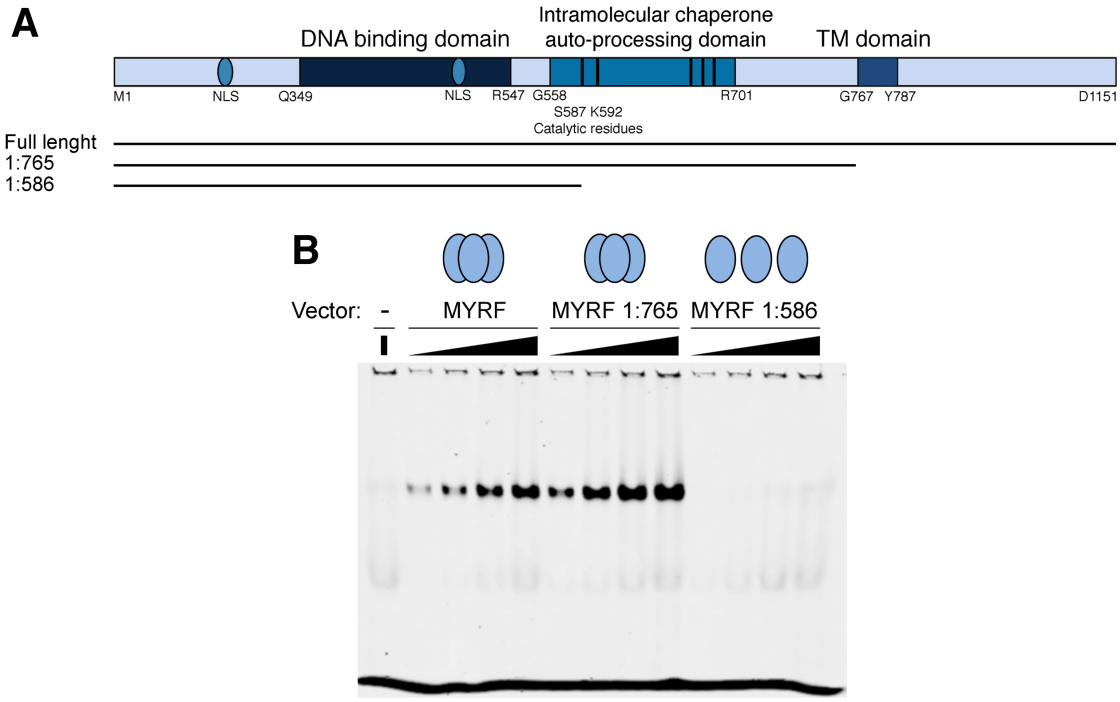


Figure 14: MYRF capability of binding requires proper protein folding and trimerization. A) Schematic representation of MYRF protein structure with indication of the fragments used in the subsequent experiment. B) The probe with MYRF motif was incubated with nuclear lysates overexpressing different MYRF variants. MYRF: wild type protein. MYRF 1:765: MYRF portion that lacks the transmembrane domain but is able to trimerize and self-cleave. MYRF 1:586: monomeric MYRF

While the trimerizing fragment was able to bind the DNA similarly to the wild type protein (Figure 14), the monomeric MYRF completely lost this capability. This was not due to a difference in protein expression or stability, since Western analysis on equal amounts of lysate showed comparable levels of expression of the different proteins (Figure 15).

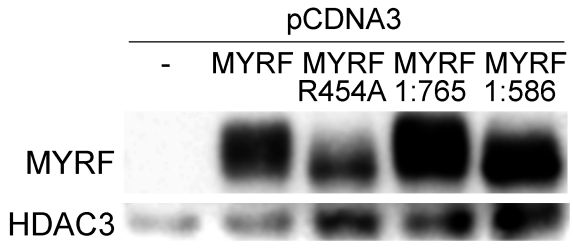


Figure 15: Effects observed in EMSA are not due to differences in protein stability. WB on the same nuclear extracts used in the aforementioned EMSA experiments. HDAC3 is shown as loading control.

This result demonstrates that trimerization is essential for MYRF binding to the DNA. Since trimeric MYRF recognizes a palindromic sequence, different modes of interaction are possible (Figure 16). Only one monomer might contact the DNA sequence; alternatively, each part of the palindrome might be recognized by two different monomers inside the same trimer. Finally, it is also possible that each monomer inside the trimer interacts with a different motif in the genome, with the intervening sequence being looped out. The latter possibility could also be supported by the presence of multiple motifs underneath several MYRF peaks (Figure 12).

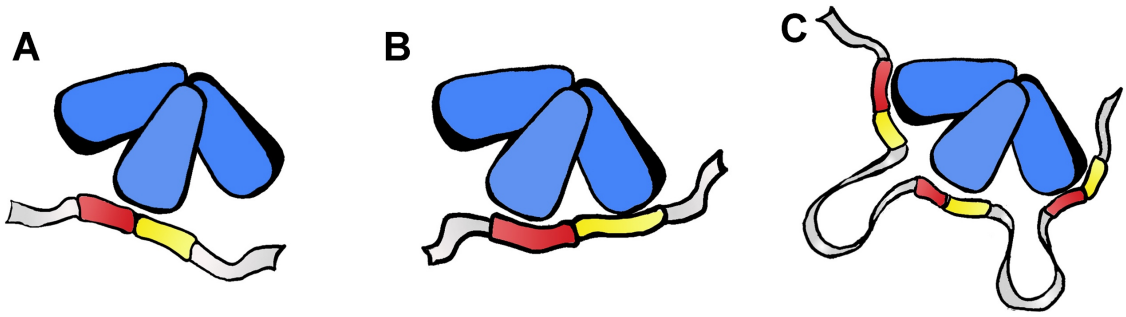


Figure 16: Different models of MYRF-DNA interaction. A) A single monomer inside the trimer directly contacts DNA. B) The two parts of the DNA palindromic sequence are recognized by two different monomers inside the same trimer. C) Each monomer contacts a different motif.

To distinguish between these different possibilities, we performed an EMSA assay with a probe containing three repeated MYRF motifs. With low quantities of lysate the probe was bound by a single trimer (as demonstrated by the fact that the shift of the band is the same as the probe with a single motif). However, increasing the amount of lysate led to a further shift of the band, indicating that the probe was progressively occupied by two and then three trimers (Figure 17). This means that each trimer contacts a single motif in the genome. Nevertheless, it remains to be elucidated whether the motif is contacted by a single monomer or if the palindrome

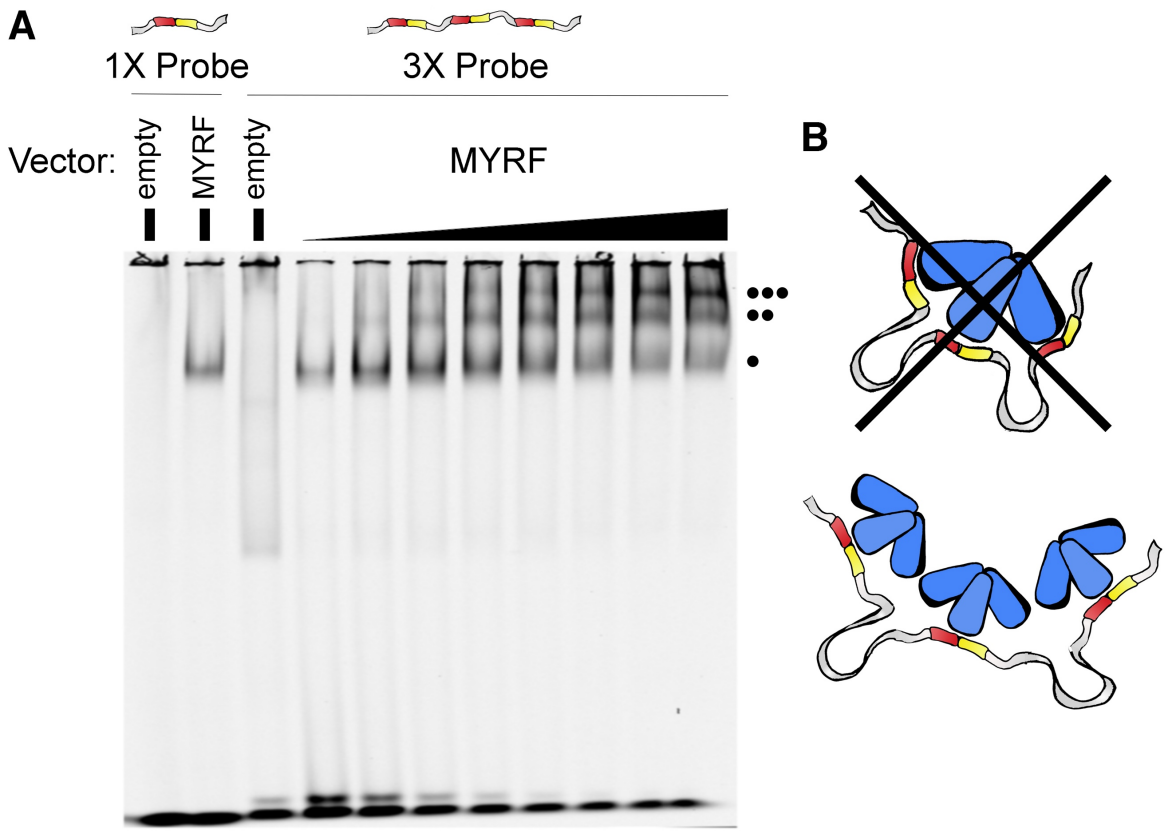


Figure 17: Each MYRF trimer recognize a single DNA motif. A) EMSA comparing the band shift between a probe with a single MYRF motif and one containing three motifs: when incubated with lysates overexpressing MYRF, the probe with three motifs is progressively occupied by three different trimers. B) Graphical representation of the two binding models between which the EMSA in panel A discriminated.

is functional for the binding of two different monomers. To answer this question we performed a competing EMSA assay incubating the wild type labeled probe with MYRF and increasing amounts of a cold probe, either wild type or mutated in one part of the palindrome. While the WT cold probe was able to compete with the labeled probe and the net result was a decrease in the signal of the shifted labeled probe (Figure 18), all the mutated probe tested completely lacked this capability, indicating that both arms of the palindrome are essential for sequence recognition by MYRF.

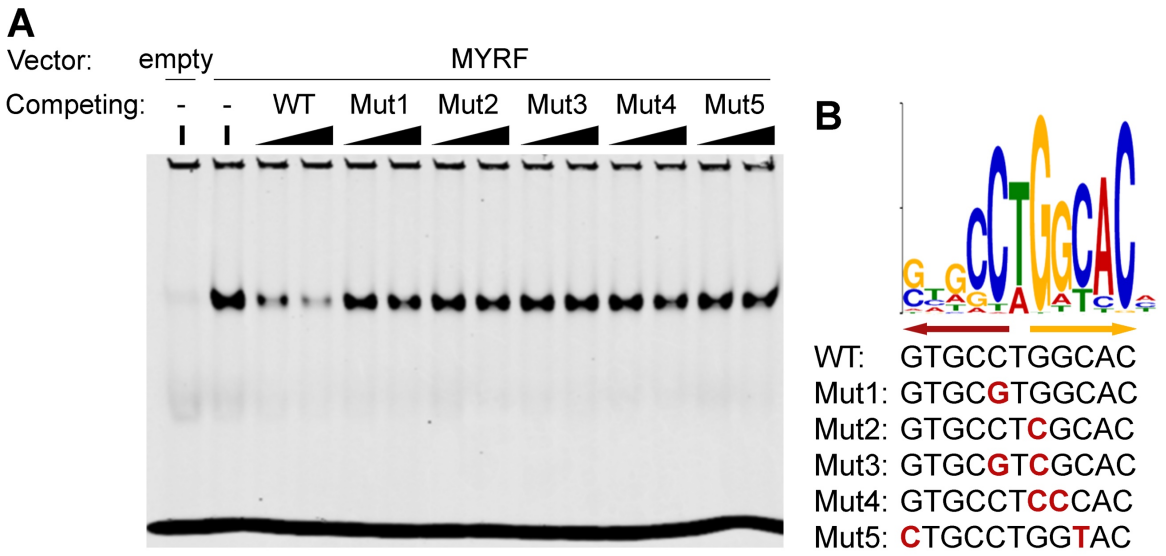


Figure 18: DNA palindromic sequence is necessary for MYRF binding to DNA. A) Competing EMSA in which the labelled probe containing MYRF single motif is incubated with MYRF overexpressing lysates and increasing amount of not-labelled competing probes. B) Graphical representation of the sequences used to design the competing probes in the previously described EMSA experiment.

We then asked if MYRF, in addition to bind the DNA, is able *per se* to activate transcription. We cloned a sequence containing a three-times repeated MYRF motif upstream to the NanoLuc gene and co-transfected this vector with different MYRF over-expression constructs. As illustrated in Figure 19, MYRF was potently able to activate the transcription of the NanoLuc. On the contrary, mutations in either the DNA binding domain (MYRF R454A) or in the catalytic residue responsible for cleavage and release of the nuclear trimer (MYRF S587A) completely abolished transcriptional activation. Thus, mirroring the EMSA results, the fragment that maintains trimerization capacity (MYRF 1-765) is able to activate transcription as well as the wild type proteins, while monomeric MYRF is not.

Taken together, these results indicate that MYRF is able to directly contact DNA and activate transcription and this capability is strictly dependent on correct protein folding and trimerization.

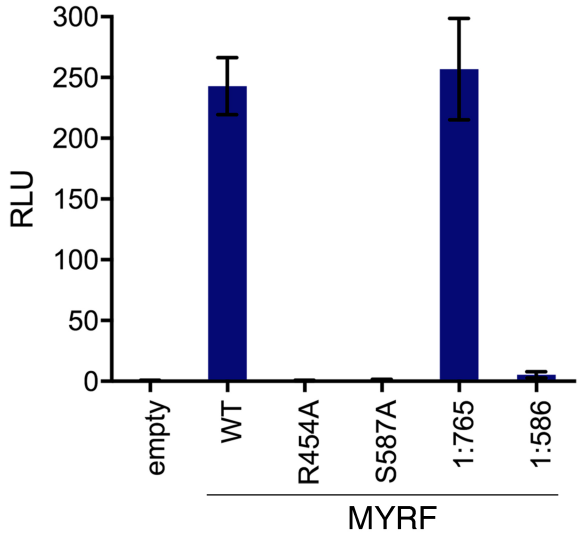


Figure 19: MYRF is a potent transcriptional activator. A) Graph reporting the results of a luciferase assay performed co-transfecting a NanoLuc vector (with NanoLuc under the control of MYRF motif and a minimal promoter) and different MYRF overexpressing constructs. Results are expressed as Relative Luminescence Unit (RLU) with the cotransfection of a Firefly luciferase expressing vector. Mean and SD from three different experiments are shown.

MYRF transcriptional program in Low-Grade PDAC cells

In order to get insights into MYRF function in Low-Grade PDAC cells, we performed CRISPR-Cas9 mediated genome editing to delete MYRF in CFPAC1 cells. Single guide RNAs (sgRNAs) were designed to target either the second or the third MYRF exon and cloned in the lentiCRISPv2 vector carrying the expression of Cas9 and puromycin resistance. After infection and selection cells were seeded in clones. Knock out clones were screened through Western Blot and three WT and three KO clones were selected (Figure 20).

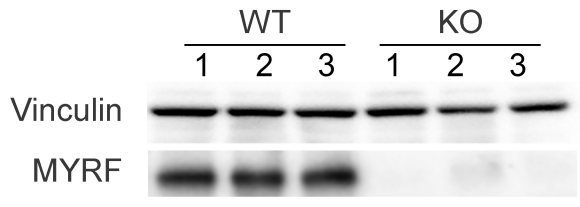


Figure 20: MYRF deletion in Low-Grade CFPAC1 cells. MYRF expression in MYRF KO clones generated by CRISPR/Cas9-mediated genome editing. Matched control clones are also shown.

We analyzed the transcriptional profile by RNA-seq using Smart-seq2 technology for cDNA synthesis and Illumina Nextera library preparation. Libraries were sequenced on a Next Generation sequencing platform. Using a high stringency cutoff, only genes with a Log_2FoldChange greater than 0.7 and a false discovery rate (FDR) less than 0.01 where selected as differentially expressed genes (DEGs). As illustrated in the Volcano, MYRF KO causes the differential expression of 253 genes, almost equally distributed between up- and down-regulated genes (Figure 21).

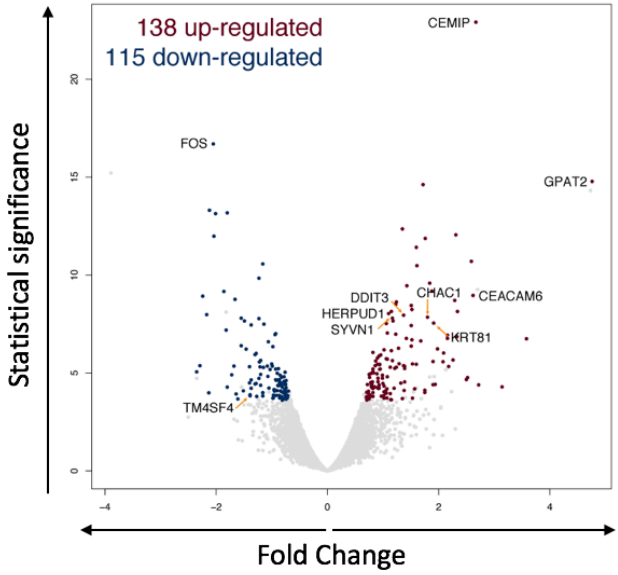


Figure 21: Impact of MYRF deletion on CFPAC1 transcriptomic profile. Volcano plot showing the effects of MYRF deletion in CFPAC1 cells. Differentially expressed genes were identified by comparing three WT with three MYRF-deleted clones. The y axis shows the -Log10 of the FDR determined by CuffDiff.

To uncover which molecular functions were impacted by MYRF in Low-Grade cells, we performed a gene ontology (GO) analysis on these differentially expressed genes. Statistically significant GO are shown in Figure 22. Down-regulated genes were related almost entirely to DNA replication and cell cycle, while up-regulated

genes were associated with inflammation and, very interestingly considering MYRF protein processing, endoplasmic reticulum stress and unfolded protein response.

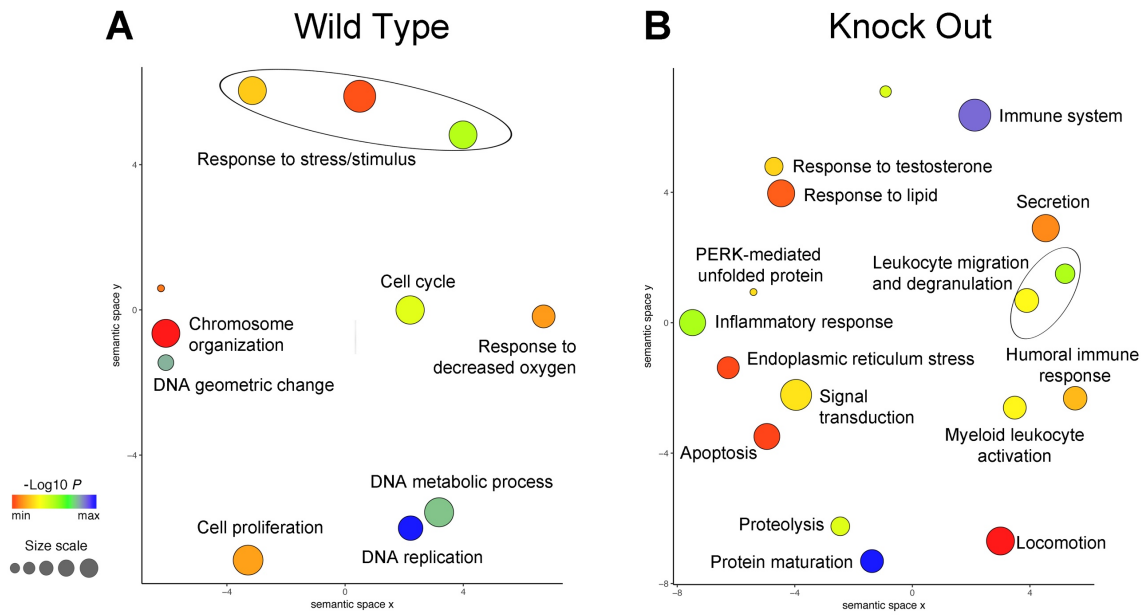


Figure 22: Gene ontology functions impacted by MYRF deletion in Low-Grade PDAC cells. Revigo representation of Gene ontology analysis on the set of differentially expressed genes in MYRF-KO cells. Bubble dimension correlates with generality of the term in the GOA database while bubble color indicates the FDR (color legend on the left).

To further confirm the functional significance of MYRF DEGs, we carried out a Gene Set Enrichment Analysis (GSEA). Ranking the entire list of DEGs, this type of analysis allows for the identification of gene sets overrepresented exclusively in up- or down-regulated genes, avoiding the recurrence of categories that could be impacted in both directions.

Futhermore, this analysis corroborates what was shown with Gorilla gene ontology analysis. Gene set enriched in WT (and consequently down regulated in KO) were related to DNA replication and DNA elongation, while gene sets enriched in KO showed terms like peptide elongation and SRP-dependent co-translational protein targeting to membrane, again linking MYRF to ER function (Figure 23).

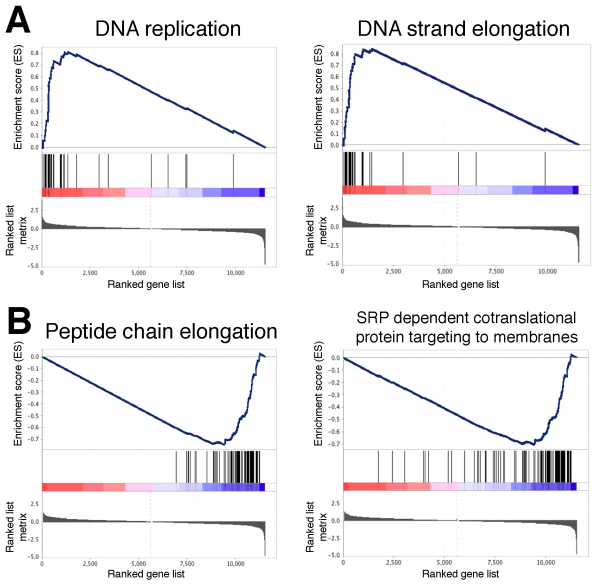


Figure 23: GSEA unveils the impact of MYRF deletion on Low-Grade PDAC cells. Representative gene sets enriched in either WT (Panel A) or MYRF KO (Panel B) CFPAC1 cells. Genes are ranked from left to right based on their relative level of expression, with genes on the left showing higher expression in WT cells.

MYRF KO causes a reduction in cell proliferation

Taken together, the MYRF peculiar protein processing and the transcriptional effects of its knock out suggest that MYRF could be synthesized as an ER membrane protein to work as a sensor of proper ER functionality. Only in a situation of normal ER function will proper MYRF folding, trimerization, self-cleavage and translocation into the nucleus occur, where it could then serve to induce the transcription of replication-related genes, thus licensing for replication cells with a normal ER function.

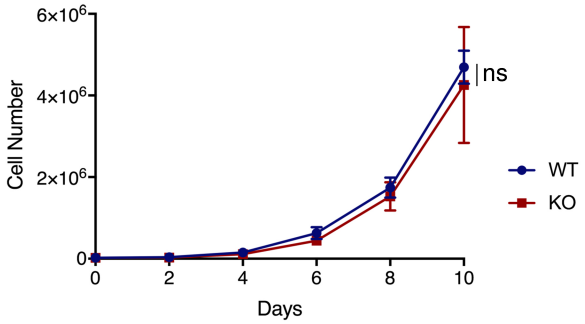


Figure 24: MYRF KO does not impact on cell proliferation in 2D culture condition. Cumulative proliferative curves of WT and MYRF-KO clones grown in 2D normal conditions. Results are expressed as mean of three different clones per condition. P value was calculated using unpaired Welch's t test.

To test this hypothesis, we asked whether MYRF KO induces proliferation defects in Low-Grade PDAC cells. Although the proliferation rate of WT and KO clones was not significantly different in normal 2D-culture condition (Figure 24), we plated these cells in a 3D-methylcellulose system and allowed them to grow as spheroids. As shown in Figure 25, MYRF KO caused defects in proliferation, with a significant reduction in sphere size.

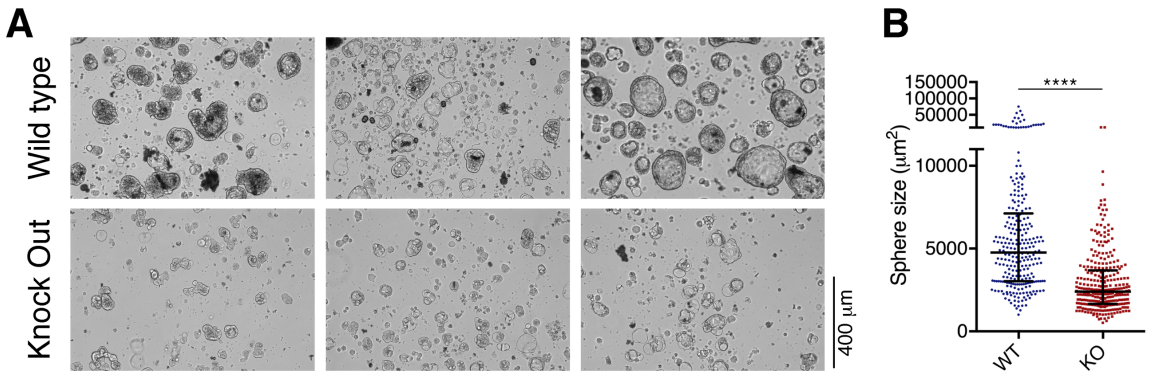


Figure 25: MYRF deletion causes a reduction in spheroids formation. A) Representative images of spheroids obtained from three different WT and three MYRF KO clones. Experiment was performed by plating cells in medium containing methylcellulose and allowing cells to grow for 10 days. B) Quantification of sphere size: around 100 sphere per clone were measured and statistical significance was analyzed using unpaired Mann Whitney t test.

To test if the proliferation defects could emerge in other challenging growth conditions, we performed a subcutaneous transplantation of MYRF clones in nude mice and allowed tumor formation. As shown in Figure 26, MYRF KO reduced the growth *in vivo*, expressed as positivity for the proliferation marker Ki67.

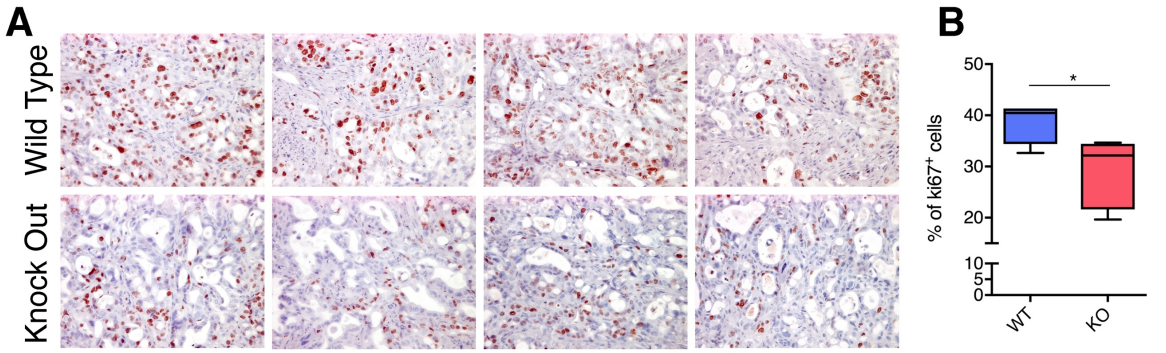


Figure 26: MYRF deletion reduces proliferation *in vivo*. A) Representative images reporting immunohistochemistry staining for the proliferation marker Ki67 in tumors originated through subcutaneous transplantation of MYRF WT and KO clones in nude mice (n=5 per condition). B) Quantification of the percentage of Ki67 tumor positive cells using QuPath software. Statistical significance was analyzed using unpaired two tailed t test.

In conclusion, MYRF KO causes the down-regulation of DNA replication and cell cycle related genes, in turn resulting in a reduction of cell proliferation *in vitro* and *in vivo*.

MYRF KO causes profound alterations in ER morphology and function

Given the association of MYRF up-regulated genes with ontologies related to ER stress and unfolded protein response, we asked whether MYRF KO causes defects in endoplasmic reticulum morphology and function. To test this hypothesis, we fixed the spheroids obtained from WT and KO clones and analyzed their intracellular morphology through electron microscopy.

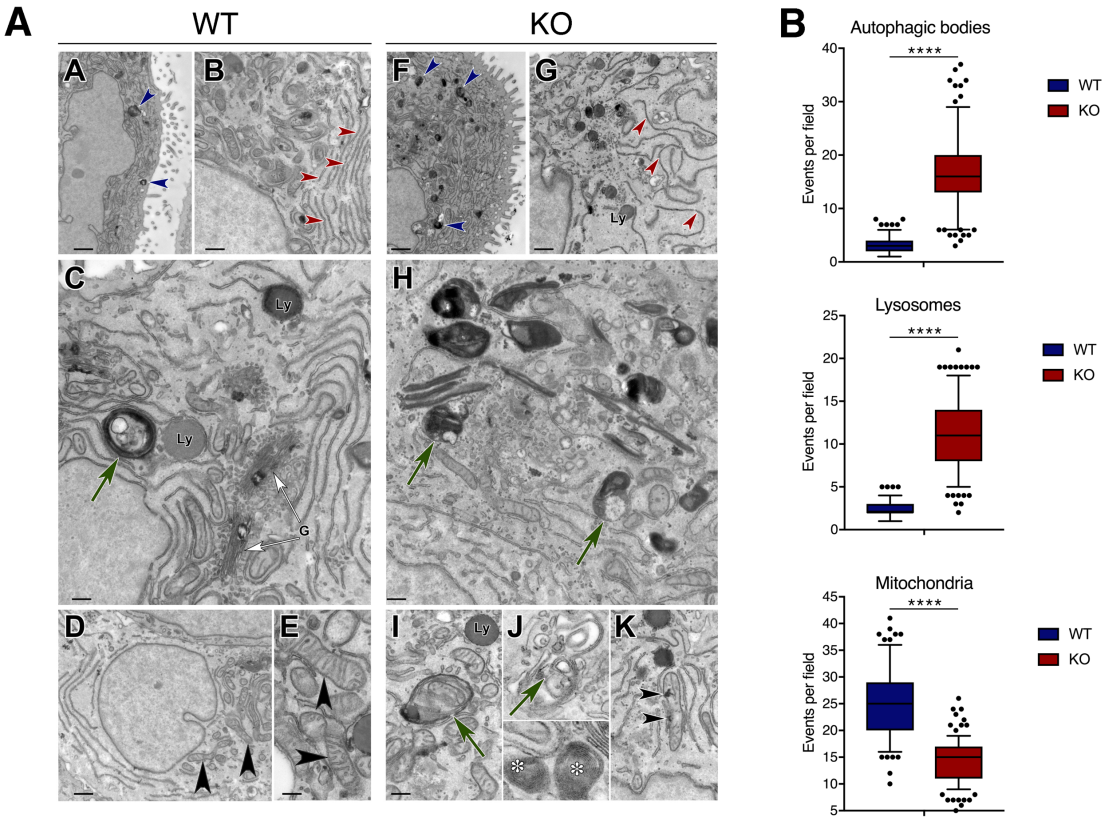


Figure 27: MYRF deletion causes multiple intracellular defects in Low-Grade PDAC cells. A) Representative images of the defects observed in MYRF-KO cells through electron microscopy imaging. Blue arrows indicate peroxisomes that have an higher frequency in MYRF KO cells. Red arrows indicate ER whose quality is dramatically altered in MYRF deleted cells. Green arrows show examples of autophagosome bodies while black arrows highlight mitochondria. B) Quantification of the different defects listed in the graphs: 200 microscopic field coming from three different clones per condition (three WT and three MYRF KO clones) were analyzed. P values were calculated using unpaired two tail t test.

As shown in a representative image (Figure 27A), while WT clones presented a normal ER morphology with thin cisternae well-organized around the nucleus, MYRF KO cells displayed an altered ER morphology, with dysmorphic and enlarged cisternae. Quantification of autophagic bodies and mitochondria number revealed also an increase in lysosomes and autophagic bodies as well as a decrease in mitochondria number and quality (Figure 27B, $p<0,0001$, students T-test with Welch's correction).

In summary, the electron microscopy analysis showed that MYRF KO causes a massive alteration in ER morphology, combined with increased autophagy and mitochondria defects.

To test if the alterations in ER morphology correlate also with an ER dysfunction in the absence of MYRF, we treated CFPAC1 cells with an inducer of ER stress, Thapsigargin, and measured the expression of several components of the three branches of the unfolded protein response (UPR). The genes tested comprise both transcripts already up-regulated in MYRF KO at a basal level, and genes not differentially expressed in the untreated condition, still representing important hubs in the UPR pathway. As shown in Figure 28A, MYRF KO increases UPR response in Low-Grade cells, and this effect spanned the entire UPR pathway (Figure 28B).

Taken together these results indicate that MYRF KO induces strong defects in ER morphology and function, which could easily be the cause for the reduction of proliferation described above.

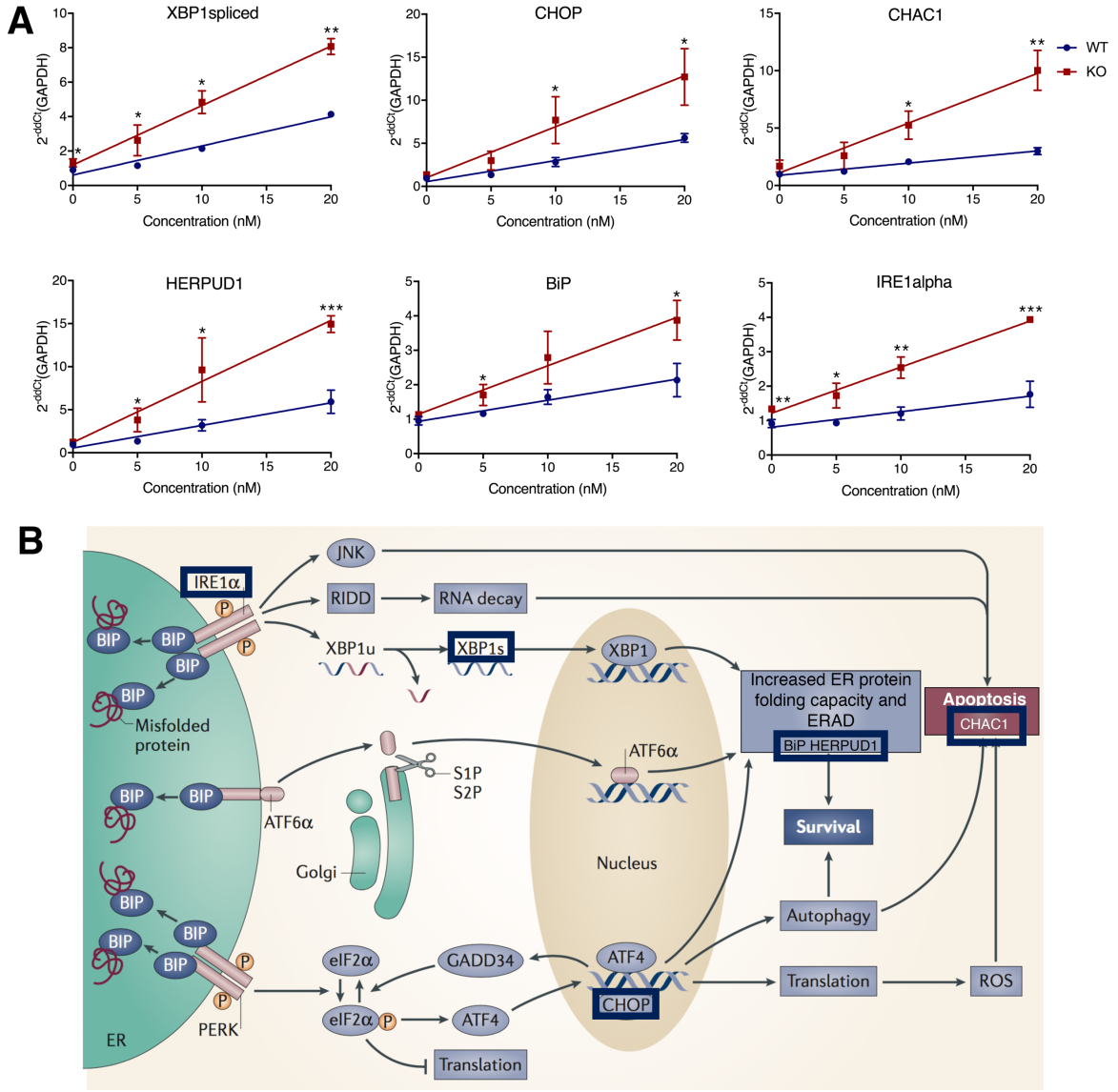


Figure 28: MYRF deletion increases UPR response in Low-Grade PDAC cells. A) Graphs reporting the qPCR results measuring the expression of different components of the UPR pathway after treatment with thapsigargin in CFPAC1 cells. Results are shown as mean of three WT and three MYRF KO clones and p values were calculated using unpaired two tail t test. B) Graphical representation of the three branches of the UPR pathway. Blue boxes highlight the gene tested in the abovementioned experiment. Image adapted from Wang et al.[4]

MYRF KO causes the accumulation of membrane and secreted proteins

We then investigated what could be underling UPR activation in the absence of MYRF. We focused our attention on the up-regulated genes in MYRF KO and observed that the majority were genes encoding for membrane or secreted proteins, as highlighted by a gene ontology analysis focused on the cellular components (Figure 29A). Moreover, when we specifically looked at the genes that were found in these categories, many of them encoded for proteins with complex folding or

those that are highly glycosylated, such as mucins and CEACAMs (Figure 29B), thus proteins that require a heavy processing in the ER. Hence, MYRF could be necessary for controlling the expression of membrane and secreted protein that in its absence are over-expressed, causing an ER overload and ER stress.

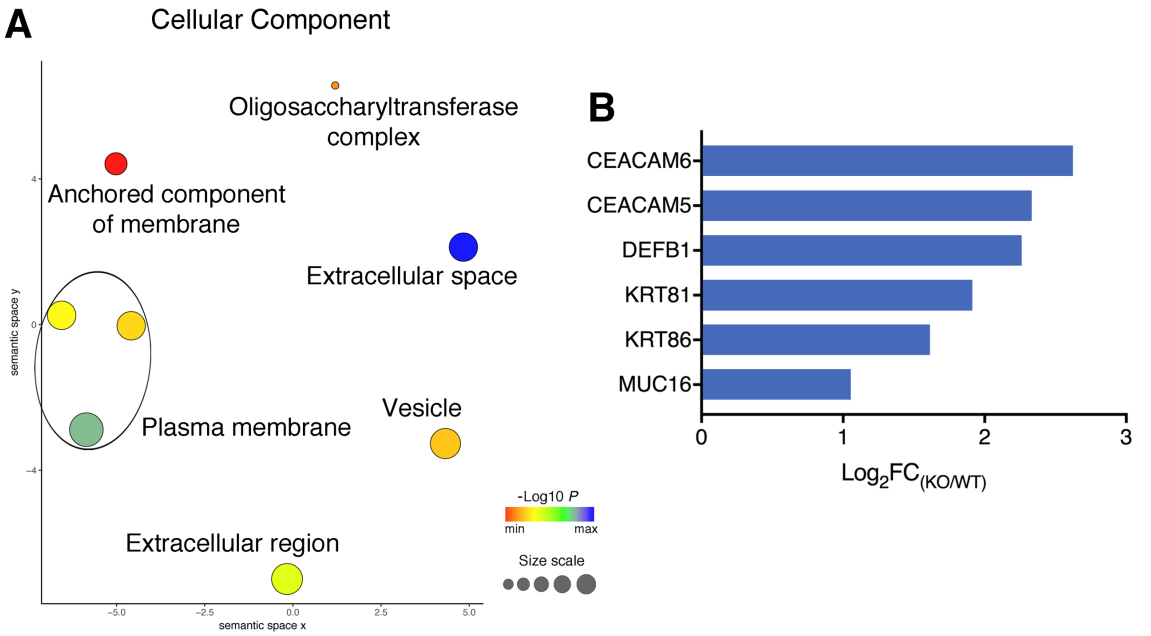


Figure 29: MYRF deletion causes the up-regulation of genes encoding for membrane and secreted proteins. A) Gene ontology analysis focused on cellular components enriched in genes up-regulated upon MYRF KO in CFPAC1 cells. Bubble dimension correlates with generality of the term in the GOA database while bubble color indicates the FDR (color legend on the right). B) Examples of genes up-regualted in MYRF KO cells: fold change from the RNA-seq experiment is reported.

To confirm this hypothesis *in vivo*, we stained the tumors generated from MYRF clones for CEACAM6, one of MYRF targets up-regulated in the 2D-context after MYRF deletion. As shown in Figure 30A, MYRF KO induced the up-regulation of CEACAM6 protein *in vivo*. Furthermore, to test if MYRF KO could cause a general dysregulation of secretion, we stained the tumors generated from MYRF WT and KO cells for Alcian Blue, a stain for all acidic polysaccharides. As illustrated in Figure 30B, MYRF deletion caused a massive increase in secretion, with the accumulation of enlarged duct-like structures filled with secreted glycosylated proteins.

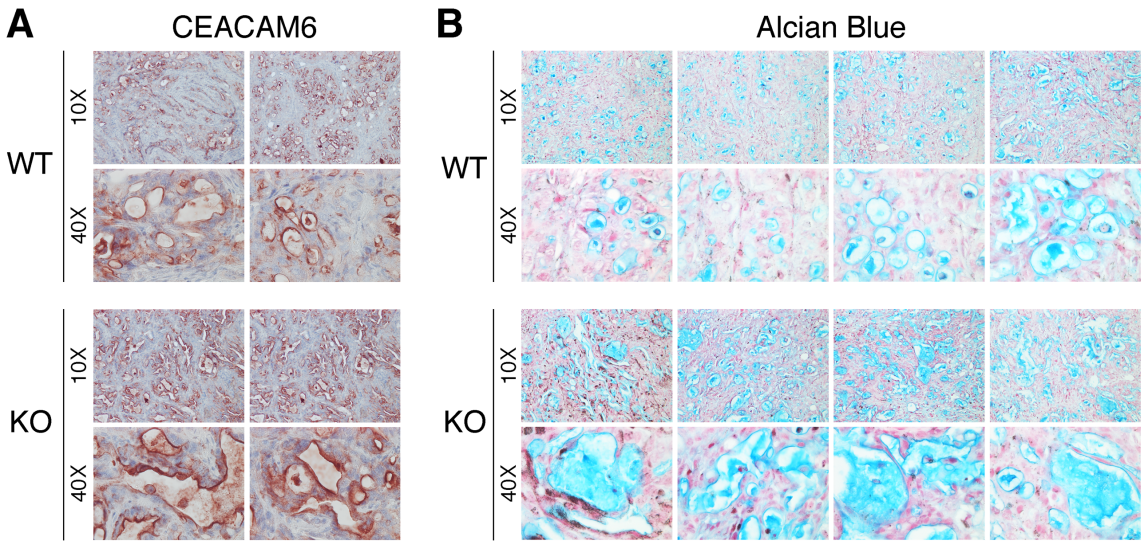


Figure 30: MYRF deletion causes the accumulation of membrane and secreted proteins *in vivo*. A) CEACAM6 IHC staining in tumors generated through subcutaneous transplantations of MYRF WT and KO clones in nude mice. Images from two representative tumors per condition are shown. B) Alcian blue staining in the same tumors described above. Images from four different tumors per condition are shown.

Taken together, these results indicate that MYRF deletion leads to the upregulation of membrane and secreted proteins that usually require a concerted processing effort in the ER, and thus may cause an ER overload, activation of UPR and inhibition of cell proliferation.

Connection between MYRF phenotype and transcriptional activity

Given our proposed model, we asked whether MYRF transcriptional activity could explain the phenotype observed. To this aim, taking into account the previously obtained transcriptomic and epigenetic data, we first investigated if MYRF could act both as a transcriptional activator and as a repressor. This would reinforce a model in which MYRF, when normally processed in the ER, could activate the transcription of replication-related genes, licensing for cell cycle progression, and in parallel dampen down expression of ER-processed genes to avoid ER overload.

Therefore, we plotted the distance between the transcription start site (TSS) of differentially expressed genes in MYRF KO and the summit of the closest MYRF

peak retrieved in ChIP-seq and found that there was not a statistically significant difference in this distance between the groups of up- and down-regulated genes, suggesting that MYRF has the same probability to directly regulate the two groups of genes (Figure 31).

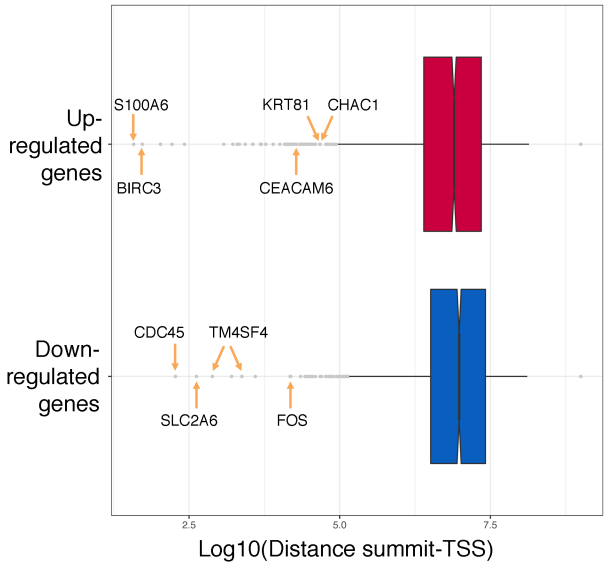


Figure 31: MYRF has an equal probability to act as transcriptional activator and repressor. A) Box plot showing the distance between the transcriptional start site (TSS) of up- and down-regulated genes in MYRF KO cells and the summit of the nearest MYRF peak in CFPAC1 cells. Median and interquartile ranges are indicated.

We then moved to a targeted approach, taking specific regions bound by MYRF in proximity of DEGs and cloning them in a luciferase system in order to verify if MYRF is able to regulate transcription through binding to these sequences.

We first selected a peak located in the promoter of TM4SF4, one of the top down-regulated genes in MYRF KO, thus putatively activated by the transcription factor (all these characteristics are elucidated in the snapshot in Figure 32A). We cloned this sequence upstream to the NanoLuc gene and performed a luciferase assay in the presence of MYRF: as shown in Figure 32B, the promoter region cloned was able to induce the transcription of the NanoLuc and the co-transfection of MYRF

further increased this induction, indicating that MYRF is potently able to activate transcription when binding to this sequence.

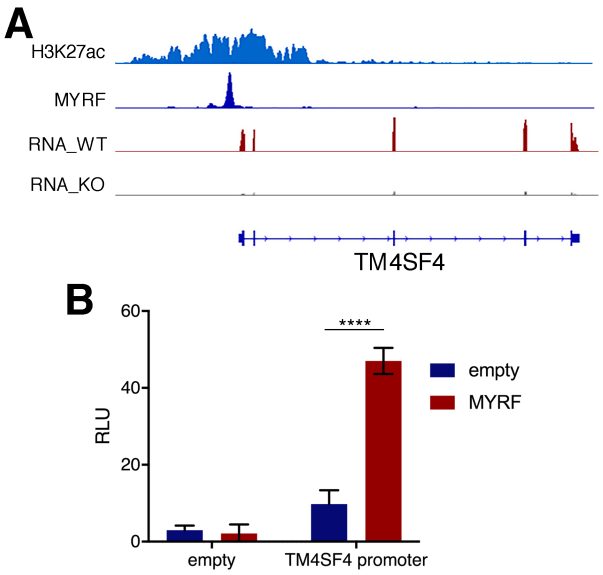


Figure 32: MYRF can directly activate the transcription. A) Snapshot showing MYRF peak retrieved from the ChIP-seq experiment. This peak is located on the promoter of TM4SF4, one of the genes down-regulated upon MYRF deletion, as highlighted by the RNA-seq tracks shown in the lower part on the snapshot. B) Histogram reporting the luciferase activity related to a construct containing the NanoLuc under the control of TF4SF4 promoter. Results are expressed as relative luminescence units (RLU) and significance was measured using unpaired two tail t test.

We then selected regions in which MYRF is putatively acting as a transcriptional repressor, meaning regions found in the proximity of genes that got activated in MYRF KO (snapshots are shown in Figure 33A and 33C). When these regions were used as described before in a luciferase assay, MYRF was not able to influence the transcription levels (Figure 33B and 33D). This suggests that either MYRF is not able to repress transcription *per se*, or the up-regulation of transcripts encoding for secreted proteins is a post-transcriptional event.

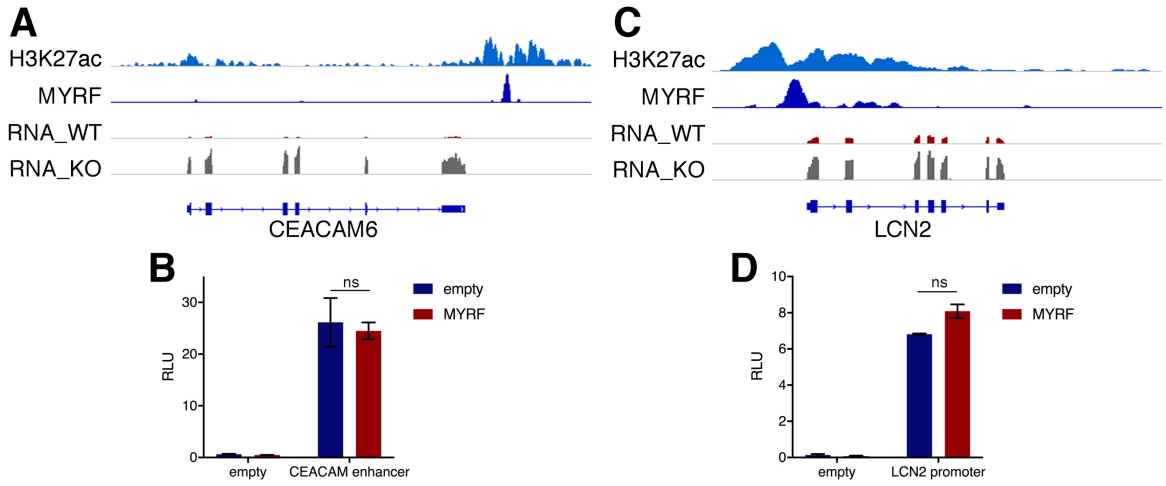


Figure 33: MYRF is not able *per se* to repress transcription in luciferase assay. A) and C) Representative snapshots reporting regions bound by MYRF in proximity of genes that are uo-regulated in MYRF deleted cells. B) and D) Histograms reporting the results of a luciferase assay performed through the cotransfection of a vector containing the NanoLuc genes under the control of the indicated genomic region (corresponding to MYRF bound region shown in snapshors above), a MYRF-overexpressing vector and a Luciferase vector for normalization. Results are expressed as relative luminescence units (RLU) and significance was measured using unpaired two tail t test.

To discern between these two possibilities, we decided to extract nascent RNA from WT and MYRF KO clones and to evaluate if the up-regulation was visible at this level or if it represented a downstream event.

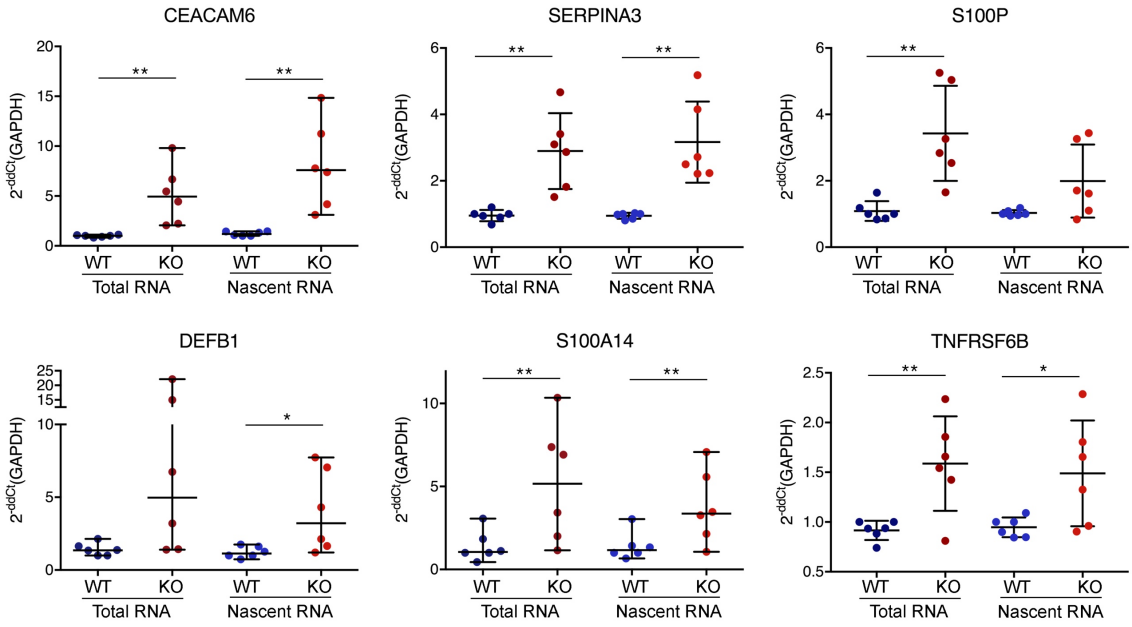


Figure 34: Genes encoding for membrane and secreted proteins are up-regulated at the transcriptional level in MYRF-deleted cells. Scatter plot showing the expression level of different genes encoding for membrane and secreted protein: expression was measured through qPCR both from total RNA and 4sU labelled, nascent RNA. Total and nascent samples were collected in parallel from three WT and three KO clones in two different experiments. Median and range are shown. Significance was measured using unpaired two tail t test.

To this aim, cells were treated with 4-thiouridine (4sU) for 30 minutes to label nascent transcripts that were then purified and analyzed through RT-qPCR. Results were compared with total RNA and reported in Figure 34. All the genes tested were up-regulated at comparable levels upon MYRF KO in the two conditions, indicating that the upregulation of these genes is a transcriptional event.

MYRF creates a feed-forward transcriptional loop with FOS

The RT-qPCR results indicate that the up-regulation of secreted proteins in MYRF KO is not due to a post-transcriptional event. And yet MYRF is nevertheless unable to repress transcription *per se*. This suggests that the transcriptional repression could be achieved through the cooperation with other repressive factors. To test this hypothesis, we selected MYRF peaks located within 100 kb from the TSS of differentially expressed genes and we ran a motif enrichment analysis to unveil which could be the transcription factors recognizing these regions.

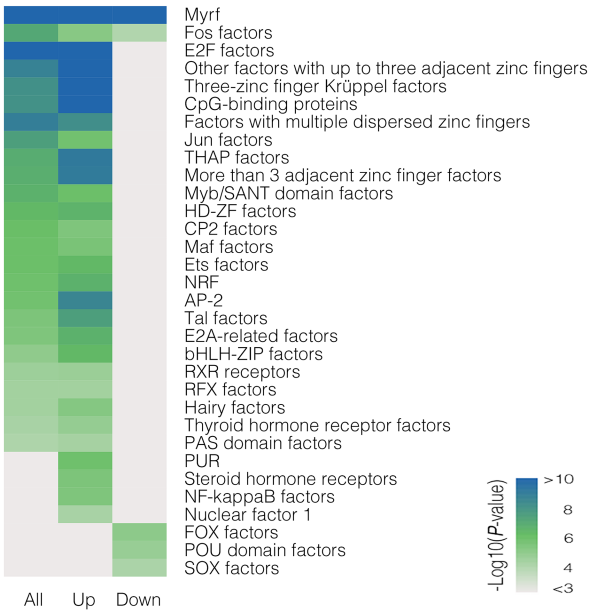
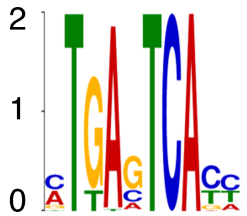


Figure 35: MYRF possible transcriptional partners. Heatmap reporting the results of the motif enrichment analysis ran on regions bound by MYRF within 100 kb from the transcription star site of differentially expressed genes. The analysis was performed either pulling all regions together or clustering them in regions associated to up-regulated and down-regulated genes. Colors indicate the significance (legend on the right).

As illustrated in the heat map in Figure 35, the MYRF motif was enriched in both peaks associated with up-regulated genes and those associated with down-regulated targets, suggesting again a direct role of this transcription factor in regulating its targets.

Furthermore, the second most enriched motif in both the groups was the FOS matrix. This is particularly interesting given that:

i) FOS and AP-1 related motifs were also enriched when the entire set of MYRF bound regions were tested in a motif enrichment analysis (Figure 36);



Sites: 350
E-value: 6,7e-160
Matches: Jun- and Fos-related factors

Figure 36: AP1 matrix is enriched under MYRF bound peaks. Second top match retrieved in MEME *de novo* motif discovery analysis run on MYRF 1000 top peaks. The matrix is associated to the best matching TF motifs through TOMTOM analysis.

ii) FOS and FOSB were down-regulated in MYRF KO cells, both in normal culture conditions and in tumors generated subcutaneously in nude mice (Figure 37A-B);

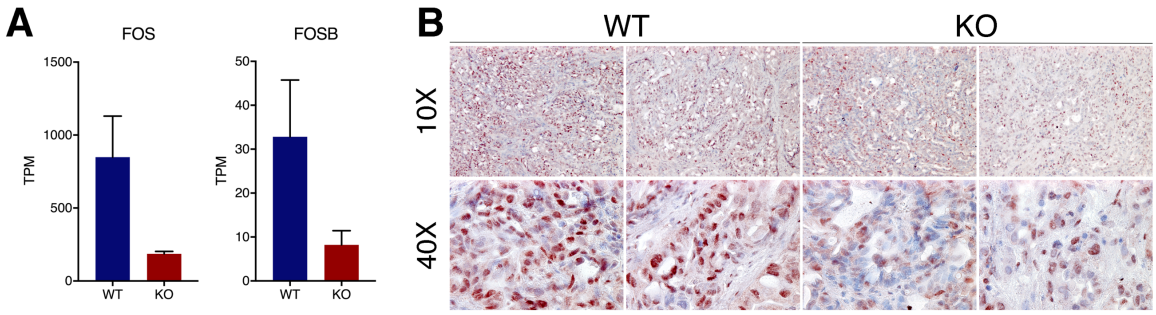


Figure 37: MYRF deletion causes the down regulation of FOS and FOSB *in vitro* and *in vivo*. A) Histograms reporting FOS and FOSB levels of expression retrieved in RNA-seq experiment. Results are expressed as mean of three WT and three MYRF-KO clones. B) FOS immunohistochemistry in xenografted tumors originated from MYRF WT and KO clones, respectively.

iii) MYRF binds the genome in proximity of both FOS and FOSB loci (Figure 38).

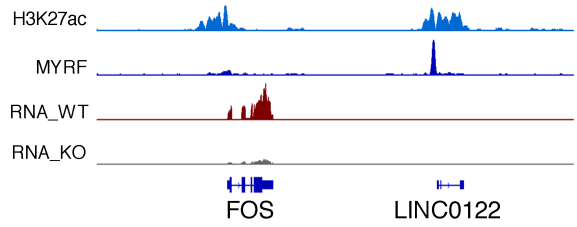


Figure 38: MYRF genomic occupancy in the proximity of FOS gene. Snapshot showing FOS genomic region: MYRF track shows the peak found upstream of FOS TSS while RNA tracks report FOS down-regulation upon MYRF deletion

All these observations point towards an interesting feed-forward loop in which MYRF controls the expression of FOS and FOSB with which it then cooperates in the regulation of its target genes.

To confirm this hypothesis, we first cloned a region bound by MYRF close to FOS gene upstream of NanoLuc and confirmed that MYRF was able to activate transcription when binding this sequence (Figure 39), corroborating the idea that MYRF controls FOS expression in our system.

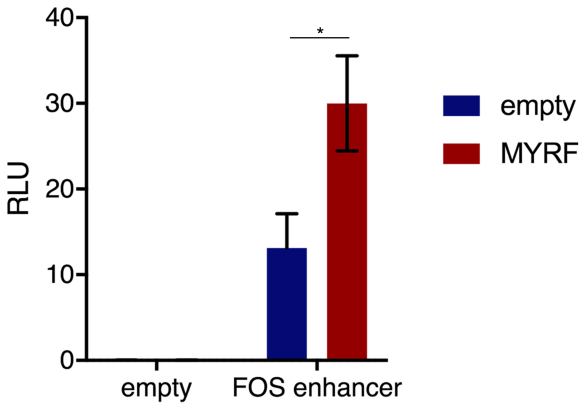


Figure 39: MYRF directly activates FOS transcription. Histogram reporting luciferase assay results obtained using a construct carrying the NanoLuc genes under the control of the genomic region bound by MYRF in the proximity of FOS gene. Luminescence values are normalized on co-transfected firefly luciferase and significance was measured using unpaired two tail t test.

Secondly, we sought to test if MYRF and FOS physically interact. After overexpression in HEK-293 cells, MYRF was immunoprecipitated and shown to co-immunoprecipitate with FOS (Figure 40).

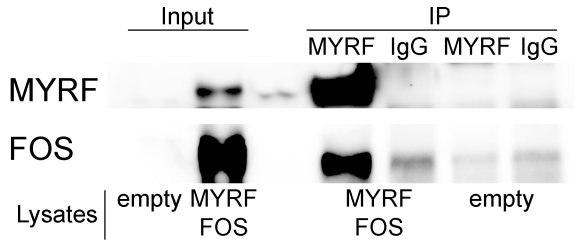


Figure 40: MYRF physically interacts with FOS. WB on immunoprecipitated MYRF shows the recovery of FOS as co-immunoprecipitated with MYRF. MYRF and FOS were co-overexpressed in HEK-293 cells and MYRF was immunoprecipitated using a specific antibody. FOS was then detected in MYRF IP through WB. Controls using species matched IgG in overexpressing cells and MYRF antibody in not-transfected cells are shown in the other lanes.

Given that these indications point to an interaction between MYRF and FOS that could be functional in regulating MYRF targets, we evaluated FOS genomic occupancy in CFPAC1 cells performing ChIP-seq of this transcription factor.

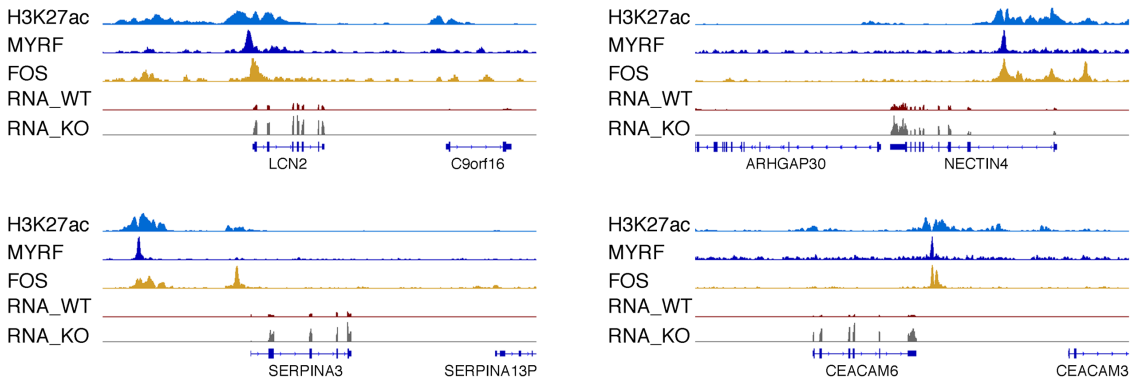


Figure 41: MYRF and FOS frequently colocalize in Low-Grade PDAC cells. Snapshots showing regions that are co-occupied by MYRF and FOS in CFPAC1 cells. RNA tracks indicate that these regions are found in proximity of genes differentially expressed in MYRF deleted cells.

As shown in representative snapshots in Figure 41, FOS and MYRF largely co-localized in Low-Grade cells. Moreover, it is important to note that these regions of co-binding were often located in proximity of genes differentially expressed in MYRF

KO, reinforcing the idea that MYRF and FOS cooperate in the regulation of MYRF target genes.

To further investigate if FOS could regulate MYRF target genes, we depleted MYRF and FOS (either alone or in combination) in Low-Grade CFPAC1 cells using a siRNA approach. Depletion efficiency was evaluated by immunofluorescence and is shown in Figure 42.

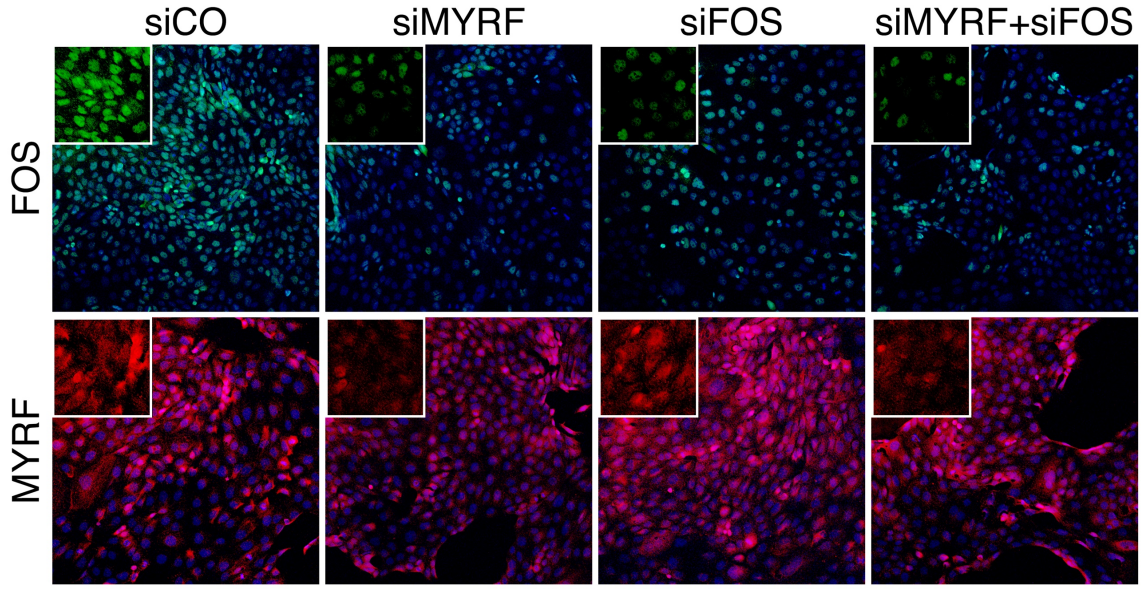


Figure 42: MYRF and FOS depletion in Low-Grade PDAC cells. MYRF and FOS immunofluorescence on CFPAC1 cells transfected either with Control siRNA and/or MYRF and FOS siRNA. The inserts represent a magnified detail showing only green or red channel. Green: FOS, Red: MYRF, Blue: DAPI.

Given that MYRF alone is able to activate but not repress transcription in a luciferase assay, we hypothesize that it needs a coordinated activity of FOS to induce transcriptional repression. We therefore focused our attention on genes up-regulated in MYRF KO and measured the impact of FOS depletion on their expression levels. FOS depletion caused an up-regulation of genes encoding for membrane and secreted proteins (Figure 43). The effect was comparable to what was observed upon MYRF deletion and the combination of MYRF and FOS depletion did not give rise to a cumulative effect, suggesting that the two TFs act on the same pathway.

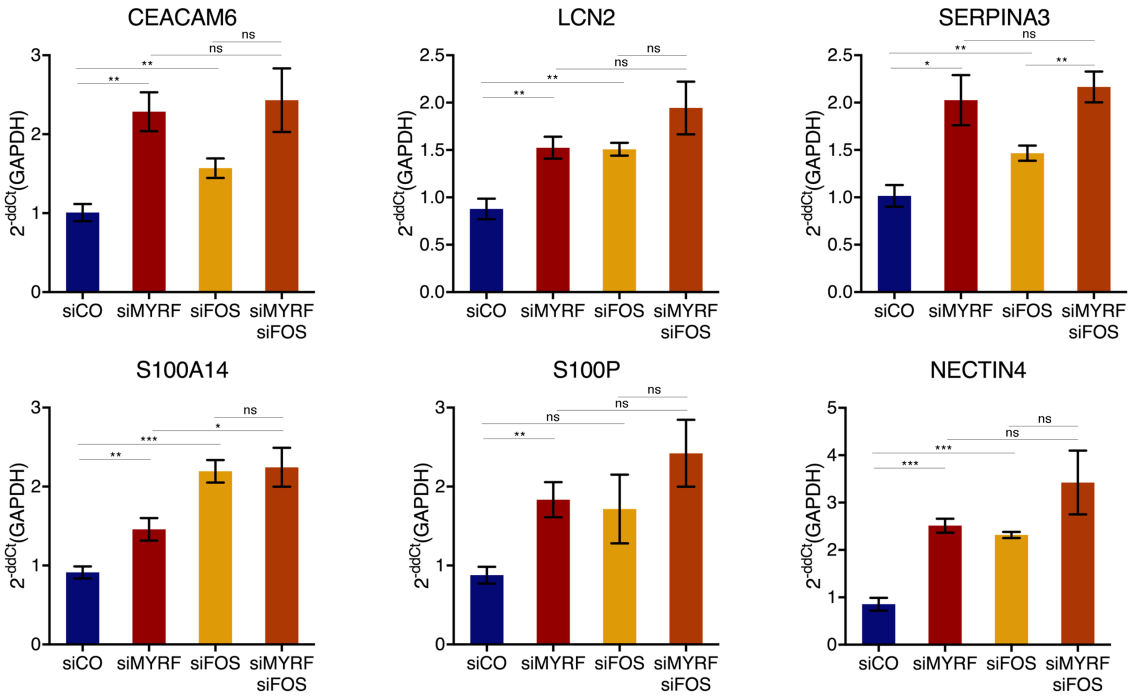


Figure 43: MYRF and FOS coregulate a subset of MYRF targets in Low-Grade PDAC cells. Histograms showing qPCR expression levels of different MYRF targets upon MYRF and/or FOS deletion. Results are expressed as mean of three experiments and p values are derived from an unpaired two tail t test.

To summarize, MYRF directly controls FOS expression in Low-Grade PDAC cells where then these two TFs often colocalize and regulate the expression of a subset of MYRF targets. MYRF therefore creates a feedforward loop in which it regulates FOS that in turn, once activated, collaborates with MYRF in the control of its target genes.

Discussion

MYRF links ER and nuclear functions

Having shown that MYRF is highly expressed in Low-Grade PDAC cell lines, we perform an in depth characterization of MYRF by epigenetic and transcriptomic profiling. We propose a model in which MYRF could serve as a sensor and gatekeeper of ER functionality.

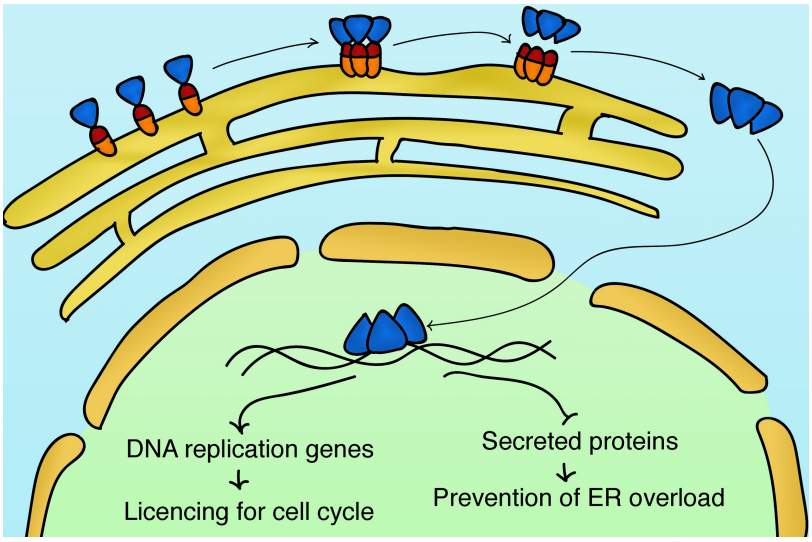


Figure 44: MYRF working model. Graphical representation of MYRF function in highly secretory cancer cells. MYRF works as a sensor of ER functionality, allowing cells with a normal ER function to proceed in the cell cycle and as a gatekeeper of ER functionality, repressing genes that could cause ER overload.

As depicted in Figure 44, in physiological ER conditions, MYRF can trimerize, auto-cleave and translocate into the nucleus where its function is bimodal. On one hand, it induces the transcription of FOS and genes involved in DNA replication, licensing for replication cells with a normal ER functionality. On the other hand, MYRF prevents ER overload and ensures optimal functionality by inhibition of genes encoding both secreted and membrane proteins.

This model is intriguing for several reasons. Firstly, because it explains why MYRF is differentially expressed in Low-Grade PDAC cells and tumors. This sub-

population is indeed characterized by the high expression of matrix and adhesion molecules[1] which have a high cysteine content and are heavily glycosylated, thus are highly depending on proper ER functionality. This dependence on ER functionality suggests that these cells could therefore be highly susceptible to MYRF inhibition and this could be exploited for therapeutic purposes.

It is important to note that protein conservation analysis has shown that MYRF and its homolog MYRFL are the only eukaryotic proteins that possess the ICA domain[43]. Thus, the development of drugs that target this domain could give rise to highly specific compounds, leading to one of the rare cases where it is possible to target a transcription factor.

MYRF transcriptional activity

In the characterization of MYRF function, we managed to obtain the first ChIP-seq of MYRF in human cells. Moreover, thanks to the production of a highly specific homemade antibody, it was also the first ChIP-seq performed on the endogenous protein in any species.

Resent re-analysis of mouse MYRF ChIP-seq revealed a novel trimeric DNA binding motif. Although highly similar to the motif that we retrieve for human MYRF, this motif presents an additional part that the authors considered necessary for the formation of a three-times repeated sequence that could be bound by the three monomers of the MYRF trimer[2], as shown in Figure 45. It is nevertheless important to note that, even increasing the nucleotide window in the *de novo* motif discovery analysis, we could not retrieve a longer motif then the one proposed here (data not shown). Second, and even more conclusively, our “dimeric” motif is sufficient for MYRF binding, as demonstrated by EMSA and luciferase assays. Thus this

supports our intriguing model in which a palindromic sequence is sufficient for the binding a trimeric transcription factor.

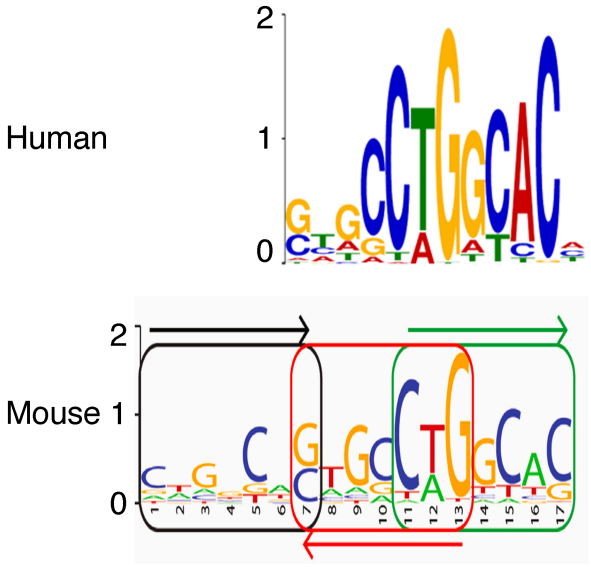


Figure 45: Comparison of MYRF human and mouse DNA binding motif. In the upper part of the figure, the DNA binding motif retrieved in this work for human MYRF is reported. In the lower part, the motif corresponding to mouse MYRF DNA binding motif is shown. This is obtained from a recent reanalysis of the published ChIP-seq of mouse MYRF over-expressed protein[2].

Moreover, the structure of human MYRF DNA binding domain has recently been solved[39] and integration of our data with this discovery could further provide insight into some unanswered questions. Firstly, the authors of this paper found that this domain crystalizes as a trimer but could not exclude that this may be due to a crystallization artifact. Here we corroborate their finding by showing that MYRF binds the DNA as a trimer.

Second, they were not able to co-crystalize MYRF DBD with DNA, possibly due to the usage of a DNA sequence retrieved from published work in mice and thus not perfectly overlapping with the human sequence. Here, we retrieve the MYRF DNA binding motif from genome-wide data and confirme a MYRF-specific interaction in molecular experiments. Further use of this motif could definitely shed light on the precise mechanism of MYRF interaction with DNA, resolving the open question of

how a trimeric TF can contact a palindromic sequence. Indeed, it is still unclear if a single monomer in the trimer contacts the DNA while the others provide conformational stability or if more than one monomer are directly binding DNA.

This clarification could also be interesting in an evolutionary perspective to explain why the fusion between the ICA domain and the DNA binding domain was stabilized in eukaryotic genome. One possibility is that this allows MYRF to work as a sensor of ER function given the strict dependence of ICA catalytic activity on protein folding. Another possibility is that this trimeric structure deriving from ICA mediated trimerization is also necessary for multimeric binding to DNA. Further analysis could help in discerning between the two alternatives.

MYRF function

In Low-Grade PDAC cells, MYRF KO causes a reduction of cell proliferation linked to ER stress and UPR activation.

Ndt80 is the yeast protein that possess the ortholog of MYRF DNA binding domain but lacks the ICA domain. This transcription factor is required for full meiotic completion and its mutants show a reduction in growth[38]. It is then possible that MYRF retained this ancestral capability to regulate the cell cycle, while acquiring the novel function of linking it to ER physiology through the fusion with the ICA domain. This could further provide an explanation of the probable gene transfer of the ICA domain from bacteriophage to eukaryotic genome in evolution.

Both in our hands and in published works[34,36], MYRF cleavage is constitutive and seems to simply depend on tethering in the ER. As such, when MYRF is overexpressed, the ratio between uncleaved and cleaved protein is almost entirely shifted to the cleaved form, indicating that, as soon as three MYRF monomers get in contact in the ER membrane, they self-cleave (data not shown). This means that

MYRF cleavage is constitutive and does not depend on external signals, differentiating MYRF from all the other ER-associated TFs that need either to be activated by upstream pathways (in the case of XBP1 and ATF4) or to be cleaved after UPR activation (ATF6)[48]. This renders MYRF processing unique and provides it with a completely novel function. While the other ER-associated TFs that are part of the UPR get activated only in response to ER stress, MYRF is a sensor of ER physiology and its constitutive cleavage is a measure of ER functionality rather than ER dysregulation.

Moreover, MYRF processing leads to the final formation of two different trimers from the same protein: while the N-terminal part has clear transcriptional activity, it remains to be elucidated if the trimer that is left on the ER membrane has a role in MYRF function. This would add additional peculiarity to MYRF function and worth investigating in the future.

MYRF and ER stress

In Low-Grade PDAC cells, MYRF KO is associated with up-regulation of membrane and secreted proteins and ER stress: this phenotype suggests different possibilities that are particularly interesting in a therapeutic perspective.

First, MYRF KO causes altered ER morphology and increased autophagy. Although we don't have a direct demonstration that in our system autophagy is a consequence of ER stress activation, several lines of evidence indicate that autophagy is activated in UPR in an attempt to attenuate ER stress and promote cell survival[80]. This means that inhibition of autophagy could be therapeutically exploited in MYRF deficient cells or more in general in Low-Grade PDAC in combination with MYRF inhibition.

Another interesting aspect of the MYRF KO phenotype is the massive accumulation of secreted proteins seen *in vivo* in subcutaneous xenografts. It remains to be determined whether these proteins are properly folded or not since this could cause formation of damage-associated molecular patterns (DAMPs). Since DAMPs can be recognized by the immune system and cause immunogenic killing of cancer cells[55], this phenomenon could represent an additional vulnerability of MYRF KO cells exploitable in cancer therapy.

MYRF interplay with FOS

With the intention of unveiling MYRF transcriptional circuitries, this thesis led to the discovery of a feed-forward loop in which MYRF controls FOS and FOSB expression, TFs that in turn contribute to MYRF transcriptional outcome.

FOS and FOSB are member of the Ap1 family of transcription factors. This family comprises the Jun subfamily (with Jun, JunB and JunD) and the FOS subfamily (including FOS, FOSB, FRA1 and FRA2). While Jun proteins can form homodimers, FOS members need to dimerize with Jun proteins and the various dimeric combinations are responsible for the variegate transcriptional responses linked to this family of TFs[81].

Additionally, AP1 proteins have been shown to be also able to act as transcriptional repressors[82]. This means that they could be recruited by MYRF to its target genes and be responsible for the repression of secreted and membrane proteins that are then up-regulated in MYRF KO where FOS and FOSB are down-regulated. This could explain why MYRF overexpression was not sufficient *per se* to repress transcription in luciferase assay, even using regions bound by MYRF in proximity of repressed genes *in vivo*.

Moreover, AP1 proteins have also been linked to ER stress and UPR. Ablation of c-Jun in mouse primary hepatocytes exacerbates ER stress and increases cell death[83]. This phenotype is highly similar to what is observed in MYRF KO cells in which indeed FOS and FOSB are down-regulated. Our model fills a gap, linking FOS activity as transcriptional repressor to the phenotype of ER stress observed in different model system upon its deletion.

In conclusion, we found that MYRF, a TF that until now was associated only with oligodendrocyte differentiation in higher eukaryotes, is expressed in Low-Grade PDAC cells, pointing to broader function of this TF then the restricted role in myelin production. MYRF is indeed expressed in highly secretory cancer cells where it serves a double function: on one hand, licensing for cell cycle cells with a normal ER function; on the other hand, preventing ER overload through inhibition of hyperproduction of secreted and membrane proteins. This opens to the possibility of an interesting dependence of Low-Grade PDAC cells on proper ER functionality that could be exploited to therapeutically target this PDAC sub-population and sheds light on the molecular functions of this novel and poorly characterized transcription factor.

References

1 Diaferia, G. R. *et al.* Dissection of transcriptional and cis-regulatory control of differentiation in human pancreatic cancer. *EMBO J* **35**, 595-617, doi:10.15252/embj.201592404 (2016).

2 Kim, D. *et al.* Homo-trimerization is essential for the transcription factor function of Myrf for oligodendrocyte differentiation. *Nucleic Acids Res* **45**, 5112-5125, doi:10.1093/nar/gkx080 (2017).

3 Walter, P. & Ron, D. The unfolded protein response: from stress pathway to homeostatic regulation. *Science* **334**, 1081-1086, doi:10.1126/science.1209038 (2011).

4 Wang, M. & Kaufman, R. J. The impact of the endoplasmic reticulum protein-folding environment on cancer development. *Nat Rev Cancer* **14**, 581-597, doi:10.1038/nrc3800 (2014).

5 Rahib, L. *et al.* Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res* **74**, 2913-2921, doi:10.1158/0008-5472.CAN-14-0155 (2014).

6 Society, A. C. https://www.cancer.org/cancer/pancreatic-cancer/detection-diagnosis-staging/signs-and-symptoms.html.

7 Collisson, E. A. *et al.* Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat Med* **17**, 500-503, doi:10.1038/nm.2344 (2011).

8 Bailey, P. *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47-52, doi:10.1038/nature16965 (2016).

9 Hruban RH, K. D., Pitman MB. Tumors of the Pancreas. Atlas of Tumor Pathology, 4th Series edn. Armed Forces Institute of Pathology: Washington, DC, 2006. (2006).

10 Whatcott, C. J., Posner, R. G., Von Hoff, D. D. & Han, H. in *Pancreatic Cancer and Tumor Microenvironment* (eds P. J. Grippo & H. G. Munshi) (2012).

11 Naqvi, A. A. T., Hasan, G. M. & Hassan, M. I. Investigating the role of transcription factors of pancreas development in pancreatic cancer. *Pancreatology* **18**, 184-190, doi:10.1016/j.pan.2017.12.013 (2018).

12 Notta, F. *et al.* A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* **538**, 378-382, doi:10.1038/nature19823 (2016).

13 Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27-40, doi:10.1016/j.cell.2010.11.055 (2011).

14 Bradner, J. E., Hnisz, D. & Young, R. A. Transcriptional Addiction in Cancer. *Cell* **168**, 629-643, doi:10.1016/j.cell.2016.12.013 (2017).

15 Spitz, F. & Furlong, E. E. Transcription factors: from enhancer binding to developmental control. *Nature reviews. Genetics* **13**, 613-626, doi:10.1038/nrg3207 (2012).

16 Campos, E. I. & Reinberg, D. Histones: annotating chromatin. *Annu Rev Genet* **43**, 559-599, doi:10.1146/annurev.genet.032608.103928 (2009).

17 Hnisz, D., Day, D. S. & Young, R. A. Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell* **167**, 1188-1200, doi:10.1016/j.cell.2016.10.024 (2016).

18 Riester, M., Stephan-Otto Attolini, C., Downey, R. J., Singer, S. & Michor, F. A differentiation-based phylogeny of cancer subtypes. *PLoS Comput Biol* **6**, e1000777, doi:10.1371/journal.pcbi.1000777 (2010).

19 Li, D. *et al.* Pathway analysis of genome-wide association study data highlights pancreatic development genes as susceptibility factors for pancreatic cancer. *Carcinogenesis* **33**, 1384-1390, doi:10.1093/carcin/bgs151 (2012).
20 Roy, N. *et al.* PDX1 dynamically regulates pancreatic ductal adenocarcinoma initiation and maintenance. *Genes Dev* **30**, 2669-2683, doi:10.1101/gad.291021.116 (2016).
21 Lane, D. & Levine, A. p53 Research: the past thirty years and the next thirty years. *Cold Spring Harb Perspect Biol* **2**, a000893, doi:10.1101/cshperspect.a000893 (2010).
22 Brosh, R. & Rotter, V. When mutants gain new powers: news from the mutant p53 field. *Nat Rev Cancer* **9**, 701-713, doi:10.1038/nrc2693 (2009).
23 Makohon-Moore, A. & Iacobuzio-Donahue, C. A. Pancreatic cancer biology and genetics from an evolutionary perspective. *Nat Rev Cancer* **16**, 553-565, doi:10.1038/nrc.2016.66 (2016).
24 Jones, P. A., Issa, J. P. & Baylin, S. Targeting the cancer epigenome for therapy. *Nature reviews. Genetics* **17**, 630-641, doi:10.1038/nrg.2016.93 (2016).
25 Stathis, A. *et al.* Clinical Response of Carcinomas Harboring the BRD4-NUT Oncoprotein to the Targeted Bromodomain Inhibitor OTX015/MK-8628. *Cancer Discov* **6**, 492-500, doi:10.1158/2159-8290.CD-15-1335 (2016).
26 Delmore, J. E. *et al.* BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **146**, 904-917, doi:10.1016/j.cell.2011.08.017 (2011).
27 Biankin, A. V. *et al.* Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491**, 399-405, doi:10.1038/nature11547 (2012).
28 Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495-501, doi:10.1038/nature14169 (2015).
29 Shain, A. H. *et al.* Convergent structural alterations define SWItch/Sucrose NonFermentable (SWI/SNF) chromatin remodeler as a central tumor suppressive complex in pancreatic cancer. *Proc Natl Acad Sci U S A* **109**, E252-259, doi:10.1073/pnas.1114817109 (2012).
30 Grimmer, M. R. & Costello, J. F. Cancer: Oncogene brought into the loop. *Nature* **529**, 34-35, doi:10.1038/nature16330 (2016).
31 Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110-114, doi:10.1038/nature16490 (2016).
32 Umer, H. M. *et al.* A Significant Regulatory Mutation Burden at a High-Affinity Position of the CTCF Motif in Gastrointestinal Cancers. *Hum Mutat* **37**, 904-913, doi:10.1002/humu.23014 (2016).
33 Emery, B. *et al.* Myelin gene regulatory factor is a critical transcriptional regulator required for CNS myelination. *Cell* **138**, 172-185, doi:10.1016/j.cell.2009.04.031 (2009).
34 Bujalka, H. *et al.* MYRF is a membrane-associated transcription factor that autoproteolytically cleaves to directly activate myelin genes. *PLoS biology* **11**, e1001625, doi:10.1371/journal.pbio.1001625 (2013).
35 Schwarzer, D., Stummeyer, K., Gerardy-Schahn, R. & Muhlenhoff, M. Characterization of a novel intramolecular chaperone domain conserved in endosialidases and other bacteriophage tail spike and fiber proteins. *J Biol Chem* **282**, 2821-2831, doi:10.1074/jbc.M609543200 (2007).
36 Li, Z., Park, Y. & Marcotte, E. M. A Bacteriophage tailspike domain promotes self-cleavage of a human membrane-bound transcription factor, the myelin regulatory factor MYRF. *PLoS biology* **11**, e1001624, doi:10.1371/journal.pbio.1001624 (2013).

37 Rudolph, M. J. & Gergen, J. P. DNA-binding by Ig-fold proteins. *Nat Struct Biol* **8**, 384-386, doi:10.1038/87531 (2001).
38 Pierce, M. *et al.* Sum1 and Ndt80 proteins compete for binding to middle sporulation element sequences that control meiotic gene expression. *Mol Cell Biol* **23**, 4814-4825 (2003).
39 Chen, B., Zhu, Y., Ye, S. & Zhang, R. Structure of the DNA-binding domain of human myelin-gene regulatory factor reveals its potential protein-DNA recognition mode. *J Struct Biol* **203**, 170-178, doi:10.1016/j.jsb.2018.04.007 (2018).
40 Choi, J. O. *et al.* Elucidating the transactivation domain of the pleiotropic transcription factor Myrf. *Scientific reports* **8**, 13075, doi:10.1038/s41598-018-31477-4 (2018).
41 McKenzie, I. A. *et al.* Motor skill learning requires active central myelination. *Science* **346**, 318-322, doi:10.1126/science.1254960 (2014).
42 Koenning, M. *et al.* Myelin gene regulatory factor is required for maintenance of myelin and mature oligodendrocyte identity in the adult CNS. *J Neurosci* **32**, 12528-12542, doi:10.1523/JNEUROSCI.1069-12.2012 (2012).
43 Senoo, H., Wang, H. Y., Araki, T., Williams, J. G. & Fukuzawa, M. An orthologue of the Myelin-gene Regulatory Transcription Factor regulates Dictyostelium prestalk differentiation. *Int J Dev Biol* **56**, 325-332, doi:10.1387/ijdb.120030jw (2012).
44 Russel, S., Frand, A. R. & Ruvkun, G. Regulation of the C. elegans molt by pqn-47. *Developmental biology* **360**, 297-309, doi:10.1016/j.ydbio.2011.09.025 (2011).
45 Meng, J. *et al.* Myrf ER-Bound Transcription Factors Drive C. elegans Synaptic Plasticity via Cleavage-Dependent Nuclear Translocation. *Dev Cell* **41**, 180-194 e187, doi:10.1016/j.devcel.2017.03.022 (2017).
46 Roy, N. & Hebrok, M. Regulation of Cellular Identity in Cancer. *Dev Cell* **35**, 674-684, doi:10.1016/j.devcel.2015.12.001 (2015).
47 Kim, H., Bhattacharya, A. & Qi, L. Endoplasmic reticulum quality control in cancer: Friend or foe. *Semin Cancer Biol* **33**, 25-33, doi:10.1016/j.semcancer.2015.02.003 (2015).
48 Ron, D. & Walter, P. Signal integration in the endoplasmic reticulum unfolded protein response. *Nat Rev Mol Cell Biol* **8**, 519-529, doi:10.1038/nrm2199 (2007).
49 Clarke, H. J., Chambers, J. E., Liniker, E. & Marciniak, S. J. Endoplasmic reticulum stress in malignancy. *Cancer Cell* **25**, 563-573, doi:10.1016/j.ccr.2014.03.015 (2014).
50 Sha, H., He, Y., Yang, L. & Qi, L. Stressed out about obesity: IRE1alpha-XBP1 in metabolic disorders. *Trends Endocrinol Metab* **22**, 374-381, doi:10.1016/j.tem.2011.05.002 (2011).
51 Hollien, J. *et al.* Regulated Ire1-dependent decay of messenger RNAs in mammalian cells. *J Cell Biol* **186**, 323-331, doi:10.1083/jcb.200903014 (2009).
52 Vattem, K. M. & Wek, R. C. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc Natl Acad Sci U S A* **101**, 11269-11274, doi:10.1073/pnas.0400541101 (2004).
53 Bobrovnikova-Marjon, E. *et al.* PERK promotes cancer cell proliferation and tumor growth by limiting oxidative DNA damage. *Oncogene* **29**, 3881-3895, doi:10.1038/onc.2010.153 (2010).
54 Verfaillie, T. *et al.* PERK is required at the ER-mitochondrial contact sites to convey apoptosis after ROS-based ER stress. *Cell Death Differ* **19**, 1880-1891, doi:10.1038/cdd.2012.74 (2012).
55 Obeid, M. *et al.* Calreticulin exposure dictates the immunogenicity of cancer cell death. *Nat Med* **13**, 54-61, doi:10.1038/nm1523 (2007).

56 Papandreou, I. *et al.* Identification of an Ire1alpha endonuclease specific inhibitor with cytotoxic activity against human multiple myeloma. *Blood* **117**, 1311-1314, doi:10.1182/blood-2010-08-303099 (2011).
57 Mahoney, D. J. *et al.* Virus-tumor interactome screen reveals ER stress response can reprogram resistant cancers for oncolytic virus-triggered caspase-2 cell death. *Cancer Cell* **20**, 443-456, doi:10.1016/j.ccr.2011.09.005 (2011).
58 Li, S., Francisco, A. B., Munroe, R. J., Schimenti, J. C. & Long, Q. SEL1L deficiency impairs growth and differentiation of pancreatic epithelial cells. *BMC Dev Biol* **10**, 19, doi:10.1186/1471-213X-10-19 (2010).
59 Gao, Y. *et al.* PERK is required in the adult pancreas and is essential for maintenance of glucose homeostasis. *Mol Cell Biol* **32**, 5129-5139, doi:10.1128/MCB.01009-12 (2012).
60 Lee, A. H., Chu, G. C., Iwakoshi, N. N. & Glimcher, L. H. XBP-1 is required for biogenesis of cellular secretory machinery of exocrine glands. *EMBO J* **24**, 4368-4380, doi:10.1038/sj.emboj.7600903 (2005).
61 Iwawaki, T., Akai, R. & Kohno, K. IRE1alpha disruption causes histological abnormality of exocrine tissues, increase of blood glucose level, and decrease of serum immunoglobulin level. *PLoS One* **5**, e13052, doi:10.1371/journal.pone.0013052 (2010).
62 Genovese, G. *et al.* Synthetic vulnerabilities of mesenchymal subpopulations in pancreatic cancer. *Nature* **542**, 362-366, doi:10.1038/nature21064 (2017).
63 Pommier, A. *et al.* Unresolved endoplasmic reticulum stress engenders immune-resistant, latent pancreatic cancer metastases. *Science* **360**, doi:10.1126/science.aao4908 (2018).
64 Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome research* **24**, 2033-2040, doi:10.1101/gr.177881.114 (2014).
65 Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-181, doi:10.1038/nprot.2014.006 (2014).
66 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
67 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
68 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:gb-2008-9-9-r137 [pii]
10.1186/gb-2008-9-9-r137 (2008).
69 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
70 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
71 Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**, W202-208, doi:10.1093/nar/gkp335 (2009).
72 Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. S. Quantifying similarity between motifs. *Genome Biol* **8**, R24, doi:10.1186/gb-2007-8-2-r24 (2007).
73 Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-578, doi:10.1038/nprot.2012.016 (2012).

74 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).
75 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).
76 McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288-4297, doi:10.1093/nar/gks042 (2012).
77 Wagner, G. P., Kin, K. & Lynch, V. J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* **131**, 281-285, doi:10.1007/s12064-012-0162-3 (2012).
78 Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800, doi:10.1371/journal.pone.0021800 (2011).
79 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
80 Ogata, M. *et al.* Autophagy is activated for cell survival after endoplasmic reticulum stress. *Mol Cell Biol* **26**, 9220-9231, doi:10.1128/MCB.01453-06 (2006).
81 Lopez-Bergami, P., Lau, E. & Ronai, Z. Emerging roles of ATF2 and the dynamic AP1 network in cancer. *Nat Rev Cancer* **10**, 65-76, doi:10.1038/nrc2681 (2010).
82 Mittelstadt, M. L. & Patel, R. C. AP-1 mediated transcriptional repression of matrix metalloproteinase-9 by recruitment of histone deacetylase 1 in response to interferon beta. *PLoS One* **7**, e42152, doi:10.1371/journal.pone.0042152 (2012).
83 Fuest, M. *et al.* The transcription factor c-Jun protects against sustained hepatic endoplasmic reticulum stress thereby promoting hepatocyte survival. *Hepatology* **55**, 408-418, doi:10.1002/hep.24699 (2012).

Acknowledgments

I am usually not sentimental, but I think that this time is definitely worthy an exception since I wouldn't have reached this result without the help of very important people that need to be thanked.

First of all, I want to thank Marta and Fulvio because without them I literally wouldn't be here. The choice of the PhD is not always rational, and the easy roots are often too attractive. Thanks for making me open my eyes on what exists outside the nice but small world of Trieste and on the real possibilities that I could have (and thanks for all the support that you gave me when starting my new life here in Milan).

Several are the people that helped me in this PhD experience. A big thank to Giuseppe, the most patient person I have ever met. With his enormous technical experience and knowledge, he is the perfect tutor for a PhD student. Thanks for the constant support and scientific discussion: I really hope you will reach the position that you deserve.

The PhD period isn't always easy and people that surround you are essential to enjoy good moments and support you in the bad ones. I want to thank all friends and colleagues that I have met in these years and in particular Gabriele, with which I shared this PhD journey. We started together and even if both of us don't like to show our feelings I think that we fitted pretty well in supporting and helping each other with friendships and without rivalry, something not always obvious in the scientific environment. Thanks for all.

And now I cannot avoid giving a big thank to Gioacchino. I started the PhD fascinated by his cleverness and the feeling remained the same throughout the all PhD: thanks for inspiring me and giving me the possibility to express myself at the maximum level.

Finally, I want to thank my family for their constant support. I know how hard it was for them to see me go away to follow my dream, but this has never stopped them from supporting me and making me feel loved. Thank you for making me how I am and allowing me to do what I like most.