DISSERTATION

Submitted to the

Combined Faculties for the Natural Sciences and for Mathematics of the
Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

Yenan Wu

born in: Hainan, China

Oral Examination: 14. 01. 2020

Identification of transcription factors that specifically bind methylated recognition sites

Referees: Prof. Dr. Stefan Wiemann

PD. Dr. Ralf Bischoff

Contents

SUMMARY

Genome-wide expression and methylation studies in patients with pancreatic adenocarcinoma (PDAC) indicate that numerous genes involved in the development of cancer are highly methylated in their promoter regions but are nevertheless strongly transcribed. The mechanisms underlying the relationship between altered DNA methylation and increased transcription, as well as the effects on cancer development remain elusive.

Recent systematic investigations have shown that many transcription factors (TFs) which lack methyl-CpG binding domains (MBDs) can also bind to methylated DNA *in vitro* and *in vivo*. As a consequence, the binding preference of such TFs to mCpG leads to the activation of gene expression, the splicing regulation, and the chromatin remodeling. Based on these observations, it's hypothesized that TFs that specifically bind to highly methylated promoters are involved in the regulation of transcription activity.

With the help of protein microarrays covering 667 DNA binding domains of TFs, the binding patterns of the methylated/unmethylated promoter together with TFs were analysed after incubating the promoter on the TF-microarray. The analysis results showed that the transcription factors NFATc1/2/3 (Nuclear Factor of Activated T cells 1/2/3) of the NFAT family preferentially bound to the methylated promoters. Afterward, NFATc1 was selected for further investigation since it was upregulated in PDAC tissues compared with healthy tissues. The viability, colony, and migration assays indicated that NFATc1 played an oncogenic role in pancreatic cancer cell lines (Panc1 and Miapaca2). To better understand how NFATc1 regulates transcription, mRNA profiling of NFATc1-knockdown cells was used to determine the down-regulated genes. The decreased expression of ALDH1A3 was confirmed further by q-PCR. Next, *in silico* analysis revealed that multiple methylated/unmethylated binding sites of NFATc1 were in the promoter region of ALDH1A3. In addition, luciferase assay and chromatin immunoprecipitation (ChIP) verified that NFATc1 directly regulated the transcription of ALDH1A3. Moreover, the *in vitro* methylation and the *ex vivo* demethylation assay also showed that NFATc1 regulated the transcription of ALDH1A3, though the promoter region was methylated.

In summary, this work reveals that NFATc1 plays an oncogenic role in pancreatic cancer cell lines (Panc1 and Miapaca2) and regulates the transcription of ALDH1A3, though DNA methylation is in its promoter region. The elucidation of the methylation-dependent binding of NFATc1 provides insights for a better understanding of methylation-mediated biological processes.

ZUSAMMENFASSUNG

Genomweite Expressions- und Methylierungsstudien an Patienten mit Pankreas-adenokarzinom (PDAC) deuten darauf hin, dass zahlreiche an der Krebsentstehung beteiligte Gene in ihren Promotorregionen zwar hochgradig methyliert sind aber dennoch stark transkribiert werden. Die Mechanismen, die der Beziehung zwischen veränderter DNA-Methylierung und erhöhter Transkriptionsrate zugrunde liegen sowie die Auswirkungen auf die Krebsentwicklung sind weitgehend unverstanden.

Neuere systematische Untersuchungen haben ergeben, dass viele Transkriptions-faktoren auch ohne Methyl-CpG-Bindungsdomäne (MBD) *in vitro* und *in vivo* an methylierte DNA binden können. Die bevorzugte Bindung an mCpG führt dann zur Aktivierung der Genexpression, der Spleißregulation und zum Umbau des Chromatins. Aufgrund dieser Beobachtungen gehen wir davon aus, dass Transkriptionsfaktoren, die spezifisch an hochmethylierte Promotoren binden, an der Regulation der transkriptionellen Aktivität beteiligt sind.

Mit Hilfe von Protein-Microarrays, die 667 unterschiedliche DNA-Bindungsdomänen von Transkriptionsfaktoren repräsentieren, wurde das Bindungsverhalten an methylierte und nicht methylierte Promotorregionen untersucht. Nach Inkubation der Arrays mit den entsprechenden DNA-Fragmenten zeigte sich, dass die Transkriptionsfaktoren NFATc1/2/3 (**N**ukleärer **F**aktor von **A**ktivierten **T**-Zellen 1/2/3) der NFAT-Familie vorzugsweise an methylierte Promotoren binden. Da NFATc1 in PDAC-Geweben im Vergleich zu gesundem Gewebe hochreguliert wird, wurde NFATc1 für die weiteren Untersuchungen ausgewählt. Der Lebensfähigkeit, der Kolonieassay sowie der Migrationstest zeigten, dass NFATc1 eine onkogene Rolle in den PDAC-Zelllinien Panc1 und Miapaca2 spielt. Um genauer zu verstehen, wie NFATc1 die Transkription reguliert, wurde durch mRNA-Profiling von NFATc1-siRNA-Knockdown-Zellen untersucht, welche Gene herunterreguliert wurden. Die verminderte Expression des dabei entdeckten ALDH1A3 wurde durch q-PCR bestätigt. Zudem ergab eine Sequenzanalyse, dass die Promotorregion von ALDH1A3 mehrfach methylierte/ nicht methylierte Bindungsstellen für NFATc1 aufweist. Durch LuziferaseAssay und Chromatin-Immunpräzipitation (ChIP) konnte die transkriptionelle Regulation durch NFATc1 weiter bestätigt werden. Schließlich zeigte der In-vitro-Methylierungs- und der Ex-vivo-Demethylierungsassay, NFATc1 die Transkription von ALDH1A3 regulierte, obwohl die Promotorregion methyliert war.

Zusammenfassend lässt sich aufgrund der vorgestellten Ergebnisse sagen, dass NFATc1 eine onkogene Rolle in Pankreaskrebszelllinien (Panc1 und Miapaca2) spielt und die Transkription von ALDH1A3 reguliert, obwohl sich die DNA-Methylierung in seiner Promotorregion befindet. Die Aufklärung der methylierungs-abhängigen Bindung von NFATc1 liefert weitere Erkenntnisse zum besseren Verständnis methylierungsvermittelter biologischer Prozesse.

1 Introduction

1.1. What is gene transcription?

1.1.1. The general process of gene transcription

Transcription is the starting point of DNA-based gene expression, in which a segment of DNA is copied into RNA molecule by RNA polymerase [1]. In the process of transcription, RNA polymerase in cooperation with general transcription factor binds to core promoter DNA and initiates the transcription. Specific transcription factors control the rate of transcription and thus make sure genes are expressed at right place, at right time, and in the right amount by recognizing multiple cis-acting regulatory elements including promoters, enhancer, silencers and insulator/boundary elements [2] in a sequence-specific manner. In general, the process of transcription can be divided into initiation, promoter escape, elongation and termination, and DNA stores all the code of controlling the transcription.

Gene transcription is the result of complexed process especially in eukaryotic organisms. As for prokaryotes, genes are organized into operons and transcribed into RNA with the modulation of a single promoter. However, eukaryotic organisms utilize much more complex mechanisms to regulate expression. Firstly, apart from the protein coding genome, around 99% genome context is non-coding part. Recent Encyclopedia of DNA Elements (ENCODE) project have revealed that a big fraction of non-coding genome are *cis*-regulatory elements. Promoters, enhancers, silencers and insulators could fine-tune the gene expression in a cell-type specific and/or biological context dependent manner [3]. Secondly, apart from the genomic aspects, the gene expression can also be modulated by the chromosome structure, and the factor of inter-chromosomal interactions comes to researchers' attention. Thirdly, the cooperativity of TFs and interactions with nucleosomes and/or TF co-factor also have effects on the transcription regulation. Additionally, epigenetic modification, including histone modification also contributes to the regulation of transcription [4]. These are

illustrated in the figure below. However, the complexity of transcription regulation is still remaining elusive.
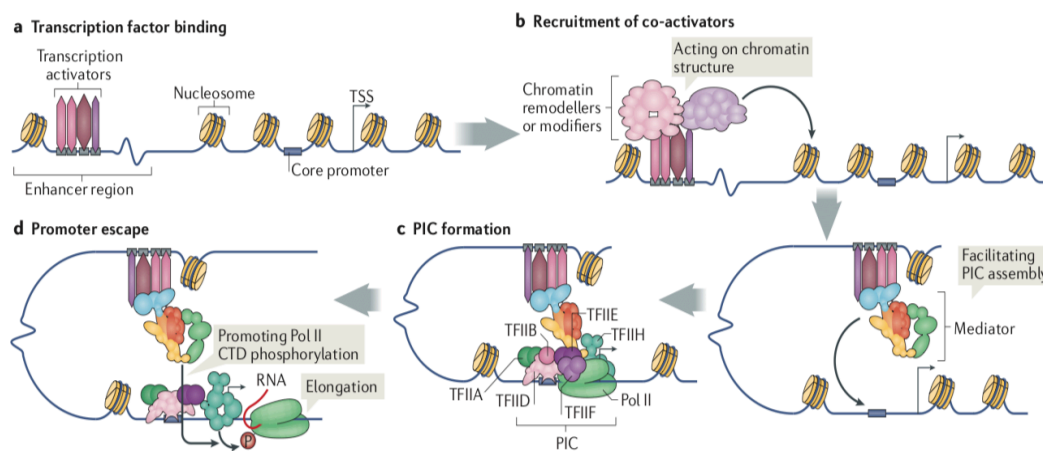
Figure 1.1 The simplified model of transcription initiation.
The simplified model of transcription initiation, figure is modified from [5]. a, transcription starts with TFs binding in the cis-regulatory element's region. b, TFs then recruit other co-activation including chromatin remodelers or modifiers to alter the structure of the chromatin and make it more accessible to the other factors. c, preinitiation complex (PIC) is formed in the core promoter region. PIC includes includes Pol II (12 subunits) and general transcription factors. d, phosphorylated Pol II escapes from the promoter and the elongation step starts.

1.1.2. The transcription *cis*-regulatory element

Coding genes are roughly distributed in 1% genome region. The function of the vast majority of the human genome still remains unclear. ENCODE project, started in 2003, has performed diverse sequencing-based methods to decode the human genome especially the non-coding sequence. Inspiringly, they have identified that over 80% genome participate in transcription-associated biochemical events in the cell line they've studied [6]. Non-coding DNA sequences that are located in or near one coding-gene region and required for proper spatiotemporal expression through binding TFs, are called cis-regulatory sequences [7]. As *cis*-regulatory elements store the information to control when, where and how the gene is expressed, the functional characterization of *cis*-regulatory elements including promoter, enhancer, insulator, silencer and super enhancer is essential to better understand how genes transcription is regulated [8].

In eukaryotic genome, three classes of promoters have been identified based on relative distance to the transcription starting sites (TSS): the core promoter, the proximal promoter and the distal promoter. The core promoter contains TSS, RNA polymerase binding site, general TF binding sites, and spaces between them, which is the essential component for transcription initiation, but not sufficient to mediate the efficient transcription by itself [9]. To achieve efficient transcription, proximal and distal promoters which contain a cluster of sequence-specific TF binding sites are required. The proximal promoter is normally located around 250bp upstream of TSS and the distal promoter resides further upstream in the same strand as TSS. The activity of these elements depends on the relative location and orientation [2].

Another type of regulatory elements containing multiple specific TF binding sites to greatly increase transcription rates are called enhancers. Enhancers are normally accommodated several kilobases or even hundreds of kilobases (kb) upstream or downstream of the target gene. At a distance, the enhancers regulate the transcription in a DNA-looping model, that a DNA loop brings activator proteins bound to distant enhancer elements into protein complexes which are associated with promoter-proximal *cis*-acting sequences [10]. In human genome, there are approximately 1 million enhancers that activate transcription in a tissue-specific manner, which can be explained by two hypotheses: one is that certain activator proteins are dominantly or exclusively expressed in specific cell types; the other explanation is that tissue-specific repressor-silencer complex is formed to block binding of TFs into the enhancer region [11].

Other regulatory elements include insulator, silencer, and super enhancer. Insulator, also name boundary element, is a class of regulatory element containing clustered binding sites for sequence-specific DNA binding proteins which mediate intra and inter-chromosomal interactions. As an inter-chromosomal interaction mediation, it blocks the cross talk between enhancers and promoter of the neighboring genes. On the other hand, it acts as a barrier to protect the active gene from the heterochromatin [12]. In contrary to the enhancers, silencers prevent genes from being expressed and are categorized into two classes: For classical silencers, the gene is actively repressed when the silencers interfere binding with general transcription factor (GTF) binding. For non-classical silencers, the mechanism is rather complex through which gene

expressions are repressed by silencers in a way of recruiting the transcription repressor to disrupt the transcription process [13]. Super enhancers are newly identified regulatory elements that control the expression of crucial genes which can determine the cell entity and cell state. Super enhancers are a cluster of enhancers and densely bound by the high levels of transcription factors and mediators. As a result, the transcription is enhanced with abundant transcripts of target genes. Furthermore, expression of genes associated with super-enhancers is particularly sensitive to perturbations, which may facilitate cell state transitions. Indeed, this has been confirmed indeed in the case of cancer. Super enhancers are bound by a cluster of key oncogenes or transcription factors which can determine the tumor cell phenotype [14]. The disease-associated DNA variants are highly enriched in the super enhancers of the disease relevant cells. Thus, the information of super enhancers can provide biomarkers for therapeutic targets [15].

1.2   What are the transcription factors (TFs)?

1.2.1 The overview of TFs

In human genome, with the updated knowledge, around 1639 known transcription factors have been identified [16]. Transcription factors are multi-domain proteins capable of binding DNA in a sequence-specific manner and regulating transcription [17] [18]. DNA binding domains (DBDs) of TFs mainly exert the function of DNA binding [19]. Additionally, most human TFs also contain additional protein domains including Trans-Activating Domain (TAD) which can interact with other proteins or protein complex [20], and Signal Sensing Domain (SSD) which senses external signal and in turn regulates the gene expression activity [20]. It is summarized that 78 TFs contain multiple homotypic or heterotypic DBDs, 713 TFs contain C2H2 zinc finger arrays, and 779 TFs contain only a single DBD . Based on the types of DNA binding domains, TFs are categorized into different families. In a numbering order, the top TF family are listed as below: C2H2-ZFs (747 TFs), Homeodomains (196 TFs), bHLH (108 TFs), bZIP (54 TFs), Forkhead (49 TFs), nuclear hormone receptor (46 TFs), HMG/Sox (30 TFs), and ETS (27 TFs). Additionally, 69 of TFs in this updated list are grouped as ''unknown family'' due to the lack of a canonical DBD [16]. Functionally, TF can be classified into 2 classes: The General Transcription Factors (GTFs) and the sequence specific transcription factors. Principally, GTFs mainly bind to the core

promoter region and recruit RNA polymerase II to initiate the transcription. They include transcription factor II A/B/D/E/F/H (TFII A/B/D/E/F/H) [20]. Additionally, the sequence specific TFs which mainly regulate the rate of transcription and are expressed spatially and temporarily.

1.2.2 The motifs of TFs

TFs play a key role in the recognition regulatory DNA, and binding motif of around three-quarters (1211) of human TFs have been identified [16]. Notably, The C2H2-ZF family is the biggest TF family, while hundreds of motifs are still not characterized yet. Additionally, more efforts are needed to identify binding motifs in multiple TFs families, including AT-hook proteins, THAP finger, BED-ZF, and those with unknown DBD [16]. In the case of C2H2-ZF family, it's not available to perform *in vitro* expression concerning large family number, and the proteins of C2H2-ZF family are relatively big and thus difficult for the *in vitro* expression [21]. In summary, within various TFs families, same or similar binding motifs are shared by many TFs and this leads to over 500 specific motif groups [17]. Furthermore, the epigenetic modification of DNA can also result in the alteration of TF binding behavior. As an example, DNA methylation is a regulator of gene expression [22]. The effect of DNA methylation has been analyzed on the binding of 542 human TFs by using methyl-SELEX. In addition to the inhibition-binding of some TFs, the mCpGs can also promote binding capacities in certain TF families, including Homeobox, POU, and NFAT families [23].

Many models have been developed to describe TF-DNA interaction such as degenerate code (IUPAC code) [24], Position Weight Matrix (PWM) model [25], binding energy model [26], transcription factor flexible model [26] and connecting matrix model [27]. Different models show different emphasis on the coverage of TF-DNA binding information. IUPAC code is the simplest model but only describes the consensus sequence to one TF prefers to bind. However, the affinity of the binding site is also needed to be considered. PWM model is introduced to cover this information based on the assumption that each position to TF binding is independent. While dinucleotide interdependency, dimer formation and other factors are also important to the TF binding. Thus, binding energy model, transcription factor flexible model and connecting matrix model are established. In general, degenerate code and PWM model are used in most cases. These two models are illustrated below.

The degenerate code is commonly used and easy to be understood. As TFs can bear variance of DNA sequence at each position, degenerate code is used to represent this variance. The detailed information is shown in the table.

Table 1.1 IUPAC code for nucleotide

Code letter	Bases covered
A	A
T	T
C	C
G	G
R	A, G
Y	C, T
W	A, T
S	C, G
M	A, C
K	G, T
B	C, G, T
D	A, G, T
H	A, C, T
V	A, C, G
N	A, C, G, T

The PWM model is used to describe the consensus sequence and the affinity information collectively. In order to achieve this, a numeric score is assigned. The PWM shows the consensus sequence with the highest score at each position. The score at each nucleotide position is the ratio of the counts of such nucleotide divided by the total counts of all nucleotides. The score for a given DNA consensus sequence is the sum of the scores of all positions for such sequence. As an example, the PWM model for GBX2 is listed below, the consensus sequence of GBX2 is CCAATTAG [23].

Table 1.2 PWM model for GBX2

A	1296	48	**7783**	**7783**	0	0	**7783**	1882
C	**2749**	**4441**	409	0	0	152	0	1608
G	2195	0	201	0	0	97	0	**2795**
T	1544	3343	0	0	**7783**	**7783**	0	1500

To understand this PWM model, as an example, the score for A at the third position is 7783 / (7783+409+201+0) = 0.93. The score for a random sequence is 0.25, and any nucleotide with the score higher than 0.25 has higher affinity than random. The score for the consensus sequence with the sum of highest score at each nucleotide position is (0.35+0.57+0.93+1+1+0.97+1+0.36=6.18), and the score for a random sequence of 8 bases is 0.25*8=2. Any sequence with the score higher than 2 is considered as a binding possibility.

1.2.3 The binding physiology of TFs

TFs bind to DNA in a sequence-specific manner. In terms of protein-DNA interaction, it is either direct or indirect contact. Directly, the amino acids of TFs' side chain interact with the binding sites by hydrogen bond or Van Der Waals forces, and thus allows the discrimination of sequence [28]. The decoding of 120 crystal structures of DNA-protein complexes indicates the hydrogen bonds formed between arginine and guanosine nucleotide and between asparagine glutamine and adenosine nucleotide are more common than the other pairs [29]. Indirectly, TFs bind with the backbone of standard B-form DNA, either with a broader or narrower major groove or being bent. These interactions are also sequence associated [30]. In terms of protein-methylated DNA interaction, the structure analysis of multiple TFs which show the preference for methyl cytosine indicates that the interaction is based on hydrophobic bonds with its 5-methyl group [23]. TF-DNA interactions are not extremely strict on the sequence-specificity. Thus, TFs could bind with a number of closely related DNA motifs, but with differential affinities showing higher specificity than random.

1.2.4 The function of human TFs in genetics and disease

TFs play central roles in biology. They are responsible for decoding the human genome, exerting the function of controlling the gene expression and thus controlling

processes that specify cell types and phenotypes [31]. In the case of cancer, diverse carcinogenesis events are controlled by TFs. As an example, EMT pathway is mainly controlled by a panel of transcription factors including members of the SNAIL, TWIST and ZEB (zinc finger E-box-binding homeobox) families [32]. Furthermore, the variants of DBDs can alter the sequence specificity. For example, the multiple variants of TP53 protein lead to the alteration of protein's activity by changing the protein interactions, reviewed in [33]. Moreover, the variants within a regulatory region control the expression of a TF and ultimately lead to the altered TF function. For instance, the mutation of a TCF7L2-binding site within an enhancer of MYC reduces the expression of MYC, and ultimately results in decreasing risk for tumorigenesis in the colon [34]. Additionally, since the main function of TFs is DNA binding, modification or mutation of regulatory DNA leads to the alteration of TF binding sites, and results in the altered expression pattern and occurrence of disease. As an example, the recurrent somatic mutations in the TERT promoter in specific types of human cancers lead to enhanced expression of telomerase[35].

1.3 Biological characteristics of pancreatic cancer

1.3.1 The overview of pancreatic cancer

The overall five-year survival rate for people with pancreatic cancer is 9% according to the latest statistics of pancreatic cancer provided by American National Cancer SEER in 2019. About half (52%) of patients are diagnosed at a distant stage which is a late stage, and the five-year survival rate of this stage is only 3%. Surgery, radiation therapy, and chemotherapy are treatment options that may extend survival and/or relieve symptoms, but still, pancreatic cancer is often considered as incurable [36].

1.3.1.1 Epidemiology and risk factors of PDAC

Pancreatic cancer refers to pancreatic ductal adenocarcinoma (PDAC) since around 95% of pancreatic cancers are classified as exocrine tumors, of which PDAC is the most common malignant neoplasms. For the anatomy of the pancreas, the pancreas mainly consists of 5 different cell types including acinar cells which secrete digestive enzyme, ductal cells which secret bicarbonate, endocrine islets which secrete hormone, centro-acinar cells which are the geographically bridge between acinar and ductal cells, quiescent and activated pancreatic stellate cells which can express matrix molecules inducing pancreatic fibrosis [37]. Other less common exocrine tumors are

acinar cell carcinoma, intraductal papillary mucinous neoplasms (IPMNs) and mucinous cystic neoplasms (MCNS). The pancreatic endocrine neoplasms are called Pancreatic Neuroendocrine Tumors (PNETs), which are rare pancreatic neoplasms [38]. PNETs grow slower compared with the other neoplasms and are often best treated surgically.

As for risk factors of pancreatic cancer, age, living habit, family inheritance, and medical history are mainly involved. Age is the determining reason for the occurrence of pancreatic cancer. Healthy lifestyle is important for preventing the pancreatic cancer. Tobacco smokers have a twofold to a threefold higher risk of neoplasm incidence than non-smokers. Some nutritional and dietary factors, including high intake of (saturated) fats, low intake of vegetables and fruits and consumption of red and processed meats, are also associated risks. Additionally, heavy alcohol consumption is also considered. In terms of disease association, obesity, chronic pancreatitis, and diabetes mellitus are risk factors for pancreatic cancer. Furthermore, a family history of pancreatic cancer and certain genetic syndromes are also the risk factors [36].

1.3.1.2 Diagnosis and therapy of PDAC

Pancreatic tumors are often considered as incurable. One of the reasons is the lack of reliable diagnostic biomarkers for early detection. Serum cancer antigen 19–9 (CA19-9) is the most widely used biomarker to monitor disease progression, recurrence and/or therapy response. However, CA19-9 can't be used for early pancreatic cancer diagnosis because of the low sensitivity and specificity [39]. Additionally, a panel of mutated genes such as mutated KARAS or mutated TP53 from circulating tumor DNA could be applied as a non-invasive early diagnostic test since this panel has been detected at the time of diagnosis in 43% of patients with localized disease [40]. Furthermore, the increasing circulating branched chain amino acids can be indicators for the pancreatic cancer of early-stage [41]. Moreover, in terms of diagnosis, clinically, imaging must be conducted. MultiDetector CT (MDCT) or even more sensitive method MRI is the commonly used imaging method. However, pathology is still the gold standard for the diagnosis when the pancreatic cancer is suspected.

Over the past decades, little improvements have been made to increase the 5-year survival rate. For the treatment, patients are classified according to the tumor stage and performance status based on Eastern Cooperative Oncology Group (ECOG) score. Different treatments for patients are suggested accordingly. In terms of resectable tumors, the surgery is followed by adjuvant chemotherapy (gemcitabine plus capecitabine). This combination chemotherapy showed an improved 5-year overall survival rate over gemcitabine monotherapy [42]. As for borderline resectable and locally advanced tumors, and unresectable tumors, neoadjuvant chemotherapy (e.g. FOLFIRINOX or radiochemotherapy) is applied. In the case of patients in the metastatic stage, FOLFIRINOX and nab-paclitaxel–gemcitabine are standard treatment options when patients are in good performance status. Besides the standard of care, the newer strategies for the treatment are embraced. Pathway-specific targeted therapies have failed since signaling in pancreatic cancer is complex and other strategies (e.g. therapies targeting and modulating the stroma and tumor microenvironment, immunotherapies, novel biomarker in early stage, multimodal imaging, and identification of druggable key signaling hubs) are in the development which might reboot the pancreatic treatment options in the future [43].

1.3.1.3   The microenvironment of pancreatic cancer
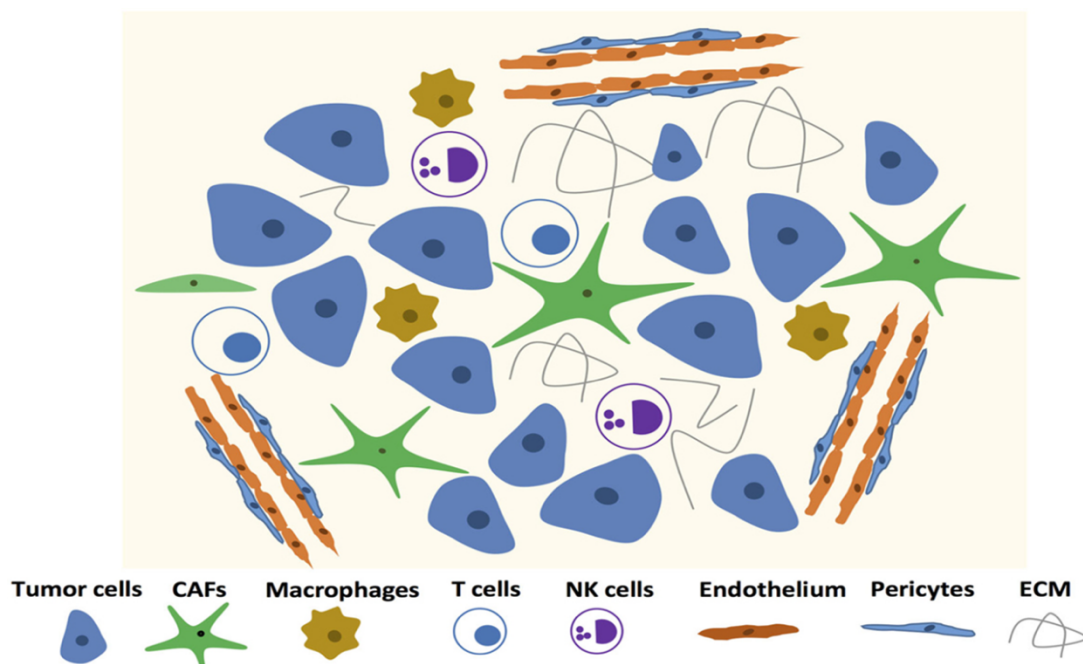
Figure 1.2 Tumor Microenvironment

It mainly includes fibroblasts, immune cells, tumor cells, and extracellular matrix (ECM). Tumor cells generate signals including dysfunction and death of immune cells, the immune cells are a source of signals promoting the activation of cancer associated fibroblasts which secrets ECM and promote the growth of tumor cells. (Figure was modified from [44].)

When PDAC is studied, the microenvironment, and heterogeneity are always challenging topics. The complexity of interactions between the microenvironment and cancer cells still remains elusive. While this complexity also provides new opportunities for the exploration of therapy.

What is the microenvironment of pancreatic cancer? In the site of lesions, the tumor cells interact with the stroma, and the tumor cells promote the development of the stroma. On the other hand, the stroma supports the growth and metastasis of the tumor cells, blocks the treatment of the tumor cells, and has an effect on the chemo-resistance as well as recurrence of the disease. The stroma consists of the cellular component (pancreatic fibroblasts, pancreatic stellate cells, vascular cells, infiltrating inflammatory/immune cells, endothelial cells and neuronal cells) and the acellular component (extracellular matrix proteins including collagenous and non-collagenous proteins; soluble proteins such as cytokines and growth factors). In terms of cellular component, the cells mainly have 2 origins, one type is called cells of hematopoietic origin which mainly are immune cells that arise in the bone marrow, the other type is called cells of mesenchymal origin which mainly include fibroblasts, myofibroblasts, mesenchymal stem cells, adipocytes and endothelial cells [45].

Pancreatic fibrosis is a characteristic pathological feature of PDAC and chronic pancreatitis that can disrupt pancreatic exocrine and endocrine function irreversibly [46]. Pancreas cancer environment is crucial for the development of pancreatic fibrosis. In the 1990s, in human pancreas, people firstly identified a new cell type displaying the expression of α- smooth muscle actin (α-SMA) and synthesis of ECM proteins which are the characteristics of activated myofibroblasts in the human pancreas [47]. The quiescent pancreatic stellate cells can be activated by soluble factors including IL-1, IL6, TNF-α, TGF-β1 and activin1 which are released from platelets, macrophages, pancreatic acinar cells and endothelial cells in the inflamed pancreas, and the activation can also be achieved by *in vitro* cell culture [48]. The activated stellate cells

lose the retinoid-containing fat droplets, change to myofibroblast-like cells expressing α-SMA, produce large amounts of the ECM proteins as well as synthesize cytokines such as TGF-β1, activin A and IL-1[48]. In the clinic, studies have shown an association between the activated PSCs and poor clinical outcome [38]. These *in vitro* and *in vivo* evidences are accumulated and indicating that the activation of PSCs plays a key role in the development of stromal cancer compartment and pancreatic fibrosis. Thus, targeting signaling pathways that play a crucial role in PSC activation is a promising therapeutic strategy that may inhibits the pancreatic fibrosis.

The crosstalk between the immune cells (T cells which mostly are CD4$^+$ positive, myeloid-derived suppressor cells, macrophages, and mast cells) and the tumor cells also becomes a focus of attention. In summary, the immune infiltrating indicates an immunosuppressive phenotype. The monocytes from the bone marrow are recruited and transformed into macrophages by the PDAC tumor cells. Macrophage polarization is differentially controlled by a complex network of signaling effectors. There are normally two types of activated macrophages, one is M1 which are classically activated by interferon γ (IFNγ) with lipopolysaccharide (LPS) or tumor necrosis factor (TNF), the other is M2 type which includes M2a subtype alternatively activated by IL-4 and IL-13, M2b subtype activated by immune complex and Toll-like receptor (TLR) ligands, and M2c subtype activated by IL-10 and glucocorticoids [49]. It was reported that M1^{high}/M2^{low} correlated significantly with longer survival period, and M1/M2 ratio can be used as a prognosticator [50]. In general, M1 macrophages are so-called tumor suppressive macrophages which normally promote T-helper-1 (Th1) responses, and M2 macrophages are namely tumor-promoting macrophages that promote T-helper-1 (Th1) responses. The phenotype state of macrophage is changing over the development of a tumor. When a tumor is initiated, M1 type is mainly abundant in chronic inflammatory sites. During the progress of the cancer, M2 type takes the dominant part [51].

In addition to harboring carcinoma cells, immune cells, and cancer-associated fibroblast, the tumor microenvironment also comprises cancer stem cells (CSCs). Even though CSCs is a minority subpopulation within tumors, it plays an important role in initiating and driving the tumor growth. CSCs have been identified in 1994, and they have been revealed in the pancreas in 2007. Afterward, the knowledge of CSC's

function is expanded. One of the explanations for the presence of CSCs is that they are induced by epithelial to mesenchymal transition (EMT) program. The EMT is induced by various signals from the tumor stroma including the ECM and secreted factors. Such signaling induces the expression of a certain set of EMT-TFs including SNAIL, TWIST, and ZEB family of transcription factors and thus changes the phenotype of the carcinoma cells. CSCs share phenotypes and characteristics of normal stem cells, including self-renewal which is an indicator of the tumor cell heterogeneity and maintains the CSC pool. Additionally, CSCs also divide and generate more CSCs or the multiple cell lineages within the tumor bulk. They are normally located near vessels in a perivascular niche or a hypoxia niche, indicating that they have close communication with the other component in the microenvironment [52]. Moreover, CSCs show the intrinsic resistance to the traditional therapeutics because of the quiescent state of stem cells e.g. chemotherapy and radiotherapy [53].

The complex microenvironment of PDAC and stroma-rich characteristic of PDAC bring challenges of illustrating the signaling of PDAC, targeting precisely, and thus escape from the chemotherapy of PDAC. The extent of tumor heterogeneity remains poorly understood. With the development of omics technology, the multi-omics level data of tumor is applied to decode the genomics and epigenomics of tumor. The complexity of omics data reveals the inter-tumor heterogeneity which is indicated by similar histology but various genomic aberrations. Additionally, the diverse tumor-harboring cell components including tumor or non-tumor cells and acellular components show the diverse and versatile intro-tumor heterogeneity. The observed intro and inter heterogeneity in tumor challenge precision medicine. To tackle the heterogeneity problem, a more comprehensive assessment is needed, which includes a better characterization of tumor samples with spatial and temporal variations to monitor overtime and identify a reliable target, the consideration of multi-omics data to identify the intra and inter-tumor heterogeneity, and involvement of non-invasive biomarker monitor and *in vivo* functional characterizations [54].

1.3.2 The general molecular biology of pancreatic cancer

Genetic and epigenetic alterations have an influence on tumor progression and chemotherapy resistance. During pancreatic cancer development, histologically, from early precursor lesions (PanIN 1-3 lesions) consequently to PDAC, the histological investigations are accompanied by infiltrating immune cells, increasing desmoplastic stromal response, multiple involved signaling pathways [55].
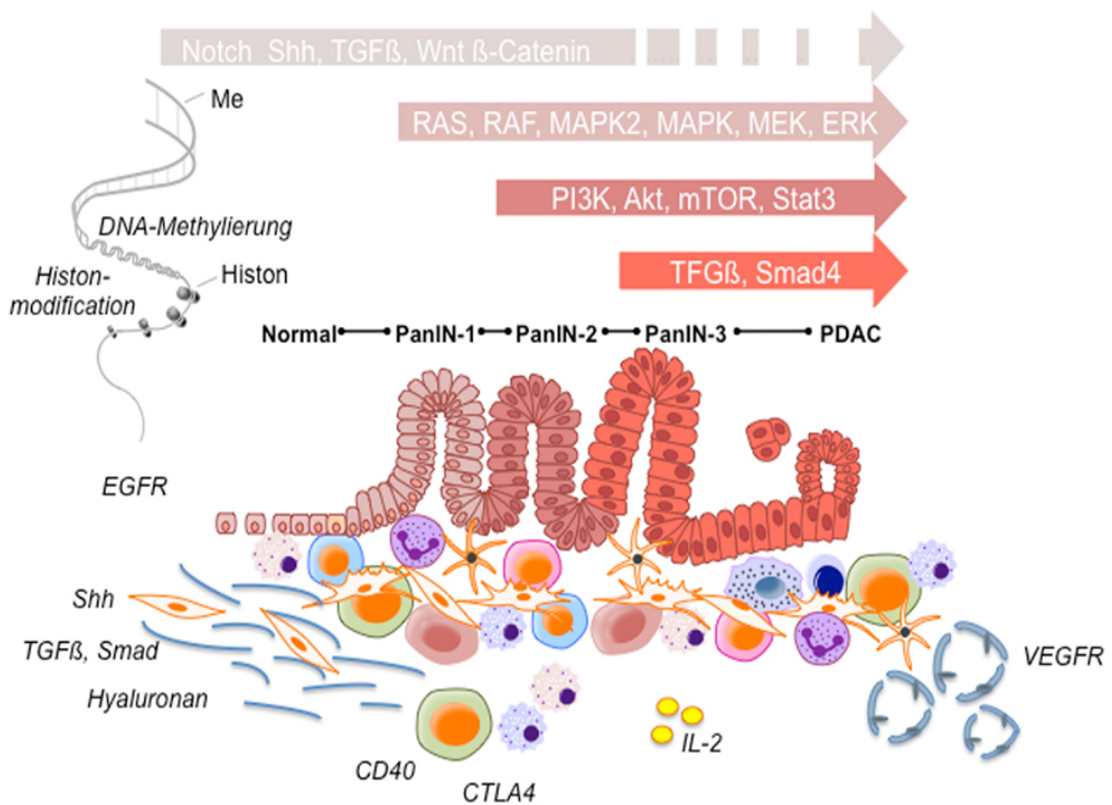
Figure 1.3 Development of pancreatic cancer

A variety of signaling pathways are involved in multiple stages of pancreatic tumorigenesis from PanINs1-3 lesions to PDAC. These histological changes are accompanied by a dynamic tumor microenvironment and multiple genetic and epigenetic signaling alterations which lead to an immune response against the tumor, tumor cell proliferation, angiogenesis and invasion of tumor cells. Plot was modified from [55].

1.3.2.1   The genetic regulation of pancreatic cancer

In terms of genomic events, KRAS, CDKN2A, TP53 and SMAD4 are the commonly four altered genes in pancreatic cancer but the encoded proteins are no favorable drug targets. Point mutations of individual genes are important for the molecular pathology of pancreatic cancer. Activating mutations in KRAS are present in over 90% of pancreatic cancers [56]. Inactivating mutations of TP53, CDKN2A and SMAD4 occur

in 50–80% of pancreatic cancers, whereas other genes, including ARID1A, MLL3 and TGFBR2, are mutated in around 10% of tumors [57]. Besides the mutation events, copy number alterations and homozygous deletions also play an important role. Analysis of data from whole-genomic sequencing and copy number variations (CNV) identified new driver genes for pancreatic cancer, such as KDM6A and PREX2. Based on the analysis, genes which are important in pancreatic cancer, including TP53, SMAD4, CDKN2A, ARID1A and ROBO2 are affected by chromosomal rearrangements leading to gene disruption [58]. Furthermore, several gene alteration (RBM10 mutations, BRAF mutations) with prognostic significance have been identified due to the analysis of whole-exome sequencing [59].

Classifying tumors is the first step towards personalized treatment. According to diverse omics data analysis, different classifications of pancreatic cancer have been suggested. The analysis of transcriptomics data revealed four subtypes of pancreatic cancer: Squamous, pancreatic progenitor, immunogenic and aberrantly differentiated endocrine exocrine (ADEX) pancreatic cancer that correlate with histopathological characteristics [60]. Three highly distinct metabolic subtypes of PDAC could be identified by the metabolite profiling: Reduced proliferative capacity, glycolytic and lipogenic [61]. Moreover, PDAC has been classified into four subtypes with potential clinical significance (termed stable, locally rearranged, scattered and unstable) based on the patterns of structural variations [58]. The identified subtypes provide a solid foundation for novel promising therapeutic strategies, targeting pancreatic cancer.

Signaling pathways of pancreatic cancer are complex due to multiple nodes and complex crosstalk. Analysis of 24 pancreatic tumors' transcriptomes with next-generation sequencing-by-synthesis technologies revealed twelve cellular signaling pathways which were altered in 67%-100% of the tumors. These core regulatory processes or pathways are: Apoptosis, DNA damage control, regulation of G1/S phase transition, hedgehog signaling, homophilic cell adhesion, integrin signaling, C-Jun N-terminal kinase signaling, KRAS signaling, regulation of invasion, small GTPase-dependent signaling (other than KRAS), TGF-β signaling and Wnt /Notch signaling [62]. Furthermore, an exome sequencing together with copy number analysis revealed diverse somatic aberrations in genes described traditionally as embryonic regulators of axon guidance. Particularly SLIT/ROBO signaling pathways are also

involved in the carcinogenesis of pancreatic cancers [63]. Moreover, an integrated genomic analysis identified 32 recurrently mutated genes which were enriched into 10 pathways: KRAS, TGF-β, WNT, NOTCH, ROBO/SLIT signaling, G1/S transition, SWI-SNF, chromatin modification, DNA repair and RNA processing [60]. Whole-exome sequencing have identified high-frequency alterations in Wnt signaling, chromatin remodeling, hedgehog signaling, DNA repair and cell cycle processes [59]. All these provide a better understanding of the molecular pathology of pancreatic cancer.

1.3.2.2 The epigenetic regulation of pancreatic cancer

The development of pancreatic cancer is not only attributed by mutations, such as activating mutations of KRAS or inactivation of the tumor suppressor genes TP53 and CDKN2A. More recently, the important role of epigenetic regulations for PDAC became more evident. Epigenetic changes including DNA methylation, histone modification and non-coding RNA alterations can lead to the alterations in gene expression without changing the DNA sequence. These changes result in silencing of important tumor suppressor genes, cell cycle checkpoints, hyperactivation of oncogenes, and growth stimuli.

DNA methylation is a process in which DNA methyltransferases (DNMTs) add a methyl group to the 5' carbon of the cytosine pyrimidine ring. This modification normally occurs in the region of cytosine-guanine dinucleotides (CpGs). The expression level of DNMTs increases in pancreatic cancer, representing a potential therapeutic target [64]. In general, DNMT inhibitors contain two different types, nucleoside analogs such as 5-Azacytidine and 5-aza-2'-dC [65]; non-nucleoside inhibitors such as RG108, SGI-1027, hydralazine [66].
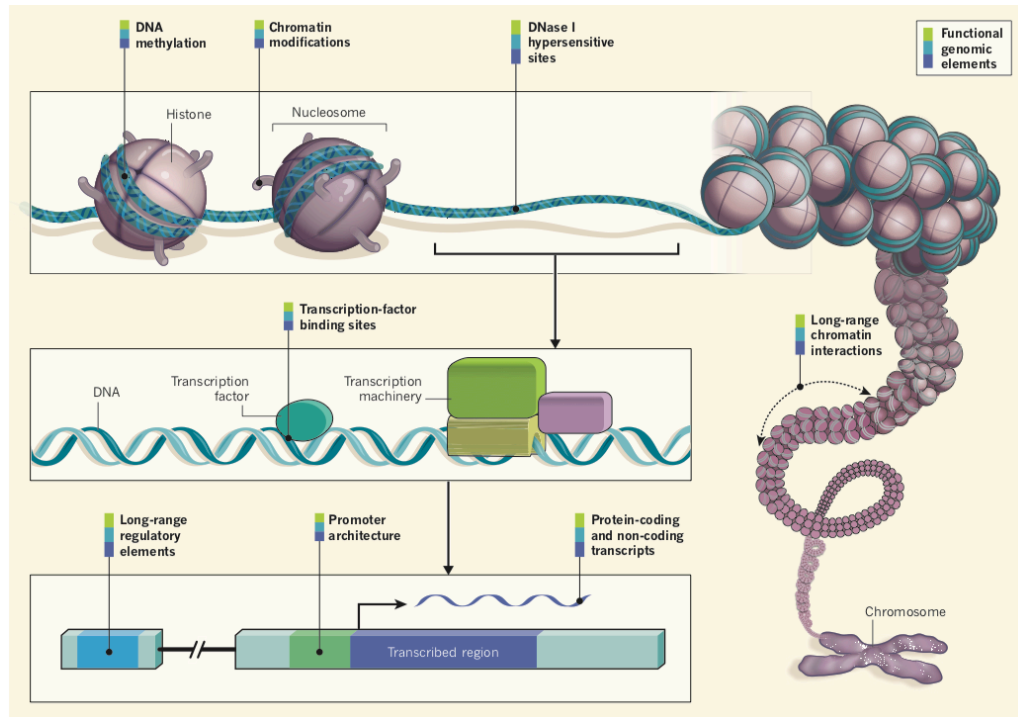
Figure 1.4 Beyond sequence

Concerning the regulation of expression, besides the DNA sequence itself, DNA methylation, chromatin interaction, and histone modification are involved in the regulation of transcription. Figure was adapted from [4].

It is known that aberrant DNA methylation is associated with transcription regulation. Generally, in the early stage of tumor, DNA hypermethylation of gene promoter CpGs downregulates the transcription of tumor suppressing genes [67]. While, DNA in cancer cells was hypomethylated in the late stage, this hypomethylation is related with genomic instability, activation of oncogene and activation of silenced transposable sequences [68, 69]. Regionally, DNA methylation in gene bodies may facilitate transcription, elongation and regulate splicing events [70].

Histone modification is more complex than DNA methylation. Nucleosome which are the fundamental unit of chromatin, is 145–147 base pairs of DNA-wrapped around a histone protein octamer (dimers of H2A, H2B, H3 and H4) [71]. DNA-associated histone modifications with different degrees of modification (e.g., mono-, di-, and trimethylation) include phosphorylation, methylation, acetylation, ubiquitination and sumoylation [72]. Modifications of histones determine the configuration of the chromatin, and transcriptionally open or closed structured chromatin changes

accessibility of transcription-associated protein with DNA, thus consequently influence the transcription of gene [67].

Histone acetyltransferases (HATs) transfer an acetyl group from the acetyl coenzyme A to the ε-amino group of lysine such as CREBBP (cAMP response element-binding protein), p300 and p300-CBP-associated factor (P/CAF). Thus, the histone gets acetylated and the gene transcription is activated. The reverse reaction is called deacetylation which is performed by histone deacetylases (HDACs). The deacetylation leads to closed chromatin [73].

In terms of histone methylation, the lysine, arginine, and histidine residues at the amino acid side chains are modified by histone methyltransferases (HMTs) and histone demethylases (HDMs). The well-characterized histone activation markers are H3K4me1/2/3, H3K9me1, H3K27me1 and H3K79me1/2/3. In contrast, the histone inactivation markers are H3K9me2/3, H3K27me2/3, H3K79me3 [74].

Another class of histone modification is histone ubiquitination. It influences the pancreatic homeostasis and pancreatic cancer tumorigenesis. This type of histone modification mainly occurs on H2A and H2B which alters nucleosome dynamics and indirectly affect other histone modifications. DNA damage can trigger H2A ubiquitination which is followed by the recruitment of downstream DNA damage repair proteins. H2B ubiquitination is crucial for the double strand break repair and deubiquitylation plays a role in transcription-coupled repair [75].

Besides the core canonical histones (H3, H4, H2A, H2B), mentioned previously, there are multiple variants of each core canonical histone representing specific structural and functional features. For example, the histone variants H3.3 and H2A.Z may be involved in histone exchange or nucleosome eviction during chromatin remodeling [76]. Additionally, the histone variants could also indicate the potential accessible regions of the genome which are potentially the regulatory element regions [77] [78] [79].

The integrative analyses of chromatin immunoprecipitation-sequencing (ChIP-seq) on multiple histone modifications, RNA-sequencing (RNA-seq), and DNA methylation profiling highlight the epigenomic landscapes for PDAC subtypes and thus, suggested

that epigenetic changes have an effect on carcinogenesis as well as heterogeneity. A clear association between the methylome, transcriptome, and chromatin state-based clustering could be shown. Moreover, the landscapes are regulated by different membrane-to-nucleus pathways and can predict the patient outcomes regarding the relative aggressiveness and survival potential [80].

All these observations demonstrate that genetic regulations together with epigenetic modifications contribute to the development of pancreatic cancer.

1.4 Why identification of TFs that specifically bind methylated recognition sites?

1.4.1 The impact of methylation in the carcinogenesis of pancreatic cancer

Alteration of methylation at promoter region is associated with expression dysregulation of many cancer-regulated genes and thus the aberrant DNA methylation is considered as one of the driving factors of carcinogenesis [81] [82]. In pancreatic cancer, the most frequently methylated promoters are promoters of APC (50%), BRCA1(46%), p16INK4a (35%), p15INK4b (35%), RARβ (35%), and p73 (33%), and the methylation of at least one gene mentioned above have showed in 94% of the pancreatic cancer cases [83]. Furthermore, CDKN2A [84] is inactivated by hypermethylation. The whole genome methylation profiling reveals that the aberrant gene methylations are enriched in the following gene loci and signaling pathways: TGF-β, WNT, integrin, Slit Guidance Ligand (SLIT)-Roundabout Guidance Receptor (ROBO) signaling, cell adhesion and stellate cell activation pathways. Additionally, the SLIT-ROBO, ITGA2 and MET signaling is epigenetically deregulated [85]. Moreover, the hypermethylation in the promoter region of miRNA that are repressed in cancer is a common mechanism in all human tumors [86].

In the microenvironment, various cell types contain almost identical copy of the genome, while the transcriptome and phenotype of each is unique. This can't be explained only by the genomic information of the cells but by integrative information of the epigenome. Different cell events in microenvironment, such as PSC activation,

macrophage polarization, stem cell renewal, differentiation, and the dynamic plasticity of the cells come to the focus of the attention. Since the reprograming is changing with response to the various stimuli in the microenvironment, the epigenetic factors including histone modification, DNA methylation and non-coding RNAs, regulate the gene expression at an appropriate level, time and place. Finally, this process determines cell fates, such as the disease phenotype or non-disease state. As reported previously [80], the epigenomic landscapes highlight the PDAC heterogeneity, predict survival, and classify the molecular pathobiology subtypes. This provide an insight into a new therapy strategy which focus on developing new drugs that target the factors involved in the epigenetic regulation.

1.4.2 Integrative analysis of expression and methylation

Most eukaryotic genes are controlled at the level of transcription. In PDAC, an integrated analysis of DNA methylation and mRNA expression data indicated that 98 genes that were silenced by DNA methylation, might exert important roles in the development of cancer [87]. Moreover, the fact that hypomethylated VAV1 promoter leads to increased gene expression, suggests roles of DNA hypomethylation in upregulation of gene expression [88]. The methylation of promoter is generally considered to repress gene expression since methylated promoter could block the binding of TFs and elicit subsequent higher binding affinity of MBDs with methylated DNA sequences [67]. Recent studies indicate the impact of cytosine methylation on the DNA binding of TFs that lack the MBDs is not always negatively [89]. Yu Liu et. al [90] systematically revealed the promoter CpG methylation-dependency of transcription regulation across 21 cancer types. The mechanism of methylation-dependent TF regulation is complex. It is not as simple as we thought before that promoter methylation is the potent repressor for the gene expression.

1.4.3 The impact of methylation on the binding of TFs

TFs act as the readers of the DNA, while how about the DNA with modification? Traditionally, only TFs with a methyl-CpG (mCpG)-binding domain (MBD) is able to recognize and bind methylated CpG dinucleotides [91] [92]. In mammals, the MBD-TF family includes 5 known proteins, MeCP2 (methyl-CpG-binding protein 2), MBD1(methyl-CpG binding domain 1), MBD2, MBD3 and MBD4. In this family, only

MBD3 does not bind to methylated DNA, while all the other MBD proteins bind to methylated DNA in a non-sequence-specific way [93] [94]. However, over the past decades, more and more evidences show that some TFs lacking MBDs are also capable of binding with methylated DNA [89]. In the past 5 years, the methylated motifs of many TFs lacking MBDs have been identified systematically by diverse approaches. Data from protein microarray suggests that 47 TF proteins lacking MBDs could bind the methylated CpG sites, and some of them could recognize both methylated and non-methylated with distinct sequence similarities [89]. Data from methyl-SELEX reveals that CpG methylation influences binding of most TFs to DNA, while the effects are either negative or positive to regulate gene expression. Global analysis with 542 TFs indicates that many developmentally important TFs (homeodomain, POU and NFAT proteins) prefer to bind with mCpG sites. This conclusion facilitates future analysis of the role of DNA methylation on cell differentiation, chromatin reprogramming, and transcription regulation [23]. Furthermore, more and more evidences have showed that TFs could recognize the methylated motif and regulate diverse biological events including activation of gene expression [95], recruiting additional TFs and cofactors [96] and regulating splicing regulation events {Maunakea, 2013 #1335}. Systematical integrative analysis pipeline has revealed the methylation-dependent regulation machinery in 21 cancer types. The coupled CpG sites and TFs could also stratify the cancer subtypes with different prognoses [90]. Taken together, the knowledge on how TFs regulate gene expression has been widely expanded. Traditionally, it's accepted that TFs bind non-methylated DNA motifs in open chromatin regions because the DNA methylation could block such interactions; while nowadays the new scenario of TFs-DNA interactions uncovers that DNA methylation alternatively provide new binding sites for TFs lacking MBDs.

1.5 Strategies for studying the regulation of TFs

1.5.1 Experimental approaches for identification of binding specificities of TFs

To answer the question that what are the biological consequences upon the binding of TFs to methylated DNA, the first step is to identify the binding motif of TFs. The binding property can be studied by a wide variety of techniques both *in vitro* and *in vivo.* Different methods are selected based on the purpose of study.

Electrophoretic Mobility Shift Assay (EMSA) [97] is used to determine the affinity of TFs to different DNA fragments with known sequences. This assay is a gel-based and comply with the different mobility between large molecules with various size and charge. For example, if the TF could recognize and bind the DNA molecule, the DNA-TF complex will move slower than the DNA or the TF molecule alone because of the bigger size. Thus, by comparing the mobility of TF-incubated DNA with the DNA molecule control alone, one can determine the TF's preference to the tested DNA. However, it cannot be used to identify the novel binding motif of TFs, and the throughput of this assay is quite low. In terms of the high throughput *in vitro* assay, SELEX and PBM are illustrated below in detail.

Systematic Evolution of Ligands by Exponential Enrichment (SELEX) [98] is used to study the TF binding sites by massively incorporating paralleled sequencings. Initially, random oligos are synthesized and double-stranded. Afterward, the TFs are immobilized and incubated with DNA with the length limit of 25bp. For each round of panning, the selected DNA is amplified by PCR and used for next cycle of selection to reduce the background. Thus, the specific DNA that TFs bind can be identified by sequencing upon the selected DNA get enriched after several rounds of panning. In recent years, the application of SELEX is broadened, it can also be used to identify the methylated-motif of TFs [23]. Moreover, the TF cooperativity is important for DNA binding specificity and effector function. In order to study this topic, SELEX method is developed into CAP-SELEX to measure the multimers [99]. However, the bias introduced by PCR is difficult to be avoid in this method since the very low affinity sites may be lost in the PCR amplification process, or the high affinity sites become saturated in the PCR amplification process.

Protein binding microarray (PBM) is an alternative high throughput method to model the TF binding specificities. It is either based on the DNA microarray or protein microarray. As for the DNA microarray [100], the double stranded DNA is generated on the array by recruiting the microarray probes followed by labeling with Cy3-dUTP. The array covers all combinations of 8-mer sequences for selection. Then, the *in vitro* expressed and purified TFs are incubated on the microarray. Consequently, the

fluorophore-conjugated antibody against the common tag of all TFs is incubated on the array to detect the TF binding. Finally, the signal intensity is extracted by reading the fluorescence from binding complex between double-strand DNA and protein antibody. The signal intensity is used to quantify binding preference of one TF toward various DNA motifs. In terms of protein microarray [101], TFs are expressed, purified and spotted on the array. DNA motifs with lengths ranging from 6-34 bp are selected and synthesized with conjugated fluorophores. Subsequently, DNA is incubated with TF microarray and followed with image scanning and signal acquisition. Besides the application in identifying DNA motif, PBM can be also used to the identification of methylated DNA motif and multimers [89]. However, DNA-microarray-based PBM has been hampered with short probe length (8 bp) in generating the PWM; While protein microarray-based PBM is limitedly employed in exploring DNA motifs with known sequences and potential binding sites.

All methods mentioned above are used for *in vitro* assays. In order to identify the binding specificity of TFs in different cell types and even in tissues, the combination of chromatin immunoprecipitation (ChIP) and next generation high-throughput sequencing (NGS), is widely used to study the genome-wide binding region of TFs. Briefly, TFs are firstly cross-linked to chromatin DNA after formaldehyde treatment and further pulled down using its specific antibody together with its bound DNA, and the non-specific bound DNAs are washed away. Next, this pull-down DNA is de-crosslinked and ready for sequencing. However, there are several concerns that might limit wide application of ChIP-seq. Firstly, from the experiment's point of view, high-quality ChIP result is highly dependent on antibody quality, while ChIP-grade antibodies are not easily available. Secondly, ChIP excludes the equilibrium binding due to the use of cross-linkers. Thirdly, ChIP also includes the indirect binding. Fourthly, binding is highly influenced by chromatin state, i.e. most TFs prefer to bind in the region of open chromatin. Finally, the biases in the sequence content of the genome could also interfere the ChIP-seq result [16].

Moreover, many other methods are designed to identify the TF specificity. The table listed below summarizes all the methods.

	Method	Description	Features			
			Capability of de novo **motif discovery** (approx. length in base pairs with high information content)	Identifies **genomic binding locations** of a TF	Can measure effect of **CpG Methylation**	Can measure **cooperative binding and/or multimers**
High-throughput (In Vitro Methods)	Protein Binding Microarray (PBM)	A GST-tagged TF is bound to a glass slide that has ~41,000 spots of short immobilized DNA sequences. Fluorescence-based detection of bound spots and k-mer enrichment analysis yields motifs.	✔ (< 12 bp)	✗	✔ Methyl-PBM	✔
	Bacterial one-hybrid	TF binding sites are selected in bacterial cells from a randomized library that is cloned in front of selectable marker genes. Can be reversed to select proteins able to bind a constant DNA sequence using a library of variant protein sequences.	✔ (< 14 bp)	✗	✗	✗
	SELEX-based methods	Systematic evolution of ligands through exponential enrichment (SELEX) involves adding TFs to a DNA pool containing many randomized sequences and selecting for binding in multiple rounds. Related methods include HT-SELEX, SELEX-seq, and Bind-n-Seq. Selection can be performed using affinity tags, or molecular trapping on a microfluidic platform (SMiLE-seq).	✔ (< 25 bp)	✗	✔ Methyl-HT-SELEX	✔ CAP-SELEX ✔ SMiLE-seq
Mid-throughput (In Vitro Methods)	DAP-seq	Single step SELEX using a library of fragmented genomic sequences. Sequence diversity is less than HT-SELEX, but genomic sequences that have co-evolved with the TF are included.	✔ Limited by skewed distribution of genomic sequences	✔ Peaks are not necessarily indicative of in vivo binding	✔ AmpDAP-seq	✗
	HiTS-FLIP	Uses an Illumina sequencer's flowcell as a PBM chip to measure binding to orders of magnitude more DNA sequences.	✔ (< 17 bp)	✗	✗	✗
	Spec-seq	Single step SELEX using a synthesized library of degenerate sequences of interest. The lower complexity library is useful for quantitatively measuring effects of binding site mutations using sequencing.	✔ Limited by number of sequences assayed	✗	✔ Methyl-Spec-seq	✗
	MITOMI	A microfluidic device is used to isolate DNA-protein complexes from free DNA instantaneously to accurately measure the relative binding affinities of TFs to ~10,000 individual sites.		✗	✗	✗
Low-throughput	EMSA	Tests if a DNA sequence is bound by a protein by observing a shift in the electrophoretic migration of DNA.	✗ Useful for validating known binding sites	✗	✔	✔ EMSA-FRET
	DNA footprinting	DNA is incubated with a TF and then degraded using DNase-I, resulting in cuts in all positions except those that were protected by the bound TF.		✗	✔	✔
	ITC, SPR, MSTP	Isothermal titration calorimetry (ITC), Surface plasmon resonance (SPR) and Microscale thermophoresis (MSTP) measure the binding affinity of TF-DNA interactions.		✗	✔	✔
(In Vivo Methods)	ChIP-based assays	Proteins are crosslinked to DNA using formaldehyde and precipitated with an antibody. Bound DNA is detected with qPCR, microarray (ChIP-chip), or sequencing (ChIP-seq). The ChIP-Exo variant incorporates exonuclease treatment to enhance resolution.	✔ Limited by skewed distribution of genomic sequences, and inability to distinguish direct from indirect binding	✔	✔ ChIP + Bisulfite-sequencing	✔ Re-ChIP
	DamID-seq	A TF is expressed in mammalian cells as a fusion to bacterial Dam-methylase. The enzyme methylates a consensus sequence in close proximity to the TF's binding sites, which can be mapped using restriction enzymes and high-throughput sequencing.		✔	✗	✔ Split DamID-seq

Figure 1.5 Experimental methods for determining and validating TF-binding specificities
Figure was adapted from [16].

Besides various experimental methods to identify the motif of transcription factors, an *in-silico* approach is also widely used to map transcription binding sites. MEME suite is a collection of tools for the discovery and analysis of sequence motifs [102]. For example, MEME-ChIP facilitates refinement in discovering motifs from raw data of ChIP-seq. If motifs are known, Find Individual Motif Occurrences (FIMO) is an ideal

tool for finding evidences of the provided motifs in a sequence. Additionally, in order to interpret the functional role of DNA binding motifs, Gene Ontology for MOtifs (GOMO) is possible to be used. In such a way, comprehensive information of motif can be acquired without experiments, which provides a direction for the next steps of projects.

1.5.2 *In vitro* expressed protein binding microarray

Protein microarrays cover hundreds to thousands of protein molecules which can be used for the protein profiling analysis and functional screening. Protein profiling analysis is normally based on the antibodies which are immobilized on the array and could specifically recognize antigens from different types of samples such as serum, cell lysates, etc. In terms of functional screening, different types of functional proteins or protein domains spotted on the microarray could be used to screen interactions with proteins, peptides, nucleic acids, intact cells, and various types of smaller ligands [103]. To produce the protein microarray, one way is to spot recombinant proteins directly on the microarray. Alternatively, protein microarrays can be generated with the help of cell-free expression system. Cell-free-based protein microarrays utilize the cell lysates from prokaryotic or eukaryotic cells which contain all the essential components for *in vitro* transcription and translation of the template DNA or RNA [104]. Multiple cell-free protein microarray technologies are available, such as the original 'Protein In Situ Array' (PISA) [105], the 'Nucleic Acid Programmable Protein Array' (NAPPA)[106], the 'DNA Array to Protein Array' (DAPA) [107] or production by the 'multiple spotting technique' (MIST) [108]. For different versions of cell-free protein microarray, in general, the production starts with templates preparation. Briefly, gene sequences are amplified by PCR to introduce the functional elements, such as promoter, ribosomal binding site (RBS), epitope tags and other experimentally useful components. The templates are spotted on the arrays and incubated with cell-free mixture. The success of the protein expression can be detected by fluorophores-conjugated antibodies that are able to recognize the epitopes in the N and C terminal of the expressed proteins [109]. Furthermore, several factors are needed to be considered for the development and optimization of cell-free protein microarray, such as surface and immobilization chemistry, different cell-free systems and mass transport limitations [103].

The complete human genome sequence was published in 2001 [110]. For the human proteome, around 20687 protein coding genes are collected in the major human genome databases, such as Ensembl, RefSeq, UNSC, and GENCODE [111]. Proteins play key roles in many cellular networks, metabolism and enzyme regulation, and signal transduction [112]. To better understand the functional role of the encoded proteins in a view of altered genome in patients, the proteome-based analysis is central. The complexity of human genome poses a significant challenge to translate genomics to proteomics at the same time in quantitative view. Cell-free protein microarray provide a complementary approach to understand the proteome. It is challenging for some proteins on the array to retain their native three-dimensional structure. However, the exposed epitopes have sufficient quality for protein interaction, and further functional validations by low-throughput gel-based methods or immunoassays are needed [103]. With the high-throughput technology, thousands of proteins can be analyzed at the same time by means of protein microarray. The cell-free expression helps to avoids the tedious work of protein expression and purification. As such, cell-free protein microarray is a powerful tool for the study of protein profiling analysis and functional screening.

1.5.3 The functional study of TFs

However, knowing the binding specificities of TFs is not sufficient to decipher the transcription regulation events and transcriptome network. The properties of *cis*-regulatory elements and transcription factors help us to understand better about the modulation of the transcription. In terms of genetic level, DNA mutation can affect TF binding and the expression level of TF itself. In terms of epigenetic level, three types of epigenomic regulation may affect TF binding, such as DNA methylation, histone modification resulting in the accessibility of chromatin and histone variant. These genetic mutations or epigenetic modification may inactivate one of the two copies of a gene, reducing the gene expression level or activate the gene/mutant gene expression level resulting in abnormal phenotype.

Promoters play central role in the regulation of gene expression; the classical way to study how the gene expression is regulated, is to identify the core promoter sequences that drive or initiate the transcription of one given gene. In order to achieve this,

identification of the TF's motif in the promoter region is only the first step in decoding the gene regulation. Additionally, the mutagenesis, and change of methylation level are involved to identify TF binding sites. Next, the manipulation approaches that control TFs expression, luciferase assay and ChIP assay are utilized to further verify the interaction between the TFs and the targeted promoter. With the expanded knowledge of *cis*-regulatory elements, multiple and or remote enhancers which contribute to the gene regulation are also taken into consideration to study the regulation of the transcription.

Breakthroughs of array and NGS-driven technologies have promoted understanding of human genome and mechanisms related to gene expression regulation. Firstly, the expression patterns of multiple genes at the same time can be achieved by the gene array and RNA-seq technology. The global gene expression profiling can reveal the effect of particular signaling pathway. Additionally, the genome-based technology coupled with chromatin immunoprecipitation (ChIP) determine the profiling of TF binding sites, TF-cofactor, histone modification patterns along with nucleosome positions [113]. The rapid development of sequencing method from microarray-based methods to next-generation sequencing (NGS) boosts the field of gene expression research. High throughput TSS sequencing, ChIP-seq, MNase and DNase I hypersensitive sequencing provide a set of evidences that *cis*-regulatory elements are widely dispersed in the mammalian genomes, such as at large distances from the TSSs of the putative targeted gene, in the intergenic regions.

The topology of chromatin loop needs to be considered since many regulatory elements are not immediately next to but physically contact the target genes via looping. Chromosome confirmation capture (3C) have been developed to study the chromatin looping and further understand how TFs regulate the transcription [114, 115]. Based on 3C, more advanced technologies 4C, 5C, 6C, Hi-C and ChIA-PET have also been developed to better understand the intrachromosomic interaction [116, 117]. With the development of NGS technology, 3C linked with NGS and ChIP resulted in the following technologies, 3C-Seq, 4C-Seq, 5C, ChIA-PET and Hi-C. These technologies produced the possibility to map the interactions about the identified *cis*-regulatory elements and promoters. Additionally, it can also reveal the organization of higher-order chromatin structure in the nucleus.

1.5.4 The strategies for this study

This study aims to identify the TFs which recognize the methylated binding sites, while knowing the binding specificities is not enough to decipher the methylation-dependent regulation events. To reveal the complexities of gene regulation, in my case, many factors need to be considered. In the following part, the strategies used in this study is summarized. In the beginning of the study, a set of data provided useful information for identifying the genes potentially regulated by the aberrant methylation in the promoter region, including methylation and expression profiling from cells and tissues and prior knowledge of promoter location. However, identifying the core promoter driving the expression of genes remains very cumbersome since prior knowledge of promoter location was not enough to decipher the complexed gene regulation. That's why I started to focus more on NFATc1 which was identified from protein microarray and showed the binding preference to methylated DNA sequence. In order to identify the targets of NFATc1, the mRNA profiling of NFATc1-knockdown cell samples has been assessed by microarray to investigate the global gene profiling which was regulated by NFATc1. In order to identify the core promoter driving the gene expression, in silico analysis with known TFBSs has been employed to analyze the pre-defined promoter region. Afterward, luciferase assay and ChIP assay have been utilized to verify the interaction between NFATc1 and the promoter of target genes. *In vitro* methylation and demethylation treatment have been used to validate the methylation's effect on the binding of NFATc1. Consequently, a link between methylated TF motifs and the target gene was constructed. With a better understanding of the effect of methylation on the transcription factor binding and the function of NFATc1, the chromatin structure and the interaction between proposed enhancer and target genes needed to be studied further to reveal the complex methylation-dependent expression regulation by NFATc1.

2 Material

2.1 Antibodies

Product	Cat. No.	Manufacturer
Mouse Anti-V5-Cy3™ Monoclonal Ab	V 4014	Sigma-Aldrich, Germany
Penta·His Alexa Fluor647 Conjugate	35370	Qiagen, Germany
Anti-NFATc1 Ab	SC-7294 X	Santa Cruz Biotechnology, Inc., Germany
NFATc1 Antibody (7A6)	MA3024	Life Technologies, USA
Anti-GAPDH Ab	G9295	Sigma-Aldrich, Germany
Normal mouse IgG	SC-2025	Vector Laboratories, USA
Control Antibodies, Mouse IgG	VEC-I-2000	Vector Laboratories, USA
Anti-mouse IgG(H+L) Peroxidase	VEC-PI-2000	Vector Laboratories, USA
Anti-rabbit IgG(H+L) Peroxidase	VEC-PI-1000	Vector Laboratories, USA

2.2 Reagents

Product	Cat. No.	Manufacturer
HotStar Taq Polymerase	203203	Qiagen, Germany
Nuclease-free water	AM9939	Thermo Fisher Scientific, USA
NEB Buffer 2	B7002S	NEB, USA
Halt™ Protease & Phosphatase inhibitor	78443	Thermo Fisher Scientific, USA
Fast SYBR® Green	4385612	Thermo Fisher Scientific, USA
QuantiTect SYBR® Green	204141	Qiagen, Germany
RIPA Lysis and Extraction Buffer	89900	Thermo Fisher Scientific, USA
Benzonase Nuclease	70746-4	Merck KGaA, Germany
4X Laemlli sample buffer	161-0737	BIO-RAD Laboratories, USA
Pierce™ ECL Western Blotting Substrate	32106	Thermo Fisher Scientific, USA
Lipofectamine® 2000	11668030	Thermo Fisher Scientific, USA

T4 polynucleotide kinase	M0201S	NEB, USA
T4 DNA Ligase buffer	B0202S	NEB, USA
Bsmbl	R0580S	NEB, USA
Quick Ligation™ Kit	M2200S	NEB, USA
One Shot™ Stbl3™ Chemically Competent E. coli	C737303	Thermo Fisher Scientific, USA
Ampicillin Natriumsalz	K029.2	Carl Rot, Germany,
TurboFect	R0531	Thermo Fisher Scientific, USA
Polybrene	TR-1003	Sigma-Aldrich, Germany
Q5® reaction system	M0491S	NEB, USA
EcoRI enzyme	R0101S	NEB, USA
BamHI enzyme	R0136S	NEB, USA
In-Fusion® HD cloning system	638909	Clontech Laboratories, Takara, Japan
Resazurin	10684882	Acros Organics by Fisher Scientific, USA
BstUI	ER0921	New England Biolabs GmbH
FastDigest BgIII	FD0084	Thermo Fisher Scientific, USA
GeneRuler 1kb ladder	SM0312	Thermo Fisher Scientific, USA
GeneRuler 100bp Plus Ladder	SM0323	Thermo Fisher Scientific, USA
GeneRuler Low Range Ladder	SM1191	Thermo Fisher Scientific, USA
M.SssI	M0226 L	New England Biolabs GmbH
OneShot PIR1 (chemically competent cells)	C1010-10	Invitrogen by Thermo Fisher Scientific Inc., USA
Quick Ligase	M2200	New England Biolabs GmbH
X-Gal	2315.1	Carl Roth GmbH & Co. KG, DE
Zeocin	R25001	Life Technologies, USA

2.3 Kits

Product	Cat. No.	Manufacturer
S30 T7 High-Yield Protein	L1110	Promega, USA

Expression Kit		
AllPrep DNA/RNA Mini Kit	80204	Qiagen, Germany,
Illustra MicroSpin G-25 Columns	27532501	GE Healthcare, USA
ProtoScript® First Strand cDNA Synthesis kit	E6300S	NEB, USA
Keratinocyte-SFM Medium kit	17005042	Thermo Fisher Scientific, USA
Pierce™ BCA Protein Assay kit	23225	Thermo Fisher Scientific, USA
NucleoTrap® kit	740584	Machery-Nagel, Germany
QIAprep Spin Miniprep kit	27104	Qiagen, Germany
PureLink® PCR Purification Kit	K310001	Life technology, USA
QIAquick Gel extraction kit	28115	Qiagen, Germany
Annexin V-Cy5 Apoptosis Detection Kit	ALX-850-254	Enzo Life Sciences, USA
Propidium iodide (PI)	P3566	Invitrogen by Life technology, USA
EpiTect Bisulfite Kit	59104	Qiagen, Germany
EpiTect PCR Control DNA Set	59568	Qiagen, Germany
SimpleChIP Enzymatic Chromatin IP Kit (Magnetic Beads)	9003	Cell Signaling Technology, USA
S30 T7 High-Yield Protein Expression Kit	L1110	Promega, USA
NucleoSpin® Plasmid Transfection-grade	740490.50	MACHEREY-NAGEL, DE
The Original TA Cloning® Kit	450030	Invitrogen by Thermo Fisher Scientific, USA
Dual-Luciferase Reporter Assay	E1910	Promega

2.4   Chemicals

Product	Cat. No.	Manufacturer

Material

dNTP set, Sodium salt	M3015.4100	Genaxxon bioscience GmbH, Germany
Agarose Standard	3810.3	Carl Roth GmbH & Co.KG, Germany
Tween® 20	P2287	Sigma-Aldrich, Germany
BSA		Carl Roth GmbH, Germany
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	HN77.4	Carl Roth GmbH, Germany
NaOH	303126.1920	AppliChem GmbH, Germany
Glutamate Potassium	49601	Fluka Analytical, Sigma-Aldrich, Germany
Magnesium Acetate	M5661	Sigma-Aldrich, Germany
Triton X-100	T8787	Sigma-Aldrich, Germany
Glycerol	G5516	Sigma-Aldrich, Germany
Dithiothreitol (DTT)	R0861	Invitrogen, USA
Herring sperm DNA	15634017	Invitrogen, USA
Phenylmethylsulfonylfluoride (PMSF)	8553	Cell Signaling Technology, USA
TRIS	A411.2	Carl Roth GmbH & Co.KG, Germany
Sodium chloride (NaCl)	S9888	Sigma-Aldrich, Germnay
TRIzol Reagent	15596-018	Invitrogen, USA

2.5   Labware

Product	Cat. No.	Manufacturer
Adhesive PCR Seal	600208	Biozyme, Germnay
384 Well Lightcycler Platte	72.1985.202	Sarstedt
Microseal 384-Well Skirted	MSP-3842	Bio-Rad Laboratories,

PCR Plates		USA
3D-Epoxy Glas Slides	10400201	PolyAn, Germany
Hybridization Cassette	AHC	ArrayIt®, USA
Nitrocellulose membrane 0.45µm	GE10600007	GE Healthcare, UK
Falcon® 5 mL Round Bottom Polystyrene Test Tube, with Cell Strainer Snap Cap	352235	Corning Science, USA
Camera	D7000	Nikon, Japan
Cell culture flask T175 red adherent cells	12649	Greiner bio one, Germany
Cell culture flasks 25cm	13640	Greiner bio one, Germany
Cell culture flasks 75cm	12667	Greiner bio one, Germany
Cell culture plates-6 well	657160	Greiner bio one, Germany
Cell culture plates-48 well	677180	Greiner bio one, Germany
Cell culture plates-96 well	655180	Greiner bio one, Germany
Cell culture flask T75 Yellow flask for better adhesion	90076	TPP Techno Plastic Products AG, Switzerland
Syringe filters 25 mm, 0,45µm	514-0063	VWR, USA
Vivaspin 20 centrifugal concentrator (MWCO 30kDa)	Z614629-12EA	Sigma-Aldrich, Germany

2.6   Equipment

Product	Manufacturer
Nano-plotter	GeSiM
Orbital shaker	NeoLab Migge GmbH, Germany

NanoDrop ND-1000	Thermo Fisher Scientific, USA
Centrifuge 5810R	Eppendorf, Germany
Tecan power scanner	Tecan Group AG, Switzerland
LifeECO thermal cycler	BioER Technology Co., Ltd., China
Centrifuge 2K15	Sigma, Germany
The Infinite® M200 plate reader	Tecan Trading AG, Switzerland
Heating block	Grant Instruments, United Kingdom
Microcomputer electrophoresis power supply	Renner GmbH, Germany
Trans-Blot® Turbo™	BIO-RAD Laboratories, United States
Bench-top rocker	Phoenix Instrument, Germany
Image Quant Luminescent Image Analyzer LAS-4000 mini	Fujifilm, Japan
FACSAria™ III machine	BD Biosciences, USA
FLUOstar Galaxy	BMG Technologies, Germany
Sarstedt TC insert for 24 well plate (8µm)	Sarstedt, Germany
Zeiss Axio Examiner. Z1	Zeiss, Germany
BD FACSCanto II	BD Bioscience, USA
BD FACSAria III	BD Bioscience, USA
Illumina iScan array scanner	Illumina, USA
Sonoplus-sonicator	BANDELIN, Germany
LightCycler 480	Roche, Germany
Vi Cell XR cell counter	Beckmann Coulter, Germany
Galaxy 170 S incubator	Eppendorf, Germany
Swing wing Centrifuge R5810	Eppendorf, Germany
Gel Imaging Workstation	Azure biosystems, USA
Vacuum concentrator	Bachofer GmbH, DE

2.7   Media

Product	Cat. No.	Manufacturer
PBS	10010056	Life Technologies (Gibco), USA

IMDM with phenol red	21980065	Life Technologies (Gibco), USA
DMEM 4.5g glucose, no phenol red	31053044	Life Technologies (Gibco), USA
RPMI 1640 with phenol red	21875091	Life Technologies (Gibco), USA
DMEM/F12 no phenol red	21041033	Life Technologies (Gibco), USA
B-27 Supplement (50X), minus vitamin A-10 mL	12587010	Life Technologies, USA
EGF Recombinant Human Protein	PHG0315	Life Technologies, USA
bFGF Recombinant Human Protein	13256029	Life Technologies, USA
PBS	10010056	Thermo Fisher Scientific, USA
PenStrep	15140122	Thermo Fisher Scientific, USA
Serum-free Opti-MEM I Reduced Serum Media	31985062	Thermo Fisher Scientific, USA
Endopan 3 kit	P04-0010K	Pan-Biotech, Germany
Trypsin (0.05%)	25300062	Life Technologies by Thermo Fisher Scientific, USA
FBS	10500064	Life Technologies

2.8   Buffers and solution

Name	Composition
1×TBST (1L)	100ml 10×TBS, 1ml Tween 20
10%APS	1g APS, 10ml H_2O
10%SDS	0.5g SDS, 50ml H_2O
10×TBS (1L)	31.52g Tris HCl, 80g NaCl, adjust pH to 7.6
5% Milk	10g fat skim milk powder, 200ml 1×TBST
Anode I buffer (1L)	36.4g Tris base, 200ml Methanol

Anode II buffer (1L)	3g Tris base, 200ml Methanol
Cathode buffer (1L)	3g Tris base, 5.2g 6-aminocaproic acid, 200ml methanol
Lysis buffer (10ml)	NP-40(20%) 500µl, Na-cholate (10%) 1000µl, ASB-14 (5%) 1000µl, 12-Maltoside (2.5%) 1000µl, Glycerol (99%) 2000µl, Bicine (0.5M, pH 8.5) 1000µl NaCl (1.50M) 1000µl, EDTA.2Na (0.02M) 1000µl PMSF (200mM) 50µl, Pro&Phosph inbihitor 100µl Benzonase 4µl, dH$_2$O 1346µl
1×PBST (1L)	8g NaCl, 0.2g KCl, 1.44g NaHPO$_4$, 0.24g KH$_2$PO$_4$ 1ml Tween 20, adjust pH to 7.4
Sammel Buffer	47.28g Tris HCl in 200ml dH$_2$O, adjust pH to 6.6 with NaOH
10×TBE Buffer (1L)	108g Tris, 55g Boric acid, 40 ml 0.5M Na$_2$EDTA, pH=8
Trenn Buffer (200ml)	36.33g Tris.Base, adjust pH to 8.8 with HCL
Western blot wet transfer buffer	3g Tris Base, 14.4g Glycine, 1gSDS, 800ml H$_2$O, 200ml methanol
LB-Medium (1 L)	10 g Tryptone/Pepton, 5 g yeast extract, 5 g NaCl, pH 7.2
LB-Agar	LB-Medium + 1.5% (w/v) Agar
PBS 10× (1 L)	80g NaCl, 2g KCl, 26.8g Na$_2$HPO4, 2.4g KH$_2$PO4, pH 7.4
TBE 10× (1 L)	108 g Tris, 55 g Boric acid, 40 ml 0.5 M Na$_2$EDTA, pH 8
TBS 10× (1 L)	50 mM Tris, 150 mM NaCl with HCl, pH 7.5
1M HEPES-KOH pH 7.5 (1 L)	238.30 g HEPES. adjust pH to 7.5 with KOH
1 M Tris-HCl pH 6-8 (100ml)	12.1 g tris base, adjust pH with HCl
0.5 M EDTA pH 8.0 (1L)	186.1 g Na$_2$EDTA.2H2O, adjust pH to 8.0 with NaOH (~20 g of NaOH pellets). EDTA dissolve at pH 8.0.
FACS sorting buffer	PBS with 2% FBS
Blocking buffer (immune-assay)	1x PBS, 0.05 % Tween, 2 % BSA

Blocking buffer (PDI)	25mM HEPES-NaOH (pH=7.9), 50mM glutamate potassium, 8mM magnesium acetate, 0.1 % Triton X-100, 10 % glycerol, 1xHalt™ Protease & Phosphatase inhibitor, 1mM Dithiothreitol, 4µg Herring sperm DNA, 2 % BSA
Washing buffer (PDI)	25mM HEPES-NaOH (pH=7.9), 50mM glutamate potassium, 8mM magnesium acetate, 0.1 % Triton X-100, 10 % glycerol
Hybridization buffer (PDI)	125nM methylated and unmethylated DNA fragments respectively in 1 ml blocking buffer (PDI)

2.9 Software and packages

Software	Company/Websites
GO Enrichment Analysis	http://geneontology.org
GenePix Pro.6.0	Molecular devices, USA
R	version 3.4.4
Package ggplot2	version 2.2.1
GEPIA	http://gepia.cancer-pku.cn
Image J software	National Institutes of Health
MultAlin	Corpet, 1988
NEB-Tm calculator	https://tmcalculator.neb.com/#!/main
OpenCFU 3.8 BETA software	open-source software developed by Quentin Geissmann
FlowJo	Ashland, Oregon-based FlowJo LLC, USA
Ingenuity pathway analysis	QIAGEN Bioinformatics, Germany
Gene set enrichment analysis	Broad institute, USA
DBTSS	https://dbtss.hgc.jp
BiQ Analyzer	Max Planck Institut Informatik, DE

2.10 Vectors, siRNA, primer

Name	Company	Note
Control scramble siRNA	Santa Cruz, USA	-
MISSION Pre-designed siRNA - 2 OD (Anti-NFATc1)	Sigma, USA	-
pL-CRISPR.EFS.GFP	Gift from Prof. Dr. med. Carstern Müller-Tidow's group	System Biosciences, USA
VSVG	Gift from Prof. Dr. med. Carstern Müller-Tidow's group	-
pLP1	Gift from Prof. Dr. med. Carstern Müller-Tidow's group	-
pLP2	Gift from Prof. Dr. med. Carstern Müller-Tidow's group	-
pCDH	Gift from Prof. Dr. med. Carstern Müller-Tidow's group	System Biosciences, USA, CD513B-1_10042017
q-PCR primer	Qiagen, Germany	HPRT1, GAPDH, ALDH1A3, MKNK2, SLC7A5
Primer	Biomers.net GmbH, Germany	-
pCpGL	Rehli's lab	

3 Methods

3.1. The mRNA and methylation profiling of patients and cell lines

3.1.1. The mRNA profiling of patients and cell lines

Expression profiling data of patients was available from previous studies [118]. In summary, the total RNA from individual samples with RNA integrity number of at least seven was analyzed on the Sentrix Human-6v3 Whole Genome Expression BeadChips (Sentrix Human WG-6; Illumina). The raw data was quantile normalized and log2 transformed. Differential expression analysis was performed using the LIMMA package by pairwise comparisons of the groups. The resulting p-values were adjusted for multiple testing using Benjamini-Hochberg's false discovery rate (FDR) method; features with FDR P value < 0.01 and absolute log2-fold change (log2FC)>0.5 were considered significant.

Additionally, Panc1, HPDE were cultured as written in section.3.3.3. RNA and DNA were extracted simultaneously using AllPrep DNA/RNA Mini Kit (DNA). For the mRNA profiling, RNA was analyzed on Illumina HT12 (Human Sentrix-12 BeadChip). The raw data was quantile normalized and log2 transformed. HPDE was considered as the healthy control cell line. The fold change was simply calculated by comparing the mRNA expression level of two different cell lines.

3.1.2. The methylation profiling of patients and cell lines

Genome-wide DNA methylation analysis of patients was performed using the Illumina Infinium 450k DNA methylation platform (Illumina) on 26 PDAC tissues, 24 normal pancreases, 12 chronic pancreatitis (CPs), and 2 cell lines which mentioned in section 3.1.1. The analysis procedure followed the manufacturer's standard workflow. The resulting raw data files were preprocessed using default RnBeads workflow [119]. Briefly, Quality control, probe filtering, background correction, and batch effect correction were performed as recommended. The preprocessed data was normalized by SWAN method. Differential methylation analysis was performed by limma-based method.

For the methylation data of tissue, after quality control assessments, differentially methylated probes were selected from the list of Infinium probes whose FDR-adjusted $p \leq 0.01$ and absolute methylation difference ≥ 0.15. Concerning the definition of promoter region, it's defined as the regions 1.5 kb up-stream and 0.5 kb downstream of transcription start sites. Differentially methylated promoters were selected by setting the criteria (FDR-adjusted $p \leq 0.05$, absolute methylation difference ≥ 0.1). For the analysis of cell line data, the fold change of beta value was simply calculated by comparing the mRNA expression level of two different cell lines.

3.1.3. The integration analysis

In order to investigate the association between expression and methylation, methylation and expression data of tissue was integrated. Two strategies are used, one was based on the probes, the other was based on the average beta value in the promoter region. Firstly, significantly differentially methylated probes or promoters were selected followed by the selection of significantly differentially expressed genes. Afterward, the methylation data of each probe or promoter corresponding with one gene was integrated with the expression data of the same gene. As such, the expression data and methylation data of tissue was integrated.

3.1.4. Gene ontology enrichment analysis

The gene ontology enrichment analysis [120-122] was performed by using PANTHER14.1 according to the manual of GO Enrichment Analysis. The reference list was set to 'humo sapiens'; 'GO biological process complete' was chosen to be the annotation data set; the test type was Fisher's exact and the correction method was to caculate False Discovery Rate.

3.2. The identification of methylation-dependent transcription factor

3.2.1. Template Generation for Protein Microarray Production

DNA templates for on-chip protein expression were constructed by 2-step PCR reaction. Template DNA for PCR was obtained from a transcription factor library which was kindly provided by Professor Jussi Taipale. The open reading frames (ORFs) of 667 DBD of transcription factors were constructed in gateway system-PDNOR223

vector. In order to generate DNA templates (5'end to 3'end) consisting of T7 promoter, untranslated region (UTR), ribosome binding sites (RBS), 6His, ORF, V5, stop codons and T7 terminator, following steps were performed. Briefly, the bacteria with the clone was inoculated in bacterial culture with 150μl 2YT which contained 50μg/mL spectinomycin in a 96 plate well for each clone in the transcription factor library. The plate was shaken at 37℃ at 150rpm overnight. The next day, the plate was centrifuged at 4000rpm for 30mins at room temperature. After the centrifugation, the supernatant was discarded, and the pellet was resuspended with 100μl sterilized water. Afterward, the plate was incubated in an oven at 75℃ for 20minutes, and the plate was centrifuged at 4000rpm for 30mins at room temperature. In the end, the supernatant was transferred into a new 96 plate, and this was the templates ready for the downstream 2-step PCR amplification. The PCR reaction system are listed as below. When cycling was done following the manufacturer's instructions, 5μl of PCR products was checked on 1.2% agarose gels.

Table 3.1 1st PCR system

1st PCR		
Component	**Final Concentration**	**Volume (μl)**
10xBuffer with $MgCl_2$(25mM)	1xBuffer	14
dNTP (10mM)	200μM	2
M13 Primer for (100μM)	400nM	0.4
M13 Primer rev (100μM)	400nM	0.4
Taq (5 units/μl)	1U	0.2
Nuclease free water		73
Template (from boiled E. coli)		10
In total		**100**

Table 3.2 2nd PCR system

2nd PCR		
Component	Final Concentration	Volume (µl)
10xBuffer with MgCl$_2$ (25mM)	1xBuffer	14
dNTP (10mM)	200µM	2
TF Primer for (100µM)	400nM	0.4
TF Primer rev (100µM)	400nM	0.4
Taq (5 units/µl)	5U	1
Betain (5M)	0.5M	10
Nuclease free water		62.2
Template (From 1st PCR)		10
In total		**100**

3.2.2. In Situ Cell-Free Protein Expression

The high-throughput spotting technique was used for on-chip protein expression. S30 T7 High yield protein expression system was used for the *in vitro* expression. Expression constructs were transferred to 384 plates. The plate was briefly spun down. Approximately 7nl of template (2 droplets) was spotted by using a non-contact Nano plotter 2 onto the epoxy coated slides surface followed by 28nl (8 droplets) of the cell-free expression. Each slide was incubated in a metal hybridization chamber which were filled into 30µl nuclease-free water in both ends. Afterward, the hybridization chamber was placed in a plastic box which was filled with wet tissue paper to keep certain humidity. The whole box was incubated in a ventilated oven at 30℃ for 2 hours followed by 37℃ incubation overnight. Slides were removed from the metal hybridization chambers and stored at –20°C for at least 24 hours before use.

3.2.3. Detection of Expressed Proteins

To determine successful expression of TFs, the spotted epoxy slide was firstly blocked with 2 ml blocking buffer for 1h on an orbital shaker at 100 rpm. Subsequently, it was

washed twice with 2 ml PBST-0.05 % Tween 20 for 5 mins at 200 rpm. Next, the epoxy slide was incubated for 1 h with 1:1000 dilution of Alexa Fluor® 647 conjugated anti-His™ antibody and Cy3™ conjugated anti-V5- antibody in 1 ml blocking buffer at 100 rpm. Before three times rinsing with VE-H2O, the slide was washed thrice with 2 mL PBST for 7 min at 200 rpm. Subsequently, the slide was centrifuged at 1000 rpm for 5 min to remove all liquids. In the end, the scan was performed using the Tecan power scanner V1.2 with a resolution of 10µm, laser intensity of 75 %, channel one wavelength of 580/30 nm and channel two wavelength of 676/37 nm. The images were analyzed with GenePix Pro.6.0. The median fluorescence intensities (MFI) was extracted for the analysis. The value of mean MFI of negative controls plus 4 standard deviation was set as the threshold for the positive expression.

3.2.4. Protein-DNA interaction analysis

When the protein microarray is ready, the slides were used for the protein-DNA interaction (PDI) assay. Firstly, the annealing of methylated and unmethylated Cy-tagged Twist1 promoter oligonucleotides was performed. The unmethylated forward strand was conjugated with Cy3, while the methylated forward strand was conjugated with Cy5. 20µM of complementary forward and reverse strands were heated in 1x NEB buffer 2 at 95 °C for 5 min using a LifeECO thermal cycler. Afterwards, the sample was cooled down to 4 °C at a rate of 1°C per second to enable annealing of complementary strands. The annealed oligos were purified by following the manufacturer's protocol of Illustra MicroSpin G-25 Columns. Subsequently, the concentration of the purified oligonucleotides was measured using the Nanodrop ND-1000.

To test interactions between TFs and methylated/unmethylated fragments, a competitive protein-DNA interaction assay was performed. Firstly, the slide was blocked with 2 ml blocking buffer for 1h on an orbital shaker at 100 rpm. Next, it was washed twice with 2mL washing buffer for 5 min at 200 rpm on an orbital shaker. 1 ml hybridization buffer was added and incubated overnight at 4 °C on an orbital shaker. Next, the slide was washed thrice with 2 ml washing buffer for 5 min at 200 rpm on an orbital shaker. Afterward, the slide was rinsed with VE-H2O thrice followed by centrifugation at 1000 rpm for 5 min to remove all liquids. The scan was performed by

using the Tecan Power Scanner V1.2 with a resolution of 10μm, laser intensity of 75 %, channel one wavelength of 580/30 nm and channel two wavelength of 676/37 nm. The images were analyzed with GenePix Pro.6.0.

3.3. The Expression pattern of NFAT family in PDAC tissue and related cell lines

3.3.1. in-house data analysis

The mRNA profiling of the tissue samples was from previous study as written in section3.1.1 {Bauer, #582}. In summary, there are 41 healthy tissue sample, 59 chronic pancreatitis and 195 PDAC. The expression of NFATc1 was analyzed with the statistical analysis of one-way ANOVA test. Next, the result was visualized in a Box-and-Whisker Plot by using R and the package of ggplot2.

3.3.2. online data analysis

In order to validate the expression of NFATc1 in an independent dataset, TCGA and GTEx data were used for validation. The online tool, GEPIA [123] was applied. Boxplots of NFATc1 were generated by using "Expression on Box Plots" module, PAAD dataset and Log Scale. Moreover, TCGA normal and GTEx data were used as normal control. Stage plot of NFATc1 was generated by using "Pathological Stage Plot" module, PAAD dataset and Log Scale.

3.3.3. Cell culture

Panc-1, BxPC3, Miapaca2, Aspc1, Capan1, Suit2 and HPDE used in this study were purchased from ATCC and authenticated by DKFZ internal service. All cells were regularly checked for mycoplasma contamination. The detailed information of each cell line was listed as below. When necessary, all the complete culture medium was supplemented with 1% pen-strep, 1% L-glutamine, and 10% FBS with normal cell culture procedures, if not specified with *. In terms of the normal cell culture prociduals, the cells were split when the confluency is around 80 %. To do so, the old medium was discarded, and cells were washed with PBS. Next, the cells were incubated with trypsin in the cell culture incubator. Afterward, the trypsinization was stopped by

adding complete media when the cells were detached from the bottom of the flask. To determine the cell density, Vi Cell XR cell counter was used.

Table 3.3 Cell culture list

Cell line	Tumor Type	Disease	Media
Panc1	Primary tumor	Epithelioid carcinoma	IMDM
BxPc3	Primary tumor	Adenocarcinoma	IMDM
Miapaca2	Primary tumor	Carcinoma	DMEM
Aspc1	Ascites (metastatic)	Adenocarcinoma	RPMI
Capan1	Liver metastasis	Adenocarcinoma	IMDM
Suit2	Liver metastasis	Carcinoma	DMEM
HPDE*	None	None	Keratinocyte-SFM Medium kit

3.3.4. mRNA expression level analysis of cell lines

To investigate the mRNA expression level of NFATc1 in pancreatic cancer cell lines, RNA from Panc-1, Miapaca2, and Aspc1 cell lines was isolated by using the AllPrep DNA/RNA kit. The isolation was performed according to the protocol of the kit. Additionally, the quality of the RNA was checked by using 1 % agarose gel. RNA concentration was measured with Nanodrop and 500 ng of RNA was used for the reverse transcription.

cDNA synthesis was performed via ProtoScript® First Strand cDNA Synthesis kit, following manufacturer's instructions. After synthesis, the cDNA was stored at -20 °C for qPCR.

To perform a quantitative real-time PCR reaction, Fast SYBR™ Green and Lightcycler480 were used. Three technical replicates were prepared. The primer, reaction system, and program used for real-time PCR were listed as below. Afterward, the data was analyzed by ΔΔCt method.

Table 3.4 Reaction system used for Realtime PCR

Gene	Component	1× run (µl)
HPRT1 NFATc1	cDNA	1
	Primer (10µM)	1
	SYBR Green Master Mix (2×)	5
	Nuclease-free water	3
	Total Volume	10

Table 3.5 Program used for real time PCR

Step	Temperature (°C)	Duration	Cycles
Polymerase activation	95	20s	Hold
Denature	95	3s	
Anneal/Extend	60	30s	40

3.3.5. Protein expression level analysis of cell lines

For protein isolation, Panc-1, BxPC3, Miapaca2, Aspc1, Capan1, Suit2, and HPDE cells in culture were washed twice with ice-cold PBS and lysed with lysis buffer. To prepare for the lysis buffer, 10ml RIPA Lysis and extraction buffer was supplemented with 100µl of 100 X Halt™ Protease and Phosphatase Inhibitor Cocktail, phenylmethylsulfonylfluoride (PMSF) and 4µl Benzonase Nuclease. 300µl of RIPA buffer was added into each well of 6 well plates, and the samples were kept on ice for

5 minutes with occasionally swirling. The collected lysate was centrifuged at 4 °C for 15 minutes at 14,000 g using Centrifuge 2K15. The supernatant containing the solubilized proteins was stored at -20 °C for further analysis. Afterward, the concentration of the protein was determined by following the instruction of the Pierce^{TM} BCA Protein Assay kit. The Infinite$^®$ M200 plate reader was used for the measurement at an absorbance of 562nm.

For western blotting analysis, 10-20µg samples mixed with 25 µl 4X laemlli sample buffer were denatured for 5 minutes at 95 °C in a heating block and loaded onto 10% SDS-PAGE gels. Samples were stacked for 30 min at 75 V and separated for one hour at 135 V in the running buffer with the microcomputer electrophoresis power supply. The semi-dry method was used to transfer the samples from the SDS-PAGE gel to a nitrocellulose membrane. The Trans-Blot$^®$ Turbo^{TM} was used to transfer for 30 minutes at 25V (standard transfer program). Then, membranes were blocked in 5% non-fat milk in TBST-0.05% Tween 20 for 1h at room temperature, and the blocking was followed by washing with TBST and incubation with first antibody (anti-NFATc1, anti-GAPDH) overnight at 4°C on a benchtop rocker. Immunodetections were performed with the corresponding secondary antibodies (HRP Horse Anti-Mouse IgG Antibody). Enhanced chemiluminescence (ECL) reagents and Image Quant Luminescent Image Analyzer LAS-4000 mini was used for visualization. The densitometric analysis was done by using ImageJ software.

3.4. The Functional study of NFATc1 in pancreatic cancer cell line

3.4.1. The knock-down of NFATc1 by siRNA transfection in cell lines

Cells were seeded in a 6-well plate 24 hours before the transfection to get a confluency of at least 50 % the next day. Complete medium without Pen-Strep was used for the siRNA transfection. According to manufacturer's protocol of Lipofectamine$^®$ 2000 Transfection Reagent for 6-well transfection, Serum-free Opti-MEM I Reduced Serum Media was used to prepare the complex for transfection. Three different concentrations of siRNA (33.3 nM, 66.6 nM, 99.9 nM) were tested for all pancreatic cancer cell lines, and the knockdown effect was compared with the effect of control scramble siRNA. Medium was changed to complete medium 7 hours after transfection,

In the end, cells were harvested 24 hours post transfection for real time quantitative PCR and 48 hours post transfection for western blot.

3.4.2. The knock-out of NFATc1 by CRISPR/Cas9 gRNA transfection in cell lines

Vector cloning: CRISPR plasmid pL-CRISPR.EFS.GFP with scramble gRNA or gRNA specifically targeting NFATc1 was generated. Firstly, 1µl 100µM Single Guide RNA (sgRNA) oligos were phosphorylated using 0.5µl of the T4 polynucleotide kinase with 1µl of T4 DNA Ligase buffer, and the reaction was filled up to 10µl with nuclease free water followed by the incubation for 45 minutes at 37 °C and 2.5 minutes at 95 °C in a thermocycler. Afterward, it was cooled down to 22°C at a speed of 0.1°C per second. Secondly, 3µg of the vector pL-CRISPR.EFS.GFP was linearized using 3µl enzyme BsmbI in a 50µl reaction system following the manufacturer's instruction. The mixture was loaded on a 1% agarose gel after the incubation and the linearized vector was extracted by using NucleoTrap® kit according to the manufacturer's instruction. Subsequently, the annealed and phosphorylated oligos were ligated with the purified linearized plasmid. To do so, 30-50 ng of the vector was mixed with 1µl of the 1:500 diluted oligo mix, 0.5µl QuickLigase enzyme and appropriate 2 X QuickLigase buffer to an overall volume of 10µl. The reaction was incubated for 5 minutes at room temperature.

Afterward, 2µl ligation product was used to transform 50µl of Stbl3 chemically competent E. coli. The bacteria were plated on pre-warmed agar plates containing 0.1% Ampicillin. After overnight incubation, positive clones were confirmed by colony PCR and inoculated in 5 mL of LB-media supplemented with 0.1 % of Ampicillin. The plasmid was extracted by using Spin Miniprep kit after another overnight, and the isolated constructs were sent for sequencing (eurofins genomics, Germany, Ebersberg) with the commonly used primer LKO.1 5'. Using the free-source program MultAlin, the sequenced DNA was scanned for the integrated sgRNA.

Lenti-virus production: 10^7 HEK cells were plated in 150cm petri dish 1 day before packing the virus. On the day of virus packing, 13ml medium was refreshed one hour

before transfection. Firstly, 48µl TurboFect was incubated with 1.5ml opti-MEM for 5 minutes, 10.4µg plasmid with insertion of interest was mixed with 4µg VSVG, 7.2µg pLP1, 2.4µg pLP2 and 1.5ml opti-MEM. The TurboFect mixture was added into the plasmid mixture followed by gently pipetting up and down and incubation of 15 minutes at room temperature. Next, this 3ml mixture was added into 15cm petri dish in a droplet manner, and the medium was refreshed 5 hours later after the transfection. The supernatant was harvested 2 days and 3 days respectively after the transfection, filtered by 0.45µm filter, ultraconcentrated collectively at 20000rpm at 4 degree for 2 hours, and discarded. The virus pellet was resuspended in 150µl cold PBS. Alternatively, the supernatant can also be concentrated by Vivaspin 20 centrifugal concentrator (MWCO 30kDa) at 4 degree. In the end, the virus was aliquoted into 20µl per tube.

Transduction: For transduction, 2-3 x 10^5 cells were seeded per well in a 6-well cell culture plate. In each well, 20µl of concentrated virus mixed with 200µl of culture medium was added in a droplet manner in a final volume of 2ml culture medium with polybrene in a final concentration of 10µg/ml. The cells were incubated with the virus overnight and washed three times with PBS the next day followed by refreshing culture medium.

FACS sorting: 72 hours after the transduction, the cells were used for the FACS sorting. Every sample was pelleted and resuspended with 250µl of sorting buffer. Afterward, it was filtered by using test tube with cell strainer snap cap. Live cells with GFP positive signal were sorted by BD FACSAria™ III machine and collected in 1 ml culture medium, the service was provided by DKFZ core facility. In the end, the sorted cells were plated into a cell culture plate. After 24 hours, the culture media was refreshed.

Knockout validation: The sorted cells were harvested until the cells reached the confluency of 80% in 6-well plate. DNA and RNA were isolated simultaneously by using AllPrep DNA/RNA kit. Protein was isolated as previously described. For the knockout-validation at DNA level, the sgRNA targeting region was amplified by touchdown PCR. The reagent system and PCR running program is listed as below. After the amplification, the PCR product was purified by using PureLink® PCR Purification Kit. The purified PCR product was sent for sequencing with the forward

PCR primer. For the knockout-validation at protein level, the Western Blot (WB) experiment was performed. Anti-NFATc1 antibody and anti-GAPDH antibody were used for the specific detection of the target protein expression.

Table 3.6 Reaction system used for PCR

Sample	Component	1× run (µl)
pL-CRISPR.EFS.GFP Sg1-sg4	10x buffer	2
	Prisg1/2_For or Prisg3_For primer (10µM)	0.8
	Prisg1/2_rev or Prisg3_rev primer (10µM)	0.8
	dNTP(10mM)	0.4
	Hot start Taq polymerase (5U/ µl)	0.2
	Template	1
	Nuclease-free water	Up to 20

Table 3.7 Program Used for Touchdown PCR

Step	Temperature (°C)	Duration	Cycles
Polymerase activation	95	15min	Hold
Denature	94	30 s	
Anneal/Extend	Tm+5	30 s	20
Elongation	72	30 s	
Denature	94	30 s	
Anneal/Extend	Tm-5	30 s	20
Elongation	72	30 s	
Final Elongation	72	5min	Hold

3.4.3. The overexpression of NFATc1 in cell lines

In order to generate the stable NFATc1-overexpression cell lines, cloning, virus production, transduction, FACS sorting and validation were performed sequentially. Firstly, NFATc1 clone (CloneID: 111759121) in pENTR223 vector was ordered from DKFZ core facility. The open reading frame was amplified by Q5® reaction system with the NFATc1-OE forward primer and NFATc1-OE reverse primer. Touchdown PCR program was performed. Then, the PCR product was purified by using PureLink® PCR purification kit according to the manufacturer's instructions. The overexpression vector pCDH was digested by EcoRI and BamHI enzymes. Subsequently, the digested vector was purified by gel extraction accordingly. Afterward, In-Fusion® HD cloning reactions was applied according to the manufacturer's instructions. The positive clone was verified by sanger sequencing. Subsequently, virus production, transduction, FACS sorting and validation of protein level were performed as previously written in 3.4.2.

3.4.4. Proliferation assay

Resazurin assay was performed to check the proliferation effect of NFATc1 on Panc1 and Miapaca2. Briefly, the transfection of NFATc1-specific siRNA and scramble RNA was performed on Panc1 and Miapaca2 as previously written in 3.4.1. The experiments were conducted in a 48-well plate and 3 biological replicates were prepared respectively for experimental and control group. At the timepoint of 24, 48, and 72 hours after the siRNA knockdown, the samples were incubated with 200µl culture medium containing 20µg/ml resazurin for 1 hour in the cell culture incubator. In the end, the fluorescent resorufin was measured by using the plate reader FLUOstar Galaxy at excitation wavelength of 544 nm and emission wavelength of 590 nm.

3.4.5. Migration assay

Sarstedt TC inserts with a pore size of 8µm for 24-well plates were used to perform the migration assay. Firstly, the transfection of NFATc1-specific siRNA and scramble RNA was performed on Panc1 and Miapaca2 as previously written in 3.4.1. The cells were harvested 24 hours after the knockdown. Next, 2000 cells in 0.2ml FBS-free medium was added to each insert. Then, 0.5 ml chemoattractant (10% FBS complete culture medium) was added to the bottom of 24 well plates. 3 biological replicates were

prepared for experimental and control group respectively. Afterward, the cell invasion chambers were incubated in cell culture incubator for 24 hours followed by crystal violet staining. Briefly, the medium from the inserts was aspirated firstly. Prior to the staining, non-migrated cells were removed from the inserts manually by gently swabbing the inside part of each insert with cotton swabs. Next, the inserts were washed with ice-cold PBS and fixed with ice-cold 100% methanol for 10minutes. After the fixation, the methanol was aspirated from the inserts and the inserts were covered by 0.5% crystal violet solution in 25% methanol for 10minutes. The inserts were washed in VE-water several times until the dye stops coming off followed by drying at room temperature. In the end, Zeiss Axio Examiner Z1 was used for taking pictures for each replicate. Under microscope, 5 different fields of view were observed, and images were captured. ImageJ was used to measure the area of crystal violet staining.

3.4.6. Colony assay

The sorted transduced Panc1 and Miapaca2 cells as previously written in 3.4.2 and 3.4.3 were used for colony assay after 5 days culture. Briefly, 3 biological replicates were prepared for the experimental and control group respectively. Initially, for the first layer, 10ml autoclaved 2 % soft SeaKem® GTG® Agarose in PBS was mixed with 30 ml pre-warmed culture medium to prepare 0.5% agar. 2 ml 0.5% agar was then plated in each well of 6-well plates. The plate was incubated at room temperature for at least 20 minutes until the agar became solid. Afterward, 3.2 ml pre-warmed culture medium, 0.4 ml FBS and 0.4ml 2% agar were mixed together to prepare 0.05% agar which was incubated at 37°C. Meanwhile, the sorted cells were harvested and counted. Then, 3300 sorted cells were resuspended in 1ml 0.05% agar and the suspension was quickly added on the top of the first layer of 0.5% agar. After the incubation of 20minutes at room temperature, 1ml complete culture medium was added on the top of agar. In the end, the plates were incubated for 3 weeks in the Galaxy 170 S incubator. After 3-weeks culture, 0.5 ml of 0.005% Crystal Violet was incubated in each well for more than 1 hour on an orbital shaker. Next, each well was washed with PBS until the background is clear. Images were taken by the camera. To count colonies, OpenCFU 3.8 BETA software was used for the image analysis. For the setting of the software, the radius for a colony was set to 5 for Miacapa2 and to 7 for Panc1 since the size of cells derived from diverse cell lines was different.

3.4.7. Apoptosis assay

Firstly, for Panc1 and Miapaca2, the transfection of NFATc1-specific siRNA and scramble RNA was performed as previously written in 3.4.1. The cells including the floating apoptotic cells were harvested and washed gently with serum-containing culture medium once 48 hours after the knockdown. 3 biological replicates were prepared for experimental and control group respectively. For each replicate, approximately, 10^6 cells/ml cells were resuspended in 200µl 1x Annexin V binding buffer containing 2µl Annexin V-Cy5 and 1µg/ml PI. Next, the samples were kept on ice and must be analyzed within an hour to avoid false positive measurements. Before the analysis of FACS, negative control cells, GFP cells, Non-GFP cells with PI staining and Non-GFP cells with Annexin V staining were used for compensation setting. In terms of FACs analysis, 3 lasers including GFP, PI, Cy5 were used. The voltage was set accordingly, and the population of interest was clear visible by making adjustment of FSC/ SSC. Finally, the results were analyzed by FlowJo software.

3.5. The analysis of NFATc1 related pathway

3.5.1. mRNA profiling of knock-down cell lines

Firstly, for Panc1, Miapaca2, Suit2, the transfection of NFATc1-specific siRNA and scramble RNA was performed as previously written in 3.4.1. Total RNA was isolated using TRIzol^{TM} Reagent 40 hours after transfection according to the manufacture's protocol. 500 ng RNA from each cell line transfected with NFATc1-specific siRNA or control scramble RNA was sent for Illumina HT12 (Human Sentrix-12 BeadChip) gene expression profiling. DKFZ genomics and proteomics core facility performed this service. Briefly, the labeled cRNA was hybridized with over 47,000 probes conjugated to beads on the array at 58 °C for 17 hours. Afterward, the microarray was scanned by the Illumina iScan array scanner according to the standard Illumina scanning protocol. The intensity value of each probe was measured accordingly. The bead-level data analysis of the gene expression microarray was done with R by core Facility. To do so, after the outlier removal, the quantile normalization was performed. Studentís t-test was used for the significance test. Additionally, for each group, the average expression value was calculated as mean of the measured expressions of beads together with the standard deviation of the beads, and p-values was calculated using averaged expression values for each sample in the group. Furthermore, Benjamini-

Hochberg correction was applied to p-values of all probe IDs on the chip for the differential expression analysis. The averaged expression level of the treated group divided by the control group was the fold change (FC). In the end, the visualization of the data was performed by R.

3.5.2. Data analysis

In order to identify enriched gene sets, involved pathways, and influenced function upon the knockdown of NFATc1. Gene set enrichment analysis was performed. Firstly, the gene set enrichment analysis was done by using Broad Institute algorithm following the standard protocol. The normalized transcriptional profiling microarray data of all 6 samples was used for the input of this analysis. Gene sets, whose Normalized Enrichment Score (NES) above 1, p-value below 0.05, and false discovery rate (FDR) Q-value below 0.25 were considered as significant enriched gene sets.

3.6. Candidate validation

3.6.1. Validation of candidates via q-PCR

Candidate selection: RNA samples from Panc1, Miapaca2, and AsPc1 were sent for gene expression profiling as previously written in section 3.5.1, in which siRNA-mediated knockdown of NFATc1 was performed. The top 30 genes downregulated candidates were taken for further considerations. Afterward, genes showing upregulation in PDAC group compared with healthy group were kept, other genes were removed from the list of candidates. Then, the promoter sequence (-2000bp/+500bp) of all these genes were extracted from DBTSS [124], and the promoter sequence was analyzed by MEME-FIMO [125]. The motif of NFATc1 [23] was used as input for the analysis. When the binding site of NFATc1 in the promoter region was predicted (P<0.01), the methylation level of the binding site was analyzed by integration of 450k array data from tissue and cell line including Panc1 and HPDE.

Validation via q-PCR: The NFATc1-overexpression, -knockdown, -knockout cell models established as previously written in section 3.4.1, 3.4.2 and 3.4.3 were employed for the candidate validation. After the cells were harvested, RNA isolation, cDNA synthesis and quantitative real-time PCR reaction were performed as written in 3.3.4. HPRT1 was used as a housekeeping gene for all the samples. Primer NFATc1

was firstly recruited to verify the establishment of cell models. Afterward, Primer ALDH1A3, SLC7A5, and MKNK2 were used to detect the mRNA regulation of the selected candidates upon the overexpression, knockdown and knockout of NFATc1. The q-PCR result was analyzed via $\Delta\Delta$Ct method.

3.6.2. Validation of candidates via methylation specific PCR

In-house data analysis: For the validation at expression level, the expression of the candidate genes was checked in the mRNA profiling of cell lines and tissues. For the validation at methylation level, the beta value of candidate genes was investigated from the 450K array data of cell lines and tissues.

Methylation-specific PCR (MSP): The DNA was isolated from Panc1 and Miapaca2 by using the AllPrep DNA/RNA Mini Kit according to manufacturer's recommendations. For each cell sample, 1μg of isolated DNA was used for bisulfite conversion with the EpiTect Bisulfite Kit according to manufacturer's recommendations, and the concentration of bisulfite converted DNA was measured by Nanodrop. Next, MSP was performed as written below in the table 3.8, 3.9. Briefly, 2.5ng of each template was amplified using methylation-specific primer pair. Methylated Bisulfite Converted control DNA (MBC) and Unmethylated, Bisulfite Converted control DNA (UBC) from the EpiTect PCR Control DNA Set were used to validate the specificity of the primers. The size of the expected products was 100bp for ALDH1A3. To visualize the result, 5μl per sample was applied to a 3 % agarose gel and separated at 500 mA and 130 V for 0.7 h.

Table 3.8 Reaction system used for MSP

Sample	Component	1× run (μl)
Bisulfite converted Panc1, Miapaca2 DNA, Control DNA	10X QIAGEN PCR buffer with Coral Load PCR Buffer (Qiagen)	1
	MSP Forward Primer (10μM)	0.4
	MSP Reverse Primer (10μM)	0.4
	dNTP(10mM)	0.2
	0.5 U of Hot Start Taq (5U / μl)	0.1
	Template (5ng / μl)	0.5
	Nuclease-free water	Up to 10

Table 3.9 Program Used for Touchdown MSP

Step	Temperature (°C)	Duration	Cycles
Polymerase activation	95	15min	Hold
Denature	94	30 s	20
Anneal/Extend	53	30 s	
Elongation	72	15 s	
Denature	94	30 s	15
Anneal/Extend	43	30 s	
Elongation	72	15 s	
Final Elongation	72	5mins	Hold

3.7. Methylation-dependent validation

3.7.1. Demethylation treatment

Demethylation optimization: In order to establish an unmethylated control, Panc1 and Miapaca2 cells were treated with the DNA methyltransferase inhibitor 5-Azacytidine. Since the agent has a negative impact on cell survival, the concentration of 5-Azacytidine was determined by using a resazurin-based cell viability assay. The optimization aimed at determining the highest possible concentration of 5-Azacytidine that did not reduce cell viability by more than 50 % at the end of 136h treatment. Firstly, a standard curve for the viability assay was established. Therefore, 0, 750, 1500, 3000, 6250, 12500, 25000, 50000 or 100000 cells per well of each cell line were seeded into 96-well plates, and the cells were cultured in the cell culture incubator for 4h. Afterward, resazurin was added to each well with a final concentration of 20µg/ml. The cells were incubated for 1 h in the cell culture incubator to allow for the reduction of resazurin to resorufin. Finally, the fluorescence was measured in triplicates at an excitation wavelength of 544 nm and an emission wavelength of 590 nm using the plate reader FLUOstar Galaxy. For the viability assay, $5*10^3$ cells of Panc1 or $6*10^3$ cells of MiaPaCa2 were seeded into each well of 96-well plates. 3 biological replicates were prepared for each measurement. After 24 h, the culture medium was replaced with complete culture medium containing 0µM, 0.5µM, 1µM, 2µM, 4µM, 6µM, 8µM or 10µM of 5-Azacytidine. Due to the instability of 5-Azacytidine, the medium was refreshed every 24 h. Finally, the cell viability was measured at the start of the treatment, 40 h, 64 h, 88 h, 112 h and 136 h after the first treatment. The number of viable cells was calculated from the relative fluorescent units referring to the standard curve. The significance was assessed using a one-sided, unpaired, heteroscedastic Welch's t-test to compare the 0µM control group with the treated groups. P-values of less than 0.05 was considered statistically significant.

Demethylation treatment and validation: To obtain DNA and RNA from demethylated cells, $2*10^5$ cells of Panc1 or MiaPaCa2 per well were seeded into 6-well plates. Complete culture medium containing 0µM, 1µM, or 2µM 5-Azacytidine for Panc1 and 0µM, 0.5µM, or 1µM 5-Azacytidine for MiaPaCa2 were refreshed every 24 h respectively. 72 h after the treatment, the DNA and RNA were isolated using the AllPrep DNA/RNA Mini Kit according to manufacturer's recommendations. For each

sample, 1µg of isolated DNA was used for bisulfite conversion with the EpiTect Bisulfite Kit. To verify the demethylation, MSP targeting ALDH1A3 region was performed as written in 3.6.2.

RNA isolation and q-PCR: 1ug RNA isolated from last step was used for the cDNA synthesis as written in 3.3.4. Quantitative real-time PCR reaction was performed as written in 3.3.4. HPRT1 was used as a housekeeping gene for all the samples. Primer ALDH1A3 was used to detect mRNA regulation upon the 5-Azacytidine treatment. The q-PCR result was analyzed via ΔΔCt method.

3.7.2. Luciferase assay

Vector cloning: To obtain the identified promoter region of ALDH1A3, a PCR using the NFATc1_promoter primer pair was performed. The PCR product was purified by using the PureLink Quick PCR Purification Kit. Afterward, 1 µg of the vector pCpGL-basic as well as the PCR product were digested by BglII and NcoI enzymes. Next, the digested products were extracted by using the QIAquick Gel Extraction Kit following the kit's instructions. After the purification, the PCR product was ligated with the pCpGL-basic plasmid using a vector to insert ratio of 1:3. In the end, 1 µL of the ligation mix were used to transform 20 µL of PIR1 chemically competent *E.colis*. 100 µL of the cells were plated on pre-warmed LB agar plates containing 0.1 % of zeocin. Subsequently, colony PCR was used to confirm the positive clones by using pCpGL_For and pCpGL_Rev primers. In the end, the correct insertion was confirmed by sanger sequencing.

In vitro methylation: The plasmid containing the *ALDH1A3* identified promoter region was methylated by incubating 2 µg plasmid DNA with 4 units of M.SssI, 1X NE Buffer 2 and 320 µM S-Adenosylmethionin at 37 °C for 3 hours. Subsequently, a 5 µL reaction mix containing 2 Units of M.SssI, 1X NE Buffer 2 and 12.8 mM of SAM were added. The reaction system was incubated at 37 °C for 3 hours followed by 65 °C for 20 minutes. The methylated DNA was purified using the PureLink Quick PCR Purification Kit.

To verify the methylation the plasmid was digested using the methylation-sensitive

enzyme BstUI which only cuts unmethylated DNA. Due to the lack of a restriction site of this enzyme in the pCpGL-basic plasmid, therefore a plasmid containing the BstUI restriction site was used as a positive control. For the digestion 200 ng of plasmid DNA, 1X Cut Smart Buffer and 5 units of BstUI were filled up with nuclease-free water to a 20 µL reaction mix which was incubated at 60 °C for one hour.

Cell experiments: In terms of the knockdown cells, cells were grown to a confluency of 30-50% before NFATc1 knockdown was performed. In terms of the normal cells, cells were grown to a confluency of 80-90% before the luciferase constructs transfection. Next, the cells were allowed to grow further for 24 h and then transfected with the luciferase constructs. After incubation for 24 h, the cells were harvested and the luciferase activity measured, using the Dual-Luciferase Reporter Assay kit (Promega).

3.7.3. Chromatin immunoprecipitation (ChIP)

Panc1 and Miapaca2 cell lines were cultured in 15cm petri dish in their respective growth medium to about 80% confluency. Meanwhile, for the demethylation control group, Panc1 and Miapaca2 cells were treated with 2µM and 0.5 µM 5-Azacytidine respectively for 10 days, and the medium was refreshed every 24hours. When cells were demethylated, the cells were fixed with 37% paraformaldehyde at a final concentration of 1% for 10 minutes. Glycine solution was incubated with fixed cells for 5 minutes to quench the fixative at a final concentration of 125mM. The cells were then washed twice with 10ml ice-cold PBS. The cells were scraped into 2ml ice-cold PBS containing 1x protease and phosphatase inhibitors. The cell pellets were collected by centrifuging at 2000g in a benchtop centrifuge at 4°C for 5 min. The ChIP was performed using the SimpleChIP Enzymatic Chromatin IP Kit (Magnetic Beads) according to manufacturer's recommendation. Briefly, approximately $4*10^6$ Panc1 cells and $8*10^6$ Miapaca2 cells were used as one ChIP reaction. After the cell nuclei was isolated, the chromatin for each reaction was digested with 0.25 µl micrococcal nuclease at 37°C for 20 min. The sonication was then carried to lyse the nuclei on the lowest power setting of the Sonoplus-sonicator with 3-6 rounds of 20s pulses and 30s breaks. The water bath was cooled with ice. After the lysate clearance by centrifugation at 9400 x g for 10 min at 4°C, 10 micrograms chromatin was diluted in

400µl of 1x ChIP buffer. Afterward, 5 micrograms of ChIP-grade anti-NFATc1 antibody was added and incubated with chromatin at 4°C overnight; as a control for unspecific binding, a complex mixture of mouse IgGs was applied. Protein-DNA complexes were captured with ChIP-grade protein-G magnetic beads for 4h; unbound protein was washed away. Immunoprecipitants were eluted from the beads in 150 µl ChIP elution buffer for 30 min at 65°C. The crosslinking was reversed with 200 mM NaCl at 65°C for 2 hours, the DNA purified with spin columns and stored at -20°C. Enrichment at the ALDH1A3 promoter was determined by qPCR. The ALDH1A3_ChIP primer pair was used. PCR was carried out on a LightCycler 480. Three technical PCR replicates were carried out for each sample. The percent input method was used for normalization.

3.7.4. Statistical analysis

All experiments were done three or more times. If no extra information was indicated, data were shown as mean ± standard error mean. Two-tailed Student's t test was used to make comparisons between 2 different groups. Significant difference was defined differently based on different experiments. Statistical analysis was done using GraphPad Prism.

4 Results

4.1 The identification of hypermethylated-overexpressed genes

4.1.1 The integration analysis of tissue data

The methylation profiling and the expression profiling of tissue were integrated to identify the hypermethylated and upregulated genes. For this purpose, the mRNA profiles of 195 PDAC patients and 41 healthy donors were analyzed, and 6068 significantly regulated transcripts were chosen. The methylation profiling of 26 PDAC patients and 24 healthy donors were analyzed, and 2917 significantly differentially methylated promoters were chosen. To calculate the methylation value of promoter regions, the methylation value of probes in the defined regions was averaged. After integration, 421 transcripts showed the results. 132 transcripts showed hypermethylation and downregulation in the expression level; 119 transcripts showed upregulation in the expression level and hypermethylation. To select the candidates for the next step, genes which showed upregulation in expression and had most probes indicating hypermethylation in the defined promoter region were chosen.
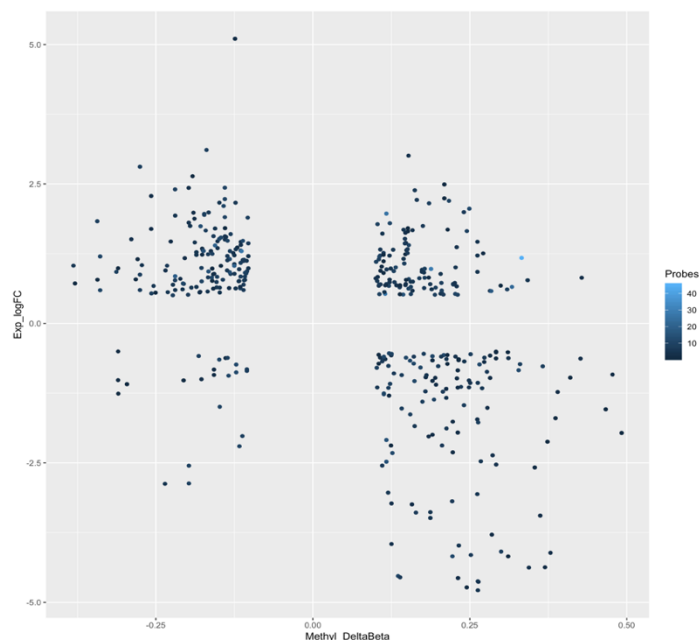
Figure 4.1 Integration of expression and methylation profiling

Each dot represents one transcript with the expression and methylation value. X-axis: The log fold change of PDAC-patients group relative to healthy-donors group. Y-axis: The mean delta beta value in the defined promoter region (-1500/+500bp) when comparing PDAC group with healthy group. The shades of color indicate the number of probes.

4.1.2 The GO enrichment analysis of hypermethylated and upregulated genes

In order to understand the functional role of genes which showed hypermethylation and upregulation in PDAC-patients group compared with healthy-donors group, the list of 119 genes were used as input for the gene ontology (GO) term enrichment analysis. Top 10 overrepresented biological process were shown here.
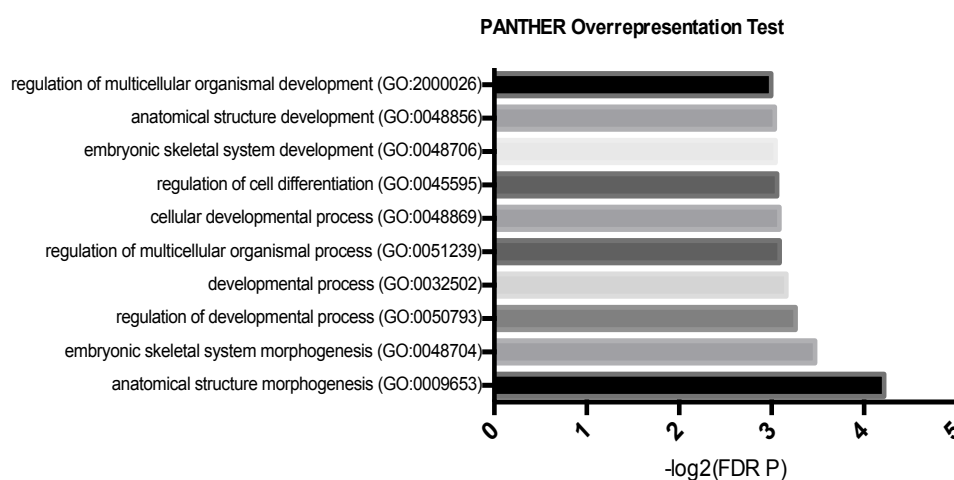
Figure 4.2 Pathway analysis of upregulated and hypermethylated genes

Gene ontology term enrichment analysis indicated that hypermethylated and upregulated genes were significantly enriched in multiple biological process (top 10 overrepresented process were shown here). The threshold was set to False discovery rate (FDR) adjusted P value which was equal with 0.05.

4.2 The identification of methylation-dependent TFs

4.2.1 Protein microarray

TF protein microarrays were generated, which covered 4 full-length TFs and 667 DNA binding domains from over 12 different TF families (Fig.4.3A). In the end, the full expression of proteins was assessed by the immunostaining which used fluorescent antibodies against 6xHis-tag of N-terminus and V5-tag of C-terminus. If the fluorescent signal intensity of C-terminus was above the threshold, the proteins were considered

as successfully expressed proteins. As such, more than 97% TFs were successfully expressed (Fig.4.3B, C).
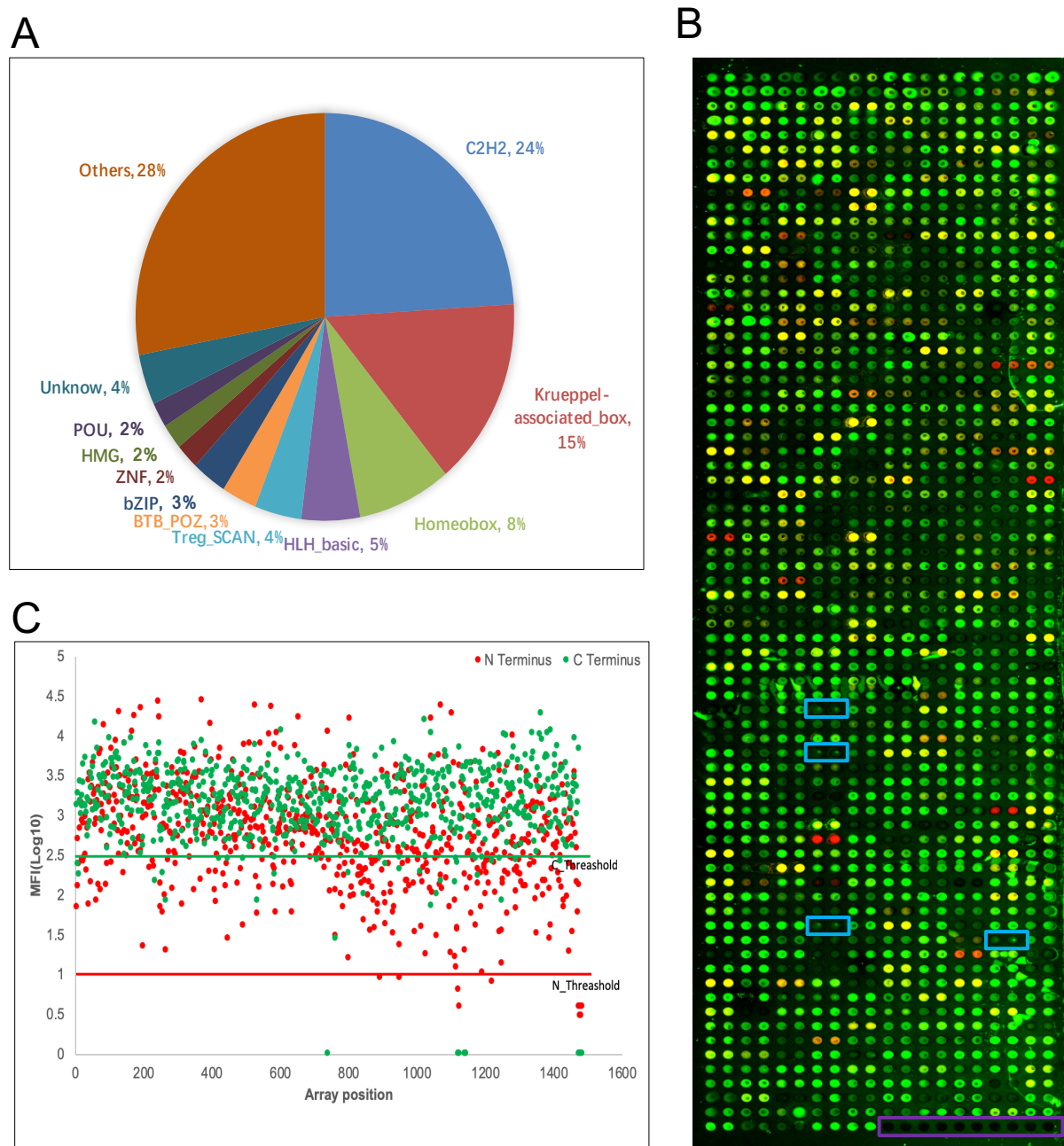
Figure 4.3 Determination of on-chip protein expression

(A) The expressed DBDs were listed according to the family they represent. (B) Proteins were immunoassayed with fluorescently labelled antibodies that targeted terminus tags. Green: Anti-V5 Ab, C-terminus. Red: Anti-His Ab, N-Terminus, Yellow: Both (anti-His and anti-V5) Abs. Spots in the blue rectangle: PCR negative control, Spots in the purple rectangle: blank control. (C) Read out: Median Fluorescence Intensities (MFI). C/N-Threashold: mean MFI (negative controls) + 4 SD (negative controls).

4.2.2 Protein-DNA interaction on microarray

To identify the methylation-specific transcription factors, the generated protein microarray was applied for the protein-DNA interaction. Methylated and unmethylated DNA-fragments of 55bp were synthesized and labelled fluorescently with Cy5 and Cy3 respectively. The methylated and unmethylated fragments were mixed in a ration of 1:1, and the competition binding screening was performed on the protein microarray. It's revealed that TFs showed different binding ability to DNA with different methylation state (Fig.4.4A). As such, TFs which bind preferentially to methylated DNA are called methyl-plus TFs, and TFs which bind preferentially to unmethylated DNA are called methyl-minus TFs. Next, candidates including top 15 methyl-plus TFs and top 3 methyl-minus TFs were identified based on the signal ratio which was calculated by dividing the methylated (Cy5) and unmethylated (Cy3) signal intensity (Fig.4.4B). As for the list of methyl-minus TFs, ELK3, FLI1 and EPAS1 were the top candidates with highest signal ratio of the unmethylated signal divided by the methylated signal. EPAS1 was excluded since the expression was not successful (Fig.4.4C). Afterward, the binding of each candidates was checked individually (Fig.4.4E). As for the top 15 methyl-plus TFs with highest signal ratio of the methylated signal divided by the unmethylated signal, NFATc1, NFATc2 and NFATc3 were the binders which showed the highest Cy5 signal intensity (Fig.4.4D). Furthermore, all these 3 TFs belong to NFAT TF family. Additionally, they were all successfully expressed (Fig.4.4C).

 Figure 4.4 Protein-DNA interaction on microarray

(A) Incubation of a 55bp fragment resembling the methylated (red) and the unmethylated (green) promoter identified specific binding; a merger of the images is shown in yellow. (B) The signal ratio was produced by dividing the methylated (methyl-Cy5) and unmethylated (unmethyl-Cy3) signal intensity. The red dots or green dots showed the binding candidates which preferred binding with methylated fragment or unmethylated fragment. (C)The C-terminus and N-terminus expression of all the candidates. (D)The methylation and unmethylation signal intensities of top 15 methyl-plus TFs shown in plot B. NC refers to negative control. (E)The methylation and unmethylation signal intensities of top 3 methyl-minus TFs shown in plot B.

A

Methylated promoter | Merged data | Unmethylated promoter

Methylation-specific TFs | Unspecific TFs | Unmethylation-specific TFs

B Binding

C Expression

D Binding_Unmethy_Hits

E Binding_Methy_Hits
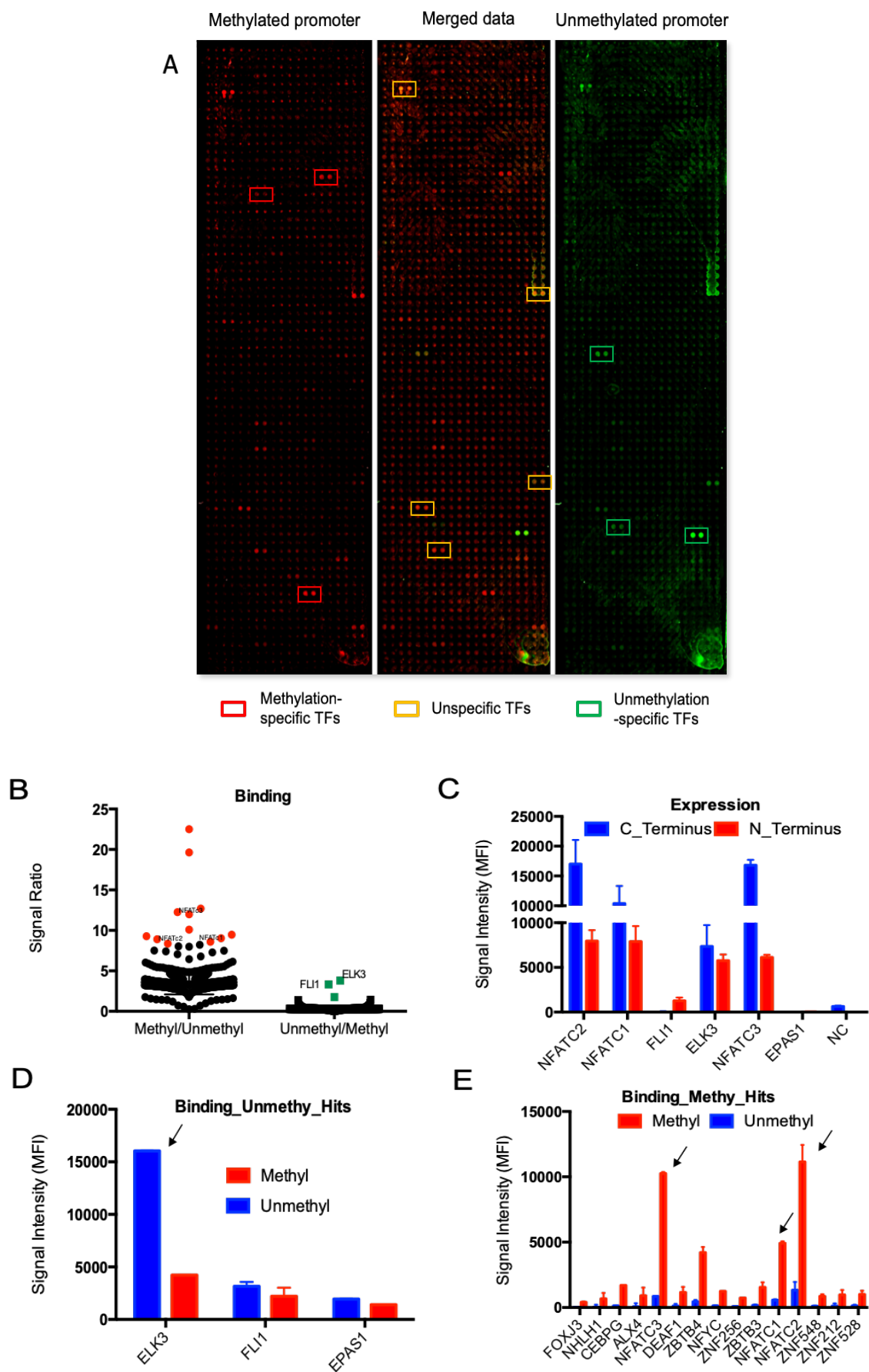
4.2.3 *In-silico* analysis of promoter sequence

In order to verify the protein-DNA interaction events, the sequence of the tested DNA fragments was analyzed by FIMO algorithm. Based on the predicted motif of TFs identified by methyl-SELEX from Yin's study [23] and the sequence of the DNA fragments used on the protein microarray (Fig.4.5A), FIMO algorithm has identified the matched TFs. Top 50 most matched TFs were shown here according to the P value, NATc1/2/3 were in the top candidate list (Fig.4.5B).
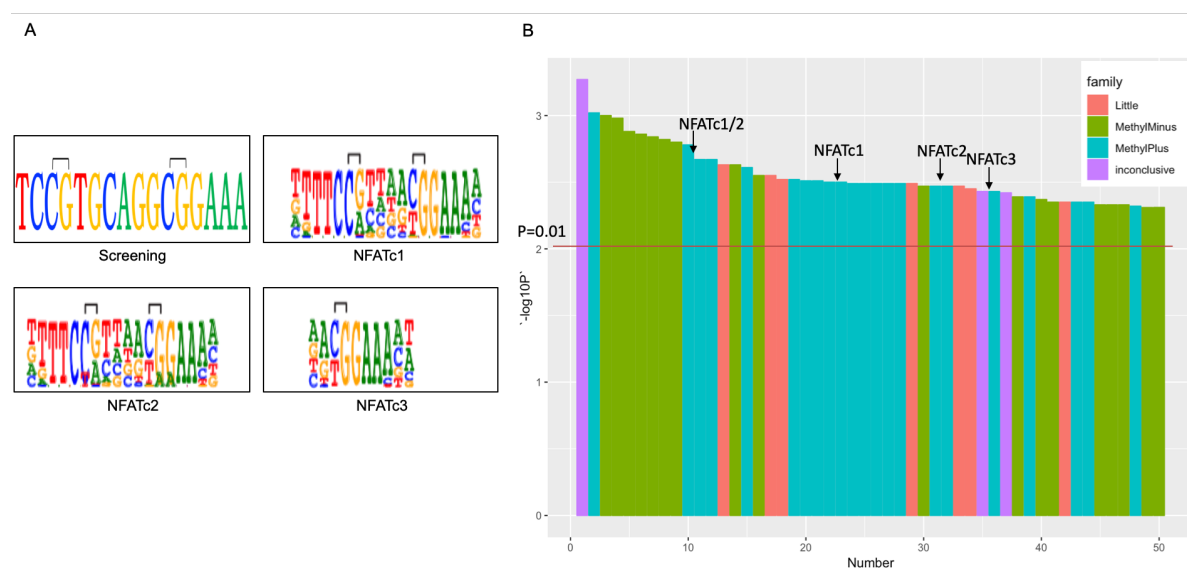
Figure 4.5 In-silico analysis verified the binding of NFATc1/2/3 to the promoter

(A) Motifs of top 3 methyl-plus transcription factors (identified by methyl-SELEX assay[23]) are shown in comparison with the sequence of screening DNA fragments. (B)Top matched TFs identified by FIMO using all the motifs of TFs identified by methyl-SELEX. Red line indicates the P value which is 0.01. NFATc1/2/3 are shown in the figure. Little/MethylMinus/MethylPlus/inconclusive refers to the little/blocking/promoting/inconclusive effect of methylated cytosine on the binding of TFs according to the data from methyl-SELEX.

4.3 The expression pattern of NFATc1 in PDAC and pancreatic cancer cell lines

As known from the 4.2 section, NFATc1, NFATc2, NFATc3 were identified as methyl-plus TFs. Consequentially, the expression pattern of 3 TFs was explored in pancreatic cancer tissues and relevant cell lines.

4.3.1 Tissue data analysis

In order to determine the expression pattern of NFATc1, NFATc2 and NFATc3 which are the methyl-plus TFs, the mRNA profiling of pancreatic tissues was firstly explored. On the Sentrix Human WG-6 array, there are 4 probes representing NFATc1, 2 probes representing NFATc2 and 5 probes representing NFATc3. However, NFATc1 was the only one to show the significant dysregulation (FDR P value < 0.01, log2FC>0.5). NFATc1 was upregulated in PDAC-patients and CP-patients groups relative to healthy-donors group (Fig.4.6A). When analyzing the TCGA data and GTEx data, NFATc1 was upregulated in PDAC-patients group (red bar) compared with healthy control group (grey bar) (Fig.4.6B). Furthermore, in-house data (Fig.4.6C) and online TCGA data (Fig.4.6D) both showed that NFATc1 was insignificant differentially expressed in different tumor stages.

Figure 4.6 NFATc1 is upregulated in PDAC

(A) The in-house mRNA profiling data showed that the mRNA expression level of NFATc1 was significantly upregulated in groups of CP and PDAC compared with N group. N refers to control, CP refers to chronic pancreatitis, PDAC refers to pancreatic adenocarcinoma; Y axis shows the normalized NFATc1 gene expression value originated from the analysis of 195 PDAC, 58 CP and 41 normal tissue samples. **** $P \leq 0.0001$. (B) The online TCGA data showed the consistent result when comparing the PDAC group with N group containing the GTEX data and the control data from TCGA. The red column indicates the PDAC group containing 179 tumor samples, the grey column indicates N group containing 171 healthy samples from TCGA and GTEx. * $P \leq 0.05$ (C) Expression dynamics of NFATc1 in different PDAC stages. According to the in-house mRNA microarray data, the mRNA expression level of NFATc1 didn't show difference in 5 different stages of PDAC. The online TCGA data (D) showed the same result.
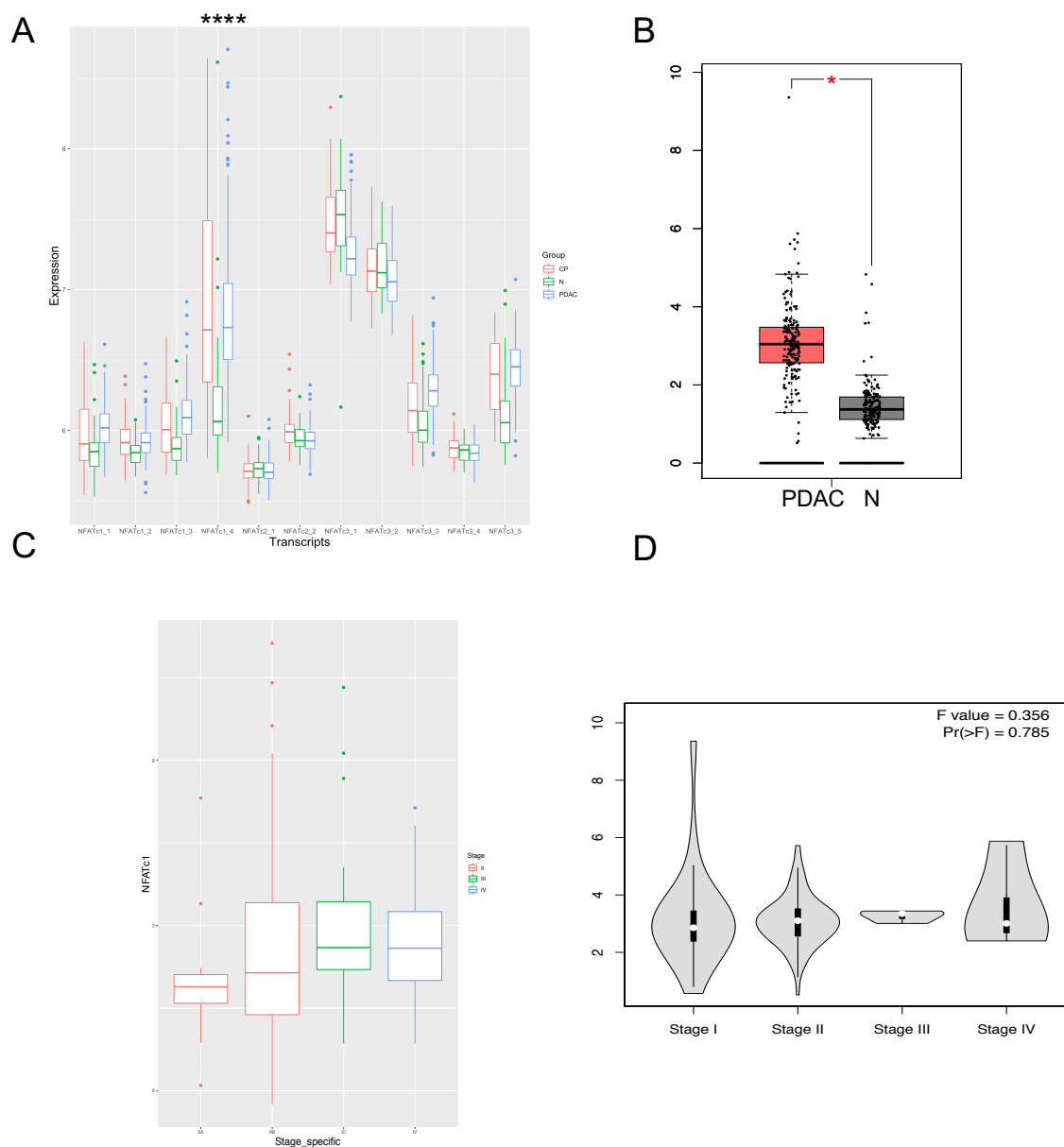
4.3.2 The expression of NFATc1 in pancreatic cancer cell lines

The mRNA and protein expression of NFATc1 were explored in 6 pancreatic cancer cell lines including Panc1, BxPC3, Miapaca2, AsPC1, Capan1, and Suit2, as well as one control healthy cell line HPDE. The q-PCR result indicated that the expression of NFATc1 was significantly higher in Panc1 when comparing to HPDE (P< 0.0001, Fig.4.7A). RNA was isolated from two batches of cells, and the experiments were repeated twice independently. The western blot result indicated that the expression in protein level was consistent with the RNA level (Fig.4.7B).

A

B
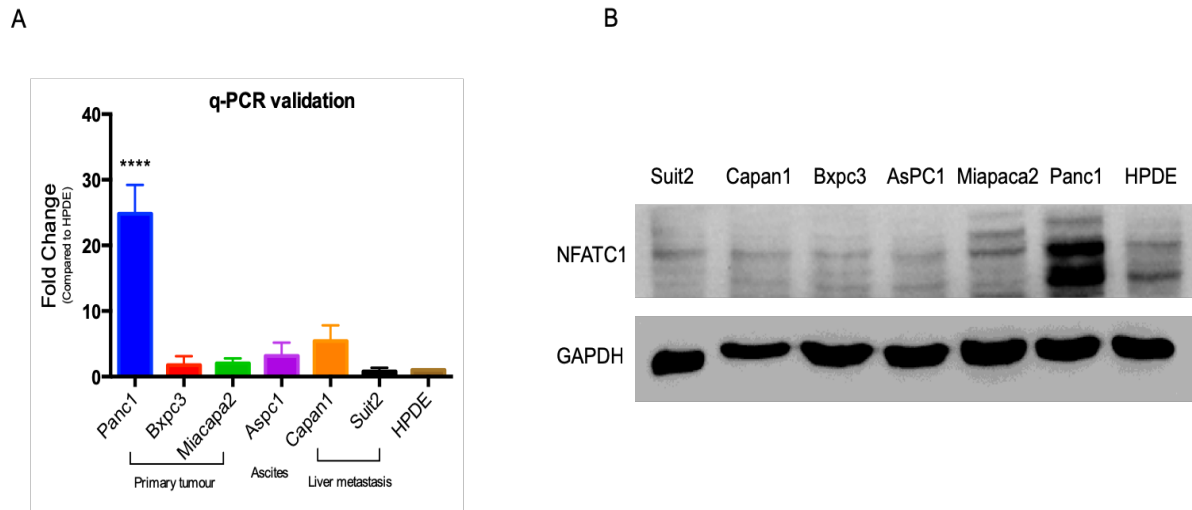
Figure 4.7 NFATc1 expression analysis in PDAC cancer cell lines

(A) Quantification of NFATc1 in mRNA level. The mRNA expression of NFATc1 in 6 pancreatic cancer cell lines and the healthy control cell line HPDE. The q-PCR result showed the expression of NFATc1 in Panc1, Miapca2, Aspc1, Capan1, and suit2, when compared with HPDE, ****$P < 0.0001$. (B) Quantification of NFATc1 in protein level. The protein expression of NFATc1 was analyzed in 6 pancreatic cancer cell lines and HPDE. The western blot result showed the consistent result as the q-PCR result.

4.4   The Functional study of NFATc1 *ex vivo*

As known from result 4.3, NFATc1 was significantly upregulated in PDAC and CP groups relative to the control group. Moreover, NFATc1 was significantly upregulated in Panc1 cell line compared with HPDE cell line and it was also expressed in Miapaca2 cell line. As such, viability assay, migration assay, colony assay, and apoptosis assay were performed in Panc1 and Miapaca2 in order to understand the functional role of NFATc1 in pancreatic cancer development.

4.4.1 NFATc1 downregulation inhibited the cell viability

The siRNA-mediated knockdown (KD) cell model was applied for the viability assay. The result of western blot confirmed the knockdown of NFATc1 in Panc1 and Miapaca2 (Fig.4.8A). The cell viability was studied using resazurin assay. Via

determining relative fluorescence units (RFUs) between control and KD treatment groups, the results showed that the downregulation of NFATc1 inhibited the cell viability significantly and this inhibition effect was increasing in the process of culturing time (Fig.4.8B).
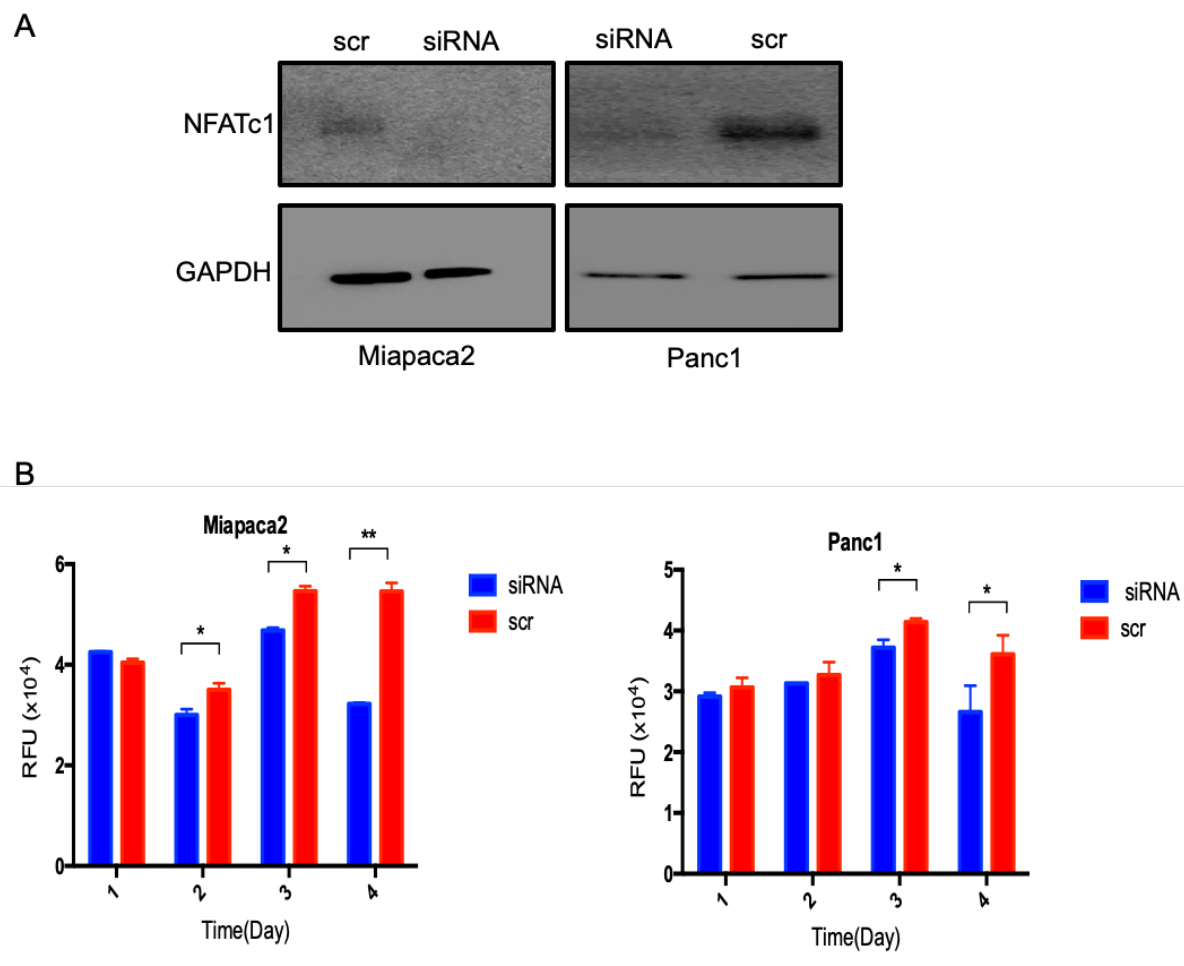
Figure 4.8 NFATc1 downregulation decreased the cell viability

(A)The knockdown of NFATc1 in Panc1 and Miapaca2 cell lines. The western blot result showed that the expression of NFATc1 was reduced upon the siRNA-mediated KD of NFATc1. (B) The viability assay result in NFATc1-KD Panc1 and Miapaca2 cell lines. The data was collected in a range of 4 days. The result showed that the viability was inhibited upon the knockdown of NFATc1. $*P \leq 0.05$, $**P \leq 0.01$, and scr indicates the scrambled siRNA.

4.4.2 NFATc1 downregulation decreased the cell migration

To investigate the effects of siRNA-mediated knockout of NFATc1 in Panc1 and Miapaca2 cell lines on the migration ability of cells, trans-well plates were used in the study. After the crystal violet staining, it was obviously shown that the NFATc1 downregulation decreased the cell migration (Fig.4.9A). All the images were analyzed by image J and the result indicated that NFATc1 significantly promoted the cell migration in Panc1 and Miapaca2 cell lines (Fig.4.9B).
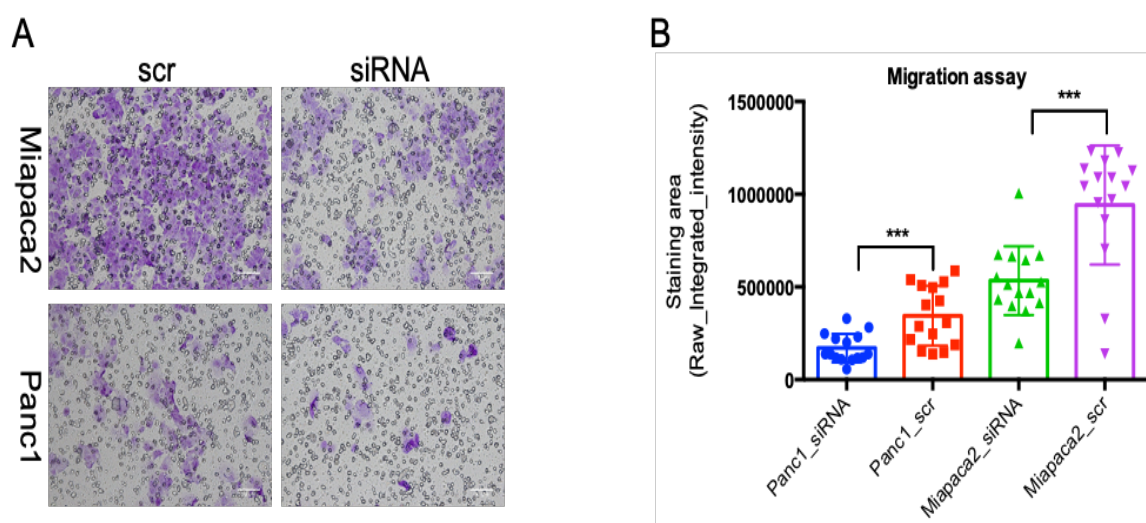
Figure 4.9 NFATc1 downregulation decreased the migration assay

(A)The crystal violet staining of transwell with siRNA-mediated cells. siRNA refers to siRNA-mediated knockdown against NFATc1 in Panc1/Miapaca2 cells, and scr refers to the negative control which are the cells transfected by scrambled siRNA. Scale bar=50µm. (B) The migration ability was decreased significantly upon the KD of NFATc1. The staining area was analyzed and each dot in the bar represented the staining (quantified by raw integrated intensity indicated in image J). The result of image analysis showed that ***$P \leq 0.0001$.

4.4.3 Apoptosis assay

The siRNA-mediated knockdown of NFATc1 in Panc1 and Miapaca2 was examined, and the effect of NFATc1 on the apoptosis was studied by AnnexinV-PI assay. Cells which showed the signals of AnnexinV-positive and PI-negative were in the stage of the early apoptosis, while cells which showed AnnexinV-positive and PI-positive

signals were in the stage of the late apoptosis. The cell events in different stages were counted respectively (Fig.4.10A). The results showed that in Panc1, the knockdown of NFATc1 significantly increased the percentage of cells in both early and late apoptosis stage. While in Miapaca2, the knockdown of NFATc1 only increased the percentage of cells in late apoptosis stage significantly (Fig.4.10B).
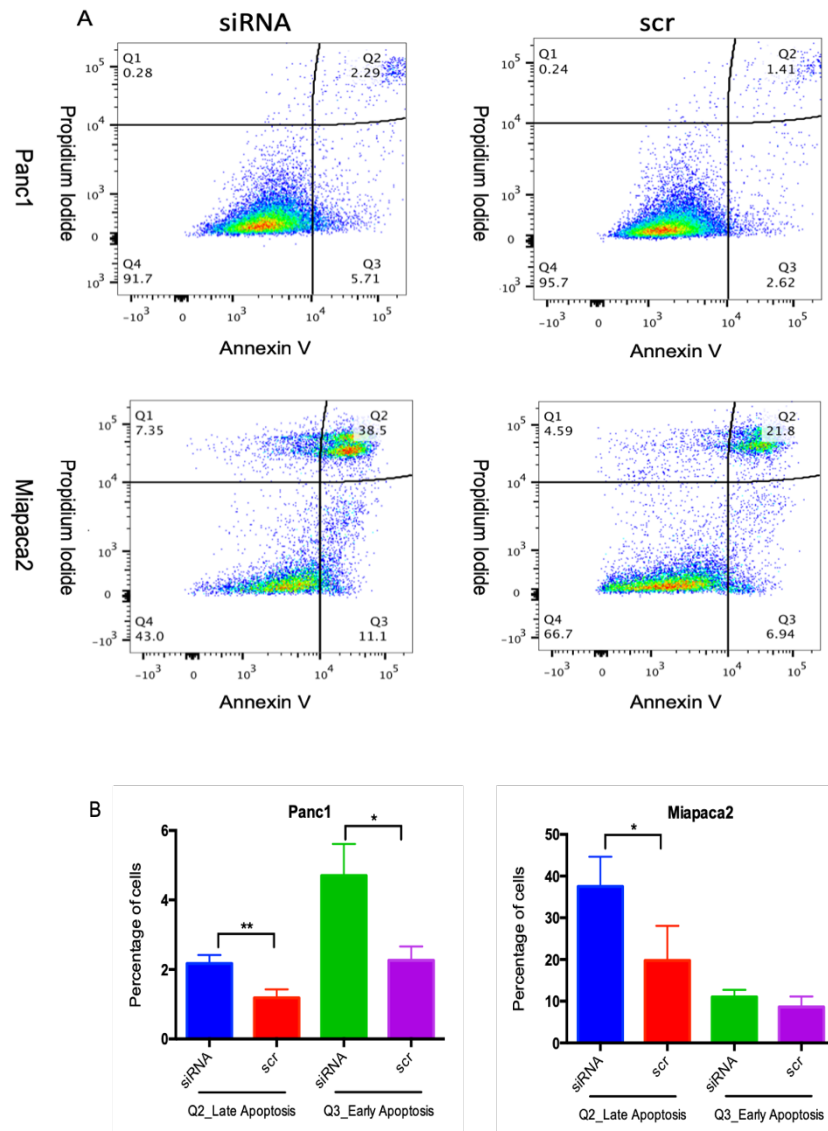
Figure 4.10 Apoptosis assay

(A) FACS analysis of apoptosis assay. Representative flow cytometry scatter plots of propidium iodide (PI) (Y axis) vs Annexin-FITC (X axis). Cells shown in Q1/Q2/Q3/Q4 were in the dead /late apoptosis /early apoptosis /live stage. (B) The apoptosis assay was performed on Panc1 and Miapaca2 cell lines with siRNA-mediated knockdown of NFATc1. The percentage of cells in late or early apoptosis stage was calculated respectively and siRNA group was compared with the control group. *$P \leq 0.05$, **$P \leq 0.01$.

4.4.4 Colony assay

The stable overexpression (OE) and knockout (KO) of NFATc1 in Panc1 and Miapaca2 were studied in colony assay. Western blot confirmed the reliability of both KO and OE (Fig.4.11A, B). The colonies were stained by crystal violet and counted well by well. The results showed that the colony formation ability was decreased upon the knockout of NFATc1 in Panc1 and Miapaca2 when compared with the control group (Fig.4.11C). In terms of compensation experiments, the number of colonies was increased upon the OE of NFATc1 when compared with the control group (Fig.4.11D).
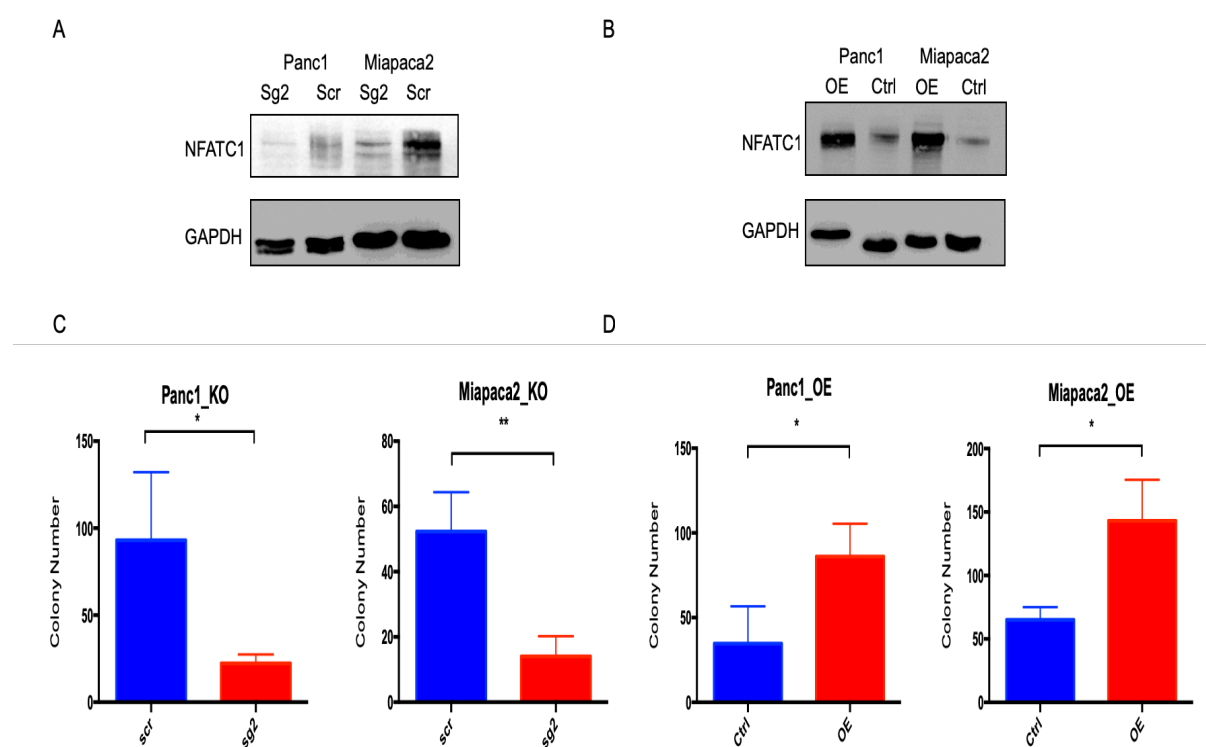
Figure 4.11 Colony assay

(A) CRISPR-Cas9 mediated KO of NFATc1 was confirmed by western blotting. (B) The NFATc1-OE cell lines were established by using the lentivirus tool and the stable OE cell lines were confirmed by western blotting. The colony assay was performed using (C) the stable KO cell lines and (D) stable OE cell lines. The crystal violet staining was performed 3 weeks after the seeding. The colonies were counted and shown in the bar blot, *P ≤ 0.05, **P ≤ 0.01.

4.5   The pathway and candidate analysis of NFATc1

As indicated in the 4.4 section, the downregulation of NFATc1 inhibited the cell viability, decreased the cell migration, blocked the ability of colony formation, and increased the percentage of apoptotic cells in Panc1 and Miapaca2 cell lines. It's concluded that NFATc1 exerted an oncogenic function and showed the importance in the process of carcinogenesis. Subsequently, in order to investigate the involved pathway of NFATc1 and identify the targets of NFATc1, a genome-wide microarray-based transcriptomic profiling was conducted in three PDAC cell lines (Panc1, Miapaca2 and AsPC1) that were pre-transfected with siRNA against NFATc1 and scramble RNA.

4.5.1 Pathway analysis

A gene set enrichment analysis was performed on the basis of the Illumina HT12 transcriptomic profiling collected from control cells and NFATc1-knockdown cells. When the analysis was referred to the 50 hallmark gene sets (Fig. 4.12A), multiple gene sets which are associated with Myc targets, P53 pathway, E2F targets, G2M checkpoint, and DNA repair were significantly enriched in the control cells compared with NFATc1-knockdown cells. When the analysis was referred to the KEGG pathway (Fig. 4.12B), various metabolism related pathways, including drug metabolism (cytochrome P450 and other enzymes), purine metabolism and pyrimidine metabolism were highly enriched in control cells compared with NFATc1-KD cells. Besides, pathways associated with transcription, DNA replication, spliceosome, mismatch repair, none small cell lung cancer, cell cycle, ribosome, acute myeloid leukemia, ERBB signaling pathway and proteasome were also detected with higher enrichment

in control cells. The result of pathway analysis confirmed the oncogenic role of NFATc1 in pancreatic cancer cell lines.
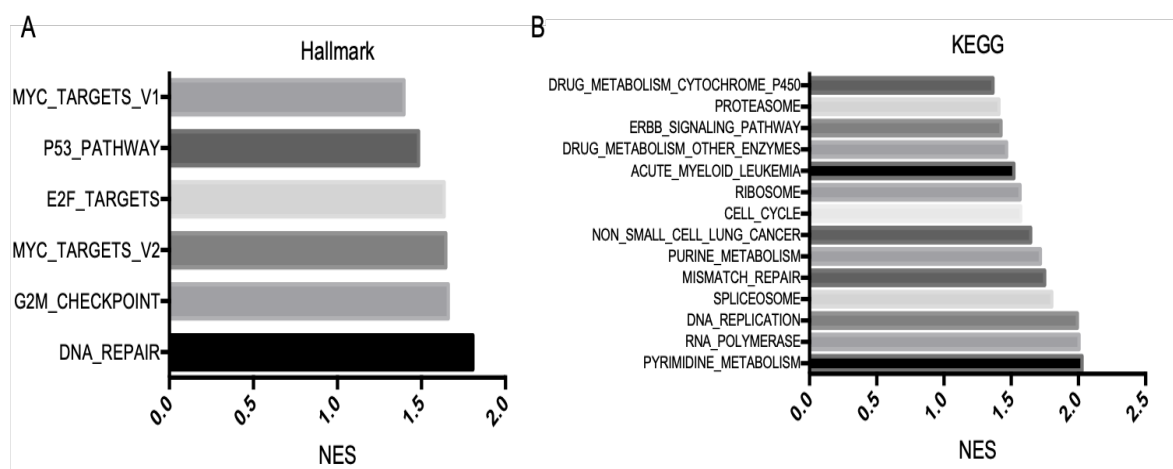
Figure 4.12 Pathway analysis

(A) Hallmark gene set enrichment analysis and (B) KEGG gene set enrichment analysis based on the transcriptomic profiling. Both analysis approaches predicted pathways and activities that were significantly enriched in control cells compared with NFATc1 knockdown cells. The normalized enrichment score (NES) was used to compare analysis results across gene sets.

4.5.2 Identification of target genes

To further investigate the targets of NFATc1 and methylation-dependent transcriptional regulation of NFATc1, differential expression analysis was primarily performed between control cells and NFATc1-knockdown cells. Top 30 candidates that showed consistently downregulated expression upon the KD of NFATc1 in Panc1, Miapaca2 and AsPC1 cell lines were listed in the heatmap (Fig.4.13A). Afterward, it's identified that 3 genes (ALDH1A3, ALDH3A1, AGR2) out of 30 downregulated genes were upregulated with hypermethylated probes in promoter regions in PDAC tissues compared with healthy tissues. (Fig.4.13B). Subsequently, methylated/unmethylated motif of NFATc1 and sequence of promoter(-2000bp/500bp) of 3 candidate genes were matched based on FIMO algorithm to identify potential binding sites. In the promoter region of ALDH1A3 (-1035bp/893bp), multiple methylated and unmethylated binding sites of NFATc1 were identified (Fig.4.13C). Moreover, the methylation data of PDAC tissues was also taken into consideration. Ultimately, ALDH1A3 was identified as a potential methylation-dependent target of NFATc1.
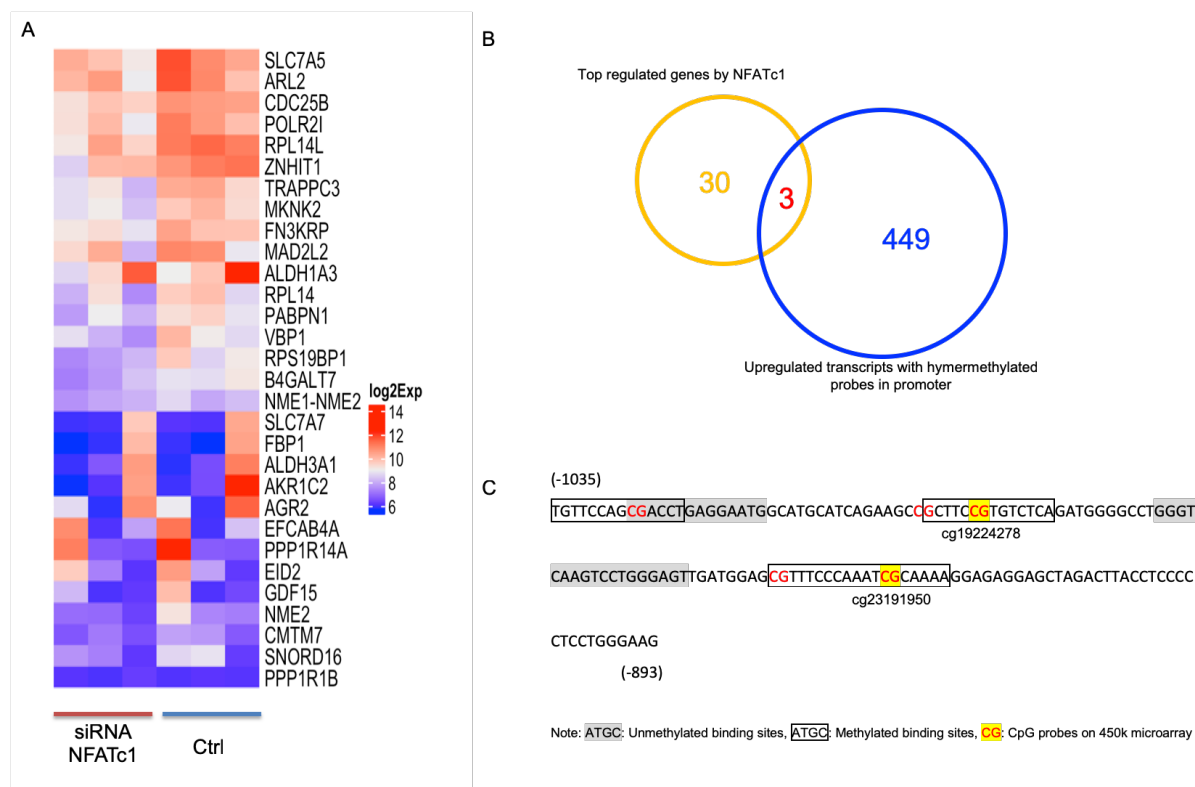
Figure 4.13 ALDH1A3 was identified as a target of NFATc1

(A) Heatmap showing the genes that were consistently down-regulated upon the knockdown of NFATc1 in 3 pancreatic cancer cell lines. Each column represents the results obtained in Panc1, Miapaca2 and AsPC1 (left to right). In addition, transfection with an siRNA of scrambled sequence are shown here as control (Ctrl). (B) The overlap between the top upregulated genes upon the knockdown of NFATc1 in 3 pancreatic cancer cell lines and upregulated transcripts with hypermethylated probes in promoter region from PDAC tissues. (C) FIMO identified significantly matched methylated/unmethylated NFATc1 binding sites in the promoter region of ALDH1A3. The CpG sites were colored in red.

4.6   Validation of ALDH1A3

It is shown in the 4.5 section, multiple cancer associated pathways were enriched and ALDH1A3 was selected as a candidate which was downregulated upon the knockdown of NFATc1. In addition, the potential methylated binding sites of NFATc1 were identified in the promoter region of ALDH1A3. It's hypothesized that ALDH1A3 was upregulated by NFATc1 in pancreatic cancer tissue cell lines due to the binding of NFATc1 in the promoter region of ALDH1A3 though the binding sites were methylated. To follow-up this hypothesis, the mRNA expression and methylation level of ALDH1A3 in tissue and cell lines (Panc1 and Miapaca2) was analyzed. Moreover, the expression of ALDH1A3 was assessed upon the regulation of NFATc1.

Furthermore, the cells were treated with 5-Azacytidine and the expression of ALDH1A3 was checked. Next, the promoter activity of the predicted region was analyzed by luciferase assay with the methylated or unmethylated construct. Subsequently, chromatin immunoprecipitation was applied to check if NFATc1 could bind in the promoter region of ALDH1A3 *ex vivo*.

4.6.1 Validation of ALDH1A3 in mRNA level

In order to validate ALDH1A3 as a target of NFATc1 in Miapaca2 and Panc1 cell lines, NFATc1 knockout (KO), knockdown (KD), overexpression (OE) cell models were applied to study the expression of ALDH1A3 in mRNA level. As shown in Fig.4.14 A, ALDH1A3 was downregulated when NFATc1 was suppressed in Panc1 and Miapaca2. While only in Panc1, ALDH1A3 was upregulated when NFATc1 was overexpressed. That means that NFATc1 positively regulated ALDH1A3. Additionally, in tissue samples, ALDH1A3 was upregulated in PDAC group (Fig.4.14B). While in cell line samples, ALDH1A3 was upregulated in HPDE cell line comparing with Miapaca2 and Panc1 (Fig.4.14C).
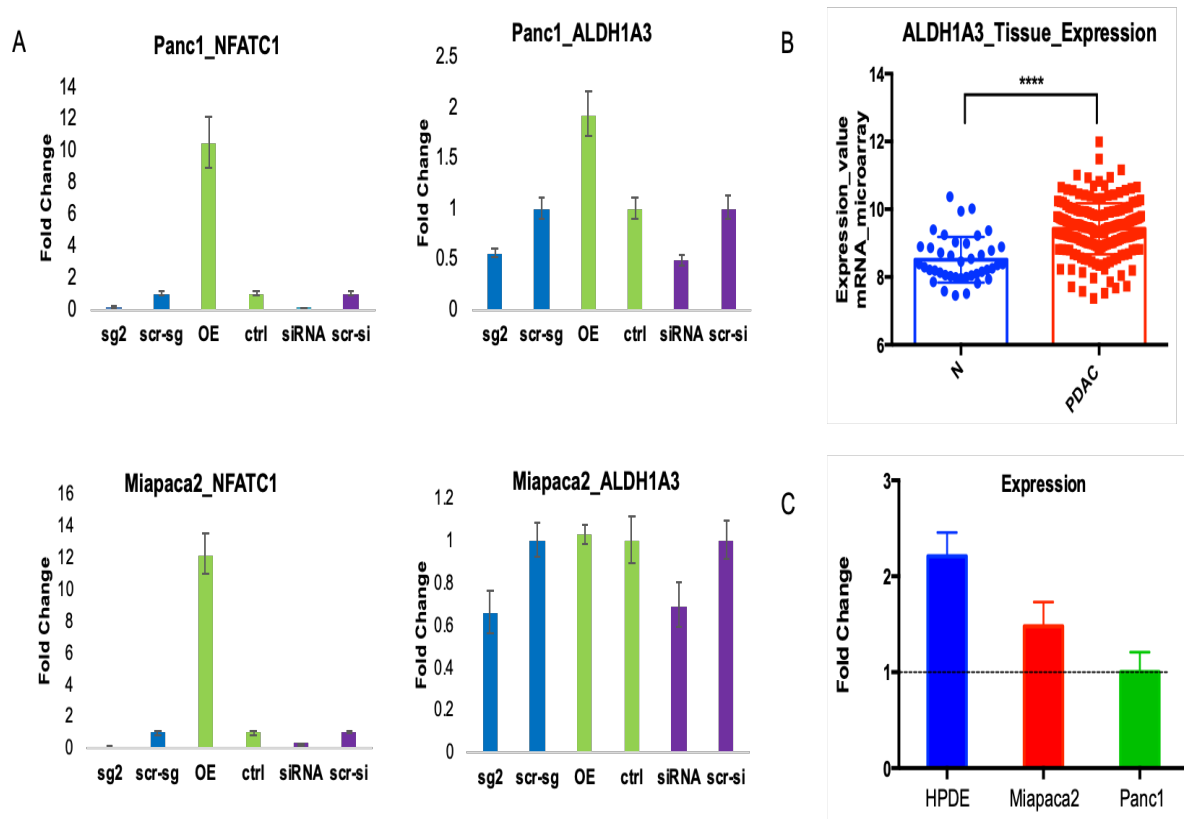
Figure 4.14 The expression of ALDH1A3 was regulated by NFATc1 in Panc1 and Miapaca2

(A) NFATc1 positively regulated the expression of ALDH1A3. q-PCR results showed the expression of NFATc1 and ALDH1A3 in knockout (blue bar), knockdown (purple bar) and overexpression (green bar) cell models. Sg2 refers to the sgRNA against NFATc1, scr-sg refers to the scramble sgRNA; OE refers to the overexpression of NFATc1, ctrl refers to the empty construct; siRNA refers to the siRNA against NFATc1, scr-si refers to the scramble siRNA. (B) The mRNA expression of ALDH1A3 in PDAC and N group, ****$P \leq 0.0001$. (C) Q-PCR results presented the mRNA expression of ALDH1A3 in HPDE, Miapaca2 and Panc1 cell lines.

4.6.2 Validation of methylation level

In order to validate the methylation level of NFATc1 binding sites in the promoter region of ALDH1A3, 450k microarray data of tissues and cell lines was analyzed. The promoter of ALDH1A3 was deciphered in Fig.4.15A. For the two CpG sites in the promoter region of ALDH1A3, the results showed that beta value was significantly higher in PDAC group and it's also significantly higher in Panc1 cell line (Fig.4.15 B, C).
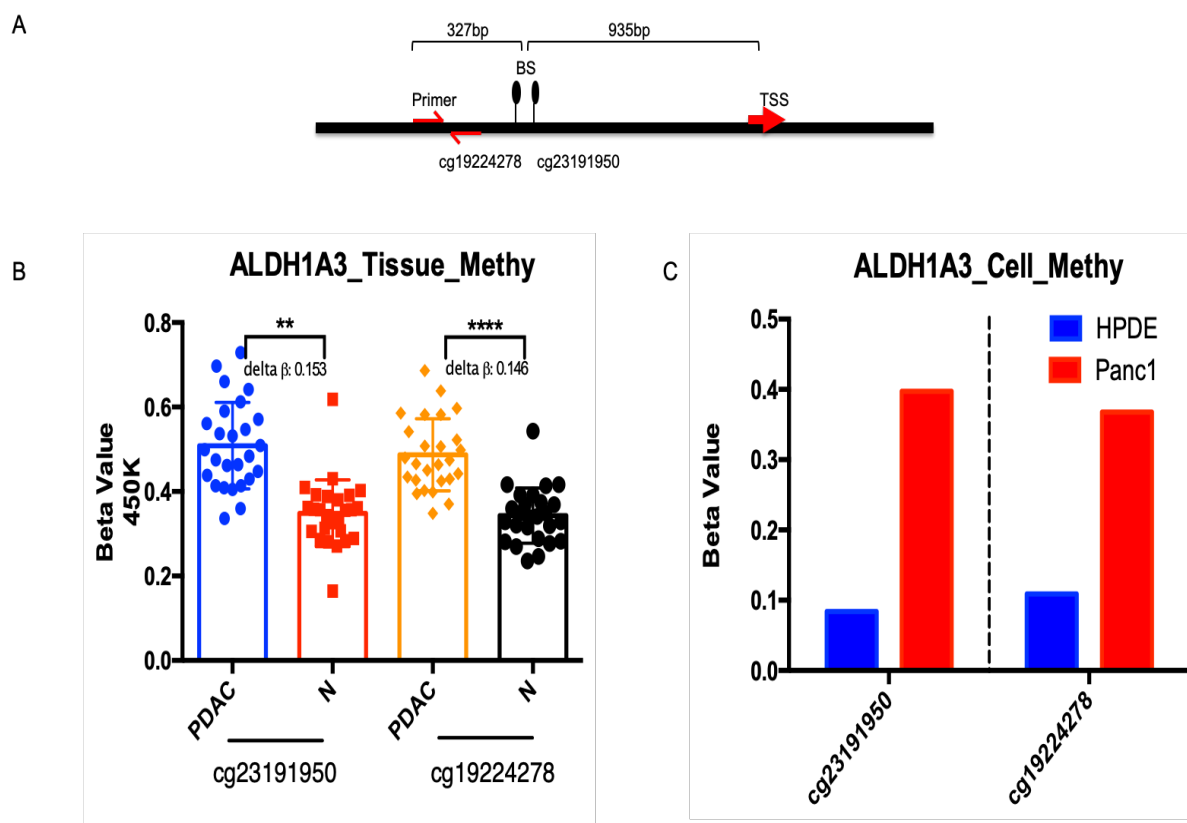
Figure 4.15 The promoter of ALDH1A3 was hypermethylated in tissue and cell lines

(A) The promoter illustration of ALDH1A3. BS refers to the predicted methylated binding sites of NFATc1. cg19224278 and cg23191959 are the CpG probes covered by 450 microarrays and the predicted binding sites were located in the position of these 2 CpG sites. TSS refers to the transcription starting site, and Primer indicated the site of ChIP-PCR primer. (B) cg19224278 and cg23191959 was hypermethylated in PDAC group. ****P ≤ 0.0001, **P ≤ 0.01. (C) The beta value of cg19224278 and cg23191959 was almost 4 folds in Panc1 cell line.

4.6.3 Analysis of demethylated cell samples

To investigate the methylation effect on the expression of ALDH1A3 and the binding of NFATc1, the demethylation in pancreatic cancer cell lines was performed by treating the cells with 5-Azacytidine drug. For the 5-Azacytidine treatment, the concentration was optimized. The growth curve of Panc1 treated by different concentration of 5-Azacytidine indicated that the optimal dose for the Panc1 was 2μM (Fig.4.16A), and the optimal dose for Miapaca2 was 0.5 μM(Fig.4.16B). After 3-day treatment of 5-Azacytidine drug, the demethylation effect in Panc1 treated with 2μM 5-Azacytidine

and Miapaca2 treated with 0.5µM 5-Azacytidine was confirmed by methylation specific PCR (Fig.4.16C). Moreover, for Miapaca2, the expression of ALDH1A3 was upregulated in drug treatment group compared with control group, while in Panc1, ALDH1A3 was downregulated upon demethylation in Panc1(Fig.4.16D).
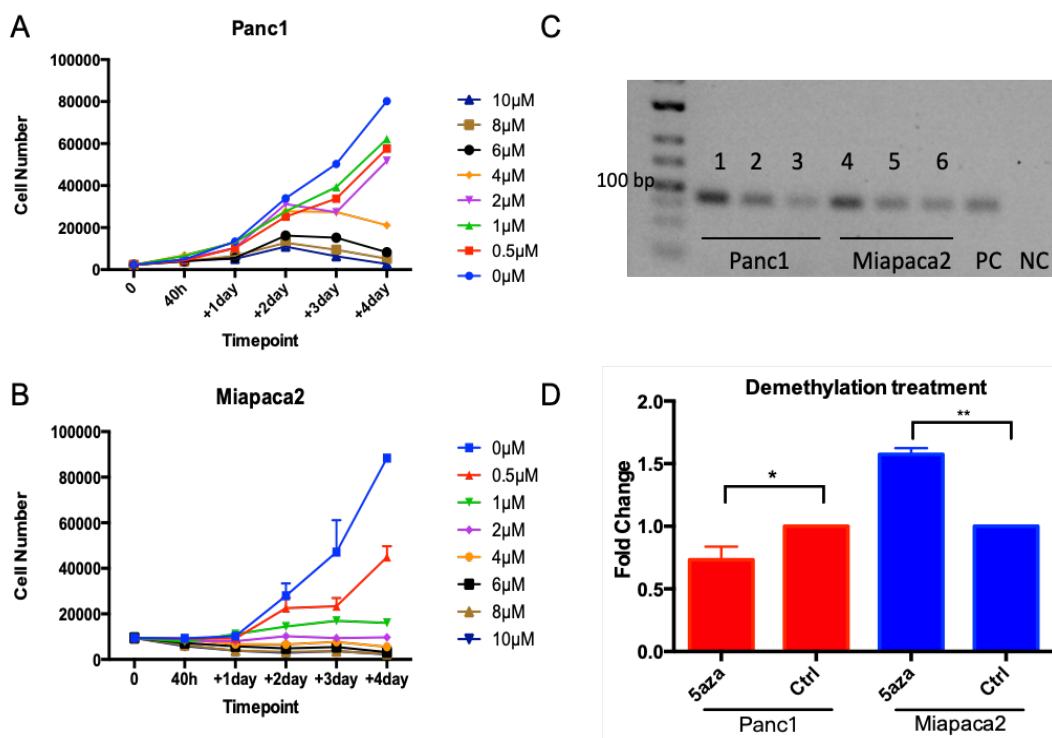
Figure 4.16 Analysis of demethylated cell samples

(A-B) The number of viable cells was determined by means of a resazurin assay. The Panc1 cells (A) and Miapaca2 cells (B) were either treated with 0 µM, 0.5 µM, 1 µM, 2 µM, 4 µM, 6 µM, 8 µM or 10 µM of the demethylating agent 5-Azacytidine. The cell viability was assessed 0 h, 40 h, 64 h, 88 h, 112 h and 136 h after the treatment. (C) Methylation specific PCR. 1: Untreated Panc1, 2: Panc1 treated with 2 µM 5-Azacytidine for 3 days, 3: Panc1 treated with 4 µM 5-Azacytidine for 3 days; 4: untreated Miapaca2, 5: Miapaca2 treated with 0.5 µM 5-Azacytidine for 3 days, 6: Miapaca2 treated with 1 µM 5-Azacytidine for 3 days; PC: positive methylated converted sample; NC: negative control. (D) The mRNA expression of ALDH1A3 was measured in cells treated with and without 5-Azacytidine by q-PCR method. **$P \leq 0.01$, *$P \leq 0.05$.

4.6.4 Luciferase assay

To investigate the promoter activity of the predicted promoter region of ALDH1A3 (-1035bp/893bp), luciferase assay was performed accordingly. The predicted promoter region was cloned into the luciferase construct followed by the *in-vitro* methylation.

Firstly, the assay was studied in NFATc1-high expression cell line-Panc1 and low expression cell line-Miapaca2. Comparing with Miapaca2, Panc1 showed significantly higher promoter activity. Additionally, the methylated promoter construct still showed significantly higher promoter activity compared with the empty promoter construct, while the methylation of the promoter decreased the promoter activity (Fig.4.17A). Next, the assay was investigated in NFATc1-knockdown Panc1 and Miapaca2 cell lines. The activity of unmethylated promoter was decreased upon the knockdown of NFATc1 in Panc1 and Miapaca2 cell lines. In terms of the methylated promoter, the activity was decreased upon the knockdown of NFATc1 in Panc1, but not in Miapaca2 (Fig.4.17B).
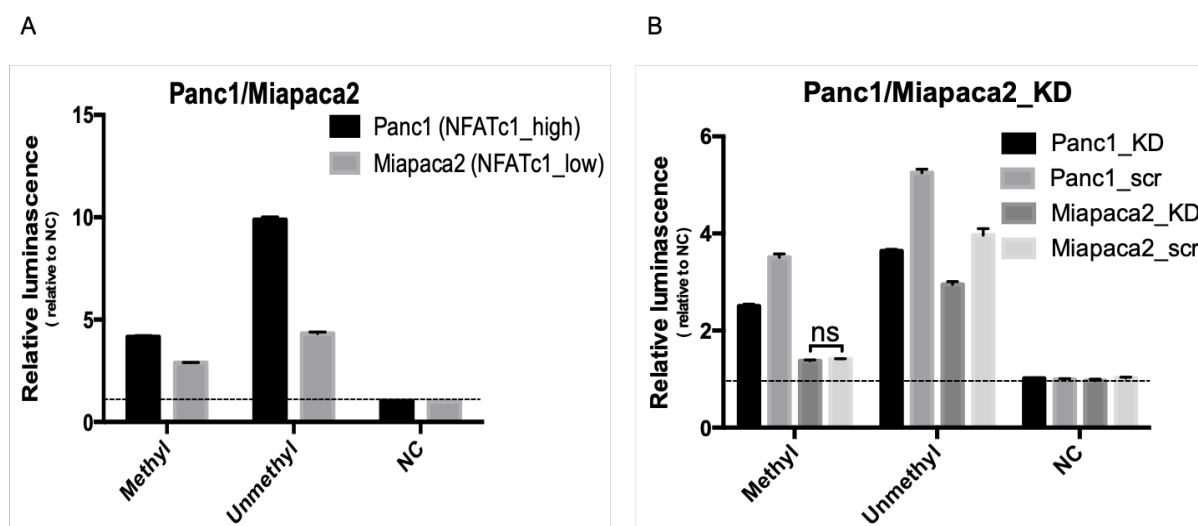
Figure 4.17 Luciferase assay

(A) The relative luciferase activity in Panc1 and Miapaca2 cell lines. After cloning ALDH1A3 promoter sequence (from -1035bp to -893bp) into a construct encoding for the luciferase gene. The luminescence was measured as an indicator of promoter activity (black and grey bars). Compared with the empty construct (NC), both *in-vitro* methylated (Methyl) and unmethylated (Unmethyl) promoter constructs showed high promoter activity in Panc1 and Miapaca2 cell lines. Moreover, there was a clear increase of activity in Panc1 in which NFATc1 expression was higher compared with Miapaca2 cell line. (A) The relative luciferase activity in Panc1 and Miapaca2 cell lines upon the knockdown of NFATc1. The activity of unmethylated promoter (Unmethyl) was decreased upon the knockdown of NFATc1 in both cell lines. KD refers to knockdown of NFATc1, and scr refers to the scramble siRNA control. As for the *in-vitro* methylated promoter (Methyl), there was a clear decrease of promoter activity in Panc1 upon the knockdown of NFATc1, while not in Miapaca2 cell line.

4.6.5 Chromatin immunoprecipitation

The direct interaction between NFATc1 and ALDH1A3 was confirmed by chromatin immunoprecipitation (ChIP). Additionally, ChIP of 5-Azacytidine treated cell samples were used to validate the methylation-dependency of NFATc1. Two primer sets were used for the ChIP-PCR. Promoter primer set was used to detect the binding of NFATc1 in the target region, and -5kb primer set located 5kb upstream of the transcription starting site was used as a negative control for ChIP PCR. The ChIP-PCR result indicated that NFATc1 showed the binding signal in the target region of ALDH1A3 in Panc1 samples, and demethylation inhibited this binding event. While in Miapaca2, no binding signal was shown (Fig.4.17).
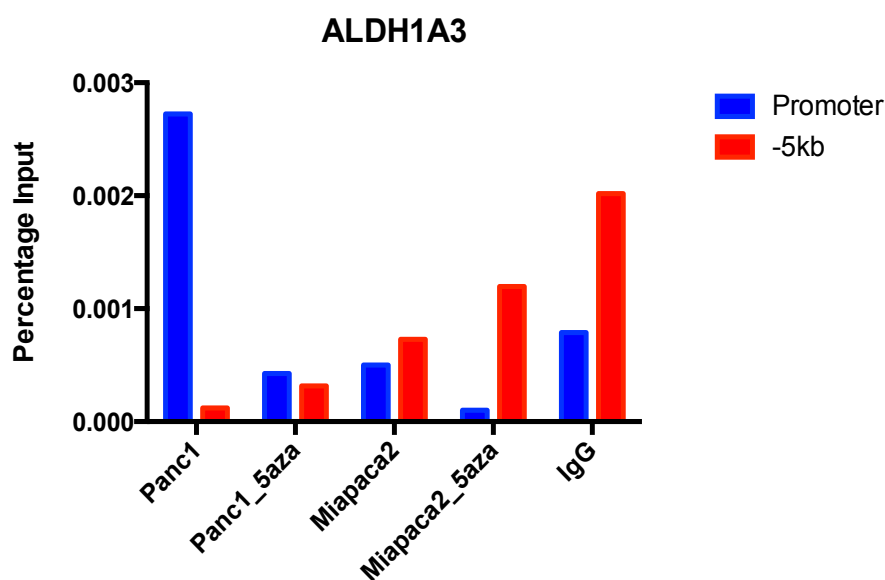
Figure 4.18 Chromatin immunoprecipitation

The binding of NFATc1 with the methylation binding sites in the promoter region of ALDH1A3 was determined by ChIP-PCR. The cells with and without 5-Azacytidine treatment were used for the ChIP. The result was indicated by percentage input. IgG was the negative control of ChIP. Promoter primer set (Blue bar) was used for ChIP-PCR, -5kb primer set (red bar) was the negative control of ChIP-PCR.

5 Discussion

5.1. Integrative analysis of methylation and expression profiling

DNA methylation is one of the key epigenetic modifications. It plays an important developmental role in multiple cancers, including pancreatic adenocarcinoma [126], colorectal cancer [127] and breast cancer [128]. The aberrant 5-methylcytosine in cancer could introduce the mutation in tumor suppressor gene, and it also induces chromosomal instability [129]. Moreover, it is widely known that methylation of cytosines is associated with a repressed chromatin state and gene repression [22]. However, more evidences suggest another scenario that the hypermethylation in the promoter region is not always associated with gene repression [130]. In order to better understand the role of DNA methylation as an epigenetic modification in regulating gene expression during the development of pancreatic adenocarcinoma, an integrative study combining global gene expression and methylation profiles is essential. The log-fold change of expression level and the mean delta beta value in the defined promoter region (-1500/+500bp) when compared PDAC group with healthy group were integrated. The integrative analysis indicates that 132 hypermethylated genes are down-regulated, while 119 genes with hypermethylation are up-regulated conversely.

For the genes which showed hypermethylation-related down-regulation, the mechanism of gene transcription regulation can be explained by the classical model, that methyl-CpG-binding domain proteins recognize methylated CpGs and recruit repressor complexes such as a histone deacetylase complex, resulting in histone modification and a more condensed chromatin structure which block the transcription [131]. However, as for these hypermethylated genes which were up-regulated, the classical model couldn't explain how this enhanced expression was modulated. Firstly, in order to understand the function of these genes, gene ontology term enrichment analysis was performed. As a consequence, genes in multiple embryonic development related processes, such as anatomical structure morphogenesis, embryonic skeletal system morphogenesis, and regulation of developmental process were enriched. Previous studies indicate that early embryo development and tumorigenesis share a

remarkable similarity [132] [133] [134]. Due to the common requirements of cellular proliferation and differentiation in oncogenic and embryonic development, epithelial-to-mesenchymal transition along with the Wnt, Hedgehog, Notch, PAP and BMP pathways are widely involved in controlling these similar processes [134]. Epigenetically, although genome-wide demethylation is observed in the development process of both embryo and cancer, the activity of DNA methyltransferases is highly intensive in embryos and tumors. Additionally, multiple genes which are highly expressed in embryonic stem cell and essential for early embryonic development are also detected in tumorigenesis, instead of in normal cells. Among 119 hypermethylated and up-regulated genes, Twist1 genes were chosen for functional protein microarray screening since they present the most intensive hypermethylation probes in the promoter areas. As for Twist1, it encodes a transcription factor involved in the regulation of organogenesis and epithelial–mesenchymal transition (EMT) pathway. In addition to embryonic development, Twist1 also plays important roles in cancer metastasis [135] and cell plasticity in breast cancer [136]. Moreover, its hypermethylation was frequently observed in pancreatic cancer and colorectal tumors [137] [138]. Concerning that multiple hypermethylated and upregulated genes exert multiple biological functions in various cancers, it's inspiring to further study whether and how their up-regulated expressions are correlated with enhanced hypermethylation during oncogenic development. To do so, it's hypothesized that certain transcription factors could recognize the methylated promoter of the candidates and thus regulate the transcription.

In terms of the method for integrative analysis, in this study, I utilized the strategy simply integrating the expression data of one gene with methylation value of its promoter. In this strategy, the methylation value is presented by the mean delta beta value in the defined promoter region (-1500/+500bp) between PDAC group and healthy group. It's noteworthy that the methylation level of promoters could vary when different strategies are adopted. Instead of calculating the mean methylation level of defined promoter region, the number of methylated probes which represent CpG sites in the defined promoter region including 1st exon, 5'UTR, TSS200 and TSS1500 could represent the methylation level of promoters. In a word, one strategy called probe integration is based on the individual probe, the other strategy called region integration is based on the mean methylation level in the defined promoter region. The region

integration method comprises a group of neighboring methylation sites, and thus indicate a more persistent methylation-based modulation across a defined region. While this method might average the methylation of all the probes in the defined promoter region and neglect the important individual probes which was associated to crucial phenotype. However, the probe integration might have a less predictive feature. In this study, the focus is to study the effect of methylation on transcription regulation, while promoter is known to be crucial to the transcription initiation of one particular gene. Therefore, the most straightforward approach is to use the predefined promoter region.

This integrative analysis has identified multiple genes which are upregulated and hypermethylated. Moreover, the functional analysis uncovered the crucial role of these genes in embryo development. Although recent studies have revealed that pathways involved in cancer progress share the embryological characteristics, the regulation mechanisms underlying the causal link between hypermethylation and up-regulated transcription remain elusive.

5.2. The identification of methylation-dependent TFs

To reveal how methylation-dependent transcription activation happens, it's hypothesized that TFs could recognize the methylated binding sites and thus regulate the transcription. TFs are widely believed to recognize and bind to specific regulatory sites on a genome-wide level. Emerging evidences indicate that DNA methylation has an impact on the specificities of TF binding and regulates diverse transcription events [130]. Our integrative analysis has revealed a number of genes showing hypermethylation and upregulation in PDAC. To identify the impact of DNA methylation on the gene transcription, the knowledge of TF-DNA (methylated and unmethylated DNA) binding comprising the aspects of biochemical affinity, genomic context for TF-DNA binding and protein-protein interaction and synergy is required. As the first step, the generation of protein microarray covering over 670 different DBDs was followed by the protein-DNA interaction screening. In this study, promoter region of Twist 1 was applied in the TF protein microarray assay.

Firstly, the quality of protein expression on microarray was examined. More than 97% DBDs were successfully expressed. In this study, successful expression event on the microarray is based on epitope signals against fusion tags on N- and C-terminus, especially V5-tag that is fused in the C-terminus of DBDs to guarantee the complete translation. However, in some cases, either only one terminus signal or differential signal intensities between C-terminus and N-terminus was observed. As reported previously [109], the successful tag epitope-antibody bindings are restricted with protein folding on the array and the attachment of tagged DBDs to the solid support. Moreover, in most cases, the interaction between the protein and DNA is based on the interaction between the epitope of the protein and DNA, while not the complete structure of the protein.

Secondly, the result of protein-DNA interaction was analyzed. Top 15 methyl-plus TFs and top 3 methyl-minus TFs were selected for detailed check. This result supports the hypothesis that DNA methylation has an impact on the specificities of TF binding and TFs lacking the methyl-binding domains are also able to bind with methylated motif. ELK3, FLI1 and EPAS1 are identified as methyl-minus TFs which preferentially bind with un-methylated motif. Interestingly, ELK3 and FLI1 both are in ETS family. Another methyl-minus TF EPAS1 contains a basic-helix-loop-helix (bHLH) domain, which is also named HIF-1alpha-Like Factor or HIF2a. Additionally, it's reported to induce genes expression by oxygen [139]. Comparing with the DNA sequences loaded on the microarray, 5'-TCCGTG-3' is identified with high similarity as the core DNA sequence of EPAS1, which was 5'-TACGTG-3' within the hypoxia response element (HRE) of promoter. Functionally, HIF1a is reported to increase invasiveness and metastasis by directly regulating Twist1 [140], and moreover, HIF1a and EPAS1 are structurally similar in binding motifs. A recent study from the methyl-SELEX has consistently identified that the binding of ETS-family and bHLH family TFs are generally blocked by mCpG [23]. On the contrary to the methyl-minus TFs, methyl-plus TFs, including NFATc1, NFATc2 and NFATc3 preferentially bind with methylated motif. NFATc1,2,3 share homology with each other and all belong to nuclear factor of activated T-cells (NFAT) family, which plays a crucial role in inducible gene transcription during immune response. In contrast to methyl-minus TFs, the methyl-SELEX reveals that NFAT family TFs prefers to bind with mCpG-containing

sequences [23]. Additionally, FIMO analysis has identified the consensus sequence of NFAT family TFs in the promoter region of Twist1.

In recent years, multiple methods have been applied to study the impact of methylation on the TF binding. In 2013, it's revealed that DNA methylation presents distinct binding sites for human transcription factors by the means of protein microarray [89]. They have identified numerous human TFs across various subfamilies that showed mCpG- and sequence-dependent binding activity. This is the first sophisticated study after prevalent phenomenon was discovered [141]. However, there are some limitations in this work. Firstly, only 150 CpG-containing motifs with known sequences are surveyed in the study, and none of novel TF motifs are identified. Secondly, the motif used in this study are with short length and thus no TF heterodimers are formed properly. Thirdly, this protein microarray strategy produces ~27% of false positives and certain false negative. Fourthly, methylated CHG, CHH sequences or hemimethylated DNA motifs that are available *in vivo* are not studied. In 2017, the impact of cytosine methylation on DNA binding specificities of human transcription factors has been systematically studied by means of methylation sensitive SELEX [23]. This is the most systematical study aiming to discover the motif of human TFs with the impact of cytosine methylation, and the coverage of individual TFs is considerably higher than previous systematic studies [142] [27]. In their study, methylation-sensitive SELEX is used to explore the binding affinity of 542 human TFs. Compared to the other methods, SELEX is able to detect longer binding motif and identify the novel motifs based on the high complexity of the input library. However, some limitations should be considered. Firstly, the method mainly measures enrichment of sequences but not the affinity of binding. Additionally, there might be bias introduced by PCR in each round of enrichment because of different DNA context.

Briefly, in this study, I utilized the strategy of combining the cell-free based protein expression with the functional screening on the protein microarray. In previous protein microarray-based study, it is a tedious work to express, purify and spot multiple proteins separately. In this study, more than a thousand protein could be expressed on the microarray simultaneously by means of cell-free expression system. This high-throughput expression platform highly boosts the screening efficiency of functional protein microarray. Novel DNA-binding activities could be discovered by this method.

Furthermore, unlike DNA microarray-based protein-DNA interaction assay and SELEX assay, the DNA fragments used in our study have the same sequence as one piece of genome. In this way, the protein-DNA interaction could partly mimic the *in vivo* DNA binding activities. However, this method also has some limirations: Firstly, the functional protein microarray couldn't quantify the affinity of a protein to double-stranded DNA (dsDNA), and the screening results only indicate whether the proteins could bind with DNAs. In this study, MeCP2 and MBD1 are used as positive controls on the microarray for the methylation-dependent TF bindings. As a consequence, the very week binding signals in the competition screening are probably due to the low affinity of these proteins to methylated dsDNA on the microarray. Secondly, this method lacks the ability to discover novel TF motifs as that SELEX demonstrates, and the genomic DNA fragments used exclusively here were from the integrative analysis of methylation and expression profiles. Despite the advantage of supporting long DNA fragments on the microarray, this approach fails to localize the binding site in the long DNA fragment since DNA of only 8-mer is enough for the TF binding.

My findings demonstrate that methylation has an impact on the binding of TFs. Consistent with the results from Yin et al. in 2017[23], It's discovered that NFATc1, NFATc2 and NFATc3 are methyl-plus TFs which preferentially bind with methylated motifs, while ELK3, FLI1 and EPAS1 are identified as methyl-minus TFs which preferentially bind with unmethylated motifs. This study provides a reliable framework to better understand the binding activities of TFs. Furthermore, efforts are put on the biological significances of methylation-dependent NFAT family TFs since methyl-SELEX-based systematic investigations have revealed the binding motif of 542 human TFs that are cytosine methylation-dependent, while the mechanisms of DNA methylation-dependent regulation are still not clear yet.

5.3. NFATc1 exerts the oncogenic role in pancreatic cancer cell lines

Nuclear Factor of Activated T Cells 1, NFATc1 or NFAT2, was firstly identified in activated T cells which acts as an inducible nuclear factor via binding the iterleukin-2(IL-2) promoter [143]. With more emerging evidences, it's revealed that genes expression in NFAT family are not limited in T cells. Even though NFAT proteins are

principally characterized in immune cells, they function in diverse cell types including non-T cells of immune system and non-immune cells [144] [145]. In this study, new evidences of the functional role of NFATc1 in pancreatic cancer cell lines are provided. From the functional protein microarray screening, NFAT family TFs including NFATc1,2,3 were identified as methyl-plus TFs. The NFAT family consists of five members: NFAT1,2,3,4,5, and NFAT1-4 are regulated by calcium signaling. All NFAT proteins have a highly conserved DNA-binding domain, which explains that NFATc1-3 present binding signals to the same methylated promoter. During T-cell activation, most cytokine genes are regulated by NFAT proteins. Moreover, it is clearer now that NFAT proteins are also involved in many other signaling pathways and their target genes control cell-cycle progression and activation-induced cell death [146]. However, more evidences are needed. On the basis of NFAT family characteristic of binding with methylated motif, the identification of methylated targets of NFATc1 will lead to a full understanding on how NFAT family TFs function in non-immune cells.

Initially, the expression of NFAT family TFs was investigated in pancreatic cancer tissues, chronic pancreatic tissues, healthy pancreatic tissues and also pancreatic cancer cell lines. The in-house data indicates that the expression of NFATc1 is up-regulated in PDAC and CP tissues, either than healthy tissues. However, no PDAC stages-related expression variations is observed. Concerning pancreatic cancer cell lines, NFATc1 is detected in the cell lines originating from primary tumor (Panc1, BxPC3 and Miapaca2), ascites (AsPC1) and from liver metastasis (Capan1 and Suit2). Interestingly, compared with the healthy cell line HPDE, the expression of NFATc1 is significantly upregulated in Panc1. Notably, NFATc1 is comprised of five mRNA transcript variants known to encode three different protein isoforms. The mRNA encoding isoform C (mRNA variant 3) was the most expressed followed by isoform B (mRNA variant 5). While, Isoform A (mRNA variant 1) and mRNA variants 2 and 4 made up less than 1% of the total NFATc1 expressed [147]. From the results of western blot, no single sharp band was observed. Instead, multiple protein isoforms were detected. In recent studies, besides inflammatory cell lines, NFATc1 is widely expressed in various tumor cells and promotes the invasion of many tumor cells from different cancer type. In Tsukasa's study, NFATc1 is detected in a small proportion of tumor cells in human carcinoma specimens, and overexpression of NFATc1 has promoted cancer cell invasion and caused associated changes in cell morphology

[148]. In colon cancer and pancreatic cancer, NFATc1 induces the expression of c-Myc and cyclinD, which facilitates TGFβ-promoted cell growth [149] [150]. On the contrary, inhibition of NFATc1 in human and mouse colon cancer cells results in decreased invasiveness *in vitro,* and also downregulation of metastasis-related network genes. Similar with the conclusion from human carcinoma specimens, overexpression of NFATc1 has significantly increased the metastatic potential of colon cancer cells [151]. Moreover, loss of NFAT1 expression in breast cancer cell lines using small interfering RNA leads to attenuated transcription of COX-2 and reduced invasion level as in other cases [152].

In this study, I investigate the function of NFATc1 in the overexpression, sgRNA-mediated knockout, and siRNA-mediated knockdown cell models. For Panc1 and Miapaca2, suppression of NFATc1 inhibits the cell viability, reduces the cell migration, and promotes the apoptosis. As for the overexpression, isoform A of NFATc1 was overexpressed. Modulations of NFATc1 expression level via knockout, knockdown and overexpression suggest that the expression of NFATc1 is positively associated with colony formation. Functional assays support oncogenic roles of NFATc1 in Panc1 and Miapaca2. In pancreatic cancer, it's reported that oncogenic mutations in KRAS contribute to the development of PDAC but are not sufficient to initiate carcinogenesis. Additionally, inflammation-induced signaling via the epidermal growth factor receptor (EGFR) and expression of SOX9, is essential for the tumor formation. Furthermore, exocrine pancreatic cells affect the tumor progression by changing the cellular phenotype. Chen et al. reveal that EGFR signaling induces expression of NFATc1 and Sox9, and furthermore leads to acinar cell trans-differentiation and initiation of pancreatic cancer [153]. Moreover, EZH2 positively regulates the oncogenic activity of NFATc1, which is an important mechanism of pancreatic cell plasticity [154].

This and previous studies have revealed the oncogenic function of NFATc1 beyond inflammation function. While the targets of NFATc1 is not revealed in pancreatic cancer. In order to further understand the mechanism of oncogenic initiation, the NFATc1-associated pathways and the targets of NFATc1 are investigated. The mRNA profiling is conducted in siRNA-mediated knockdown cell lines including Panc1, Miapaca2 and AsPC1. When the analysis was referred to 50-hallmark gene sets, multiple gene sets associated with Myc targets, P53 pathway, E2F targets, G2M

checkpoint, and DNA repairment pathways are significantly enriched in the control cells compared to NFATc1-knockdown cells. When the analysis was referred to KEGG pathways, various metabolism related pathways, including drug metabolism (cytochrome P450 and other enzymes), purine metabolism and pyrimidine metabolism are highly enriched in control cells compared with NFATc1-KD cells. These results suggest the oncogenic role of NFATc1.

Next, the targets of methylation-dependent transcription factor NFATc1 are identified by integrating data of the knock-down profiling, and methylation profiling of PDAC tissues, Panc1 and HPDE cell line. Although functional protein microarray data indicates that NFATc1 is a reader of methylation promoter of Twist1, Twist1 is not regulated by NFATc1. Interestingly, the expression of Twist1 is not further detected in Panc1 and Miapaca2 cell lines, even though the expression of Twist1 is upregulated in PDAC and the microarray screening demonstrates the potential causal link between its hypermethylation and upregulation in PDAC tissues. The heterogeneity in tissues might explain this phenomenon. Ultimately, ALDH1A3 is identified as a potential target of NFATc1 since the transcription of ALDH1A3 is regulated positively by NFATc1, and multiple methylated NFATc1 binding sites at ALDH1A3 promoter region are speculated. Additionally, MKNK2 is also a potential target of NFATc1 since multiple unmethylated NFATc1 binding sites at ALDH1A3 promoter region are identified. Subsequently, further functional evidences validate ALDH1A3 and MKNK2 as the targets of NFATc1.

5.4. Revealed model of transcription regulation by NFATc1

Notably, the expression of ALDH1A3 is positively regulated upon knockdown, knockout and overexpression of NFATc1. ALDH1A3 is one of the most important aldehyde metabolic enzyme in human cells. Metabolism reprogramming has been linked with the initiation, metastasis and recurrence of cancer. ALDH1A3 expression is associated with the development, progression, and prognosis of cancers. In addition, ALDH1A3 can act as a marker for cancer stem cells [155]. From the analysis of knockdown profiling, various metabolism related pathways are highly enriched in control cells compared with NFATc1-KD cells, which indicates that NFATc1 might regulate ALDH1A3 to influence the metabolism related pathways and leads to the

progress of malignancies. Compared with the healthy tissue group, ALDH1A3 was highly upregulated in PDAC group. However, upregulation of ALDH1A3 was not observed in Panc1 and Miapaca2 in comparison with HPDE. The enhanced expression of ALDH1A3 in PDAC tissues might be explained with probes in its promoter region were hypermethylated compared with the healthy tissues. Based on the methylation/unmethylation-motif of NFATc1, *in silico* FIMO analysis uncovered multiple NFATc1 methylated binding sites in the promoter region of ALDH1A3. According to the result of MSP, the candidate binding sites of NFATc1 were highly methylated in Panc1 and Miapaca2. In order to verify the effect of methylation on the expression of ALDH1A3, Panc1 and Miapaca2 were treated with the optimal concentration of 5-Azacitidine drug for 3 days followed with examination of expression level of ALDH1A3. Upon the treatment of 5-Aza, the expression of ALDH1A3 was downregulated in Panc1 and upregulated in Miapaca2. As known from previous experiments, NFATc1 was highly expressed in Panc1 compared to Miapaca2, which might explain why the expression of ALDH1A3 is downregulated in Panc1 upon demethylation but not in Miapaca2.

Next, in order to verify that ALDH1A3 is a direct target of NFATc1, luciferase was performed accordingly. The result indicates that predicted promoter region of ALDH1A3 is active. Meanwhile, the promoter activity is positively associated with the expression level of NFATc1. When the promoter is methylated, the promoter is still showing the activity. Though the methylation isn't increasing the promoter activity, the methylation doesn't block the transcription. Together with the evidence that NFATc1 preferentially bind with methylated motif, it's concluded that ALDH1A3 is regulated by NFATc1 directly. After performing the luciferase assay, ChIP together with the demethylation treatment was performed to validate the methylated NFATc1 binding sites that were predicated in the promoter region of ALDH1A3. ChIP assays have successfully detected the binding signal of NFATc1 in Panc1, but not in Miapaca2, de-methylated Panc1 and Miapaca2. These results indicate NFATc1 regulates the expression of ALDH1A3 directly in Panc1, and this regulation is methylation-dependent. However, the failure of binding singal of NFATc1 in Miapaca2 might be resulted from the low expression of NFATc1 in this cell line. Additionally, systematic reliability has to be considered because high background signals in negative control of IgG via western blot of the lysates have been observed. ChIP is widely used to

dissect out the association of regulatory molecules to specific promoters and histone modifications *in ex-vivo*. While certain technical limitations exist in ChIP experiments [156], the reliability of ChIP result is highly dependent on the antibody quality. Moreover, it may cause false positive results due to fixation of transient proteins to chromatin, and inefficient chromatin recovery that might be acquired from target protein epitope disruption. To more precisely validate the binding of NFATc1 in the ALDH1A3 promoter region, luciferase assays of the predicted promoter region together with demethylation treatment are urgently needed. Since NFATc1 plays crucial roles in the progress of carcinogenesis, ChIP-BS-seq is needed to systematically discover more targets of NFATc1 to reveal the oncogenic mechanism associated with NFATc1.

Taken together, ALDH1A3 is positively regulated by NFATc1, and it's the first time to reveal the biological function of NFATc1 as a methyl-plus TF in PDAC. Moreover, ALDH1A3 is identified as a direct target of NFATc1 for the first time. In contrast to the traditional scenario that hypermethylation in the promoter region of a gene represses the expression of this gene, this study provides a new understanding that hypermethylation in the promoter region is not always blocking the transcription of the gene.

How is the gene transcribed when the promoter is hypermethylated? In the examination of predicted primary structure of NAFT, NFATc1 is found to interact with CBP/p300 at the N-terminus transactivation domains (TAD) [157]. CBP/p300 are endowed with histone acetyltransferase (HAT) activity, which transfers an acetyl group from acetyl-CoA to form ε-N-acetyllysine and results in the activation of transcription and chromatin remodeling [158]. Although NFAT members share conserved domains, the highly variable TADs regions might have a critical role in NFAT function. One proposed model might be that NFATc1 firstly recognizes and stabilizes the methylated promoter as an anchor for the other co-activator. For example, CBP/p300 could recognize the N-terminus transactivation domains of NFATc1, acetylate the histone, remodel the histone, and co-activate the transcription. The working model of NFATc1 is illustrated below.
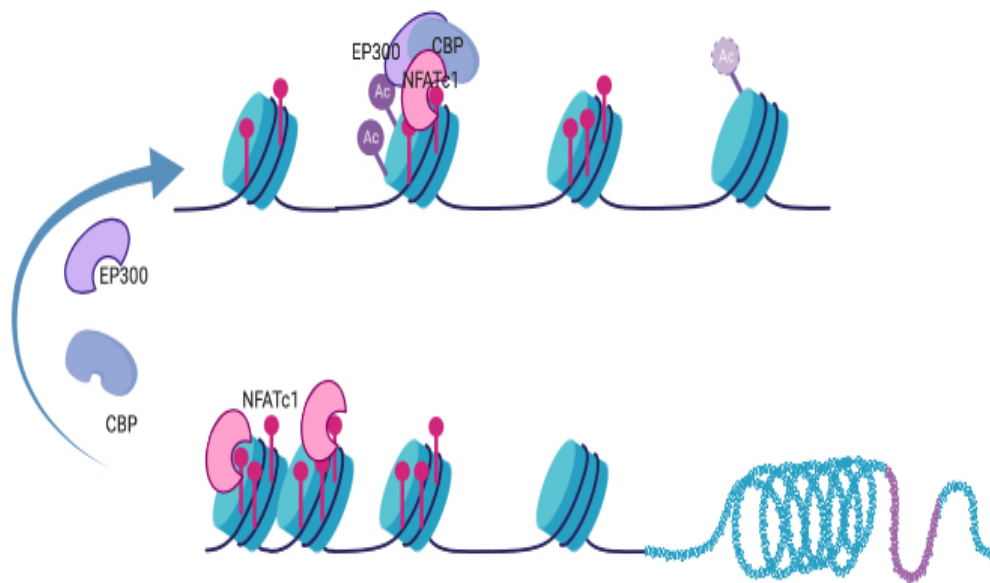
Figure 5.1 Working model of NFATc1

NFATc1 binds to methylated cis-regulatory elements, followed by chromatin remodeling and transcription activation.

In a summary, using a new experimental paradigm, this study has revealed that NFATc1 regulates the transcription of ALDH1A3, and thus exerts oncogenic role. Additionally, NFATc1 activates the transcription of ALDH1A3 via binding to its methylated regions. The work demonstrates such gene activation mechanism can mediate physiological functions in biologically relevant events. It's demonstrated that mCpG has roles in regulating TFs binding, histone modifications and gene activations in a sequence-specific manner. Most importantly, this study has provided a new notion that TFs can act as a new class of DNA methylation effectors that drive gene transactivation in biological proc

Index: Primers and Oligonucleotides

Primer name	Sequence (5'-3')	Note
M13_for	GTAAAACGACGGCCAGT	PCR primer
M13_rev	CAGGAAACAGCTATGAC	PCR primer
TF_for	GAAATTAATACGACTCACTATAGGGAG ACCACAACGGTTTCCCTCTAGAAATAA TTTTGTTTAAGAAGGAGATATACATAT GCATCATCATCATCATCATACTTTGTA CAAAAAAGTTGGCATG	PCR primer
TF_rev	CTGGAATTCGCCCTTTTATTACGTAGA ATCGAGACCGAGGAGAGGGTTAGGG ATAGGCTTACCTAATGCCAACTTTGTA CAAGAAAGCTG	PCR primer
Twist1 promoter oligonucleotides	AGTTGGG*CG*AGAGCTGCAGACTTGGA GGCTCTTATACCTC*CG*TGCAGG*CG*GA AAG	PDI oligo
NFATc1_sg1_for	*CACCG*CCCGTATGAGCTTCGGATTG	sgRNA sequence
NFATc1_sg1_rev	*AAAC*CAATCCGAAGCTCATACGGG*C*	sgRNA sequence
NFATc1_sg2_for	*CACCG*CGGAGGACACCCATCGTGC	sgRNA sequence
NFATc1_sg2_rev	*AAAC*GCACGATGGGGTGTCCTCCG*C*	sgRNA sequence
NFATc1_sg3_for	*CACCG*CTCCCGAAGACCGCAGCCGC	sgRNA sequence
NFATc1_sg3_rev	*AAAC*GCGGCTGCGGTCTTCGGGAG*C*	sgRNA sequence
Scramble_for	*CACCG*ATATCCGGAATTCGCGCGAT	sgRNA sequence

Scramble_rev	*AAAC*ATCGCGCGAATTCCGGATAT*C*	sgRNA sequence
LKO.1 5'	GACTATCATATGCTTACCG	PCR primer
Prisg1/2_for	TCCATCTTAGAGAACTGGCC	PCR primer
Prisg1/2_rev	TAACCACGACAGAGCATTC	PCR primer
Prisg3_for	GAGACTCAGAGGCTCCGAAC	PCR primer
Prisg3_rev2	GGCAGAGGAGACACCTATTG	PCR primer
NFATc1-OE for	TAGAGCTAGCGAATTCATGCCAAGCACCAGCTTTC	PCR primer
NFATc1-OE rev	TCGCGGCCGCGGATCCTCAGAAAAAGCACCCCACGCGC	PCR primer
MSP for	ATGTATTAGAAGTCGTTTTCGTG	PCR primer
MSP rev	CTCCTTTTACGATTTAAAAAACGC	PCR primer
ALDH1A3_ChIP for	TCGCCAGTGTTAGCCAGCCGATAT	PCR primer
ALDH1A3_ChIP rev	AAAGGTCTTGTGCTGTTATGGCCT	PCR primer

List of abbreviations

°C	Degrees Celsius
ab	antibody
amp	ampicillin
bp	Base pairs
BSA	Bovine serum albumin
cm	centimeter
Ctrl	Control
C-terminus	Carboxy-terminus
ddH2O	Double-distilled water
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTP	Desoxyribonukleosidtriphosphate
DTT	1,4-Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetate
e.g.	Exempli gratia
ELISA	Enzyme-linked immunosorbent assay
et al.	Et alii
g	gramm
h	hour
s	second
His	Histidine
HRP	Horse radish peroxidase
i.e.	Id est
Ig	Immunoglobulin
INF-γ	interferon-γ
kD	kilo Dalton
kV	kilovolt
LB	Lysogeny broth
LPS	Lipopolysaccharide
µl	microliter
MFI	Median fluorescence intensity
min	minute
MIST	Multiple spotting technique
ml	milliliter
mm	millimeter
mM	millimolar
NC	Negative control
nl	nanoliter
nm	nanometer
N-terminus	Amino-terminus

OD	Optical density
ORF	Open reading frame
PBS	Phosphate-buffered saline
PBS-T	PBS-Tween
PC	Posive control
PCR	Polymerase chain reaction
PISA	Protein in situ Arrays
RBS	Ribosome binding site
RNA	Ribonucleic acid
rpm	Rounds per minute
TEMED	N,N,N',N'-Tetramethylethylendiamin
5mC	5-methylcytosine
ALDH1A3	aldehyde dehydrogenase 1 family member A3
ChIP	chromatin immunoprecipitation
CpG	5'-C-Phosphate-G-3'
CRISPR/Cas	clustered regularly interspaced short palindromic repeats/CRISPR associated (protein)
DNMT	DNA methyltransferases
KRAS	kirsten rat sarcoma viral oncogene homologue
MBC	methylated, bisulfite converted control DNA
MBD	methyl-CpG-binding domain
MSP	methylation-specific polymerase chain reaction
NFAT	nuclear factors of activated T-cells
NFATc1	nuclear factor of activated T-cells 1
PDAC	pancreatic ductal adenocarcinoma
TF	transcription factor
UBC	unmethylated, bisulfite converted control DNA
FDR	false discovery rate
CP	chronic pancreatitis
FC	fold change
GO	gene ontology
SD	standard deviation
scr	scramble
NES	normalized enrichment score
KO	knockout
KD	knockdown
OE	overexpression
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
PDI	Protein DNA interaction

List of Figures

List of Tables

Acknowledgements

The whole thesis work was conducted in the laboratory of Dr. Jörg D. Hoheisel in the department of functional genome analysis, German cancer research center, Germany.

How time flies, I've been in Dr. Jörg D. Hoheisel's research group for almost five years since Dec. 2014. I still remember that when I firstly came to the lab. I was almost like a piece of white paper in the field of science in the beginning. With Joerg's supervision, many persons' kind helps and supports, I learned how to propose a project, how to manage projects, how to do science properly, how to present scientific work, and how to guide students. And I managed to finish several projects and publish results in journals. Even though, my PhD project was not always running smoothly, thanks to this experience, I became stronger. And none of my achievements could have been reached without Joerg's help and all the other's supports. I also feel so lucky to be in this lab with all the people having a golden heart. I acquired not only the scientific thoughts during these years but have also gained lots of nice memories.

Above all, I'd like to express my greatest gratitude and respect to my supervisor **Dr. Jörg D. Hoheisel**, who gave me the opportunity to pursue my Ph.D in his division. You opened window for my scientific views. Whenever I needed help, your door was always open to me. your guidance and advice helped brighten my career to be an independent scientist. More importantly, I learned from you that as an independent individual, you should always be brave to pursue what you are enjoying and be responsible for your own life.

It gives me great pleasure in acknowledging the support and help of **Prof. Dr. Stefan Wiemann, Prof. Dr. jeroen krijgsveld, Prof. Dr. Frank Lyko, Prof. Dr. Ilse Hofmann, Prof. Dr. Brors Benedikt and PD Dr. Ralf Bischoff**. I am sincerely thankful for your guidance, useful comments and discussions on this research work.

My great gratitude will also be delivered to **Dr. Damjana Kastelic** and **Dr. Smiths Lueong** who guided me in the first year in the lab. Without you, I could imagine a much more troublesome PhD life in the very beginning. I also would like to show my profound

gratitude to **Dr. Andrea Bauer** for suggesting an immense number of invaluable comments for my project, especially for the spotting and clinical samples.

I am grateful to **Melanie Bier, Stefanie Kutschmann** for your technical assistance; **Anke Mahler** and **Marie-Christine Leroy-Schell** kindly helped for all sorts of administrative issues.

My respect and thanks will be dedicated to my colleagues, **Dr. Patrick Kunz, Dr. Lizhen Liu, Dr. Katrin Hufnagel, Dr. Adriana Spalwisz, Yi Pan, Beiping Miao, Anna Chiara Pirona, Risky Oktriani, Henning Boekhoff**. Thank you very much for your help with my experiments and sharing the life experience. Apart from science, we also have had so much good time together in this fantastic lab, parties, retreats, pasta club, dinners. And I also thank those that are not listed here due to the limited space.

I express my deep gratitude to my students, **Daniel Blösel, Lea Kröller, Isabelle Lander, Marie-Therese Stiegler, Elisa Schiele**, I feel so lucky working with so many awesome persons. Without your helps, I can't finish my projects. Additionally, I would like to thank **Google** and **core facility** in DKFZ. Whenever I had questions, you could always answer my queries.

I would like to thank **Prof. Dr. Jussi Taipale, Dr. Arttu Jolma, Dr. Yimeng Yin** for the help of transcription factor library collection.

My deepest gratitude goes to my entire family: especially to **my parents** for their supports and encouragements throughout my studies. Words cannot express my love and gratefulness to my parents and my sister **Zhuoan Wu.** My beloved boyfriend **Dr. Liwei Zhang** was always there at all times to support me from personal to scientific aspects. You inspired me during difficult times when I needed words of encouragement. I can't finish my PhD study without your supports and encouragements.

Finally, I am especially thankful to the **CSC** for the first four years and **DKFZ** for the last 10 months who financed my study here in Germany. Without this financial support, I can't imagine how can I finish the work of PhD.

References

1. contributors, W., *Transcription (biology)*. Wikipedia, The Free Encyclopedia.

2. Maston, G.A., S.K. Evans, and M.R. Green, *Transcriptional Regulatory Elements in the Human Genome.* Annual Review of Genomics and Human Genetics, 2006. 7(1): p. 29-59.

3. Wittkopp, P.J. and G. Kalay, *Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence.* Nature Reviews Genetics, 2012. 13(1): p. 59.

4. Ecker, J.R., et al., *Genomics: ENCODE explained.* Nature, 2012. 489(7414): p. 52.

5. Soutourina, J., *Transcription regulation by the Mediator complex.* Nature Reviews Molecular Cell Biology, 2018. 19(4): p. 262.

6. Consortium, E.P., *The ENCODE (ENCyclopedia of DNA elements) project.* Science, 2004. 306(5696): p. 636-640.

7. Consortium, E.P., *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.* Nature, 2007. 447(7146): p. 799.

8. Tjian, R. and T. Maniatis, *Transcriptional activation: a complex puzzle with few easy pieces.* Cell, 1994. 77(1): p. 5-8.

9. Vannini, A. and P. Cramer, *Conservation between the RNA Polymerase I, II, and III Transcription Initiation Machineries.* Molecular Cell, 2012. 45(4): p. 439-446.

10. Griffiths, A., et al., *Transcription: an overview of gene regulation in eukaryotes.* an introduction to genetic analysis. 7th edition. New York: WH Freeman, 2000.

11. Ong, C.T. and V.G. Corces, *Enhancer function: new insights into the regulation of tissue-specific gene expression.* Nature Reviews Genetics, 2011. 12(4): p. 283-293.

12. West, A.G., M. Gaszner, and G. Felsenfeld, *Insulators: many functions, many mechanisms.* Genes Dev, 2002. 16(3): p. 271-88.

13. Ogbourne, S. and T.M. Antalis, *Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes.* Biochem J, 1998. 331 (Pt 1): p. 1-14.

14. Whyte, W.A., et al., *Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes.* Cell, 2013. 153(2): p. 307-319.

15. Shin, H.Y., *Targeting Super-Enhancers for Disease Treatment and Diagnosis.* Mol Cells, 2018. 41(6): p. 506-514.

16. Lambert, S.A., et al., *The Human Transcription Factors.* Cell, 2018. 172(4): p. 650-665.

17. Fulton, D.L., et al., *TFCat: the curated catalog of mouse and human transcription factors.* Genome Biology, 2009. 10(3).

18. Vaquerizas, J.M., et al., *A census of human transcription factors: function, expression and evolution.* Nat Rev Genet, 2009. 10(4): p. 252-63.

19. Mitchell, P.J. and R. Tjian, *Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins.* Science, 1989. 245(4916): p. 371-8.

20. Warnmark, A., et al., *Activation functions 1 and 2 of nuclear receptors: molecular strategies for transcriptional activation.* Mol Endocrinol, 2003. 17(10): p. 1901-9.

21. Stubbs, L., Y. Sun, and D. Caetano-Anolles, *Function and Evolution of C2H2 Zinc Finger Arrays.* Subcell Biochem, 2011. 52: p. 75-94.

22. Lim, D.H.K. and E.R. Maher, *DNA methylation: a form of epigenetic control of gene expression.* The Obstetrician & Gynaecologist, 2010. 12(1): p. 37-42.

23. Yin, Y., et al., *Impact of cytosine methylation on DNA binding specificities of human transcription factors.* Science, 2017. 356(6337).

24. Nomencl, I.-I.C.o.B., *Abbreviations and symbols for nucleic acids, polynucleotides, and their constituents.* Biochemistry, 2002. 9(20): p. 4022-4027.

25. Stormo, G.D., et al., *Use of the 'Perceptron' algorithm to distinguish translational initiation sites in E. coli.* Nucleic Acids Res, 1982. 10(9): p. 2997-3011.

26. Zhao, Y., et al., *Improved Models for Transcription Factor Binding Site Identification Using Nonindependent Interactions.* Genetics, 2012. 191(3): p. 781-U204.

27. Jolma, A., et al., *DNA-binding specificities of human transcription factors.* Cell, 2013. 152(1-2): p. 327-39.

28. Garvie, C.W. and C. Wolberger, *Recognition of Specific DNA Sequences.* Molecular Cell, 2001. 8(5): p. 937-946.

29. Luscombe, N.M., R.A. Laskowski, and J.M. Thornton, *Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level.* Nucleic Acids Research, 2001. 29(13): p. 2860-2874.

30. Rohs, R., et al., *The role of DNA shape in protein-DNA recognition.* Nature, 2009. 461(7268): p. 1248-53.

31. Lee, T.I. and R.A. Young, *Transcriptional regulation and its misregulation in disease.* Cell, 2013. 152(6): p. 1237-51.

32. Pattabiraman, D.R. and R.A. Weinberg, *Tackling the cancer stem cells - what challenges do they pose?* Nat Rev Drug Discov, 2014. 13(7): p. 497-512.

33. Muller, P.A. and K.H. Vousden, *Mutant p53 in cancer: new functions and therapeutic opportunities.* Cancer Cell, 2014. 25(3): p. 304-17.

34. Sur, I. and J. Taipale, *The role of enhancers in cancer.* Nat Rev Cancer, 2016. 16(8): p. 483-93.

35. Vinagre, J., et al., *Frequency of TERT promoter mutations in human cancers.* Nature communications, 2013. 4: p. 2185.

36. Society, A.C., *Cancer Facts & Figures* 2019.

37. Omary, M.B., et al., *The pancreatic stellate cell: a star on the rise in pancreatic diseases.* J Clin Invest, 2007. 117(1): p. 50-9.

38. Kleeff, J., et al., *Pancreatic cancer.* Nature Reviews Disease Primers, 2016. 2.

39. E Poruk, K., et al., *The clinical utility of CA 19-9 in pancreatic adenocarcinoma: diagnostic and prognostic updates.* Current molecular medicine, 2013. 13(3): p. 340-351.

40. Bettegowda, C., et al., *Detection of circulating tumor DNA in early-and late-stage human malignancies.* Science translational medicine, 2014. 6(224): p. 224ra24-224ra24.

41. Mayers, J.R., et al., *Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development.* Nature medicine, 2014. 20(10): p. 1193.

42. Semaan, A. and A. Maitra, *Pancreatic cancer in 2017: Rebooting pancreatic cancer knowledge and treatment options.* Nat Rev Gastroenterol Hepatol, 2018. 15(2): p. 76-78.

43. Neoptolemos, J.P., et al., *Therapeutic developments in pancreatic cancer: current and future perspectives.* Nature Reviews Gastroenterology & Hepatology, 2018. 15(6): p. 333-348.

44.   Kuninty, P.R., et al., *MicroRNA targeting to modulate tumor microenvironment.* Frontiers in oncology, 2016. 6: p. 3.

45.   Feig, C., et al., *The pancreas cancer microenvironment.* Clin Cancer Res, 2012. 18(16): p. 4266-76.

46.   Jaster, R., *Molecular regulation of pancreatic stellate cell function.* Mol Cancer, 2004. 3: p. 26.

47.   Saotome, T., et al., *Morphological and immunocytochemical identification of periacinar fibroblast-like cells derived from human pancreatic acini.* Pancreas, 1997. 14(4): p. 373-82.

48.   Masamune, A., et al., *Roles of pancreatic stellate cells in pancreatic inflammation and fibrosis.* Clin Gastroenterol Hepatol, 2009. 7(11 Suppl): p. S48-54.

49.   Martinez, F.O. and S. Gordon, *The M1 and M2 paradigm of macrophage activation: time for reassessment.* F1000Prime Rep, 2014. 6: p. 13.

50.   Ino, Y., et al., *Immune cell infiltration as an indicator of the immune microenvironment of pancreatic cssssioancer.* Br J Cancer, 2013. 108(4): p. 914-23.

51.   Mielgo, A. and M.C. Schmid, *Impact of tumour associated macrophages in pancreatic cancer.* BMB reports, 2013. 46(3): p. 131.

52.   Batlle, E. and H. Clevers, *Cancer stem cells revisited.* Nature medicine, 2017. 23(10): p. 1124.

53.   Phi, L.T.H., et al., *Cancer stem cells (CSCs) in drug resistance and their therapeutic implications in cancer treatment.* Stem cells international, 2018. 2018.

54.   Seoane, J. and L. De Mattos-Arruda, *The challenge of intratumour heterogeneity in precision medicine.* J Intern Med, 2014. 276(1): p. 41-51.

55.   Wormann, S.M. and H. Algul, *Risk factors and therapeutic targets in pancreatic cancer.* Front Oncol, 2013. 3: p. 282.

56.   Waters, A.M. and C.J. Der, *KRAS: the critical driver and therapeutic target for pancreatic cancer.* Cold Spring Harbor perspectives in medicine, 2018. 8(9): p. a031435.

57.   Maitra, A. and R.H. Hruban, *Pancreatic Cancer.* Annual Review of Pathology: Mechanisms of Disease, 2008. 3(1): p. 157-188.

58. Waddell, N., et al., *Whole genomes redefine the mutational landscape of pancreatic cancer.* Nature, 2015. 518(7540): p. 495-501.

59. Witkiewicz, A.K., et al., *Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets.* Nat Commun, 2015. 6: p. 6744.

60. Bailey, P., et al., *Genomic analyses identify molecular subtypes of pancreatic cancer.* Nature, 2016. 531(7592): p. 47-52.

61. Daemen, A., et al., *Metabolite profiling stratifies pancreatic ductal adenocarcinomas into subtypes with distinct sensitivities to metabolic inhibitors.* Proceedings of the National Academy of Sciences of the United States of America, 2015. 112(32): p. E4410-E4417.

62. Jones, S., et al., *Core signaling pathways in human pancreatic cancers revealed by global genomic analyses.* Science, 2008. 321(5897): p. 1801-1806.

63. Gohrig, A., et al., *Axon guidance factor SLIT2 inhibits neural invasion and metastasis in pancreatic cancer.* Cancer Res, 2014. 74(5): p. 1529-40.

64. Gao, J., et al., *Aberrant DNA methyltransferase expression in pancreatic ductal adenocarcinoma development and progression.* Journal of Experimental & Clinical Cancer Research, 2013. 32(1): p. 86.

65. Christman, J.K., *5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy.* Oncogene, 2002. 21(35): p. 5483.

66. Valente, S., et al., *Selective non-nucleoside inhibitors of human DNA methyltransferases active in cancer including in cancer stem cells.* Journal of medicinal chemistry, 2014. 57(3): p. 701-713.

67. Newell-Price, J., A.J. Clark, and P. King, *DNA methylation and silencing of gene expression.* Trends in Endocrinology & Metabolism, 2000. 11(4): p. 142-148.

68. Ehrlich, M., *DNA hypomethylation in cancer cells.* Epigenomics, 2009. 1(2): p. 239-259.

69. Fernandez, A.F., C. Huidobro, and M.F. Fraga, *De novo DNA methyltransferases: oncogenes, tumor suppressors, or both?* Trends Genet, 2012. 28(10): p. 474-9.

70. Yang, X.J., et al., *Gene Body Methylation Can Alter Gene Expression and Is a Therapeutic Target in Cancer.* Cancer Cell, 2014. 26(4): p. 577-590.

71. Eickbush, T.H. and E.N. Moudrianakis, *The histone core complex: an octamer assembled by two sets of protein-protein interactions.* Biochemistry, 1978. 17(23): p. 4955-4964.

72. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell research, 2011. 21(3): p. 381.

73. Ropero, S. and M. Esteller, *The role of histone deacetylases (HDACs) in human cancer.* Molecular oncology, 2007. 1(1): p. 19-25.

74. Greer, E.L. and Y. Shi, *Histone methylation: a dynamic mark in health, disease and inheritance.* Nature Reviews Genetics, 2012. 13(5): p. 343.

75. Cao, J. and Q. Yan, *Histone ubiquitination and deubiquitination in transcription, DNA damage response, and cancer.* Frontiers in oncology, 2012. 2: p. 26.

76. Mizuguchi, G., et al., *ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex.* Science, 2004. 303(5656): p. 343-348.

77. Jin, C., et al., *H3. 3/H2A. Z double variant–containing nucleosomes mark'nucleosome-free regions' of active promoters and other regulatory regions.* Nature genetics, 2009. 41(8): p. 941.

78. Henikoff, S., et al., *Genome-wide profiling of salt fractions maps physical properties of chromatin.* Genome research, 2009. 19(3): p. 460-469.

79. Schones, D.E., et al., *Dynamic regulation of nucleosome positioning in the human genome.* Cell, 2008. 132(5): p. 887-898.

80. Lomberk, G., et al., *Distinct epigenetic landscapes underlie the pathobiology of pancreatic cancer subtypes.* Nat Commun, 2018. 9(1): p. 1978.

81. Jones, P.A., *Functions of DNA methylation: islands, start sites, gene bodies and beyond.* Nature Reviews Genetics, 2012. 13(7): p. 484.

82. Chatterjee, R. and C. Vinson, *CpG methylation recruits sequence specific transcription factors essential for tissue specific gene expression.* Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 2012. 1819(7): p. 763-770.

83. Guo, M.Z., et al., *Epigenetic Changes Associated with Neoplasms of the Exocrine and Endocrine Pancreas.* Discovery Medicine, 2014. 17(92): p. 67-73.

84. Tang, B., et al., *Clinicopathological Significance of CDKN2A Promoter Hypermethylation Frequency with Pancreatic Cancer.* Scientific Reports, 2015. 5.

85.   Nones, K., et al., *Genome-wide DNA methylation patterns in pancreatic ductal adenocarcinoma reveal epigenetic deregulation of SLIT-ROBO, ITGA2 and MET signaling.* International Journal of Cancer, 2014. 135(5): p. 1110-1118.

86.   Lopez-Serra, P. and M. Esteller, *DNA methylation-associated silencing of tumor-suppressor microRNAs in cancer.* Oncogene, 2012. 31(13): p. 1609-1622.

87.   Cancer Genome Atlas Research Network. Electronic address, a.a.d.h.e. and N. Cancer Genome Atlas Research, *Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma.* Cancer Cell, 2017. 32(2): p. 185-203 e13.

88.   Huang, P.-H., et al., *TGFβ promotes mesenchymal phenotype of pancreatic cancer cells, in part, through epigenetic activation of VAV1.* Oncogene, 2017. 36(16): p. 2202.

89.   Hu, S., et al., *DNA methylation presents distinct binding sites for human transcription factors.* eLife, 2013. 2.

90.   Liu, Y., et al., *Dependency of the cancer-specific transcriptional regulation circuitry on the promoter DNA methylome.* Cell reports, 2019. 26(12): p. 3461-3474. e5.

91.   Hendrich, B. and S. Tweedie, *The methyl-CpG binding domain and the evolving role of DNA methylation in animals.* TRENDS in Genetics, 2003. 19(5): p. 269-277.

92.   Hendrich, B. and A. Bird, *Identification and characterization of a family of mammalian methyl CpG-binding proteins.* Genetics Research, 1998. 72(1): p. 59-72.

93.   Zhang, Y., et al., *Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation.* Genes & development, 1999. 13(15): p. 1924-1935.

94.   Saito, M. and F. Ishikawa, *The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2.* Journal of Biological Chemistry, 2002. 277(38): p. 35434-35439.

95.   Wan, J., et al., *Methylated cis-regulatory elements mediate KLF4-dependent gene transactivation and cell migration.* Elife, 2017. 6: p. e20068.

96.   Soufi, A., et al., *Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming.* Cell, 2015. 161(3): p. 555-568.

97.   Garner, M.M. and A. Revzin, *A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system.* Nucleic Acids Res, 1981. 9(13): p. 3047-60.

98.   Jolma, A., et al., *Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities.* Genome Research, 2010. 20(6): p. 861-873.

99.   Jolma, A., et al., *DNA-dependent formation of transcription factor pairs alters their binding specificity.* Nature, 2015. 527(7578): p. 384-388.

100.   Berger, M.F. and M.L. Bulyk, *Protein binding microarrays (PBMs) for rapid, high-throughput characterization of the sequence specificities of DNA binding proteins.* Methods Mol Biol, 2006. 338: p. 245-60.

101.   Hu, S., et al., *Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling.* Cell, 2009. 139(3): p. 610-22.

102.   Bailey, T.L., et al., *MEME SUITE: tools for motif discovery and searching.* Nucleic Acids Res, 2009. 37(Web Server issue): p. W202-8.

103.   S Lueong, S., *Protein Microarrays as Tools for Functional Proteomics: Achievements, Promises and Challenges.* Journal of Proteomics & Bioinformatics, 2014. 07(04).

104.   Gregorio, N.E., M.Z. Levine, and J.P. Oza, *A User's Guide to Cell-Free Protein Synthesis.* Methods and Protocols, 2019. 2(1).

105.   He, M. and M.J. Taussig, *Single step generation of protein arrays from DNA by cell-free expression and in situ immobilisation (PISA method).* Nucleic Acids Res, 2001. 29(15): p. E73-3.

106.   Ramachandran, N., et al., *Self-assembling protein microarrays.* Science, 2004. 305(5680): p. 86-90.

107.   He, M., et al., *Printing protein arrays from DNA arrays.* Nat Methods, 2008. 5(2): p. 175-7.

108.   Angenendt, P., et al., *Generation of high density protein microarrays by cell-free in situ expression of unpurified PCR products.* Mol Cell Proteomics, 2006. 5(9): p. 1658-66.

109.   Syafrizayanti, et al., *Personalised proteome analysis by means of protein microarrays made from individual patient samples.* Sci Rep, 2017. 7: p. 39756.

110. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. 291(5507): p. 1304-51.

111. Kim, M.S., et al., *A draft map of the human proteome.* Nature, 2014. 509(7502): p. 575-81.

112. Braun, P. and A.C. Gingras, *History of protein-protein interactions: from egg-white to complex networks.* Proteomics, 2012. 12(10): p. 1478-98.

113. Daniel, B., G. Nagy, and L. Nagy, *The intriguing complexities of mammalian gene regulation: How to link enhancers to regulated genes. Are we there yet?* FEBS letters, 2014. 588(15): p. 2379-2391.

114. Dekker, J., et al., *Capturing chromosome conformation.* science, 2002. 295(5558): p. 1306-1311.

115. Klionsky, D.J., et al., *Guidelines for the use and interpretation of assays for monitoring autophagy (3rd edition).* Autophagy, 2016. 12(1): p. 1-222.

116. Simonis, M., et al., *Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C).* Nature genetics, 2006. 38(11): p. 1348.

117. Dostie, J., et al., *Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements.* Genome research, 2006. 16(10): p. 1299-1309.

118. Bauer, A.S., et al., *Transcriptional variations in the wider peritumoral tissue environment of pancreatic cancer.* Int J Cancer, 2018. 142(5): p. 1010-1021.

119. Assenov, Y., et al., *Comprehensive analysis of DNA methylation data with RnBeads.* Nat Methods, 2014. 11(11): p. 1138-1140.

120. Consortium, G.O., *The Gene Ontology resource: 20 years and still GOing strong.* Nucleic acids research, 2018. 47(D1): p. D330-D338.

121. Mi, H., et al., *PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements.* Nucleic acids research, 2016. 45(D1): p. D183-D189.

122. Ashburner, M., et al., *Gene ontology: tool for the unification of biology.* Nature genetics, 2000. 25(1): p. 25.

123. Tang, Z., et al., *GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses.* Nucleic Acids Res, 2017. 45(W1): p. W98-W102.

124. Suzuki, A., et al., *DBTSS/DBKERO for integrated analysis of transcriptional regulation.* Nucleic acids research, 2017. 46(D1): p. D229-D238.

125. Grant, C.E., T.L. Bailey, and W.S. Noble, *FIMO: scanning for occurrences of a given motif.* Bioinformatics, 2011. 27(7): p. 1017-1018.

126. Thompson, M.J., et al., *Pancreatic cancer patient survival correlates with DNA methylation of pancreas development genes.* PLoS One, 2015. 10(6): p. e0128814.

127. Ashktorab, H. and H. Brim, *DNA Methylation and Colorectal Cancer.* Curr Colorectal Cancer Rep, 2014. 10(4): p. 425-430.

128. Bediaga, N.G., et al., *DNA methylation epigenotypes in breast cancer molecular subtypes.* Breast Cancer Res, 2010. 12(5): p. R77.

129. Jones, P.A. and M.L. Gonzalgo, *Altered DNA methylation and genome instability: a new pathway to cancer?* Proceedings of the National Academy of Sciences, 1997. 94(6): p. 2103-2105.

130. Zhu, H., G. Wang, and J. Qian, *Transcription factors as readers and effectors of DNA methylation.* Nat Rev Genet, 2016. 17(9): p. 551-65.

131. Nan, X., et al., *Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex.* Nature, 1998. 393(6683): p. 386-9.

132. Ma, Y., et al., *The relationship between early embryo development and tumourigenesis.* J Cell Mol Med, 2010. 14(12): p. 2697-701.

133. Cofre, J. and E. Abdelhay, *Cancer Is to Embryology as Mutation Is to Genetics: Hypothesis of the Cancer as Embryological Phenomenon.* ScientificWorldJournal, 2017. 2017: p. 3578090.

134. Kelleher, F.C., D. Fennelly, and M. Rafferty, *Common critical pathways in embryogenesis and cancer.* Acta Oncol, 2006. 45(4): p. 375-88.

135. Qin, Q., et al., *Normal and disease-related biological functions of Twist1 and underlying molecular mechanisms.* Cell Res, 2012. 22(1): p. 90-106.

136. Xu, Y., et al., *Breast tumor cell-specific knockout of Twist1 inhibits cancer cell plasticity, dissemination, and lung metastasis in mice.* Proc Natl Acad Sci U S A, 2017. 114(43): p. 11494-11499.

137. Sen-Yo, M., et al., *TWIST1 hypermethylation is observed in pancreatic cancer.* Biomedical Reports, 2013. 1(1): p. 31-33.

138. Okada, T., et al., *TWIST1 hypermethylation is observed frequently in colorectal tumors and its overexpression is associated with unfavorable outcomes in patients with colorectal cancer.* Genes Chromosomes Cancer, 2010. 49(5): p. 452-62.

139. Tian, H., S.L. McKnight, and D.W. Russell, *Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells.* Genes Dev, 1997. 11(1): p. 72-82.

140. Yang, M.H., et al., *Direct regulation of TWIST by HIF-1alpha promotes metastasis.* Nat Cell Biol, 2008. 10(3): p. 295-305.

141. Karlsson, Q.H., et al., *Methylated DNA recognition during the reversal of epigenetic silencing is regulated by cysteine and serine residues in the Epstein-Barr virus lytic switch protein.* PLoS pathogens, 2008. 4(3): p. e1000005.

142. Badis, G., et al., *Diversity and complexity in DNA recognition by transcription factors.* Science, 2009. 324(5935): p. 1720-3.

143. Shaw, J.-P., et al., *Identification of a putative regulator of early T cell activation genes.* Science, 1988. 241(4862): p. 202-205.

144. Rao, A., C. Luo, and P.G. Hogan, *Transcription factors of the NFAT family: regulation and function.* Annual review of immunology, 1997. 15(1): p. 707-747.

145. Hogan, P.G., et al., *Transcriptional regulation by calcium, calcineurin, and NFAT.* Genes & development, 2003. 17(18): p. 2205-2232.

146. Macian, F., *NFAT proteins: key regulators of T-cell development and function.* Nature Reviews Immunology, 2005. 5(6): p. 472.

147. Day, C.J., et al., *NFAT expression in human osteoclasts.* Journal of cellular biochemistry, 2005. 95(1): p. 17-23.

148. Oikawa, T., et al., *Acquired expression of NFATc1 downregulates E-cadherin and promotes cancer cell invasion.* Cancer Res, 2013. 73(16): p. 5100-9.

149. Buchholz, M., et al., *Overexpression of c-myc in pancreatic cancer caused by ectopic activation of NFATc1 and the Ca2+/calcineurin signaling pathway.* The EMBO journal, 2006. 25(15): p. 3714-3724.

150. Singh, G., et al., *Sequential activation of NFAT and c-Myc transcription factors mediates the TGF-β switch from a suppressor to a promoter of cancer cell proliferation.* Journal of Biological Chemistry, 2010. 285(35): p. 27241-27250.

151. Tripathi, M.K., et al., *Nuclear factor of activated T-cell activity is associated with metastatic capacity in colon cancer.* Cancer Res, 2014. 74(23): p. 6947-57.

152. Yiu, G.K. and A. Toker, *NFAT induces breast cancer cell invasion by promoting the induction of cyclooxygenase-2.* Journal of Biological Chemistry, 2006. 281(18): p. 12210-12217.

153. Chen, N.-M., et al., *NFATc1 links EGFR signaling to induction of Sox9 transcription and acinar–ductal transdifferentiation in the pancreas.* Gastroenterology, 2015. 148(5): p. 1024-1034. e9.

154. Chen, N.-M., et al., *Context-dependent epigenetic regulation of nuclear factor of activated T cells 1 in pancreatic plasticity.* Gastroenterology, 2017. 152(6): p. 1507-1520. e15.

155. Duan, J.J., et al., *ALDH1A3, a metabolic target for cancer diagnosis and therapy.* Int J Cancer, 2016. 139(5): p. 965-75.

156. Gade, P. and D.V. Kalvakolanu, *Chromatin immunoprecipitation assay as a tool for analyzing transcription factor activity*, in *Transcriptional Regulation*. 2012, Springer. p. 85-104.

157. Mognol, G., et al., *Cell cycle and apoptosis regulation by NFAT transcription factors: new roles for an old player.* Cell death & disease, 2016. 7(4): p. e2199.

158. Chan, H.M. and N.B. La Thangue, *p300/CBP proteins: HATs for transcriptional bridges and scaffolds.* Journal of cell science, 2001. 114(13): p. 2363-2373.