

Deciphering the Role of LINC00261 and FOXA2  
in Epithelial-Mesenchymal Transition and  
Metastasis of Pancreatic Cancer Cells

**Dissertation**

zur Erlangung des

Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der

Naturwissenschaftlichen Fakultät I

der Martin-Luther-Universität

Halle-Wittenberg,

vorgelegt

von Frau Agnes Dorn

Gutachter: Prof. Dr. Stefan Hüttelmaier

Prof. Dr. Monika Hämmerle

Prof. Dr. Martin Pichler

Tag der Verteidigung: 19.08.2024



# TABLE OF CONTENTS

I. INTRODUCTION.....	1
1. The pancreas .....	1
1.1. Structure and function .....	1
1.2 Development.....	2
2. Pancreatic cancer .....	2
2.1 Epidemiology .....	2
2.2 Types.....	2
2.3 Pathogenesis.....	2
2.4 Stages and grades .....	3
2.5 Transcriptional landscape .....	4
2.6 Diagnosis, prognosis and treatment.....	5
3. Non-coding RNAs .....	5
3.1 Background.....	5
3.2 Classification.....	6
3.3 Long non-coding RNAs.....	6
3.3.1 Evolution of long non-coding RNAs .....	6
3.3.2 Classification of long non-coding RNAs .....	7
3.3.3 Functions of long non-coding RNAs .....	8
3.3.4 Roles of long non-coding RNAs in pancreatic cancer.....	8
3.3.5 The long non-coding RNA LINC00261 .....	9
4. The transcription factor FOXA2 .....	10

4.1 Role of FOXA2 in pancreatic development .....	10
4.2 Role of FOXA2 in PDAC.....	10
5. Epithelial-to-mesenchymal transition (EMT).....	11
5.1 Subtypes of EMT.....	11
5.2 Markers of EMT.....	11
5.3 EMT signaling pathways involved in metastasis of PDAC .....	13
5.3.1 TGF $\beta$ pathway.....	13
5.3.2 c-Src pathway .....	15
6. The aim of the study.....	16
II. MATERIALS.....	18
1. Patient samples .....	18
2. Animals.....	18
3. Cell lines.....	18
4. Bacteria.....	18
5. Instruments.....	18
6. Consumables.....	20
7. Chemicals and reagents.....	20
8. Kits and systems.....	21
9. Plasmids .....	22
10. Cloning reagents .....	23
11. Oligonucleotides .....	24
12. Antibodies .....	25

13. Buffers and solutions .....	26
14. Software .....	28
15. Online tools and databases .....	28
III. METHODS.....	29
1. Cell biology methods.....	29
1.1 Cell culture.....	29
1.2 Treatment with TGF $\beta$ and TGF $\beta$ type 1 receptor inhibitor .....	29
1.3 Treatment with Src kinase inhibitors .....	29
1.4 siRNA transfection .....	30
1.5 CRISPR interference .....	30
1.6 CRISPR/Cas9 .....	30
1.7 2D cell proliferation assay.....	30
1.8 Clonogenic assay.....	31
1.9 Cell migration and invasion assays .....	31
2. Animal work .....	31
2.1 Orthotopic xenograft mouse model.....	32
2.2 <i>In vivo</i> imaging .....	32
2.3 Histopathology .....	32
3. Molecular biology methods.....	33
3.1 Cloning.....	33
3.2 Genomic DNA and total RNA extraction followed by qRT-PCR .....	33
3.3 RNA extraction from PDAC and normal pancreas tissue samples .....	34

3.4 <i>In vitro</i> transcription .....	34
3.5 Protein pulldown with biotinylated RNA .....	34
3.6 Protein extraction and western blot analysis.....	35
3.7 Reverse phase protein array (RPPA).....	35
3.8 Subcellular fractionation .....	35
3.9 Chromatin immunoprecipitation (ChIP).....	36
3.10 Luciferase reporter assay .....	36
4. Bioinformatics .....	36
4.1 Analysis of LINC00261 expression in PDAC samples.....	36
4.2 RNA-seq data analysis.....	37
5. Statistics.....	38
IV. RESULTS.....	39
1. Subtype-specific gene expression analysis of PDAC samples.....	39
2. Expression of LINC00261 in PDAC .....	39
2.1 Expression in PDAC patient's samples .....	39
2.2 Correlation with PDAC stage, grade and patient survival .....	41
2.3 Expression in PDAC cell lines .....	41
3. Characterization of LINC00261 .....	42
3.1 Conservation of the LINC00261 locus.....	42
3.2 Coding potential of LINC00261 .....	43
3.3 Localization of LINC00261 .....	43
4. Cellular function of LINC00261 .....	44

4.1 CRISPR-based knockdown of LINC00261 .....	44
4.2 Stable overexpression of LINC00261 .....	47
5. Molecular function of LINC00261 .....	47
5.1 Gene set enrichment analysis (GSEA) of Bailey's PDAC samples .....	48
5.2 RNA-seq and GSEA of the CRISPR-based knockdown cells.....	48
5.3 Analysis of deregulated genes in CRISPR-based knockdown cells.....	49
5.3.1 Analysis of genes related to EMT .....	49
5.3.2 Analysis of FOXA2 gene expression .....	51
6. Analysis of the LINC00261-FOXA2 regulatory circuit.....	52
6.1 Correlation of LINC00261 and FOXA2 gene expression .....	52
6.2 Analysis of the patient survival time considering FOXA2 expression .....	53
6.3 Molecular regulation of LINC00261 expression by FOXA2 .....	54
7. Cellular analysis of the interdependencies between LINC00261 and FOXA2 .....	55
7.1 CRISPR-based knockout of FOXA2 in WT and LINC00261 <sup>low</sup> cells.....	55
7.2 Stable overexpression of FOXA2 .....	58
8. Molecular analysis of the interdependencies between LINC00261 and FOXA2 .....	59
8.1 FOXA2-dependent and -independent regulation of genes by LINC00261 .....	59
8.1.1 FOXA2-independently regulated genes.....	61
8.1.1.1 Analysis of FOXA2-independently regulated genes and proteins .....	61
8.1.1.2 Regulation of the FAK/c-Src/PXN axis by LINC00261 .....	62
8.1.1.3 The role of the FAK/c-Src/PXN axis in cell migration and invasion .....	63
8.1.2 FOXA2-dependently regulated genes .....	67



8.1.3 Identification of LINC00261-binding proteins.....	68
8.2 LINC00261-independent regulation of genes by FOXA2 .....	69
9. Importance of LINC00261 and FOXA2 for <i>in vivo</i> tumor growth and metastasis .....	75
10. Tabular overview of the results of this study.....	77
V. DISCUSSION.....	78
1. Deregulated and subtype-specific lncRNAs in PDAC .....	78
2. Expression of LINC00261 in PDAC .....	78
3. Characterization of LINC00261 .....	79
4. Role of LINC00261 in EMT of PDAC cells.....	80
5. The LINC00261-FOXA2 regulatory circuit .....	82
6. Interdependencies between LINC00261 and FOXA2.....	83
6.1 FOXA2-independent LINC00261 functions.....	84
6.2 Mutual LINC00261 and FOXA2 functions .....	86
6.3 LINC00261-independent FOXA2 functions.....	87
7. The role of LINC00261 and FOXA2 for <i>in vivo</i> tumor growth and metastasis .....	89
8. The impact of EMT for <i>in vivo</i> tumor growth and metastasis.....	91
VI. SUMMARY .....	93
VII. REFERENCES .....	95
VIII. APPENDIX.....	112
1. Supplementary Figures .....	112
2. Supplementary Tables .....	114
3. List of Figures .....	131

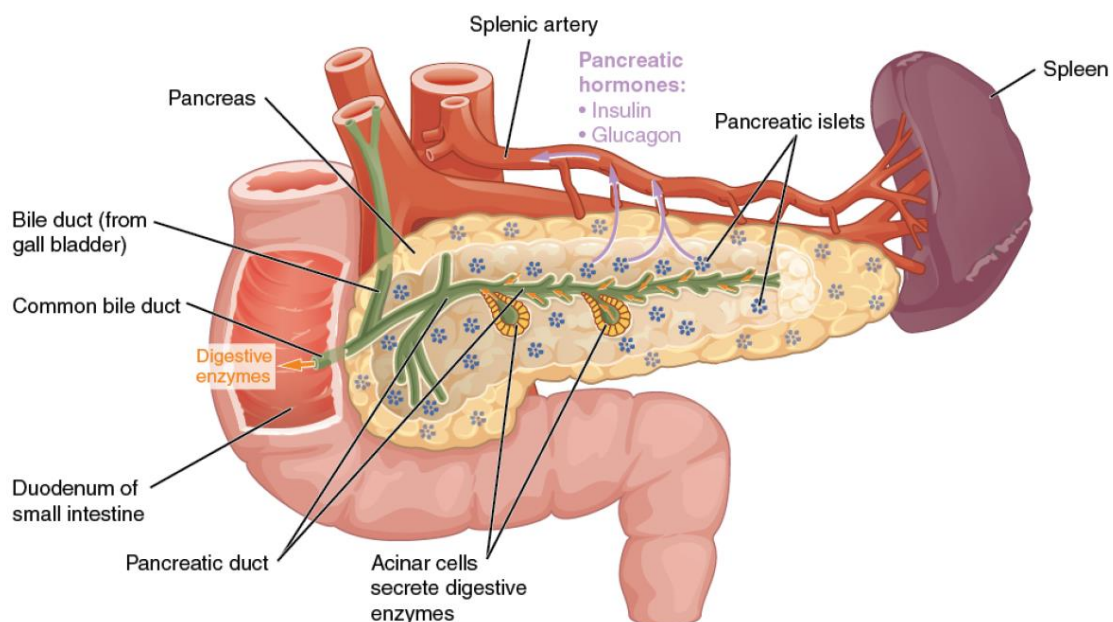
4. List of Tables .....	134
5. List of Abbreviations.....	135
EIDESSTÄTTLICHE ERKLÄRUNG .....	138
DANKSAGUNG .....	139
CURRICULUM VITAE.....	140
LIST OF PUBLICATIONS AND PRESENTATIONS.....	141

# I. INTRODUCTION

## 1. The pancreas

### 1.1. Structure and function

The pancreas is a flat, elongated gland located in the upper abdomen between the stomach and the spine. Anatomically, the pancreas consists of four parts: the head, neck, body, and tail (Figure 1). The pancreas has two major functions in the body, digestive (exocrine) and hormonal (endocrine). The endocrine activity is executed by the Langerhans islets (or pancreatic islets), which are distributed throughout the pancreas. These islets produce hormones, such as insulin and glucagon that regulate blood glucose levels and glandular secretion. The exocrine function of the pancreas involves the production of pancreatic juice, which contains various enzymes that facilitate the digestion of food in the intestines. These enzymes include trypsin and chymotrypsin for protein digestion, amylases for carbohydrate digestion, and lipases for fat breakdown. The exocrine tissue constitutes about 95% of the organ's mass and is composed of ducts formed by duct cells and cell clusters called acini. The acinar cells release their components for pancreatic juice into the center of the acini, from where the juice is drained into the pancreatic duct. The pancreatic juice then flows into the duodenum, the first section of the small intestine<sup>1</sup>.



**Figure 1: Structure and main components of the pancreas.**

The pancreas is a glandular organ situated in the retroperitoneum between the duodenal curvature and the hilum of the spleen. It comprises two functional structures: the exocrine acinar glands and the endocrine pancreatic islets. The exocrine glands produce digestive enzymes that are secreted into the duodenum, while the endocrine islets produce hormones which are released into the bloodstream. The figure was modified from Betts et al.<sup>2</sup>.

## 1.2 Development

During early embryonic development, three germ layers - endoderm, mesoderm, and ectoderm - are formed. The endoderm is the innermost of these layers and gives rise to several organs, including the pancreas, the epithelial lining of the digestive and respiratory tracts, the thyroid, the liver, and the gall bladder. Pancreatic development begins with the formation of one dorsal bud in the duodenal region, followed by the formation of two ventral buds. The left ventral bud regresses, while the right ventral bud fuses with the dorsal bud. In the final anatomic arrangement, the majority of the pancreas originates from the dorsal bud. Solely, the inferior part of the head originates from the ventral bud<sup>3</sup>.

## 2. Pancreatic cancer

### 2.1 Epidemiology

As of 2020, pancreatic cancer (PC) is the seventh leading cause of cancer-related deaths worldwide. In fact, in countries with a very high human development index (HDI), such as Germany and the United States of America, it is the third leading cause of cancer-related deaths<sup>4</sup>. The significant disparity in mortality rates may be attributed to the unequal quality of data on cancer incidence and mortality provided by developed and developing countries<sup>5</sup>. Also, studies propose that risk factors such as tobacco smoking<sup>6</sup>, obesity<sup>7</sup> and diabetes<sup>8</sup> contribute to the geographic variation.

### 2.2 Types

Neoplasms arising from the endocrine component of the pancreas (endocrine neoplasms) constitute a minor fraction of pancreatic neoplasms (<2%). Exocrine tumors comprise the majority of pancreatic neoplasms, dominated by ductal pancreatic adenocarcinomas (PDACs), accounting for over 90% of all exocrine neoplasms<sup>9</sup>.

### 2.3 Pathogenesis

Clinicopathological studies suggest that pancreatic cancer primarily originates from pancreatic intraepithelial neoplasia (PanIN), which is a classic pre-neoplastic lesion, although it may also arise from other precursor lesions such as intraductal papillary mucinous neoplasms (IPMNs) or mucinous cystic neoplasms (MCN)<sup>10</sup>. Progression from low-grade PanINs to high-grade PanINs is characterized by the accumulation of genetic alterations, usually starting with telomere shortening and activating mutations of the oncogene kirsten rat sarcoma virus (KRAS), followed by mutations of the tumor suppressor gene cyclin-dependent kinase inhibitor 2A (CDKN2A/P16) and subsequently tumor protein P53 (TP53) and sma- and mad-related protein 4 (SMAD4). IPMNs and MCNs also give rise

to pancreatic cancer by accumulating gene alterations. In established PDACs, KRAS mutations are found in over 90% of all cases, CDKN2A/P16 mutations in 95%, TP53 in 50-80% and SMAD4 in 30-60%<sup>11</sup>.

## 2.4 Stages and grades

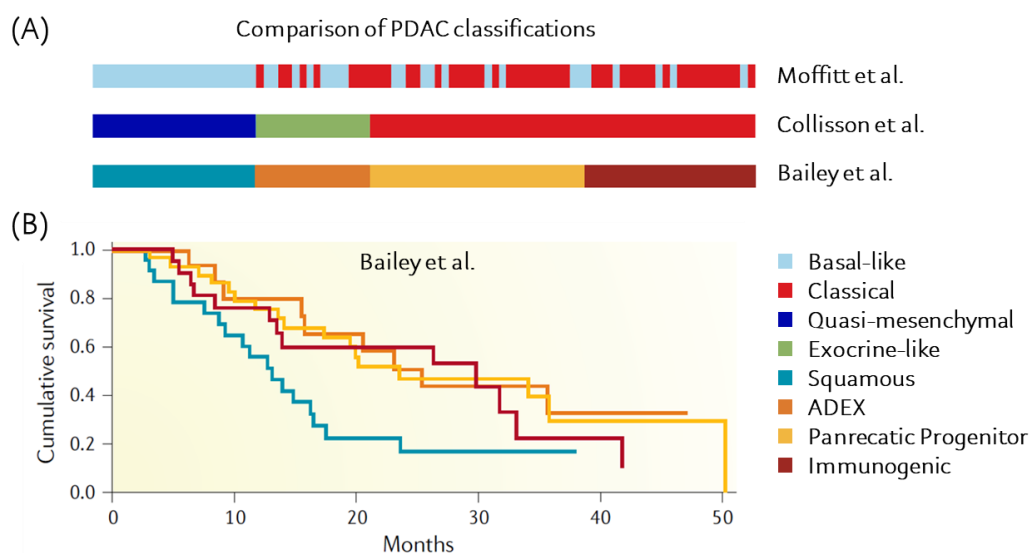
Cancer staging is applied to categorize malignant tumors based on their extent of growth and spread. The TNM system, developed by the AJCC (American Joint Committee on Cancer), is a globally accepted standard used to characterize (1) the local extent of primary tumor (T), (2) the degree of spread to nearby lymph nodes (N), and (3) the presence of metastases to distant sites (M)<sup>12</sup>. An overview of pancreatic cancer stages is shown in Table 1<sup>13</sup>. Additionally, pancreatic tumors can be classified into four tumor grades based on their degree of differentiation. Generally, the higher the grade, the more pronounced the dissimilarity between cancer cells and healthy cells, and the more rapid the growth and spread of cancer cells. Grade G1 tumors are well-differentiated, grade G2 tumors are moderately differentiated, grade G3 tumors are poorly differentiated, and grade G4 tumors are undifferentiated<sup>14</sup>.

Table 1: Pancreatic cancer stages

Stage	TNM grouping	Description
IA	T1, N0, M0	<ul style="list-style-type: none"> <li>Tumor is <math>\leq 2</math> cm in greatest dimension (T1)</li> <li>Tumor has not spread to nearby lymph nodes (N0)</li> <li>Tumor has not spread to distant sites (M0)</li> </ul>
IB	T2, N0, M0	<ul style="list-style-type: none"> <li>Tumor is <math>&gt; 2</math> cm and <math>\leq 4</math> cm in greatest dimension (T2)</li> <li>Tumor has not spread to nearby lymph nodes (N0)</li> <li>Tumor has not spread to distant sites (M0)</li> </ul>
IIA	T3, N0, M0	<ul style="list-style-type: none"> <li>Tumor is <math>&gt; 4</math> cm in greatest dimension (T3)</li> <li>Tumor has not spread to nearby lymph nodes (N0)</li> <li>Tumor has not spread to distant sites (M0)</li> </ul>
IIB	T1/2/3, N1, M0	<ul style="list-style-type: none"> <li>Tumor has spread to 1-3 nearby lymph nodes (N1)</li> <li>Tumor has not spread to distant sites (M0)</li> </ul>
III	T1/2/3, N2, M0	<ul style="list-style-type: none"> <li>Tumor has spread to <math>\geq 4</math> nearby lymph nodes (N2)</li> <li>Tumor has not spread to distant sites (M0)</li> </ul>
	T4, any N, M0	<ul style="list-style-type: none"> <li>Tumor is growing outside the pancreas and into nearby major blood vessels or nerves (T4)</li> <li>Tumor has not spread to distant sites (M0)</li> </ul>
IV	any T, any N, M1	<ul style="list-style-type: none"> <li>Tumor has spread to distant sites (M1)</li> </ul>

## 2.5 Transcriptional landscape

Multiple transcriptomic studies provided helpful insights into the development and progression of pancreatic cancer and identified disease subtypes that exhibit variations in overall survival and treatment response. These studies have proposed at least three distinct subtype classifications based on transcriptomic data (Figure 2A). For instance, Bailey and colleagues utilized RNA-seq and microarray data from pancreatic cancer samples and defined four molecular subtypes: Squamous, pancreatic progenitor, immunogenic, and aberrantly differentiated endocrine exocrine (ADEX). These subtypes were found to correlate with histopathological characteristics<sup>15</sup>. Another study by Moffitt et al. described two subtypes; a basal-like subtype, generally associated with poorer outcomes, and a classical subtype<sup>16</sup>. Collisson et al., on the other hand, suggested three subtypes: classical, quasi-mesenchymal and exocrine-like<sup>17</sup>. Notably, the gene signature derived from Collisson et al.'s classical subtype exhibited substantial overlap with Moffitt et al.'s classical subtype<sup>16</sup>. Despite variations in the methodology and input materials, one subgroup was consistently identified in all three analyses<sup>18</sup>. This subgroup, referred to as either 'squamous'<sup>15</sup>, 'quasi-mesenchymal'<sup>17</sup> or 'basal-like'<sup>16</sup>, is strongly associated with poorer overall and progression-free survival compared to all other identified groups. Moreover, this subtype is characterized by a gene signature related to the loss of endodermal identity and oncogenic signaling, contributing in part to the process of epithelial-to-mesenchymal transition (EMT). The Kaplan-Meier survival analysis of the subtypes described by Bailey et al.<sup>15</sup> is shown in Figure 2B.



**Figure 2: Comparison of transcriptomic PDAC subtypes.**

(A) Comparison of transcriptomic datasets of subtypes described by Moffitt et al.<sup>16</sup>, Collisson et al.<sup>17</sup> and Bailey et al.<sup>15</sup>; (B) Kaplan-Meier analysis of PDAC subtypes of Bailey et al.<sup>15</sup>. The figures were modified from Collisson et al.<sup>18</sup>.

## 2.6 Diagnosis, prognosis and treatment

In spite of the significant variation in incidence and mortality rates of pancreatic cancer between countries with high and low HDI, the 5-year survival rate for patients with this disease is consistently low worldwide, ranging from 2% to 9%<sup>19</sup>. There are several factors contributing to poor prognosis. Firstly, prevention strategies are limited, and early detection is challenging due to the absence of symptoms and a lack of sensitive tumor markers. Consequently, at the time of diagnosis, less than 20% of patients have a resectable disease or have already developed distant hematogenous metastases. Of those who undergo resection followed by adjuvant therapy, approximately 80% will relapse and eventually die of their disease. Moreover, pancreatic cancer is highly resistant to existing treatment options, including chemotherapy, radiotherapy, and molecularly targeted therapies. Nevertheless, some chemotherapies showed modest survival benefits and are now widely used to improve the patient's quality of life<sup>10</sup>. In particular, gemcitabine, an antimetabolite drug that inhibits the synthesis of deoxyribonucleic acid (DNA) during the S phase of the cell cycle, is applied as mono- or combination therapy with nab-paclitaxel or capecitabine to treat various stages of pancreatic cancer. Furthermore, a combination of 5-fluorouracil, folinic acid, oxaliplatin, and irinotecan (FOLFIRINOX) has shown significant survival benefits for advanced pancreatic cancer<sup>20</sup>. However, these conventional chemotherapies have only marginally improved the survival rate of pancreatic cancer patients over recent decades, and the number of deaths is expected to increase substantially in the coming decade<sup>19</sup>.

## 3. Non-coding RNAs

### 3.1 Background

For many years, the central dogma of molecular biology assumed that ribonucleic acid (RNA) acts as an intermediate, transferring information from a DNA sequence to the encoded protein. These RNAs are referred to as messenger RNAs (mRNAs). In recent years, however, advances in molecular biology technologies, particularly sequencing technologies, have enabled detailed analysis of both the genome and the transcriptome. These analyses have revealed that while roughly 85% of the human genome is transcribed into RNA, less than 3% of the resulting RNA transcripts are ultimately translated to protein products<sup>21</sup>. Consequently, most RNA transcripts do not encode proteins, and are therefore referred to as non-coding RNA (ncRNA). Generally, transcripts are classified as non-coding if they lack an open reading frame (ORF) that exceeds 100 amino acids (aa)<sup>21</sup>.

### 3.2 Classification

NcRNAs can be categorized into two main groups: housekeeping ncRNAs and regulatory ncRNAs. Housekeeping RNAs are usually small, constitutively expressed, and essential for cell viability. Examples include transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs). Regulatory ncRNAs, on the other hand, are typically transcribed in a site- and time-dependent manner and can be divided into two subgroups based on their length: short non-coding RNAs and long non-coding RNAs (lncRNAs). Short ncRNAs are less than 200 nucleotides (nts) in length and include micro RNAs (miRNAs) and piwi-interacting RNAs (piRNAs), which negatively regulate gene expression by inducing degradation of target mRNA or inhibiting its translation. Regulatory ncRNAs greater than 200 nts in length are referred to as lncRNAs, which include a subclass of circular RNAs (circRNAs) that are covalently closed RNA molecules. Table 2 provides an overview of the length and known functions of the most important types of non-coding RNAs<sup>22,23</sup>.

**Table 2: Classification of ncRNAs**

RNA Type			Length (avg. nts)	Function			
coding RNA		mRNA	1700	template for protein synthesis <sup>24,25</sup>			
non-coding RNA	housekeeping ncRNA		rRNA	6900	protein synthesis <sup>25</sup>		
			tRNA	<100	protein synthesis <sup>26</sup>		
			snRNA	100–200	intron splicing, RNA processing <sup>27</sup>		
			snoRNA	200	rRNA/snRNA processing <sup>28</sup>		
	regulatory ncRNA		short ncRNA	miRNA	18-22	RNA interference, protein translation regulation <sup>29</sup>	
				piRNA	26–31	regulation of transposable elements <sup>30</sup>	
			lncRNA		lncRNA	1000	see chapter 3.3
					circRNA	500	miRNA decoy, protein regulation <sup>31</sup>

### 3.3 Long non-coding RNAs

#### 3.3.1 Evolution of long non-coding RNAs

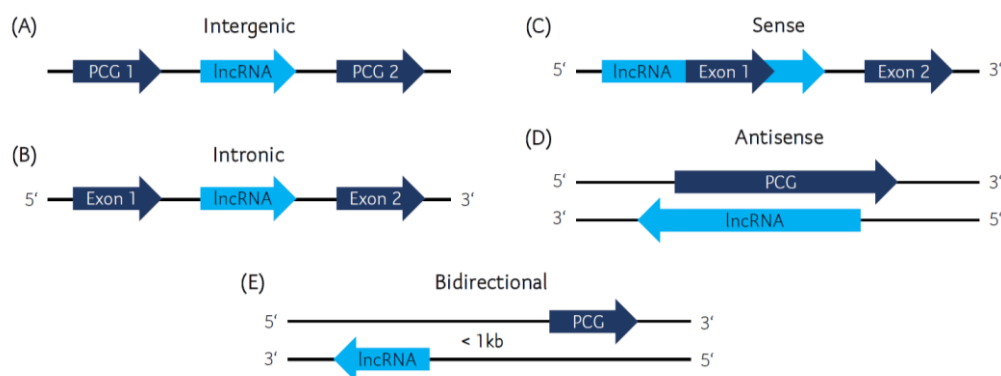
In recent years, there has been a significant increase in the number of identified lncRNAs in the human genome<sup>21</sup>. In 2021, the GENCODE project annotated nearly 18,000 lncRNA genes in the human genome, producing over 48,000 distinct transcripts<sup>32</sup>. The abundance of annotated lncRNAs is not surprising, considering the pervasive transcription of the



human genome. lncRNAs are typically expressed at lower levels than mRNAs, but are highly tissue-specific<sup>22,33,34</sup>. Most of the non-coding genome has evolved without any discernible selection, such as sequence conservation, leading to the assumption that most of the ncRNAs produced are non-functional<sup>35</sup>. However, complex organisms, as the humans, have small effective population sizes and, consequently, evolution of their genomes is a relatively weak subject to selection and dominated by genetic drift. Palazzo and Koonin propose that this weak selection regime promotes the production of excess genomic material, providing the raw material for the evolution of various lncRNAs and contributing to the high complexity of the organism<sup>35</sup>. It is therefore not astonishing, that 80% of annotated human lncRNAs originated during primate evolution, and only 3% are conserved in more distantly related species such as chicken or frog<sup>36</sup>. Palazzo and Koonin assume that most lncRNAs originate from junk sequences due to transcription events of surrounding genes (e.g. chromatin regulation, enhancer function). Another, less common source of lncRNAs may be mRNAs that have lost their coding capacity during evolution, such as the well-studied lncRNA X-inactive specific transcript (XIST)<sup>35</sup>. In summary, many annotated lncRNAs may be non-functional, but lncRNAs do not necessarily require strong sequence conservation between distant species to be functional. The secondary/tertiary structure or the syntenic locus of lncRNAs may be more important than their sequence<sup>37,38</sup>.

### 3.3.2 Classification of long non-coding RNAs

lncRNAs can be categorized based on their length, with about 58% of the lncRNAs being between 200-950 nts long (small lncRNAs), 40% between 950-4800 nts long (medium lncRNAs), and only 2% longer than 4800 nts (large lncRNAs)<sup>39</sup>. Additionally, lncRNAs can be classified based on their genomic location and orientation relative to adjacent protein-coding genes. They can be intergenic, intronic, sense, antisense, or bidirectional, as illustrated in Figure 3<sup>39,40</sup>.



**Figure 3: lncRNA classification based on their genomic location.**

(A) Intergenic lncRNAs are located between two protein-coding genes (PCGs); (B) Intronic lncRNAs are positioned within the intronic region of a PCG; (C) Sense lncRNAs are transcribed from the same strand as a PCG with sequence overlaps; (D) Antisense lncRNAs are transcribed from the antisense strand of a PCG with

sequence overlaps; (E) Bidirectional lncRNAs are transcribed from the antisense strand of a PCG without sequence overlaps and sequences less than 1 kb apart from each other.

### 3.3.3 Functions of long non-coding RNAs

Like mRNAs, lncRNAs are generally transcribed by RNA polymerase II and post-transcriptionally processed by 5'-capping, 3'-polyadenylation and partially splicing. However, lncRNAs are a highly heterogeneous class of molecules and unlike mRNAs, they have diverse regulatory functions in various biological processes and can localize to different cellular compartments<sup>41</sup>. It has been shown that lncRNAs can directly interact with a range of molecules within cells, including DNA, RNA, and proteins. Through binding to these molecules, lncRNAs can function as scaffolds, guides, decoys, or signals<sup>42</sup>.

Nuclear lncRNAs often participate in pre-transcriptional gene regulation by influencing chromatin structure either through direct interaction with chromatin-modifying complexes or by regulating chromatin-modifying enzymes<sup>43,44</sup>. Furthermore, they can modulate gene transcription by forming RNA-DNA-triplexes or by interacting with transcription factors leading to either inhibition or activation of gene transcription<sup>45,46</sup>. Nuclear lncRNAs can also act as post-transcriptional regulators by affecting the mRNA splicing, for instance, by binding to pre-mRNA and hindering spliceosome function<sup>47</sup>. Cytoplasmic lncRNAs have been implicated in post-transcriptional regulation by acting as miRNA sponges (also known as competing endogenous RNAs (ceRNAs)) or by regulating mRNA stability<sup>48,49</sup>. Additionally, they can modulate the translation efficiency by directly binding to the target mRNA or by recruiting translation-regulating proteins<sup>50</sup>. Certain lncRNAs have been reported to bind to proteins, serving either as transporters to relocate them to other cell compartments or as scaffolds that bring two or more proteins into proximity, thereby facilitating protein complex formation<sup>51,52</sup>.

The term "cis-acting lncRNAs" generally refers to lncRNAs whose activity is restricted to and dependent on the loci from which they are transcribed. These lncRNAs have been demonstrated to activate, repress, or otherwise modulate the expression of their adjacent target genes. In particular, genes that are involved in transcriptional regulation, such as transcription factors (TFs) and chromatin remodelers, tend to be surrounded by lncRNAs that modulate their expression in a cis-acting manner. In contrast, "trans-acting lncRNAs" exert their function elsewhere in the cytoplasm or nucleus, independently of their transcription site<sup>53</sup>.

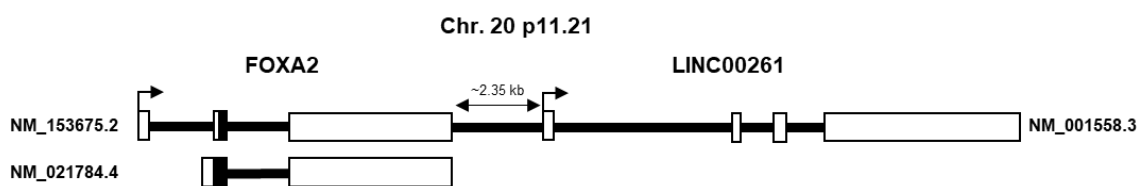
### 3.3.4 Roles of long non-coding RNAs in pancreatic cancer

The deregulation of lncRNAs has been observed in several types of cancer, suggesting their critical role as regulators and potential biomarkers for cancer diagnosis, prognosis, or

treatment<sup>54</sup>. In pancreatic cancer, several lncRNAs with both oncogenic or tumor-suppressive roles have been discovered<sup>55</sup>. Interestingly, the deregulation of many lncRNAs is not only observed between normal and tumor tissue, but also between different subtypes of PDAC, highlighting their significance in the progression of PDAC, response to therapy, and patient survival. For instance, integrated genomic characterization of PDAC has revealed differential expression of lncRNAs such as EVADR, GATA6-AS1, and LINC00261 between the classical and basal-type PDAC subtypes<sup>56</sup>.

### 3.3.5 The long non-coding RNA LINC00261

The LINC00261 gene (also known as DEANR1 or FALCOR) is 31,293 base pairs (bp) long, comprises four exons, and is located on the minus strand of the short arm of chromosome 20, adjacent to the transcription factor forkhead box protein A2 (FOXA2). This gene encodes a long intergenic non-coding RNA (lincRNA), which exists in several splice variants. Transcript abundance analysis suggests that the longest splice variant, which is 4912 bp in length, is the most biologically significant<sup>57</sup>. The genomic location and relative position of the LINC00261 lncRNA to the adjacent FOXA2 gene are depicted in Figure 4.



**Figure 4: Genomic locus of LINC00261 and FOXA2 on chromosome 20.**

Approximately 2.35 kilobases (kb) lie between the last exon of FOXA2 and the first exon of LINC00261. The figure was modified from Dorn et al.<sup>58</sup>

The lncRNA LINC00261 has gained increasing attention in recent years, as evidenced by a PubMed® search for publications with the search terms 'LINC00261', 'DEANR1' or 'FALCOR' in the title (Figure 5A). Out of 63 publications with these terms in the title, 51 also contained the term 'cancer' in the text, indicating that most of these publications examined the role of LINC00261 in tumorigenesis. However, the first publication describing a function of LINC00261 reported its importance in human endoderm differentiation, where it activates the expression of the FOXA2 gene<sup>59</sup>. Subsequently, LINC00261 has predominantly been described as a tumor suppressor in various cancers, involved in multiple cellular processes, as reviewed by Zhang et al.<sup>60</sup>. Interestingly, only ten out of these 63 publications mentioned the adjacent transcription factor FOXA2, including our own publication from 2020, which investigated the role of LINC00261 in PDAC (Figure 5B)<sup>58</sup>. In conclusion, most studies did not explore a possible cis regulation between LINC00261 and the adjacent transcription factor FOXA2.

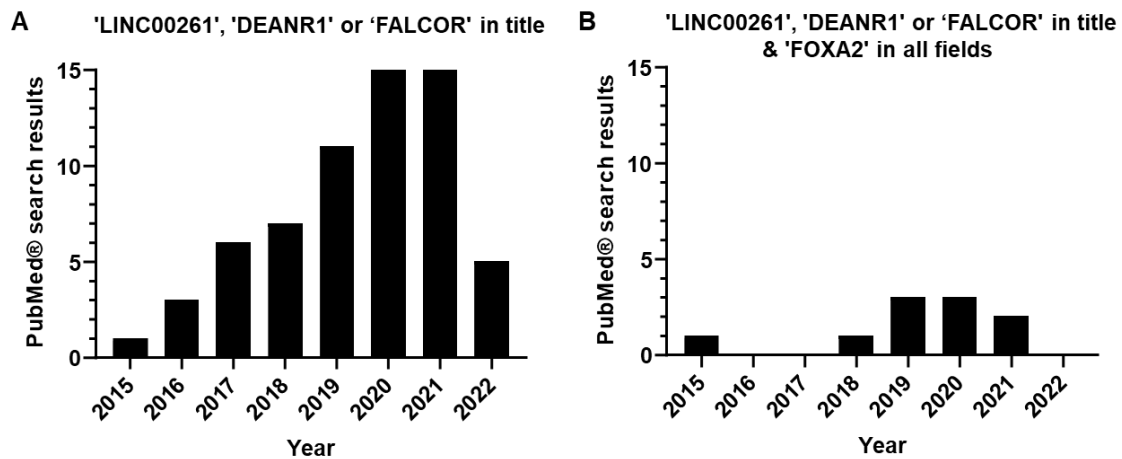


Figure 5: Search results in the PubMed® database.

(A) The number of publications per year containing the search terms 'LINC00261', 'DEANR1' or 'FALCOR' in the title; (B) The number of publications per year containing the search terms 'LINC00261', 'DEANR1' or 'FALCOR' in the title and 'FOXA2' in any field (as of 12.10.2022).

#### 4. The transcription factor FOXA2

The FOXA2 gene is 4,493 bp in length and comprises three exons. It is located around 2.35 kilobases (kb) upstream of LINC00261 (Figure 4). The gene produces two mRNAs with lengths of 2,538 bp and 2,401 bp, respectively. These two variants encode proteins with a length of 457/463 amino acids and a homology of 98%, differing only in the initial six/twelve amino acids.

##### 4.1 Role of FOXA2 in pancreatic development

Several studies have demonstrated the essential role of FOXA2 in the differentiation of endoderm-originating organs, including the pancreas<sup>61,62</sup>. Lee et al. showed that proper chromatin remodeling, H3K4me1 deposition before enhancer activation, and recruitment of the transcription factor GATA6 to these enhancers were required for pancreatic differentiation, all of which depended on FOXA2<sup>63</sup>. Other studies have also described a critical role of FOXA2 in the transcriptional regulatory network controlling pancreatic development by acting on enhancers of other developmental transcription factors such as PDX1 and GATA4<sup>64,65</sup>. Importantly, FOXA2-null mice showed pancreatic hypoplasia, hyperglycemia, impaired acinar and islet cell content, and subsequent death<sup>66</sup>.

##### 4.2 Role of FOXA2 in PDAC

Although FOXA2 is known to regulate the normal development of endoderm-derived organs, its involvement in cancer has remained largely unexplored until recently. In recent years, several studies have suggested multiple functions of FOXA2 in cancer. Interestingly, these studies have demonstrated contrary roles of FOXA2 in cancer development and

progression. In some cancer types, FOXA2 has been defined as an epithelial marker and tumor suppressor<sup>67-71</sup>, whereas in other studies, it has been suggested that FOXA2 may have oncogenic functions and drives migration, invasion, and EMT<sup>72-75</sup>. Similarly, opposing roles of FOXA2 have been described in pancreatic cancer<sup>76-78</sup>. The conflicting observations might be attributed to distinct differentiation grades of the tumor models. For instance, Milan et al. found that FOXA2 displayed different genomic distributions and regulated distinct gene expression programs dependent on the differentiation grade of PDAC. These grade-specific functions of FOXA2 relied on its interactions with transcription factors whose expression varied depending on the differentiation grade<sup>79</sup>. Additionally, the type of pathogenesis, including mutational events, may be relevant to the role of FOXA2. Li et al., for instance, found that the presence of mutated KRAS locks in a proto-oncogenic transcriptional program in pancreatic progenitor cells, in which FOXA2 plays a critical part<sup>80</sup>.

## 5. Epithelial-to-mesenchymal transition (EMT)

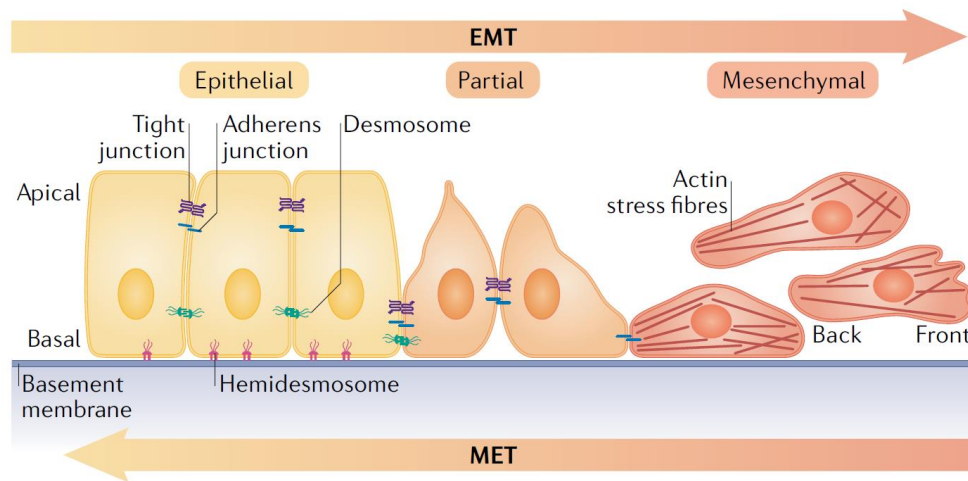
Epithelial-to-mesenchymal transition (EMT) is a cellular process in which epithelial cells undergo multiple biochemical changes resulting in the loss of apical-basal polarity, basal membrane interaction, and cell-cell adhesion. The cells acquire characteristics of mesenchymal cells, such as enhanced migratory capacity, invasiveness, and increased production of components of the extracellular matrix (ECM)<sup>81</sup>.

### 5.1 Subtypes of EMT

EMT occurs in three distinct biological settings with varying functional consequences. Type 1 EMT is associated with embryonic development and organ formation. It is necessary for the formation of the endoderm and mesoderm, as well as for the migration of neural crest cells. During Type 1 EMT, primitive epithelial cells transition into mesenchymal cells, which subsequently undergo mesenchymal-epithelial transition (MET) to form secondary epithelial cells. Type 2 EMT is associated with tissue regeneration, wound healing, and organ fibrosis. This EMT program generates fibroblasts or other related cells to repair tissues following trauma or inflammatory injury. Type 3 EMT occurs in the context of cancer progression and metastasis. Cancer cells undergo EMT to obtain invasive properties that enable them to move into blood vessels and spread to distant organs<sup>81</sup>.

### 5.2 Markers of EMT

Various molecular processes are involved in the initiation and completion of an EMT. Transcription factors are activated, specific cell surface proteins are expressed, ECM-degrading enzymes are produced, cytoskeletal proteins are reorganized, and many more (Figure 6).



**Figure 6: Process of epithelial-to-mesenchymal transition.**

Epithelial cells exhibit an apical-basal polarity and interact with the basement membrane through hemidesmosomes. They also display strong cell-cell adhesion facilitated by tight junctions, adherens junctions, and desmosomes<sup>82</sup>. During EMT, gene expression changes result in the loss of epithelial cell characteristics and acquisition of mesenchymal cell features, such as a front-to-back polarity and a highly reorganized actin cytoskeleton. Thereby, mesenchymal cells obtain migratory and invasive capabilities. Mesenchymal cells can transform back to epithelial cells by undergoing mesenchymal-epithelial transition (MET)<sup>82</sup>. The figure was modified from Dongre and Weinberg<sup>82</sup>.

These observed alterations are frequently used as indicators of an ongoing or completed EMT of cells<sup>83</sup>. A well-studied marker of EMT is the loss of E-cadherin (CDH1), a protein that is a main component of the adherens junctions, that connect neighboring epithelial cells and stabilize cell shape by interacting with the actin cytoskeleton. Transcription factors that repress CDH1 expression have been proposed as EMT-inducing transcription factors and thereby EMT markers themselves. These transcription factors include the zinc-finger proteins SNAI1 and SNAI2 (*Drosophila melanogaster* homologues Snail 1 and 2), the zinc-finger E-box-binding homeobox proteins ZEB1 and ZEB2 (Zinc finger e-Box binding homeobox 1 and 2), and the basic helix-loop-helix (bHLH) factor TWIST1 (twist family bHLH transcription factor 1). Meanwhile, several more transcription factors have been shown to induce EMT, such as forkhead box protein C2 (FOXC2)<sup>84</sup> and krueppel-like factor 8 (KLF8)<sup>85</sup>. The role of all these TFs in EMT is not limited to the repression of CDH1. They also decrease the expression of other genes relevant for cell-cell adhesion, such as the components of tight junctions (e.g. occludin, claudins) and desmosomes (e.g. desmoplakin, plakophilins), or components of epithelial intermediate filaments (cytokeratins). Moreover, they activate the expression of genes that define the mesenchymal phenotype. These include genes encoding components of mesenchymal intermediate filaments (e.g. vimentin), components of the actin cytoskeleton (e.g. actins, myosins), components of focal adhesions (e.g. integrins), proteins of the ECM (e.g. fibronectin, vitronectin, collagens, fibulins), and matrix metalloproteases (MMPs).

Additionally, many intracellular signaling pathways have been shown to be crucial for EMT<sup>86</sup>. Notably, not a single but several parallel processes in the cell lead to its mesenchymal phenotype. Liberzon et al. have built a collection of 200 genes that define a mesenchymal phenotype. This gene set is widely used as “EMT hallmark gene set” in gene set enrichment analyses (GSEAs) to identify occurring EMT in transcriptomic datasets<sup>87</sup>. Importantly, guidelines for research on EMT on behalf of the EMT international association (TEMTIA) recommend that the EMT status should not be assessed only on the basis of a small number of molecular markers but with a combinatorial approach and in conjunction with changes in cellular characteristics, such as migration or invasion properties<sup>83</sup>.

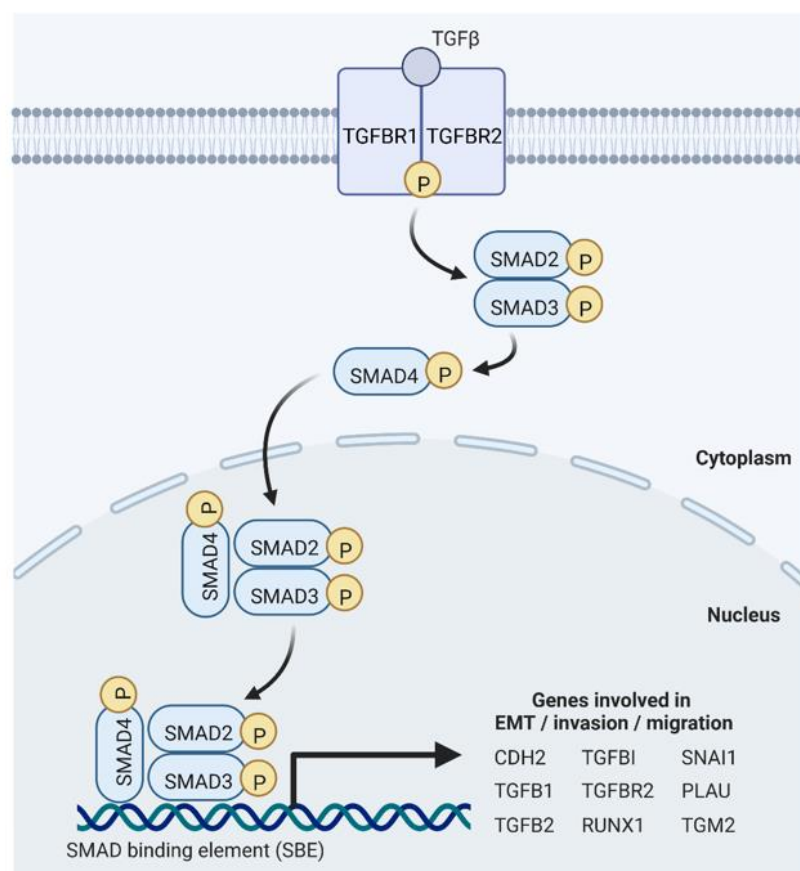
### 5.3 EMT signaling pathways involved in metastasis of PDAC

Multiple studies have shown that EMT is a crucial process in the metastatic cascade of PDAC<sup>88,89</sup>. Various extracellular signaling molecules such as transforming growth factor  $\beta$  (TGF $\beta$ ), hepatocyte growth factor (HGF), and bone morphogenetic proteins (BMPs) have been reported to induce EMT in pancreatic cancer cells<sup>90</sup>. These signaling molecules activate intracellular signaling cascades by binding to cell membrane receptors. While these signaling molecules are commonly associated with the activation of specific receptor families, it is now assumed that they can also initiate a range of parallel intracellular signaling pathways that involve the transactivation of receptors. This can result in downstream signaling cascades that are not directly linked to the initially activated receptor. In PDAC, EMT is known to involve various pathways, including the TGF $\beta$ /SMAD, rat sarcoma virus (Ras)/mitogen-activated protein kinase (MAPK), janus kinase (JAK)/signal transducer and activator of transcription proteins (STAT), and proto-oncogene tyrosine-protein kinase Src (c-Src)/focal adhesion kinase 1 (FAK) pathways<sup>90-92</sup>. As two EMT-associated signaling pathways play a significant role in this project, they will be discussed in greater detail below.

#### 5.3.1 TGF $\beta$ pathway

The TGF $\beta$  family is a multifunctional cytokine family comprising three TGF $\beta$  isoforms. To initiate signaling, TGF $\beta$  family ligands bind and assemble a heterotetrameric receptor complex of two TGF $\beta$  type I receptors (TGFBR1s) and two TGF $\beta$  type II receptors (TGFBR2s) on the cell plasma membrane, thereby activating it. Both receptor components contain cytoplasmic serine/threonine domains that phosphorylate SMAD2 and SMAD3 proteins at serine-containing sequences at their carboxy-termini. The phosphorylation enables the binding of a SMAD4 protein to the acidic tails of SMAD2 and SMAD3, resulting in a trimeric functional unit that translocates to the nucleus and binds to SMAD-binding elements (SBEs) of the DNA, where it regulates the transcription of several genes involved in EMT, cell migration, and invasion<sup>93</sup>. Both binding and transcriptional activity are supported by

partner transcription factors and chromatin modifiers, such as the SWI/SNF nucleosome positioning complex and the histone acetyltransferases p300 and CBP (cyclic AMP response element-binding protein). Fine-tuning of this pathway is accomplished by transcriptional and post-transcriptional mechanisms, epigenetic modifications and non-coding RNA-mediated regulation<sup>94</sup>. The described signaling pathway involving the proteins SMAD2, 3, and 4 is the main pathway activated by TGF $\beta$  and is therefore referred to as the "canonical TGF $\beta$  pathway" (Figure 7). However, it has been shown that TGF $\beta$  can also activate other signaling pathways, such as the MAP kinase pathway, Rho-like GTPase pathways, and the phosphoinositide 3-kinase/AKT serine/threonine kinase 1 (PI3K/AKT) pathway, which are referred to as "non-canonical TGF $\beta$  pathways"<sup>93</sup>.



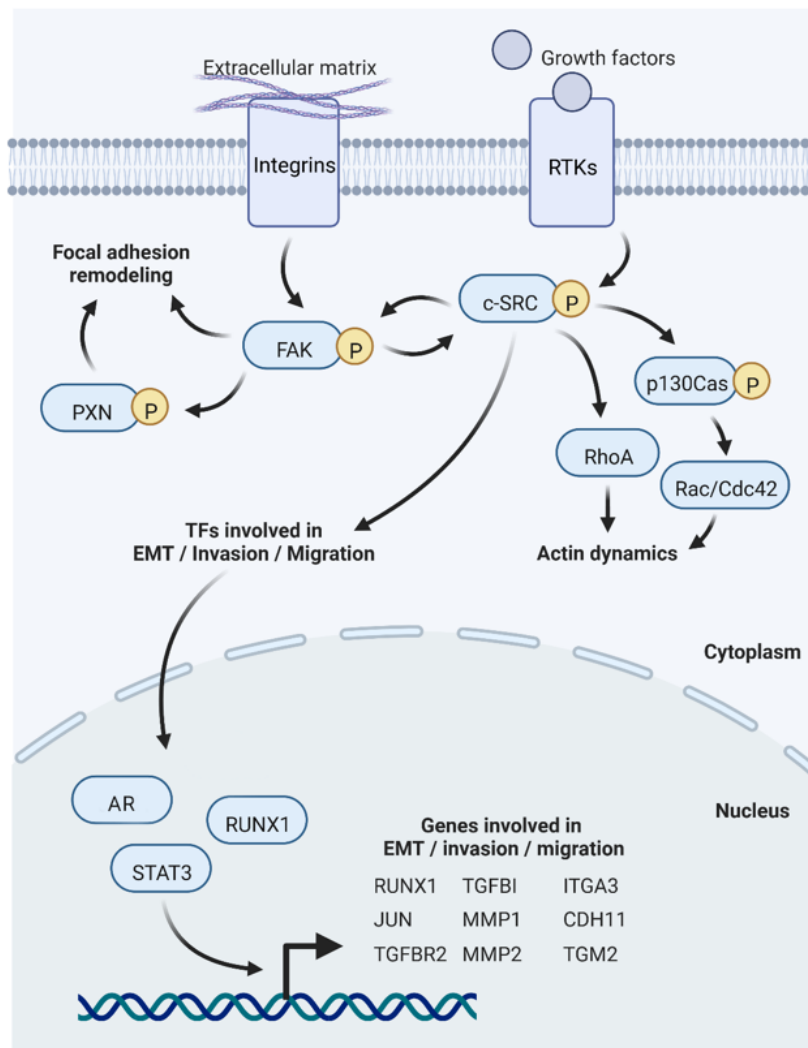
**Figure 7: Canonical TGF $\beta$  signaling.**

TGF $\beta$  binds to a heterotetrameric receptor complex composed of two TGF $\beta$  type I receptors and two TGF $\beta$  type II receptors located on the cell plasma membrane. This binding event triggers the activation of SMAD2 and SMAD3 proteins, which then form a trimeric functional unit with activated SMAD4. The resulting complex translocates to the nucleus, where it regulates gene expression by binding to specific DNA sequences known as SMAD binding elements (SBE)<sup>93</sup>. The figure was created with BioRender.com.



### 5.3.2 c-Src pathway

The process of EMT has been demonstrated to be regulated by another pathway involving the proto-oncogene tyrosine-protein kinase c-Src, a non-receptor tyrosine kinase belonging to the Src family kinases (SFKs) (Figure 8).



**Figure 8: c-Src signaling pathways.**

The c-Src signaling pathway can be activated by various stimuli such as receptor tyrosine kinases (RTKs), integrins, and other signaling proteins including focal adhesion kinase (FAK). Activation of c-Src leads to the remodeling of focal adhesions, alterations in actin dynamics, and changes in gene expression through the regulation of transcription factors which control genes involved in EMT, cell migration, and invasion<sup>56,92,95-107</sup>. The figure was created with BioRender.com.

SFKs play a crucial role in the progression, invasion, and metastasis of various cancer types<sup>92,96</sup>. The activation of c-Src signaling can occur through receptor tyrosine kinases (RTKs) such as the epidermal growth factor receptor (EGFR), which is primarily activated by epidermal growth factor (EGF), and downstream signaling molecules that phosphorylate and activate c-Src<sup>92</sup>. Additionally, c-Src acts downstream of focal adhesions, multi-protein

complexes that link the extracellular matrix to intracellular actin filaments. These cell-matrix connections are facilitated by transmembranous integrins that interact with both extracellular matrix proteins (e.g. fibronectin, collagen, laminin) and intracellular actin filaments via adaptor proteins (e.g. talin,  $\alpha$ -actinin, filamin, vinculin). In addition to structural proteins, focal adhesions also contain signal proteins such as the adaptor protein paxillin and the tyrosine kinases c-Src and FAK, which are phosphorylated upon stimulation. Particularly, c-Src is autophosphorylated at tyrosine residue 416, and FAK at the tyrosine residue 397. FAK and c-Src are interaction partners that can activate each other bidirectionally. Binding of c-Src to FAK at tyrosine 397 leads to subsequent phosphorylation of the tyrosines 407, 576, 577, 861 and 925 of FAK. Changes in the activation of these signaling molecules lead to remodeling of focal adhesions and activation of intracellular signaling cascades<sup>97</sup>. For instance, activated c-Src has been shown to stimulate Ras homolog family member A (RhoA), Rac Family Small GTPase (Rac) and cell division cycle 42 (Cdc42), three members of the Rho family of GTPases, which regulate intracellular actin dynamics and thereby alter migratory behavior of cells. Although c-Src has no intrinsic transcriptional activity, it can phosphorylate and activate the transcriptional activity of transcription factors such as runt-related transcription factor 1 (RUNX1), androgen receptor (AR) and signal transducer and activator of transcription 3 (STAT3)<sup>95,98-103,108</sup>. These transcription factors activate the expression of genes involved in EMT, cell migration, and invasion<sup>100,104-107</sup>.

## 6. The aim of the study

Pancreatic cancer has a dismal prognosis due to late diagnosis, frequent occurrence of local and distant metastasis, and high degree of resistance to therapies. Despite advances in medical science, the survival rate of patients with pancreatic cancer has not increased significantly in recent decades, and there is a projected substantial increase in the number of deaths over the next decade<sup>19</sup>. Therefore, it is critical to identify molecular mechanisms of tumorigenesis in order to understand potential pathways responsible for progression to invasive cancer. Integrative genomic analyses in PDAC provided valuable insights into pancreatic carcinogenesis and identified distinct disease subtypes that have prognostic and biological relevance and are associated with differences in therapy response<sup>15-17</sup>. LncRNAs, which are expressed in a tissue-specific manner and have been associated with various diseases, are considered promising diagnostic biomarkers and targets for therapies<sup>22,33,34</sup>. Indeed, comprehensive bioinformatic analysis identified lncRNAs whose expression was associated with different PDAC subtypes, suggesting that these lncRNAs may have differential effects within these subtypes<sup>109</sup>. Hence, defining subtype-specific expression patterns of RNAs beyond protein-coding transcripts and characterizing associated signaling pathways may help to identify unique targets for the development of personalized treatments for each individual PDAC subtype.

Thus, the objective of this doctoral research project was to identify a lncRNA that is involved in the development of the most aggressive subtype of PDAC and to investigate the underlying molecular mechanism.

The aim of this project can be sub-divided into the following objectives:

### **1. Identification of deregulated and subtype-specific lncRNAs in PDAC**

The squamous/basal-like subtype of PDAC is considered the most aggressive subtype, which has been associated with the poorest prognosis. Therefore, the objective of this study was to identify lncRNA candidates that could potentially play a role in shaping the molecular landscape of this challenging subtype.

### **2. Basic characterization of the deregulated candidate lncRNA**

After selecting the candidate lncRNA, comprehensive evaluations, including expression profiling in diverse human PDAC cell lines, and analyses of coding potential and subcellular localization, should be performed to obtain a general idea of potential roles of the lncRNA.

### **3. Characterization of cellular and morphological functions of selected lncRNA**

Elevated cell proliferation, migration, and invasion are critical in PDAC progression and metastasis. Therefore, the functional relevance of the selected lncRNA should be characterized in terms of its impact on these cellular processes.

### **4. Investigation of molecular mechanisms of selected lncRNA**

In case significant changes in the cellular assays were observed, the molecular mechanism of the selected lncRNA should be explored by analyzing its influence on the transcriptomic landscape and cellular signaling cascades.

### **5. Exploration of potential cis regulation of adjacent protein-coding genes**

If the lncRNA candidate is located near a protein-coding gene, its possible involvement in the molecular functions of the lncRNA should be investigated to gain insight into possible cis regulation.

### **6. Investigation of the lncRNA's importance for *in vivo* tumor growth and metastasis**

To further examine the role of the lncRNA and its potential adjacent protein-coding gene within a more complex biological setting, *in vivo* models should be employed.

## II. MATERIALS

### 1. Patient samples

Formalin-fixed paraffin-embedded (FFPE) blocks containing human pancreatic normal and cancer tissues were acquired from the Institute of Pathology, Martin Luther University Halle-Wittenberg, following the approval of the Ethics Committee of the Medical Faculty, Martin Luther University Halle-Wittenberg (approval No. 2015-016 and No. 2017-81).

### 2. Animals

Immunodeficient NSG™ mice (strain: #005557, NOD.Cg-Prkdc<sup>scid</sup> Il2rg<sup>tm1Wjl</sup>/SzJ) were procured from Jackson Laboratory (Bar Harbor, USA).

### 3. Cell lines

The human pancreatic cancer cell lines AsPc-1, Bxpc3, Capan-1, Colo-357, MiaPaca2, Su.86.86, Panc-1 and the human embryonic kidney (HEK)293T cells were obtained from American Type Culture Collection (ATCC, Manassas, USA). The human pancreatic cancer cell lines Pa-Tu-8988S (PATU-S) and Pa-Tu-8988T (PATU-T) were purchased from the German Collection of Microorganisms and Cell Cultures GmbH (DSMZ, Braunschweig).

### 4. Bacteria

Escherichia coli MACH1™ strain, obtained from Thermo Fisher Scientific, Waltham (USA), was used for cloning purposes. Bacteria were cultured in lysogeny broth (LB) medium comprising of 1% (w/v) tryptone, 0.5% (w/v) yeast extract, and 1% (w/v) sodium chloride. To generate LB agar plates, 1.5% (w/v) agar was added to the LB medium. The LB medium was supplemented with 100 µg ampicillin/ml for the selection of recombinant clones.

### 5. Instruments

**Table 3: Instruments**

Instrument	Name	Company
Automated electrophoresis platform	Bioanalyzer 2100 system	Agilent Technologies, Santa Clara (USA)
Blotting system	Trans-Blot® Turbo™ Transfer System	Bio-Rad Laboratories, Hercules (USA)
Cell counter	TC20™ Automated Cell Counter	Bio-Rad Laboratories, Hercules (USA)
Cell sorter	FACSMelody	BD Biosciences, Franklin Lakes (USA)

Cell sorter	BD FACS Influx cell sorter	BD Biosciences, Franklin Lakes (USA)
Centrifuge	Biofuge™ Primo	Heraeus, Hanau
Centrifuge	Centrifuge 5804R	Eppendorf®, Hamburg
Centrifuge	Eppendorf miniSpin	Eppendorf®, Hamburg
Chamber system for gel electrophoresis (Agarose)	Wide Mini-Sub Cell GT Systems	Bio-Rad Laboratories, Hercules (USA)
Chamber system for gel electrophoresis (SDS-PAGE)	Mini-PROTEAN® Tetra Cell	Bio-Rad Laboratories, Hercules (USA)
Detection system for western blots	Odyssey Infrarot Scanner	LI-COR Biosciences, Lincoln (USA)
Electronic multipipette	Pipet-Lite E4-XLS	Mettler Toledo, Columbus (USA)
Heating block	Grant Instruments™ QBD2	Thermo Fisher Scientific, Waltham (USA)
Heating block	Biometra TS1 ThermoShaker	Analytik Jena, Jena
<i>In vivo</i> imaging system (IVIS)	IVIS® Spectrum	PerkinElmer, Inc., Waltham (USA)
Incubator	Heracell™	Thermo Scientific, Waltham (USA)
Live cell analysis imaging system	IncuCyte	Sartorius AG, Göttingen
Magnetic stirrer	IKA RCT basic	IKA®-Werke GmbH & Co. KG, Staufen
Microplate reader	Infinite® M Plex	Tecan Group, Männedorf (Switzerland)
Microscope	Primovert	Carl Zeiss, Oberkochen
Multipipette	Multipipette plus	Eppendorf®, Hamburg
Pipettes	Research® plus	Eppendorf®, Hamburg
Pipetting aid	Pipetboy acu	Integra Biosciences GmbH, Biebertal
Platform shaker	IKA KS 125	Heidolph Instruments GmbH & Co. KG, Schwabach
qPCR system	LightCycler® 480 II	Roche Holding AG, Basel (Switzerland)
Rotary microtome	Leica RM2235	Leica Camera AG, Wetzlar
Thermocycler	Thermocycler Rotor-Gene™ 6000	Biometra GmbH, Göttingen
Tissue processor	Leica ASP300S	Leica Camera AG, Wetzlar
Ultrasonic lab homogenizer	UP200S Lab Homogenizer	Hielscher Ultrasonics, Teltow
Voltage source	EV3020 Consort	Carl Roth, Karlsruhe
Vortexer	VF2	IKA®-Werke GmbH & Co. KG, Staufen

## 6. Consumables

**Table 4: Consumables**

Material	Name	Company
Adhesion slides	SuperFrost® Plus	Menzel GmbH und Co. KG, Braunschweig
Cell culture plates, flasks	CELLSTAR® Cell Culture	Greiner Bio-One GmbH, Frickenhausen
Closure foil	PARAFILM M®	Bemis Company, Inc., Neenah (USA)
Cover slips	Cover Glasses for Microscopy	Glaswarenfabrik Karl Hecht GmbH, Sandheim
Filter paper	Whatman® paper	GE Healthcare Life Sciences, Darmstadt
Nitrocellulose membrane	Amersham™ Protran™ Premium NC Nitrocellulose Membranes	Amersham Biosciences, Amersham (United Kingdom)
PCR tubes with lids	PCR 8er-SoftStrips, 0.2 ml	Biozym Scientific GmbH, Hessisch Oldendorf
Reaction tubes	Safe Seal Reaktionsgefäße	Sarstedt AG & Co, Nümbrecht
Transwell™ membrane inserts	Corning™ 3422	Corning Inc., Corning, New York (USA)

## 7. Chemicals and reagents

**Table 5: Chemicals and reagents**

Chemical/Reagent	Company
Agarose	Biozym Scientific GmbH, Hessisch Oldendorf
Ampicillin	Carl Roth GmbH & CO. KG, Karlsruhe
APS (ammonium peroxodisulfate)	Carl Roth GmbH & CO. KG, Karlsruhe
Biotin-16-UTP	Jena Bioscience, Jena
Blasticidin	Santa Cruz, Dallas (USA)
BSA (Bovine serum albumin)	Sigma-Aldrich® Chemie GmbH, Steinheim
DAPI (4',6-Diamidino-2-phenylindole)	Thermo Fisher Scientific, Waltham (USA)
Dasatinib	Sigma-Aldrich® Chemie GmbH, Steinheim
D-Luciferin Potassium Salt IVISbrite	PerkinElmer, Inc., Waltham (USA)
DMEM (Dulbecco's modified eagle medium)	Gibco™, Lifetechnologies™, Carlsbad (USA)
DMSO (Dimethyl sulfoxide)	Sigma-Aldrich® Chemie GmbH, Steinheim
EDTA (Ethylenediaminetetraacetic acid)	Merck KGaA, Darmstadt
Ethanol	Sigma-Aldrich® Chemie GmbH, Steinheim
FBS (Fetal bovine serum)	Gibco™, Lifetechnologies™, Carlsbad (USA)
Glycerol	AppliChem GmbH, Darmstadt
HCl (Hydrochloric acid)	Merck KGaA, Darmstadt
Isopropanol	Sigma-Aldrich® Chemie GmbH, Steinheim
Lipofectamine RNAiMax	Thermo Fisher Scientific, Waltham (USA)
Lysogeny broth	Sigma-Aldrich® Chemie GmbH, Steinheim
Matrigel	Corning Inc., Corning, New York (USA)

Mercaptoethanol	AppliChem GmbH, Darmstadt
Methanol	Sigma-Aldrich® Chemie GmbH, Steinheim
Mounting medium (Entellan)	Sigma-Aldrich® Chemie GmbH, Steinheim
NaCl (Sodium chloride)	AppliChem GmbH, Darmstadt
Na-deoxycholate	Sigma-Aldrich® Chemie GmbH, Steinheim
NaOH (Sodium hydroxide)	Carl Roth GmbH & CO. KG, Karlsruhe
Paraformaldehyde	Sigma-Aldrich® Chemie GmbH, Steinheim
PBS (Phosphate-buffered saline)	Gibco™, Lifetechnologies™, Carlsbad (USA)
Penicillin/Streptomycin	Gibco™, Lifetechnologies™, Carlsbad (USA)
Phosphatase inhibitor	Thermo Fisher Scientific, Waltham (USA)
Ponceau S	Carl Roth GmbH & CO. KG, Karlsruhe
PP2	Abcam, Cambridge (UK)
PP3	Abcam, Cambridge (UK)
Protease inhibitor cOmplete Tablets EASYpack	Roche Holding AG, Basel (Switzerland)
Puromycin	Thermo Fisher Scientific, Waltham (USA)
RepSox	Selleckchem, Houston (USA)
Ribolock RNase inhibitor	Thermo Fisher Scientific, Waltham (USA)
Ribonucleosid vanadyl complex	New England Biolabs Inc., Ipswich (USA)
Rotiphorese® NF-Acrylamid/Bis-Lösung 40%	Carl Roth GmbH & CO. KG, Karlsruhe
RPMI (Roswell Park Memorial Institute)-1640	Gibco™, Lifetechnologies™, Carlsbad (USA)
SDS (Sodium dodecyl sulfate)	Carl Roth GmbH & CO. KG, Karlsruhe
TEMED (Tetramethylethyldiamin)	Carl Roth GmbH & CO. KG, Karlsruhe
TGFβ-1 (Transforming growth factor β-1)	PeproTech, Rocky Hill (USA)
TRIS (Tris(hydroxymethyl)aminomethane)	Carl Roth GmbH & CO. KG, Karlsruhe
Triton X-100	AppliChem GmbH, Darmstadt
Trypsin-EDTA	Gibco™, Lifetechnologies™, Carlsbad (USA)
Tryptone	Carl Roth GmbH & CO. KG, Karlsruhe
TurboFect transfection reagent	Thermo Fisher Scientific, Waltham (USA)
Tween 20	Carl Roth GmbH & CO. KG, Karlsruhe
USER enzyme	New England Biolabs Inc., Ipswich (USA)
Yeast extract	Carl Roth GmbH & CO. KG, Karlsruhe
Yellow sample buffer	Thermo Fisher Scientific, Waltham (USA)

## 8. Kits and systems

**Table 6: Kits and systems**

Kit	Name	Company
Chromatin immunoprecipitation kit	SimpleChIP® Enzymatic Chromatin IP Kit #9003	Cell Signaling Technology, Danvers (USA)
Cloning kit for sequencing	TOPO™ TA cloning™ kit	Thermo Fisher Scientific, Waltham (USA)
Cluster generation kit	PE Cluster Kit cBot-HS Kit	Illumina, San Diego (USA)
Coomassie staining kit	Colloidal Blue Staining Kit	Thermo Fisher Scientific, Waltham (USA)

Gel purification kit	GeneJET Gel Extraction Kit	Thermo Fisher Scientific, Waltham (USA)
Genomic DNA purification kit	ReliaPrep™ gDNA Tissue Miniprep System	Promega Corporation, Fitchburg (USA)
Library preparation kit	NEBNext® Ultra™ Directional RNA Library Prep Kit	New England Biolabs Inc., Ipswich (USA)
Luciferase assay system	Dual-Glo™ System or Nano-Glo™ System	Promega Corporation, Fitchburg (USA)
<i>In vitro</i> transcription kit	MEGAscript™ T7 Transcription Kit	Thermo Fisher Scientific, Waltham (USA)
Midiprep kit	PureLink™ HiPure Plasmid Midiprep Kit	Thermo Fisher Scientific, Waltham (USA)
Miniprep kit	PureLink™ Quick Plasmid Miniprep Kit	Thermo Fisher Scientific, Waltham (USA)
PCR kit	Phusion® High-Fidelity PCR	Thermo Fisher Scientific, Waltham (USA)
PCR purification kit	GeneJET PCR Purification Kit	Thermo Fisher Scientific, Waltham (USA)
PCR purification system	AMPure XP system	Beckman Coulter, Beverly (USA)
Protein assay kit	Pierce® BCA Protein Assay Kit	Pierce, Rockford, Illinois (USA)
qPCR master mix	primaQUANT qPCR SYBR Green Master Mix	Steinbrenner Laborsysteme GmbH, Wiesenbach
Reverse transcriptase kit	M-MLV Reverse Transcriptase	Promega Corporation, Fitchburg (USA)
RNA purification kit	RNeasy FFPE Kit	Qiagen N.V., Hilden
Three-step stain set	Richard-Allan Scientific	Thermo Fisher Scientific, Waltham (USA)
Western blotting standard	Precision Plus Protein™ Western C Standard	Bio-Rad Laboratories Inc., München

## 9. Plasmids

Table 7: Plasmids

Plasmid	Cat. No.	Company
EF1a_FOXA2_P2A_Hygro_Barcode	120439	Addgene, Watertown (USA), gift from Prashant Mali
LBid.nC.LINC00261.SFFV.mCMV.eGFP.P2A		This study
LBid.nC.MCS.SFFV.mCMV.eGFP.P2A		Gift from Jan-Henning Klusmann <sup>110</sup>
Lenti-dCas9-KRAB-blast	89567	Addgene, Watertown (USA), gift from Gary Hon
lentiGuide-Puro	52963	Addgene, Watertown (USA), gift from Feng Zhang
Luciferase-GFP		Gift from Kunal Rai, MD Anderson Cancer Center, Houston (USA)
pCDH-CMV-FOXA2-EF1-Puro		This study



pCDH-CMV-MCS-EF1-Puro	CD510B-1	System Biosciences, Palo Alto (USA)
pGL3-Basic vector	E1751	Promega Corporation, Fitchburg (USA)
pcDNA™3.1	V79020	Thermo Fisher Scientific, Waltham (USA)
pcDNA™3.1-LINC00261		This study
pGL3-CDH1		This study
pGL3-LINC00261		This study
pGL3-minCMV		This study
pL-CRISPR.EFS.GFP	57818	Addgene, Watertown (USA), gift from Benjamin Ebert
pL-CRISPR.EFS.GFP-sgLINC00261		This study
pMD2.G	12259	Addgene, Watertown (USA), gift from Didier Trono
pNL[NlucP/minP/Hygro]	CS188006	Promega Corporation, Fitchburg (USA)
pNL[NlucP/SBE/Hygro]	CS177101	Promega Corporation, Fitchburg (USA)
pRL-SV40	E2231	Promega Corporation, Fitchburg (USA)
pRL-TK	E2241	Promega Corporation, Fitchburg (USA)
psPAX2	12260	Addgene, Watertown (USA), gift from Didier Trono
pX330-Cas9-P2A-mCherry	98750	Addgene, Watertown (USA), gift from Jinsong Li
pX330-Cas9-P2A-mCherry-sgFOXA2		This study
pX330-Cas9-P2A-mCherry-sgLINC00261		This study

## 10. Cloning reagents

The restriction enzymes, ligases, and buffers used for the purpose of cloning were obtained from Thermo Fisher Scientific, located in Waltham, USA.

**Table 8: Cloning reagents**

Reagent	Cat. No.
FastAP Thermosensitive Alkaline Phosphatase (1 U/μl)	EF0651
FastDigest AgeI/BshTI	FD1464
FastDigest BamHI	FDO054
FastDigest Buffer (10x)	B64
FastDigest Esp3I/ BsmBI	FD0454
FastDigest Green Buffer (10x)	B72
FastDigest HindIII	FD0504
FastDigest NheI	FD0974
FastDigest XhoI	FD0694
Rapid Ligation Buffer (5x)	K1423
T4 DNA Ligase (5 U/μl)	EL0011
T4 Polynucleotide Kinase (10 U/μl)	EK0031

## 11. Oligonucleotides

The oligonucleotides were acquired from Eurofins Genomics in Ebersberg.

**Table 9: Oligonucleotides for cloning**

Name	Forward Sequence (5' to 3')	Reverse Sequence (3' to 5')	Restriction sites
CDH1 promoter -770 +92	ATTACTCGAGTTGCAGTGAGC CGAGATCGT	ATTAAGCTTCCGGGTGCGGT CGGGT	XhoI forward, HindIII reverse
c- <i>Src</i>	ATTAGCTAGCATGGGGAGCA GCAAGAGCAAG	ATTAGGATCCCTATAGGTTCT CTCCAGGCTGGTACTG	NheI forward, BamHI reverse
FOXA2	ATTAGCTAGCATGCTGGGAG CGGTGAAGATGGAAG	ATTAGGATCCTTAAGAGGAGT TCATAATGGGCCGGGAG	NheI forward, BamHI reverse
LINC00261	ATTAACCGGTGAAATGGCATC AAGATGGTT	ATTACTCGAGTATACTTAATA ATTTTATTA	AgeI forward, XhoI reverse
LINC00261	ATTAAGCTTGAATGGCATC AAGATGGTT	ATTAGCGGCCGCTATACTTAA TAATTTTATTA	HindIII forward, NotI reverse
LINC00261 promoter -1000 +100	ATTAGCTAGCAGACCTGGAG ACTGTCTTTGA	ATTACTCGAGCTGCCGAGCG TCCAGCT	NheI forward, XhoI reverse
Minimal CMV	TCGAGGGTAGGCGGTACGG TGGGAGGTCTATATAAGCAG AGCA	AGCTTGCTCTGCTTATATAGA CCTCCACCGTACACGCCTA CCC	XhoI forward, HindIII reverse

**Table 10: Oligonucleotides for qRT-PCR**

Name	Forward Sequence (5' to 3')	Reverse Sequence (3' to 5')
CDH1	CGGGAATGCAGTTGAGGATC	AGGATGGTGTAAAGCGATGGC
CDH1 promoter 1	AGGAGAGTCTCTTGAACCCGG	GCCTCCCAAAGTGCTAGGATTT
CDH1 promoter 2	AGCTTGGGTGAAAGAGTGAGAC	TTGCTAGGGTCTAGGTGGGTTA
CDH1 promoter 3	GGGCATCCGTAGAAATAAGGC	GTACCCCACTTTCCTTAGACCG
CDH2	AAGTGGCAAGTGGCAGTAAAAT	CCAGTCTCTTCTGCCTTTGT
FN1	GAGCTGAGTGAGGAGGGAGA	CAGGCGCTGTTGTTTGTGAA
FOXA2	CAGAACTCCATCCGCCACTC	AACATGTTGCCCGAGTCAGG
GAPDH	CTGGTAAAGTGGATATTGTTGCCAT	TGGAATCATATTGGAACATGTAAACC
LINC00261	ATAGGCCAGAGAGCAACCT	ACCACTACCCAGCATTGTG
LINC00261 promoter 1	AAAACACTCCGAAAGCCTGGA	GTTAGGATGGTCAAGAAGCCC
LINC00261 promoter 2	ACTGATCCCGCCGATAAGATA	ACACAAGAAGCACAGAAAAGCC
LINC00261 promoter 3	AGTGCATGACTTGAAGGATGA	CTCTCAGATCGAATCCCCAGAC
MALAT1	GAATTGCGTCATTTAAAGCCTAGTT	GTTTCATCCTACCACTCCCAATTAAT
NEAT-1	CCAGTTTTCCGAGAACCAAA	ATGCTGATCTGCTGCGTATG
PLAU	TCCACCTGTCCCCGCAG	TTTGGAGTCGCTCACGACC
PPIA	GTCAACCCACCGTGTCTT	CTGCTGTCTTTGGGACCTTGT
RUNX1	ACTGTGATGGCTGGCAATGAT	GACTTGGGTTGGGTTTGTG
SNAI1	CAATCGGAAGCCTAACTACAGC	GACAGAGTCCCAGATGAGCATT
SNAI2	GAACTGGACACACATACAGTGAT	ACTCACTCGCCCCAAAGATG
TGFB1	CCAATATTGCTTCAGCTCCAC	AGTTGGCATGGTAGCCCTTG
TGFB2	CAACAGCACCAGGGACTTG	AGACAGTTTCCGAGGGGAAG

TGFBI	TAACGGCCAGTACACGCTTT	G TTCAGCAGGTCTCTCAGGG
TGFBR2	TCTGGACCCTACTCTGTCTGTG	CATAATCTTTTACTTCTCCCACTGC
TGM2	CACCCACACCTACAAATACCCA	GTCAAAGTCACTGCCCATGTTC
VIM	ATGCGTGAAATGGAAGAGAACT	TGTAGGTGGCAATCTCAATGTC

Table 11: Small interfering RNAs (siRNAs)

Name	Forward Sequence (5' to 3')	Reverse Sequence (3' to 5')
FOXA2 siRNA 1	GCCGGGCCGGCCUCCGAGA [dT][dT]	UCUCGGAGGCCGGCCCGGC [dT][dT]
FOXA2 siRNA 2	CUGACUCGGGCAACAUGUU [dT][dT]	AACAUGUUGCCCGAGUCAG [dT][dT]

Table 12: Single guide RNAs (sgRNAs)

Name	Forward Sequence (5' to 3')	Reverse Sequence (3' to 5')
FOXA2 CRISPR/Cas sgRNA	CACCGATGAACATGTCGTCGTACGT	AAACACGTACGACGACATGTTTCATC
LINC00261 CRISPR/Cas sgRNA 1	CACCGCAAACCCCCCTCAAGCGCGT	AAACACGCGCTTGAGGGGGGTTTGC
LINC00261 CRISPR/Cas sgRNA 2	CACCGACTCGCCTCTTAGAAAAGCGG	AAACCGCCTTTCTAAGAGGCGAGTC
LINC00261 CRISPRi sgRNA 1 (i1)	CACCGGCGCGCTCCCTACCTGCGG	AAACCCGAGGTAGGGAGCGCGCC
LINC00261 CRISPRi sgRNA 2 (i2)	CACCGTGGGCGCGCTCCCTACCTG	AAACCAGGTAGGGAGCGCGCCAC

## 12. Antibodies

Table 13: Primary antibodies

Antibody	Species	Dilution	Cat. No.	Company
c-Src	rabbit	1:1000	2109	Cell Signaling Technology, Danvers (USA)
E-cadherin (CDH1)	rabbit	1:1000	3195	Cell Signaling Technology, Danvers (USA)
FOXA2	rabbit	1:1000	8186	Cell Signaling Technology, Danvers (USA)
GAPDH	rabbit	1:5000	G8795	Sigma-Aldrich® Chemie GmbH, Steinheim
Histone H3	rabbit	-	4620	Cell Signaling Technology, Danvers (USA)
IgG	rabbit	-	2729	Cell Signaling Technology, Danvers (USA)
N-cadherin (CDH2)	rabbit	1:1000	13116	Cell Signaling Technology, Danvers (USA)
Phospho-SMAD2	rabbit	1:1000	18338	Cell Signaling Technology, Danvers (USA)
Phospho-SMAD3	rabbit	1:1000	9520	Cell Signaling Technology, Danvers (USA)
Phospho-c-Src	rabbit	1:1000	2101	Cell Signaling Technology, Danvers (USA)
RPL7	rabbit	1:5000	A400-741A	Biomol GmbH, Hamburg
SMAD2	rabbit	1:1000	5339	Cell Signaling Technology, Danvers (USA)
SMAD3	rabbit	1:1000	9523	Cell Signaling Technology, Danvers (USA)
Vimentin	rabbit	1:1000	5741	Cell Signaling Technology, Danvers (USA)

Table 14: Secondary antibodies

Antibody	Antigen	Dilution	Cat. No.	Company
mouse-IRDye 680	Mouse IgG	1:10,000	926-68072	LI-COR Biosciences, Lincoln (USA)
mouse-IRDye 800	Mouse IgG	1:10,000	926-32212	LI-COR Biosciences, Lincoln (USA)
rabbit-IRDye 680	Rabbit IgG	1:10,000	926-68073	LI-COR Biosciences, Lincoln (USA)
rabbit-IRDye 800	Rabbit IgG	1:10,000	926-32213	LI-COR Biosciences, Lincoln (USA)

### 13. Buffers and solutions

Unless otherwise specified, the solutions and buffers mentioned were prepared using double-distilled water (ddH<sub>2</sub>O).

Table 15: Cell culture media

Medium	Receipt
DMEM with 0% FBS (pH 7.3-7.4)	<ul style="list-style-type: none"> <li>• DMEM, 4.5 g/l Glucose</li> <li>• 1% (v/v) Penicillin/Streptomycin</li> </ul>
DMEM with 10% FBS (pH 7.3-7.4)	<ul style="list-style-type: none"> <li>• DMEM, 4.5 g/l Glucose</li> <li>• 10% (v/v) FBS</li> <li>• 1% (v/v) Penicillin/Streptomycin</li> </ul>
DMEM for transfection (pH 7.3-7.4)	<ul style="list-style-type: none"> <li>• DMEM, 4.5 g/l Glucose</li> <li>• 10% (v/v) FBS</li> </ul>

Table 16: Lysis and wash buffers

Medium	Receipt
PBS (Phosphate-buffered saline)	<ul style="list-style-type: none"> <li>• 137 mM NaCl</li> <li>• 2.7 mM KCl</li> <li>• 10 mM Na<sub>2</sub>HPO<sub>4</sub></li> <li>• 2 mM KH<sub>2</sub>PO<sub>4</sub></li> </ul>
RIPA lysis buffer	<ul style="list-style-type: none"> <li>• 50 mM Tris-HCl, pH 8.0</li> <li>• 150 mM NaCl</li> <li>• 1% (w/v) IGEPAL CA-630</li> <li>• 0.5% (v/v) Na-deoxycholate</li> <li>• 0.1% SDS</li> </ul>
RSB buffer	<ul style="list-style-type: none"> <li>• 10 mM Tris, pH 7.4</li> <li>• 10 mM NaCl</li> <li>• 3 mM MgCl<sub>2</sub></li> </ul>

RSBG40 buffer	<ul style="list-style-type: none"> <li>• 10 mM Tris, pH 7.4</li> <li>• 10 mM NaCl</li> <li>• 3 mM MgCl<sub>2</sub></li> <li>• 10% Glycerol</li> <li>• 0.5% Nonidet P-40</li> <li>• 0.5 mM Dithiothreitol</li> </ul>
TRIZOL	<ul style="list-style-type: none"> <li>• 0.8 M Guanidinium thiocyanate</li> <li>• 0.4 M Ammonium thiocyanate</li> <li>• 0.1 Sodium acetate, pH 5.0</li> <li>• 5% Glycerol</li> <li>• 48% Roti-Aqua-Phenol</li> </ul>

Table 17: Western blot buffers

Buffer	Receipt
10x Blotting buffer	<ul style="list-style-type: none"> <li>• 250 mM Tris</li> <li>• 1.92 M Glycine</li> </ul>
10x SDS running buffer	<ul style="list-style-type: none"> <li>• 0.25 M Tris</li> <li>• 1.92 M Glycine</li> <li>• 1% (w/v) SDS</li> </ul>
10x TBS buffer	<ul style="list-style-type: none"> <li>• 247 mM Tris</li> <li>• 1.37 M NaCl</li> <li>• 26.8 M KCl</li> </ul>
1x TBS-T buffer	<ul style="list-style-type: none"> <li>• 1 ml Tween 20</li> <li>• 100 ml TBS buffer 10x</li> </ul>
4x Laemmli sample buffer	<ul style="list-style-type: none"> <li>• 250 mM Tris-HCl, pH 6.8</li> <li>• 30% (w/v) Glycerol</li> <li>• 0.03% (w/v) Bromphenol Blue 1% (w/v) stock</li> <li>• 8% (w/v) SDS</li> <li>• 10% (w/v) <math>\beta</math>-Mercaptoethanol</li> </ul>
BSA in TBS-T	<ul style="list-style-type: none"> <li>• 5% BSA</li> <li>• TBS-T buffer 1x</li> </ul>
Milk in TBS-T	<ul style="list-style-type: none"> <li>• 5% Milk powder</li> <li>• TBS-T buffer 1x</li> </ul>
Ponceau S	<ul style="list-style-type: none"> <li>• 0.1% (w/v) Ponceau</li> <li>• 5% Acetic Acid</li> </ul>

Resolving gel buffer (pH 8.8)	<ul style="list-style-type: none"> <li>• 1.5 M Tris</li> <li>• 0.4% (w/v) SDS</li> </ul>
Stacking gel buffer (pH 6.8)	<ul style="list-style-type: none"> <li>• 0.5 M Tris</li> <li>• 0.4% (w/v) SDS</li> </ul>

## 14. Software

Table 18: Software

Software	Developer
Citavi	Swiss Academic Software GmbH, Wädenswil (Switzerland)
FlowJo™	BD Biosciences, Franklin Lakes (USA)
GraphPad Prism 9	GraphPad Software Inc., San Diego (USA)
GSEA (Gene Set Enrichment Analysis)	Broad Institut of MIT and Harvard, University of California, San Diego (USA)
ImageJ (Fiji)	Wayne Rasband, NIH (USA)
Image Studio™ Acquisition Software	LI-COR Biosciences, Lincoln (USA)
IncuCyte Analysis Software	Sartorius AG, Göttingen
Living Image® Software	Caliper Life Science, Hopkinton (USA)
Microsoft Office	Microsoft Corporation, Redmond (USA)

## 15. Online tools and databases

Table 19: Online tools and databases

Tool / Database	Provider
Broad Institute CRISPR design tool	Broad Institut of MIT and Harvard, Cambridge (USA)
ENCODE project	Stanford University, Stanford (USA)
Genomic Data Commons (GDC) data portal	National Institutes of Health (USA)
NCBI BLAST (Basic Local Alignment Search Tool)	National Library of Medicine, Bethesda (USA)
R2 Genomics Analysis and Visualization Platform	University of Amsterdam, Amsterdam (Netherlands)
RNAseam	University of Campinas, São Paulo (Brazil)
TCGA (The Cancer Genome Atlas) and GTEx (Genotype-Tissue Expression) projects	National Cancer Institute and National Human Genome Research Institute (USA)
UCSC Genome Browser	University of California, Santa Cruz (USA)
Venn Diagram Tool	VIB/UGent, Gent (Belgium)

## III. METHODS

Parts of the text presented in this chapter are revised versions of the text published in the original research article “LINC00261 Is Differentially Expressed in Pancreatic Cancer Subtypes and Regulates a Pro-Epithelial Cell Identity”, authored by Dorn et al. and published in *Cancers* (2020)<sup>58</sup>.

Unless otherwise indicated, all commercially available reagents and kits were used according to manufacturer’s instructions.

### 1. Cell biology methods

#### 1.1 Cell culture

Capan-1, Colo-357, HEK293T, MiaPaca2, Panc-1, PATU-S and PATU-T cells were cultivated in DMEM medium supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin (P/S). The pancreatic cancer cell lines AsPc-1, BxPc3 and Su.86.86 were cultured in RPMI-1640 medium supplemented with 10% FBS and 1% P/S. All cell lines were grown at 37°C in a humidified incubator with 20% O<sub>2</sub> and 5% CO<sub>2</sub>. All pancreatic cancer cell lines, except for BxPc3 (KRAS WT), harbored mutations in KRAS and p53 resembling human PDAC<sup>111</sup>, and showed metastatic potential in *in vivo* settings<sup>112,113</sup>. AsPc-1, BxPc3, Capan-1, Colo-357 cells harbored nonsense or missense mutations in SMAD4, while MiaPaca2, Su.86.86, Panc-1, PATU-S and PATU-T cells did not<sup>111,114</sup>.

#### 1.2 Treatment with TGFβ and TGFβ type 1 receptor inhibitor

For TGFβ treatment, 1.0 - 1.5x10<sup>6</sup> cells of each cell line were seeded on a 10 cm plate on the previous day, starved for 24 h using medium supplemented with 0.5% FBS and then treated with 10 ng/ml TGFβ, which was diluted in 0.5% FBS in DMEM or RPMI-1640. Protein and RNA isolation were performed after 24 h, 48 h, and 72 h of TGFβ treatment. For treatment with the TGFBR1 inhibitor RepSox (200 nM) 1.5x10<sup>5</sup> A549 and 2x10<sup>5</sup> Panc-1 cells were seeded on a 6-well plate and starved in DMEM containing 0.5% FBS for 24 h. The cells were then treated using the same medium and harvested after 24 h, 48 h, and 72 h for RNA isolation.

#### 1.3 Treatment with Src kinase inhibitors

For treatment with the Src inhibitor Dasatinib, 6x10<sup>5</sup> Panc-1 cells were plated on a 6 cm plate on the previous day and then treated with 100 nM Dasatinib or an equal volume of DMSO as a control. After an incubation period of 16 h, cells were harvested for protein and RNA isolation. For the migration and invasion assays, 100 nM Dasatinib or DMSO was

added to the cell suspension before pipetting the cells into the upper chamber of a transwell membrane. The cells were then incubated for 16 h at 37°C in a humidified incubator. The same experimental procedure was applied for the treatment of cells with the c-Src inhibitor PP2 and its negative control, PP3, both of which were used at a concentration of 10 mM.

#### 1.4 siRNA transfection

Panc-1 cells were transfected with two independent FOXA2 siRNAs (Table 11) at a final concentration of 40 nM using Lipofectamine RNAiMax according to the manufacturer's instructions. After an incubation period of 72 h, cells were harvested for RNA isolation.

#### 1.5 CRISPR interference

Lentivirus was produced in HEK293T cells ( $4 \times 10^6$  cells in a 10 cm plate). Briefly, the Lenti-dCas9-KRAB-blast plasmid (10  $\mu$ g) or sgRNA coding plasmids (10  $\mu$ g) were co-transfected with lentiviral packaging plasmids, psPAX2 (5  $\mu$ g) and pMD2.G (2.5  $\mu$ g) using TurboFect reagent according to the manufacturer's instructions. The virus was harvested 72 h after transfection. Initially, the lenti-dCas9-KRAB-blast plasmid was transduced into  $3 \times 10^5$  Panc-1 cells using a 6-well plate. Two days later, cells were treated with 10  $\mu$ g/ml Blastidicin for selection of cells that were transduced with the plasmid. The lenti-dCas9-KRAB-blast Panc-1 cells were then transduced with sgRNA coding plasmids (lentiGuide-Puro as control, CRISPRi sgRNAs targeting LINC00261 named sgRNA i1 and i2) for 48 h and selected by treating with 2  $\mu$ g/ml Puromycin.

#### 1.6 CRISPR/Cas9

$3 \times 10^5$  cells were seeded in each well of a 6-well plate with 3 ml of antibiotic-free standard growth medium 24 h prior to transfection. For transfection, 4  $\mu$ g of pL-CRISPR.EFS.GFP-sgLINC00261 and pX330-Cas9-P2A-mCherry-sgLINC00261 were mixed with TurboFect reagent according to the manufacturer's instructions. 72 h after transfection, mCherry/GFP-double positive single cells were sorted into 96-wells using FACS Melody. The cell clones were expanded, and genomic DNA (gDNA) and RNA were isolated to assess LINC00261 promoter deletions. The same experimental protocol was employed to knock out FOXA2 using the plasmid pX330-Cas9-P2A-mCherry-sgFOXA2.

#### 1.7 2D cell proliferation assay

In order to assess 2D cell proliferation, a total of  $5 \times 10^3$  cells were seeded into individual wells of a 96-well plate and incubated for a period of 24 h prior to the first confluence measurement by using the IncuCyte Live Cell Analysis Imaging System. Measurements



were performed at 6-hour intervals up to 72 h. The growth curve was determined using the IncuCyte Analysis Software, and the doubling time was calculated based on the resulting growth curve.

### 1.8 Clonogenic assay

For the clonogenic assays, 1,000 cells were seeded in six-well plates and maintained at 37°C in a humidified incubator for 21 days, with weekly medium changes. To enable analysis of colony formation, the colonies were stained with 0.01% crystal violet for 60 min, and the area occupied by the colonies was determined using the ImageJ software.

### 1.9 Cell migration and invasion assays

Transwell migration and invasion assays were performed using transwell inserts with 8  $\mu\text{m}^2$  pore size. The membranes were coated with 100  $\mu\text{l}$  migration matrix (0.1% gelatin in 0.02 M acetic acid) or invasion matrix (50  $\mu\text{g}/\text{ml}$  collagen IV, 5  $\mu\text{g}/\text{ml}$  laminin, 2 mg/ml gelatin) and incubated for 2 h at room temperature on a rotating platform. Excess liquid was removed, and the membrane was dried for 1 h in a sterile environment. Subsequently, a total of  $7.5 \times 10^4$  Panc-1 or PATU-T cells, suspended in 100  $\mu\text{l}$  serum-free medium, were seeded into the upper chamber. The lower chamber was filled with 500  $\mu\text{l}$  of complete DMEM medium that contained 10% FBS. Following an incubation period of 16 h (Panc-1) or 6 h (PATU-T), non-migrated cells present in the upper chamber were wiped away with a cotton swab, while the migrated cells on the bottom of the membrane were fixed and stained with the Richard-Allan Scientific three-step stain set. To quantify the number of migrated or invaded cells, five images were captured per transwell chamber, utilizing either  $\times 10$  (PATU-T cells) or  $\times 15$  (Panc-1 cells) magnification, and subsequently analyzed via the ImageJ Cell counter. The data were calculated by determining the average cell count per image across all five images.

## 2. Animal work

The experiments involving animals were carried out in the laboratory of Andrea Viale at MD Anderson Cancer Center, located in Houston, USA, with assistance from I-Lin Ho and Rutvi Shah. The mice were housed in a pathogen-free facility, located at the University of Texas MD Anderson Cancer Center. All procedures were performed in accordance with the guidelines and regulations established by the Institutional Animal Care and Use Committee (IACUC).

## 2.1 Orthotopic xenograft mouse model

To transfect HEK293T cells with the luciferase-GFP plasmid (10  $\mu$ g), the lentiviral packaging plasmids psPAX2 (5  $\mu$ g) and pMD2.G (2.5  $\mu$ g) were co-transfected using TurboFect reagent according to the manufacturer's instructions. After 72 h of transfection, lentivirus carrying luciferase-GFP plasmid was collected, and  $3 \times 10^6$  cells of each cell line were transduced with the virus. BD FACS Influx cell sorter was employed to sort GFP-positive cells 72 h post-transduction, with untransfected wild-type (WT) cells used to set sorting gates. To exclude dead cells, samples were treated with 1  $\mu$ g/ml 4',6-diamidino-2-phenylindole (DAPI). Flow cytometry experiments were conducted at the MD Anderson South Campus Flow Cytometry and Cell Sorting Facility, and FlowJo™ was used for data analysis, which excluded doublets and dead cells during gating. The sorted cells were reseeded into 10 cm plates and cultivated at 37°C in a humidified incubator until transplantation. For transplantation,  $4 \times 10^5$  GFP-positive cells were suspended in a mixture of FBS-free DMEM and Matrigel (1:1 ratio) and orthotopically injected into 8-week-old immunodeficient mice under anesthesia with isoflurane. For xenograft studies, n=5 mice were used for each cell line. The animals were observed daily by the animal facility staff, and all mice were sacrificed upon reaching the termination conditions of the first mice. Tumor growth was monitored weekly via bioluminescence imaging.

## 2.2 *In vivo* imaging

Six weeks after orthotopic transplantation, the termination criteria were met. Ten minutes before euthanasia, the luciferase substrate, D-luciferin, was administered to the mice. The stable integrated bioluminescent reporter in the cells produces light in the presence of a substrate, which allows visualization of the cells within the mouse. Livers and lungs of the mouse were collected, and the luminescence signal in the organs were determined by the *in vivo* imaging system IVIS® Spectrum and quantified by the Living Image® Software. The liver and lungs were subsequently fixed overnight in 4% paraformaldehyde (PFA). The primary tumors were also harvested and weighed before being stored in PFA for fixation.

## 2.3 Histopathology

After fixation in PFA, tissue specimens were transferred to 70% ethanol and underwent embedding in paraffin using a tissue processor within one week. For histopathological analysis, 10  $\mu$ m thick consecutive sections were obtained using a rotary microtome. From each series, one section was baked, deparaffinized, and stained with hematoxylin (1 min) and eosin (30 sec), followed by dehydration and a short incubation in xylene (2 min). Finally, the sections were mounted on microscopic slides using mounting medium.

### 3. Molecular biology methods

#### 3.1 Cloning

The coding sequence of FOXA2 was PCR-amplified from the EF1a\_FOXA2 vector and inserted into pCDH-CMV-MCS-EF1-Puro using NheI/BamHI restriction sites. The LINC00261 sequence was PCR-amplified from Panc-1 cDNA and inserted into the LBid.nC.MCS.SFFV.mCMV.eGFP.P2A vector using the AgeI and XhoI restriction sites, as well as into the pcDNA™3.1 vector using HindIII and NotI restriction enzymes. The sequences of the primers used are provided in Table 9. Lentivirus carrying the pCDH-CMV-FOXA2-EF1-Puro or LBid.nC.LINC00261.SFFV.mCMV.eGFP.P2A plasmid was produced in HEK293T cells, and transduced cells were selected by adding 2 µg/ml Puromycin to the culture medium. For promoter analysis, the E-cadherin promoter region from -770 to +92 and the putative promoter region of LINC00261 from -1000 to +100 were amplified using primers with NheI/XhoI (E-cadherin promoter) or XhoI/HindIII (LINC00261 promoter) restriction enzyme sites from genomic DNA of Panc-1 cells. The amplified PCR products were then inserted into the upstream region of the firefly luciferase gene of the pGL3-Basic vector. Additionally, an oligo containing the sequence of a minimal CMV promoter was inserted into the pGL3-Basic vector and used as a control plasmid for all luciferase experiments. Single guide RNA (sgRNA) sequences targeting the LINC00261 gene with CRISPRi and for cutting out the putative LINC00261 promoter were designed using the Broad Institute CRISPR design tool. For CRISPRi, two independent sgRNAs were selected and cloned into the lentiGuide-Puro plasmid. For this purpose, oligonucleotides containing the sgRNA expressing sequence and BsmBI sticky ends were synthesized, annealed, phosphorylated and ligated with the vector. MACH1™ competent cells were used for transformation. The sgRNA sequences are provided in Table 12. For removing the putative promoter of LINC00261 the two sgRNAs were cloned into the pX330-Cas9-P2A-mCherry vector and the pL-CRISPR.EFS.GFP in the same way as described above.

#### 3.2 Genomic DNA and total RNA extraction followed by qRT-PCR

Genomic DNA was extracted from cell pellets (~3x10<sup>6</sup> cells) by using the ReliaPrep™ gDNA Tissue Miniprep System according to manufacturer's instructions. Total RNA was isolated by phenol/chloroform extraction. For this, 3x10<sup>5</sup> cells in a well of a 6-well plate were harvested using 1 ml Trizol reagent and transferred to a 1.5 ml tube. The samples were mixed with 200 µl chloroform, shaken vigorously for 15 seconds, and incubated at room temperature for 3 min. For phase separation, the samples were centrifuged for 15 min at 4°C and 12,000 g. The aqueous phase containing the RNA was transferred to a new tube with 500 µl of isopropanol, mixed, and incubated at room temperature for 10 min. The samples were then centrifuged for 20 min at 4°C and 12,000 g to precipitate the RNA. The

precipitated RNA was washed with 1 ml of 70% ethanol and centrifuged at 4°C and 8,000 g for 5 min. The supernatant was removed, and the pellet was air-dried for 10 min before being dissolved in 30 µl of RNase-free H<sub>2</sub>O. Then, 1 µg of RNA was reverse transcribed into cDNA using M-MLV Reverse Transcriptase and the provided 5x Reaction Buffer, following the manufacturer's instructions. Subsequently, quantitative real-time PCR (qRT-PCR) was performed in triplicates in a 384-well plate with the LightCycler® 480 II using 6.25 ng cDNA, 0.7 µM forward and reverse primers, and primaQUANT qPCR SYBR Green Master Mix. The primers used for qRT-PCR are listed in Table 10. The amplification of GAPDH was used as a reference for qRT-PCR. Relative expression values were calculated using the  $2^{-\Delta\Delta Ct}$  method<sup>115</sup>.

### 3.3 RNA extraction from PDAC and normal pancreas tissue samples

Total RNA was extracted from 35 normal pancreas and 42 PDAC tissue blocks with tumor cell content greater than 65%. The RNA was extracted from three 10 µm paraffin sections using the RNeasy FFPE Kit according to the manufacturer's instructions. Subsequently, 1 µg of total RNA was transcribed into cDNA for qRT-PCR. The patient characteristics are provided in Supplementary Table S1.

### 3.4 *In vitro* transcription

For *in vitro* transcription and biotin-labeling the MEGAscript™ T7 Transcription Kit was used according to the manufacturer's instructions. Specifically, 1 µg of the pcDNA 3.1-LINC00261 vector, which had been linearized using the NotI restriction enzyme, served as the template for the reactions. The reactions were incubated for 16 h at 30°C, using a Biotin-16-UTP:UTP ratio of 1:25, and terminated by the addition of 1 µl Turbo DNase. RNA was precipitated with lithium chloride as described in the manufacturer's instructions. Integrity and size of the resulting RNA was verified by agarose gel electrophoresis.

### 3.5 Protein pulldown with biotinylated RNA

At first, 20 µl Streptavidin Dynabeads C1 were washed three times with B&W buffer (5mM Tris pH 7.5, 0.5 mM EDTA, 1M NaCl). Then, 80 pmol of *in vitro* transcribed RNA were added and incubated for 1 h at RT. Unbound RNA was removed with 200 µl of RIPA lysis buffer. Subsequently, 500 µl of cell lysate (3 mg of protein) was added and incubated for 30 min at RT. After incubation, the lysate was removed, and the beads were washed twice with 1 ml of lysis buffer. The beads were then resuspended in 30 µl of 4x Laemmli sample buffer and boiled at 95°C for 5 minutes. Lastly, SDS-PAGE was performed, the gel was stained using the Colloidal Blue Staining Kit, and the results were visualized with the Odyssey infrared scanner. The subsequent analysis of differential protein bands by mass

spectrometry was performed by Andrea Sinz's group at Martin Luther University Halle-Wittenberg.

### 3.6 Protein extraction and western blot analysis

Cells were washed twice with phosphate-buffered saline (PBS), lysed in RIPA lysis buffer supplemented with protease and phosphatase inhibitors, and the resulting cell debris was removed by centrifugation at 16,000 g for 30 min. The protein concentration was determined using the Pierce® BCA Protein Assay Kit. 30 µg of extracted proteins were boiled at 95°C for 5 min in 4x Laemmli sample buffer, separated by 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and then transferred onto a nitrocellulose membrane by wet-blotting. The membranes were blocked with 5% skimmed milk (or 5% BSA for phosphoproteins) in TBS-T prior to antibody incubation. Diluted primary antibodies in the blocking solution were added overnight at 4°C. The primary antibodies are listed in Table 13. Secondary antibodies (Table 14) were added for 2 h at room temperature. Antibody signals were visualized using the Odyssey infrared scanner.

### 3.7 Reverse phase protein array (RPPA)

For reverse phase protein array (RPPA),  $1 \times 10^6$  cells of each cell line were collected in triplicates, washed with PBS, flash-frozen in liquid nitrogen, and stored at -80 °C. Protein extraction and RPPA were conducted at the MD Anderson RPPA Core facility. The expression levels of 499 distinct proteins, including phosphoproteins, were assessed. All antibodies were validated by the core facility. Protein expression values were reported as the mean values in  $\log_2^{116}$ .

### 3.8 Subcellular fractionation

Subcellular fractionations were executed by Carolin Neu according to a previously described protocol<sup>117</sup>. Briefly, AsPc-1, Capan-1, and Panc-1 cells were scraped off the plate in their respective growth medium, and pelleted at 500 g and 4°C for 5 min. The cell pellets were washed once with PBS and centrifuged at 800 g and 4°C for 5 min. After washing, the cells were resuspended in 1 ml of RSB buffer, incubated for 3 min on ice, and centrifuged at 1,000 g and 4°C for 5 min. Total RNA was isolated from one-fifth of the cells. The remaining cells were resuspended with four times its volume in RSBG40 buffer containing 40 U/ml Ribolock, and 5 mM ribonucleosid vanadyl complex, and incubated on ice for 3 min. After centrifugation at 4,500 g for 3 min at 4°C, the supernatant was collected as the cytoplasmic fraction, and the nuclear pellet was resuspended in RSBG40 containing one-tenth volume of detergent (3.3% v/v Na-deoxycholate and 6.6% v/v Tween 20) and incubated on ice for 5 min. Nuclei were pelleted again at 4,500 g for 3 min at 4°C, washed

with RSBG40, and collected at 9,300 g for 5 min. RNA from both nuclear and cytoplasmic fractions was isolated using Trizol.

### 3.9 Chromatin immunoprecipitation (ChIP)

ChIP assays were performed using the SimpleChIP™ Enzymatic Chromatin IP Kit according to manufacturer's instructions. Panc-1 cells, either untreated or treated with TGFβ, underwent fixation with 1% formaldehyde for DNA and protein cross-linking. Chromatin was subsequently sheared using a UP200S Lab Homogenizer (3 cycles of sonication: 20" each, 30" rest; amplitude 30%). Next, 10 µg of the chromatin fraction were incubated with 0.5 µg of antibodies specific for FOXA2, Histone H3 (positive control) and IgG (negative control), and the complex was precipitated by Protein G magnetic beads (30 µl). The protein-DNA cross-link was then reversed, the DNA purified, and the enrichment of DNA sequences was assessed using qPCR. The primers used in this study are listed in Table 10, and the genomic locations of the analyzed regions are indicated in Figure 26E and Figure 40E, respectively.

### 3.10 Luciferase reporter assay

A total of  $6 \times 10^4$  Panc-1 cells were seeded into each well of a 24-well plate. After 24 h, the cells were transfected with 500 ng of either the pGL3-CDH1, the pGL3-LINC00261 or the control pGL3-minCMV promoter construct using TurboFect reagent. To ensure normalization, 10 ng of pRL-SV40 Renilla expression construct was included for each transfection. After 48 h, the cells were harvested, and the luciferase activity was measured by using the Dual-Glo® Assay System. The relative luciferase activity was calculated using pGL3-minCMV as control. For the measurement of SMAD-binding element (SBE) activity, either the NanoLuc® pNL[NlucP/SBE/Hygro] or the pNL[NlucP/minP/Hygro] construct (500ng) was transfected along with the Renilla control vector pRL-TK (10 ng). Luciferase activity was then measured using the Nano-Glo® Reporter Assay System.

## 4. Bioinformatics

The bioinformatic analyses were carried out by Markus Glaß from Stefan Hüttelmaier's group at Martin Luther University Halle-Wittenberg.

### 4.1 Analysis of LINC00261 expression in PDAC samples

The normalized expression values provided in the supplementary table of Bailey et al.<sup>15</sup> were used to cluster the international cancer genome consortium (ICGC) PDAC samples based on their RNA expression. Initially, the 2,000 genes displaying the highest variation in their expression values were selected using the coefficient of variation as a measure for

variability. As the normalized expression data contained negative values, the overall minimal value of these 2,000 genes was added as a constant to all expression values to obtain only positive expression values. Next, non-negative matrix factorization was applied using the R-package NMF<sup>118</sup>, using Brunet's algorithm, rank = 4 and 500 iterations. Each sample was then assigned to the cluster with the highest corresponding likelihood. For differential expression analysis between PDACs and normal pancreatic tissue samples, gene-level RNA-seq read counts of TCGA primary tumor PDAC samples and GTEx V7 normal pancreas tissue were obtained through the GDC data portal and the GTEx portal, respectively. The combination of these data provided read count information of 53,045 genes for 177 primary tumor samples and 248 normal pancreas tissue samples. Differential gene expression was assessed using R/edgeR<sup>119</sup> by applying TMM normalization. CPM transformation was used to obtain normalized expression values. The Kaplan-Meier and gene expression correlation analyses of the Bailey PDAC dataset were determined using the R2 Genomics Analysis and Visualization Platform.

#### 4.2 RNA-seq data analysis

Total RNA was extracted using the Trizol reagent as described in chapter III.3.2. RNA integrity and quantity were assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system. The library preparation and sequencing were performed by Novogene Co., Ltd., Beijing (China). In detail, 1 µg RNA per sample was used as input material for RNA sample preparation. NEBNext® Ultra™ Directional RNA Library Prep Kit was used for generating sequencing libraries, following the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. Poly-T oligo-attached magnetic beads were used for mRNA purification from total RNA. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X). First-strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNaseH-). Subsequently, second-strand cDNA synthesis was performed using DNA Polymerase I and RNase H. In the reaction buffer, dNTPs with dTTP were replaced by dUTP. The remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, NEBNext Adaptor with hairpin loop structure was ligated to prepare for hybridization. To select cDNA fragments of 250-300 bp in length, library fragments were purified with the AMPure XP system. Size-selected, adaptor-ligated cDNA was treated with 3 µl USER Enzyme at 37°C for 15 min, followed by 5 min at 95°C before PCR. PCR was performed with Phusion HighFidelity DNA polymerase, Universal PCR primers and Index(X) Primer. Finally, products were purified (AMPure XP system), and library quality was assessed on the Agilent Bioanalyzer 2100 system. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using PE Cluster Kit cBot-HS according to the manufacturer's instructions. After cluster generation, the library

preparations were sequenced on a Novaseq 6000 platform, and paired-end reads were generated. Raw data (raw reads) of FASTQ format were processed through in-house scripts, and clean data (clean reads) were obtained by removing reads containing adapter and poly-N sequences and reads with low quality from raw data. The quality of clean data, including Q20, Q30, and GC content, was calculated. All downstream analyses were based on high-quality clean data. RNA-seq datasets were analyzed using the Galaxy web platform<sup>120</sup>. First, reads with a minimum of 20 bp were aligned to human genome build GRCh38/hg38 using STAR<sup>121</sup>. Subsequently, the featureCounts tool<sup>122</sup> was used to count reads according to GRCh38.87 human gene annotation. Next, differential expression analysis was performed using the DESeq2 tool<sup>123</sup>. The list of differently expressed genes was used for Gene Set Enrichment Analysis (GSEA) to identify specifically enriched hallmark gene sets. The overlap of genes is shown with Venn diagrams, which were generated using an online available Venn tool (see Table 19).

## 5. Statistics

Unless otherwise indicated, all data were reported as standard error of the mean (SEM). Statistical analysis was performed using GraphPad Prism software 9.0, and differences were considered significant when  $p < 0.05$ . Student's t-test, Mann-Whitney's test, or one- or two-way ANOVA test were used as appropriate. Statistical significance was indicated by \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , and \*\*\*\*  $p < 0.0001$ . Experiments were generally repeated at least three times.



## IV. RESULTS

Parts of the text and figures presented in this chapter are revised versions of the text and figures published in the original research article “LINC00261 Is Differentially Expressed in Pancreatic Cancer Subtypes and Regulates a Pro-Epithelial Cell Identity”, authored by Dorn et al. and published in *Cancers* (2020)<sup>58</sup>.

### 1. Subtype-specific gene expression analysis of PDAC samples

In order to identify lncRNAs associated with PDAC subtypes *in silico* analyses were performed, using the publicly available PDAC dataset of Bailey et al.<sup>15</sup>, which defined four disease subtypes by RNA expression analysis: squamous, pancreatic progenitor, immunogenic, and ADEX (Figure 9A). The previously published NMF algorithm<sup>118</sup> was applied to the ICGC PDAC data to identify these four described disease subtypes in the dataset. A total of 25 samples were assigned to the ADEX subtype, 26 samples to the immunogenic subtype, 16 samples to the pancreatic progenitor subtype and 29 samples to the squamous subtype. It was previously shown that patients with tumors characterized by the squamous subtype had a significantly worse overall survival compared to patients with tumors of all other disease subtypes<sup>15</sup>. To identify potential disease driving mechanisms responsible for dismal patient prognosis, differently expressed RNAs in the squamous subtype versus all other subtypes were examined, which led to the identification of 2,279 RNAs ( $p < 0.05$ ). By applying an absolute fold change (FC) cut-off of 2.0 and 0.5, 438 genes were found to be downregulated, whereas 178 genes were upregulated in the squamous subtype. The R2 Genomics Analysis and Visualization Platform was utilized to assess the prognostic relevance of all 616 genes on overall survival using the median expression of each gene as a cut-off for defining high and low expression groups. This analysis identified 199 genes being significantly associated with disease survival including 19 lncRNAs (Figure 9A).

### 2. Expression of LINC00261 in PDAC

#### 2.1 Expression in PDAC patient's samples

By applying these stepwise analyses LINC00261 was identified as the lncRNA with the most significant difference between the identified groups showing a strong downregulation in the squamous subtype compared to all other published subtypes (Figure 9B). In addition to the integrative analysis of the Bailey dataset, the expression changes of lncRNAs in different PDAC subtypes were also analyzed using publicly available RNA-seq data from the Cancer Genome Atlas Pancreatic Adenocarcinoma (TCGA-PAAD) and the Genotype-Tissue Expression (GTEx) project. These analyses demonstrated that LINC00261

expression was also significantly downregulated in the basal-like PDAC subtype defined by Moffitt et al.<sup>16</sup>, which closely resembles the squamous subtype<sup>15</sup> (Figure 9C). Moreover, comparison of LINC00261 expression in PDAC and normal pancreatic tissues (NP) of these datasets revealed significantly lower expression of LINC00261 in PDAC tissues (Figure 9D). Furthermore, RNA expression analysis of formalin-fixed paraffin-embedded normal pancreas (NP) and PDAC tissue from Martin Luther University Halle-Wittenberg (Figure 9E, Supplementary Table S1) indicated significantly lower expression of LINC00261 in PDAC compared to normal pancreatic tissue.

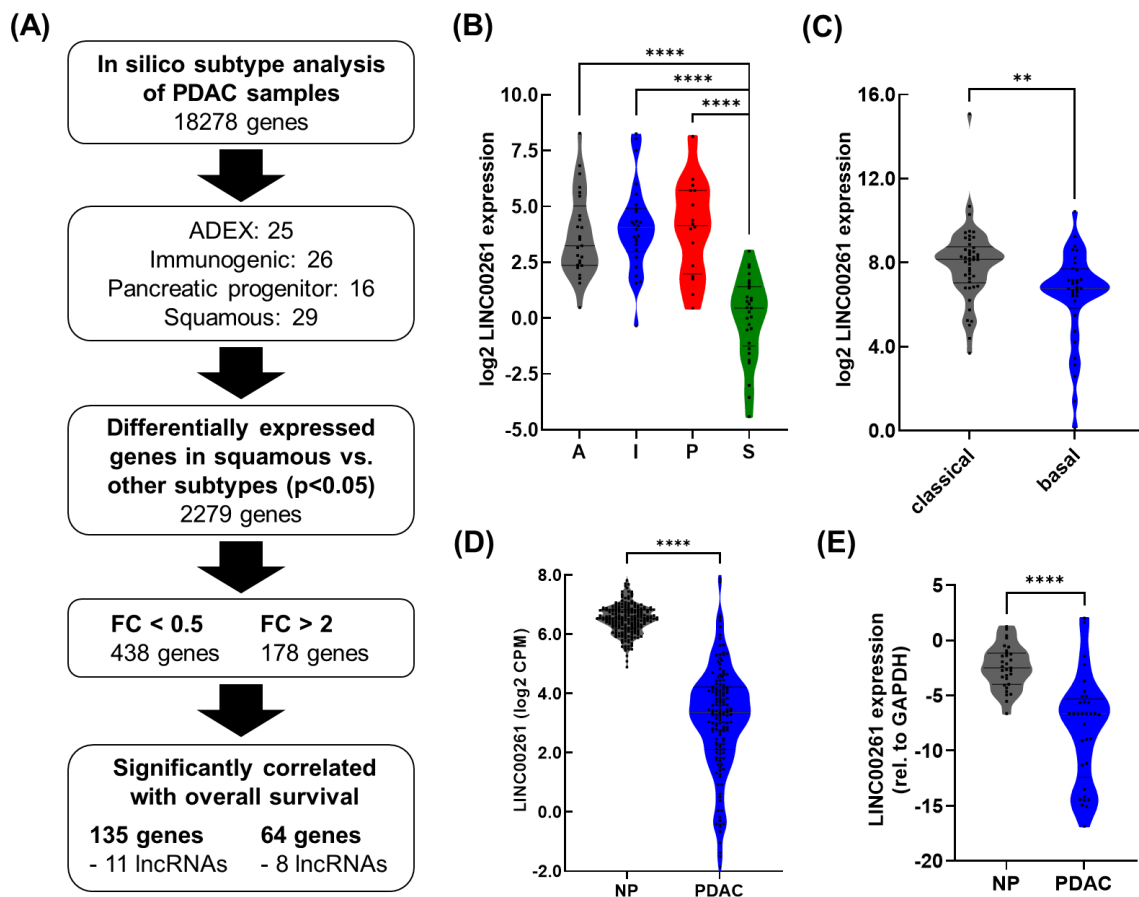
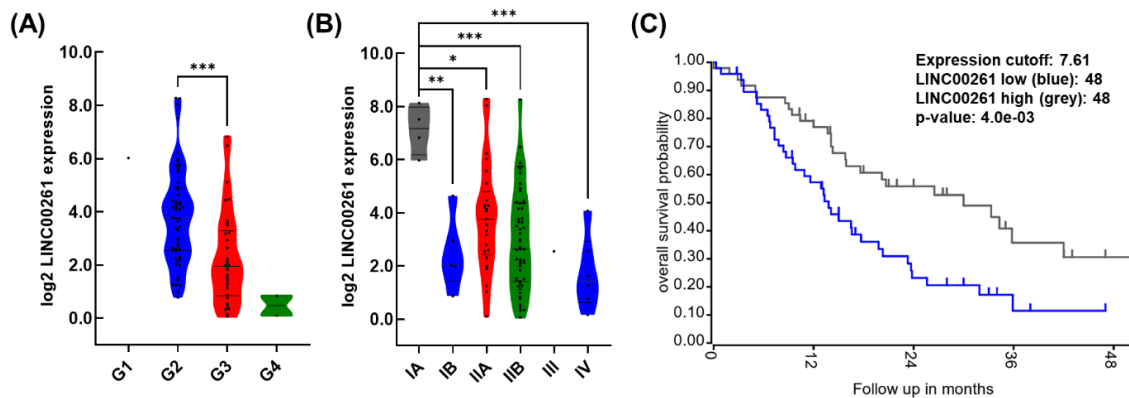


Figure 9: Analysis of LINC00261 expression in PDAC.

(A) Flowchart that outlines the overall strategy employed in this study to identify lncRNAs potentially linked to PDAC progression and patient survival; ADEX, Aberrantly Differentiated Endocrine Exocrine; FC, Fold Change; (B) Analysis of the Bailey PDAC dataset revealed a significant downregulation of LINC00261 expression in the squamous (S) compared to the pancreatic progenitor (P), immunogenic (I) and ADEX (A) subtype (\*\*\*\*  $p < 0.0001$ , one-way ANOVA); (C) Analysis of the TCGA PAAD dataset according to Moffitt's classification highlighted significant downregulation of LINC00261 expression in the basal-like compared to the classical subtype (\*\*  $p < 0.01$ , unpaired t-test); (D, E) Analysis of LINC00261 expression in publicly available TCGA and GTEx datasets (D, normal pancreas (NP):  $n = 177$ , PDAC:  $n = 248$ ) and in 35 normal pancreatic tissues and 42 PDAC tissues (E) showed significantly lower LINC00261 expression in pancreas adenocarcinoma compared to normal pancreas (log<sub>2</sub>-transformed values, \*\*\*\*  $p < 0.0001$ , Mann-Whitney U test). The figure was modified from Dorn et al.<sup>58</sup>.

## 2.2 Correlation with PDAC stage, grade and patient survival

Further in-depth analysis of LINC00261 revealed an inverse correlation between its expression and tumor grade (Figure 10A) and tumor stage (Figure 10B). Intriguingly, high expression of LINC00261 was associated with significantly better overall survival in PDAC patients using the Bailey dataset (Figure 10C).

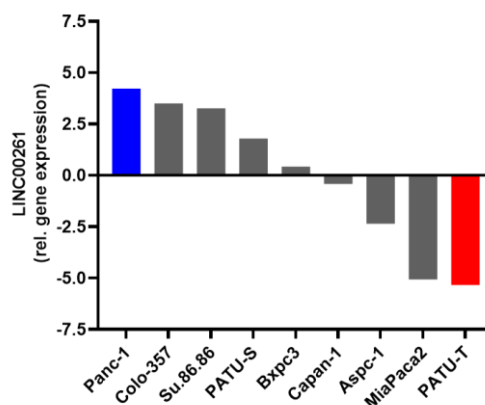


**Figure 10: LINC00261 expression correlates with PDAC stage, grade and patient survival.**

(A, B) LINC00261 expression is significantly lower in high grade (G1: n=1, G2: n=56, G3: n=34, G4: n=2) and high-stage tumors (IA: n=4, IB: n=5, IIA: n=25, IIB: n=55, III: n=1, IV: n=6), \* p<0.05, \*\* p<0.01, \*\*\* p<0.001, \*\*\*\* p<0.0001, one-way ANOVA; (C) Survival analysis for PDAC patients with low LINC00261 (blue line, n=48) versus high LINC00261 (grey line, n=48) expression (Bailey dataset, <http://r2.amc.nl>, Log rank test). The figure was modified from Dorn et al.<sup>58</sup>.

## 2.3 Expression in PDAC cell lines

The expression of LINC00261 in available PDAC cell lines was evaluated using qRT-PCR. The results demonstrated a substantial variability in the expression of LINC00261 among the tested cell lines. The Panc-1 cells exhibited the highest expression of LINC00261, while the PATU-T cells showed the lowest expression (Figure 11).



**Figure 11: Relative gene expression of LINC00261 in various PDAC cell lines.**

Gene expression of LINC00261 relative to the average of all analyzed PDAC cell lines. Panc-1 cells show the highest, while PATU-T cells show the lowest LINC00261 expression.

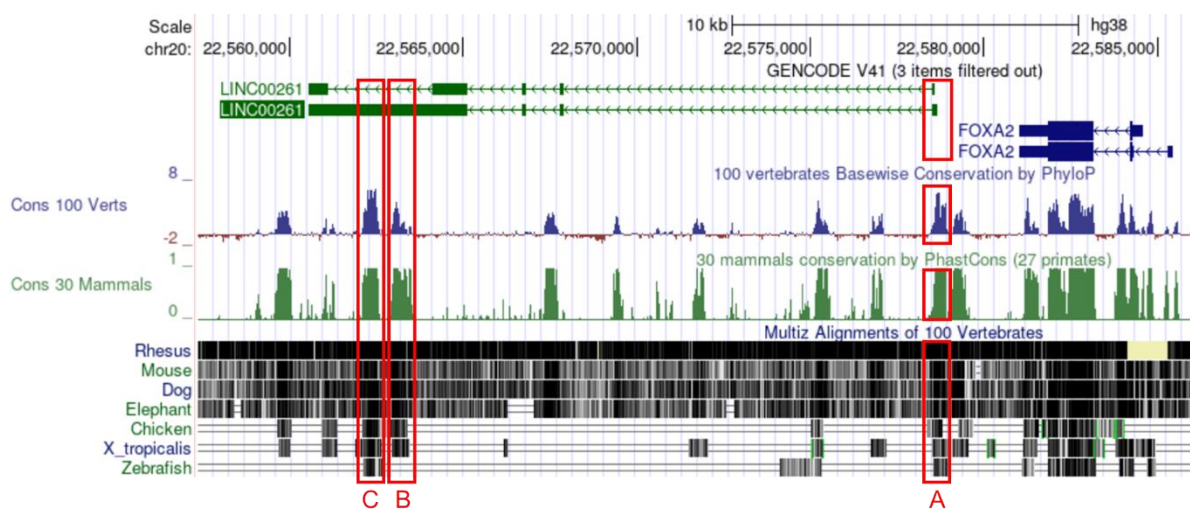
In summary, subtype-resolved gene expression analyses identified commonly deregulated lncRNAs as well as subtype-specific expression differences that may contribute to the intrinsic molecular and prognostic differences between PDAC subtypes. As a result of these comprehensive and unbiased analyses, LINC00261 was identified as a differentially expressed lncRNA between PDAC subtypes. LINC00261 emerged as an interesting candidate for follow-up studies due to its abundance, significant downregulation in the squamous/basal-like subtype, and prognostic relevance.

### 3. Characterization of LINC00261

The initial step in characterizing the candidate lncRNA involved an analysis of its sequence conservation across various species, its coding-probability, and subcellular localization.

#### 3.1 Conservation of the LINC00261 locus

To determine the evolutionary conservation of the LINC00261 locus, the UCSC genome browser's conservation analysis tools PhyloP and PhastCons were utilized. The degree of conservation in 100 vertebrate and 30 mammalian genomes was illustrated using these tools (Figure 12). It is noteworthy that the LINC00261 locus is present in all mammalian genomes, with particularly strong conservation in primates. However, outside of mammals, such as in birds, amphibians, and fish, LINC00261 conservation is restricted to a few regions, which are highlighted with red boxes in Figure 12. Specifically, exon 1 (A) and two regions within exon 4 (B and C) exhibit particularly high sequence conservation. The conservation level of regions A and C is comparable to that of the neighboring protein-coding gene FOXA2. The strong conservation of these regions of LINC00261, extending up to the zebrafish genome, suggests a physiological significance of this lncRNA.



**Figure 12: Evolutionary sequence conservation of LINC00261.**

The extract of the UCSC Genome Browser (GRCh38/hg38) shows the conserved regions of LINC00261 identified by the PhyloP (conservation across 100 vertebrates) and PhastCons (conservation across 30

mammals) conservation analysis tools. The additional track displays multiple alignments of seven vertebrate species and measurements of evolutionary conservation using the two methods above. The red boxes indicate three highly conserved regions of LINC00261 in vertebrates.

### 3.2 Coding potential of LINC00261

The coding potential of LINC00261 was analyzed using the algorithms RNAsamba<sup>124</sup> and Coding Potential Calculator 2.0<sup>125</sup>. Both algorithms classified LINC00261 as a non-coding RNA (Figure 13A). Moreover, the calculated coding probability value of LINC00261 closely resembled that of well-characterized lncRNAs, including metastasis associated lung adenocarcinoma transcript 1 (MALAT1), X-inactive specific transcript (XIST), and nuclear enriched abundant transcript 1 (NEAT1). This similarity emphasizes the non-coding nature of LINC00261.

### 3.3 Localization of LINC00261

To gain further insights into the potential functions of LINC00261, nuclear and cytoplasmic fractionation experiments were performed on pancreatic cancer cell lines to obtain information about the subcellular localization of LINC00261. The enrichment of MALAT1 and NEAT1, two well-known nuclear lncRNAs<sup>126</sup>, as well as of the cytoplasmic mRNAs of GAPDH and PPIA, were monitored to verify the purity of the respective fractions. In all cell lines examined, LINC00261 was found to be present in both the nucleus and the cytoplasm, with a higher abundance in the nucleus (Figure 13B).

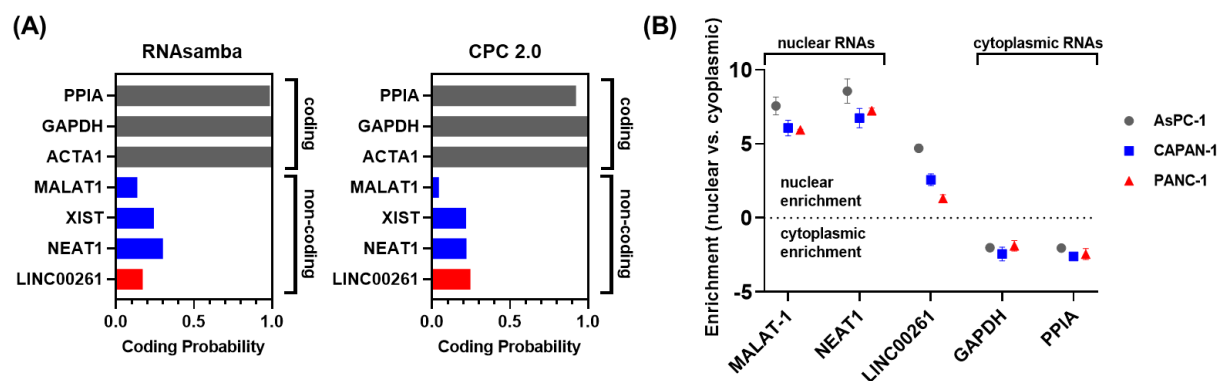


Figure 13: LINC00261 is a non-coding RNA that localizes to both the nucleus and cytoplasm.

(A) Calculation of coding probability by two different algorithms (RNAsamba<sup>124</sup> and Coding Potential Calculator (CPC) 2.0<sup>125</sup>). The bar graphs show the coding probabilities for a set of coding RNAs (PPIA, GAPDH and ACTA1) and non-coding RNAs (LINC00261, NEAT1, XIST and MALAT1); (B) Cellular fractionation in pancreatic cancer cell lines revealed a slight predominance of LINC00261 in the nucleus over the cytoplasm. Figure 13B was modified from Dorn et al.<sup>58</sup>.

## 4. Cellular function of LINC00261

In order to investigate the cellular function of LINC00261 in PDAC, manipulation of its expression was performed in PDAC cell lines. In particular, downregulation of LINC00261 was carried out in the Panc-1 cell line, which exhibited the highest expression of LINC00261, while upregulation of LINC00261 was performed in the PATU-T cell line, which exhibited the lowest expression of LINC00261 (see Figure 11).

### 4.1 CRISPR-based knockdown of LINC00261

Foremost, two CRISPR-based knockdown systems of LINC00261 were established in Panc-1 cells. Firstly, a CRISPR interference (CRISPRi) approach<sup>127</sup> was employed, utilizing two independent LINC00261-specific single guide RNAs (sgRNAs) that were stably introduced into dCas9-KRAB expressing Panc-1 cells. This approach successfully reduced the level of LINC00261, leaving only 5-6% of its expression in the cells (Figure 14A). In a second approach, the standard CRISPR/Cas9 system<sup>128</sup> was utilized to delete the potential LINC00261 promoter (~1,600 bp) applying two sgRNAs. Fluorescence-activated cell sorting was used to generate single cell clones after transient transfection of Cas9 and the two sgRNAs. Individual clones were expanded, and genomic DNA was isolated to perform a PCR-based screening for the presence of a ~250 bp fragment that was only detectable in promoter-deleted clones (Figure 14B, upper panel). Next, gene expression was analyzed by qRT-PCR showing a strong downregulation of LINC00261 in respective promoter knockout (KO) clones compared to wild-type (WT) clones. A total of three WT and three KO clones were isolated through this CRISPR/Cas9-mediated targeting strategy (Figure 14B).

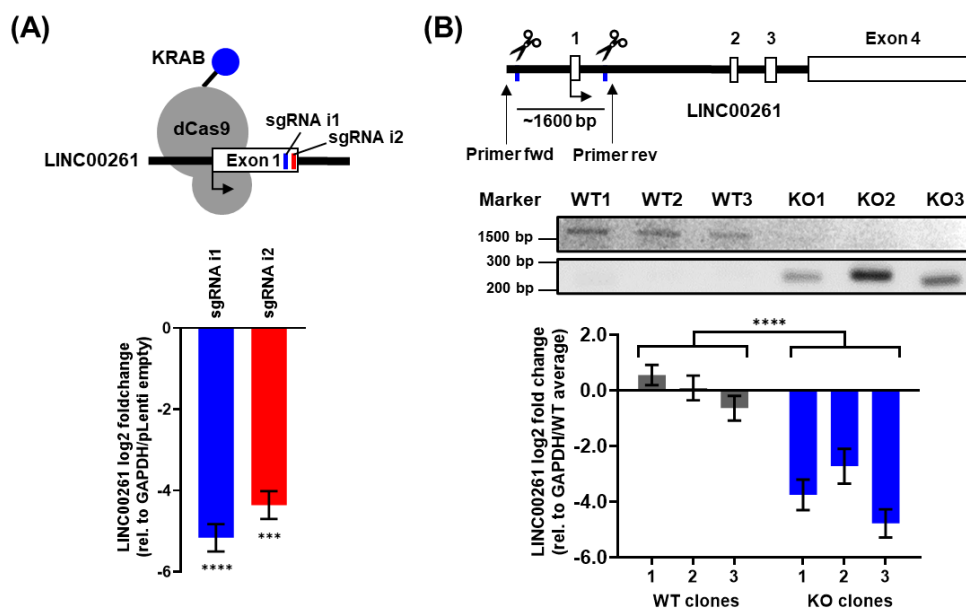


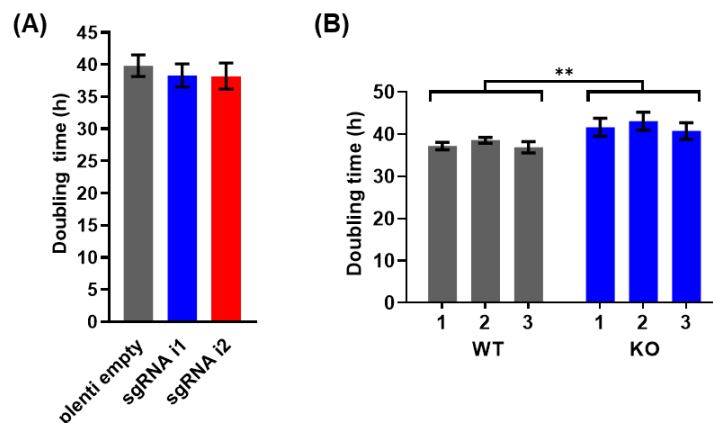
Figure 14: Downregulation of LINC00261 by using two different CRISPR approaches.

(A) Schema of CRISPRi-mediated targeting of LINC00261 and its expression levels in Panc-1 cells measured by qRT-PCR (\*\*\*)  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ , one-way ANOVA); (B) Schema of CRISPR/Cas9-mediated knockout

of the promoter region of LINC00261 using two sgRNAs. Cutting by both sgRNAs led to the removal of a genomic fragment of ~1,600 bp. PCR and gel electrophoresis using the indicated primers resulted in a ~1,600 bp product in wild-type and a ~250 bp product in knockout clones (\*\*\*\*  $p < 0.0001$ , two-way ANOVA). The figure was modified from Dorn et al.<sup>58</sup>.

Since each knockdown strategy (CRISPRi vs. CRISPR/Cas9) has its advantages and disadvantages, subsequent downstream analyses were performed using both systems in parallel.

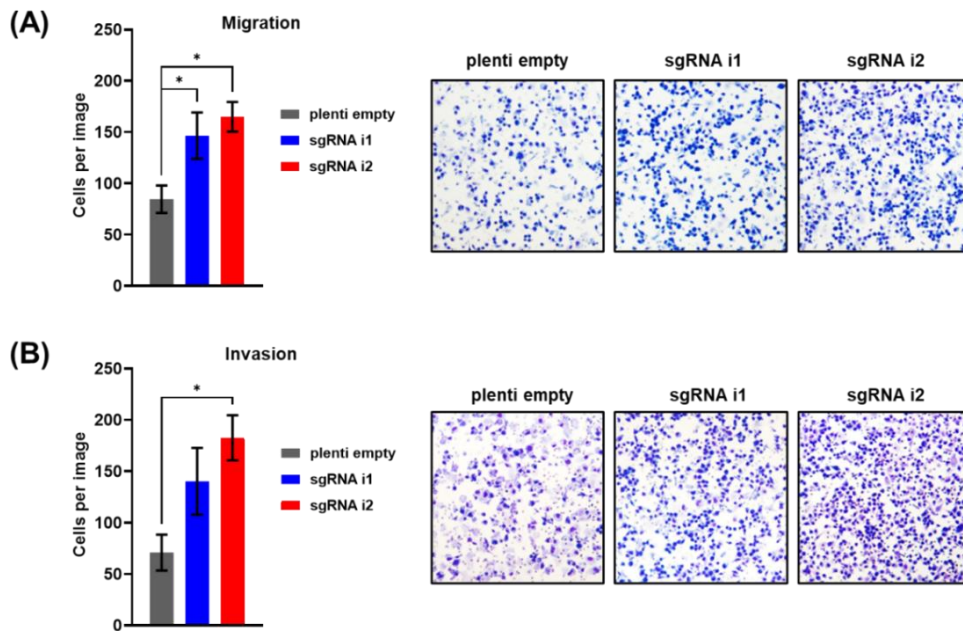
Initially, the proliferation of LINC00261-depleted cells was characterized by performing proliferation assays using the live cell analysis imaging system IncuCyte. Interestingly, the proliferative capacity of the cells remained unaltered upon knockdown of LINC00261 using the CRISPRi approach, as depicted in Figure 15A. Moreover, LINC00261 promoter knockout clones displayed only a slightly extended cell doubling time (Figure 15B).



**Figure 15: Proliferation is not substantially altered in LINC00261<sup>low</sup> cells.**

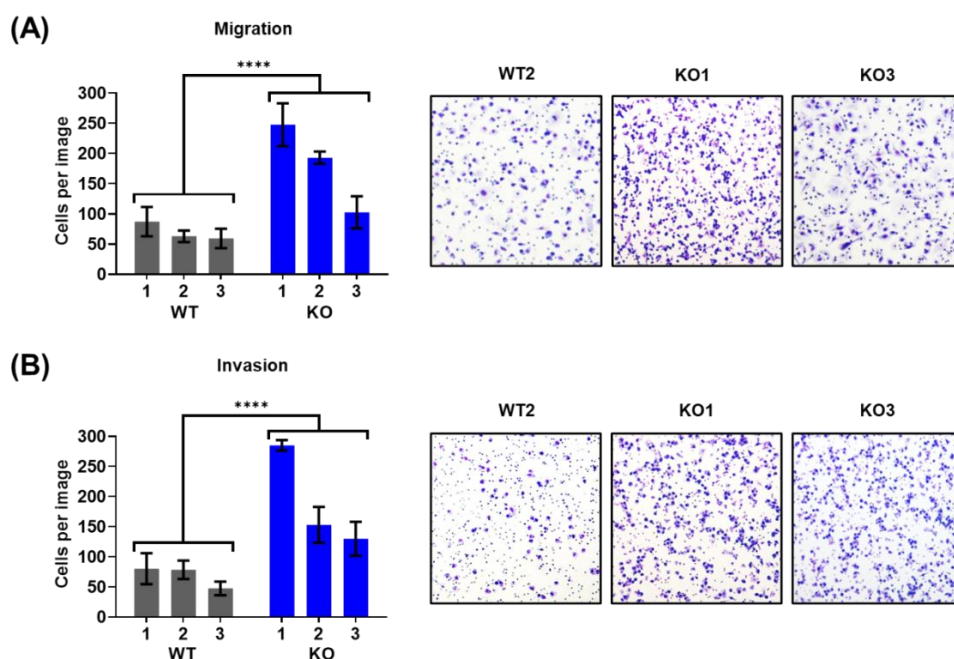
Cell doubling time in Panc-1 cells after CRISPRi-mediated LINC00261 downregulation (A, one-way ANOVA) and in three wild-type and three LINC00261 promoter knockout cell clones (B, \*\*  $p < 0.01$ , two-way ANOVA). The figure was modified from Dorn et al.<sup>58</sup>.

In contrast to the minimal and inconsistent effects observed on cell proliferation, the impact of LINC00261 expression on cell migration and invasion was notable. In detail, the CRISPRi-mediated reduction of LINC00261 resulted in a ~2-fold increase of cell migration (Figure 16A) as well as a ~2.5-fold higher invasiveness (Figure 16B).



**Figure 16: Enhanced cell migration and invasion after LINC00261 downregulation using CRISPRi.** Transwell migration (A) and invasion (B) assays with Panc-1 cells after CRISPRi-mediated LINC00261 downregulation. The quantification of migrated/invaded cells was conducted from five random fields after Eosin Y/Methylene blue staining using light microscopy (\*  $p < 0.05$ , one-way ANOVA). The figure was modified from Dorn et al.<sup>58</sup>.

These findings were confirmed using individual LINC00261 promoter knockout clones, which showed a ~2-4-fold higher cell migration rate (Figure 17A) and up to ~4-fold higher invasion capacity (Figure 17B) compared to wild-type clones.



**Figure 17: Enhanced migration and invasion after LINC00261-depletion using CRISPR/Cas9.** Transwell migration (A) and invasion (B) assays with Panc-1 cells after CRISPR/Cas9-mediated knockout of LINC00261 promoter; The quantification of migrated/invaded cells was conducted from five random fields

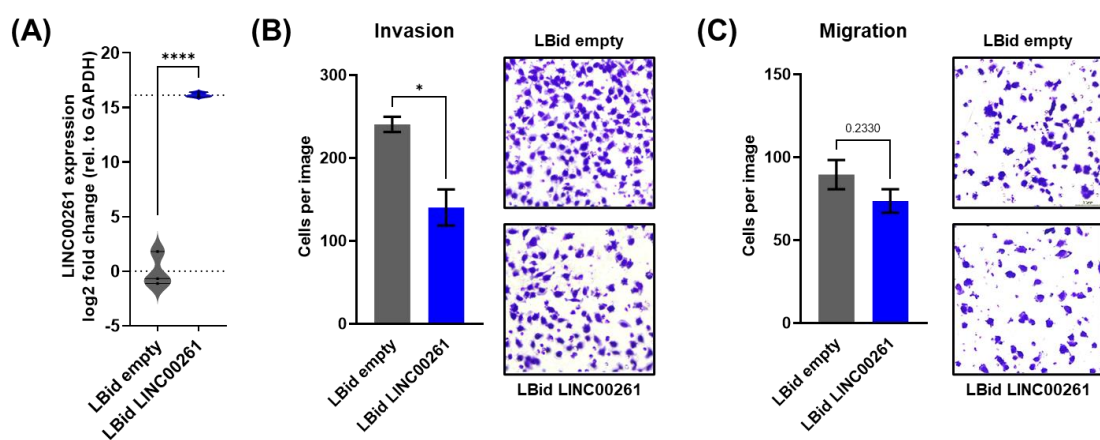


after Eosin Y/Methylene blue staining using light microscopy ( $****$   $p < 0.0001$ , two-way ANOVA). The figure was modified from Dorn et al<sup>58</sup>.

The observed effects of LINC00261 downregulation on cell migration and invasion are consistent with the herein described expression pattern of LINC00261 in PDAC samples. The decreased expression of LINC00261 in pancreatic cancer could potentially promote the invasive behavior of cancer cells and contribute to the development of a more aggressive subtype of PDAC, leading to poorer survival outcomes in patients whose tumors present a low LINC00261 expression.

#### 4.2 Stable overexpression of LINC00261

Based on the observed effects in LINC00261 KO cells, an investigation was conducted to determine if an elevated amount of LINC00261 can lower the invasion and migration of PDAC cells. To this end, LINC00261-overexpressing PATU-T cells were generated. The gene expression of LINC00261-overexpressing cells was analyzed by qRT-PCR, and a strong upregulation of LINC00261 was confirmed (Figure 18A). The PATU-T cells overexpressing LINC00261 demonstrated a significant reduction in invasion capacity (~40%, Figure 18B) and a slight decrease in migration (~15%, Figure 18C).



**Figure 18: Reduced invasion and migration of LINC00261 overexpressing PATU-T cells.**

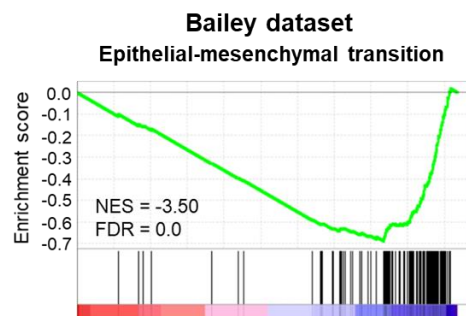
(A) LINC00261 expression was significantly upregulated by more than 15-fold in PATU-T cells expressing the lentiviral LBid LINC00261 vector ( $****$   $p < 0.0001$ , unpaired t-test); (B, C) Transwell assays were performed to evaluate the invasion (B) and migration (C) capabilities of LINC00261-overexpressing PATU-T cells. The quantification of migrated/invaded cells was conducted from five random fields after Eosin Y/Methylene blue staining using light microscopy ( $*$   $p < 0.05$ , unpaired t-test).

#### 5. Molecular function of LINC00261

The identification of signaling pathways associated with deregulated LINC00261 expression in PDAC, which may lead to the observed phenotype, was the next objective of this project.

### 5.1 Gene set enrichment analysis (GSEA) of Bailey's PDAC samples

Initially, gene set enrichment analysis (GSEA) of differentially expressed genes in PDAC samples of the Bailey dataset was performed using LINC00261 expression levels as a discriminator to define LINC00261<sup>low</sup> and LINC00261<sup>high</sup> sample groups. Here, median LINC00261 expression was used as a cut-off to assign samples to both groups. Interestingly, GSEA revealed that LINC00261 expression inversely correlates with EMT-related gene expression. (Figure 19).

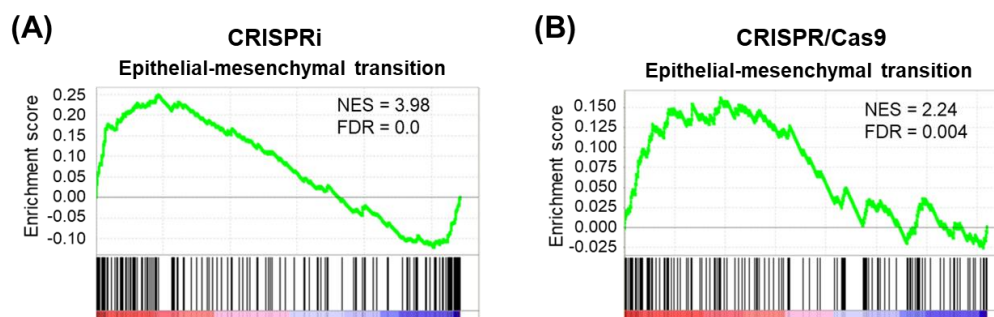


**Figure 19: LINC00261 expression inversely correlates with EMT-related gene expression.**

GSEA analysis of differentially expressed genes in PDAC patient samples revealed a significant enrichment of an EMT signature in tumors where LINC00261 expression is low ( $p < 0.0001$ ). The figure was modified from Dorn et al.<sup>58</sup>.

### 5.2 RNA-seq and GSEA of the CRISPR-based knockdown cells

To identify associated pathways in the cell systems, RNA-seq analysis was performed using both LINC00261-depleted cell systems. The complete lists of significantly deregulated genes can be found in Supplementary Table S3 (CRISPR/Cas9 system) and Supplementary Table S4 (CRISPRi system). Intriguingly, GSEA of these two datasets unraveled a significant enrichment of the hallmark gene set 'epithelial-mesenchymal transition' in both LINC00261-depleted cell systems (Figure 20A, B).



**Figure 20: RNA-Seq analysis revealed an enrichment of EMT gene set in LINC00261<sup>low</sup> Panc-1 cells.**

GSEA analysis of RNA-sequencing data revealed a significant enrichment of the EMT gene set in LINC00261-depleted cells established by both the CRISPRi knockdown (A) and the CRISPR/Cas9 promoter knockout (B) system. The figure was modified from Dorn et al.<sup>58</sup>.

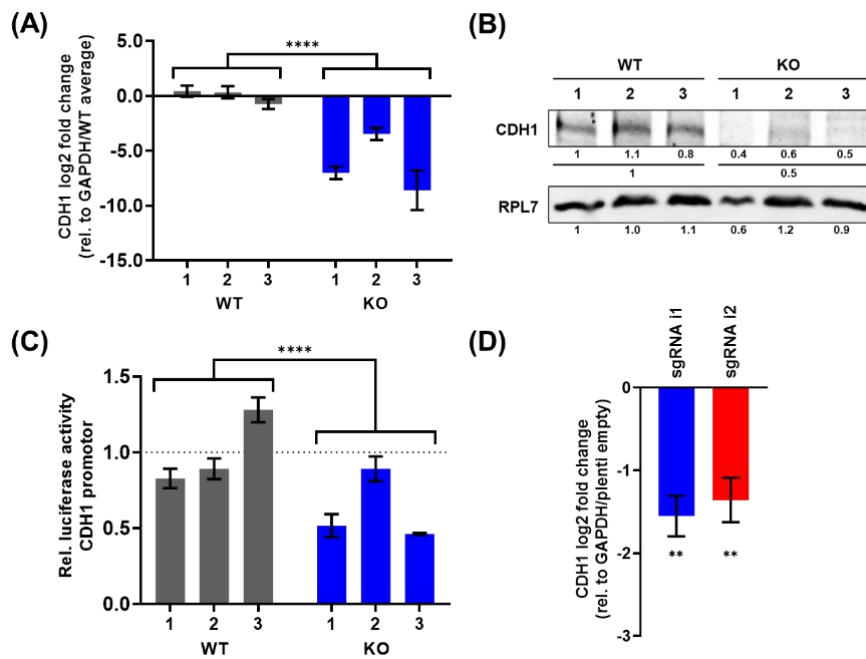
These findings align with the results derived from *in silico* analysis of PDAC patient samples (Figure 19), indicating that the downregulation of LINC00261 has functional significance. In fact, the decrease of LINC00261 may actively contribute to the establishment of an EMT signature, a characteristic feature of the squamous and basal-like subtype of PDAC.

### 5.3 Analysis of deregulated genes in CRISPR-based knockdown cells

The unbiased analysis of gene expression in the two distinct LINC00261 knockdown systems revealed a role of LINC00261 in regulating the transcriptional landscape related to EMT. However, a more detailed analysis of the RNA-seq data was required to comprehend the underlying molecular mechanisms. Therefore, an in-depth analysis of the RNA-seq data was conducted to identify the genes that could be responsible for the observed EMT signature in the LINC00261 knockdown systems.

#### 5.3.1 Analysis of genes related to EMT

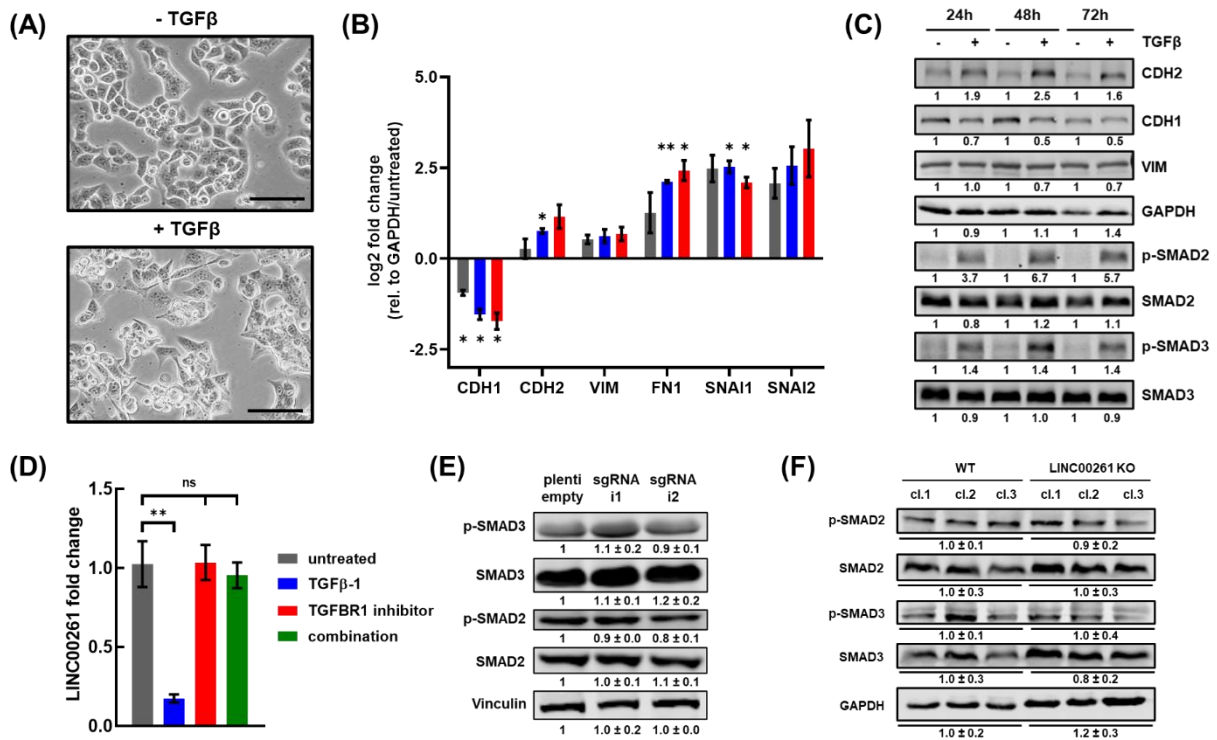
Interestingly, although enrichment of EMT hallmark genes and increased migration and invasion were observed, no significant upregulation of the typical EMT transcription factors SNAI1, SNAI2, ZEB1, ZEB2, and TWIST1 was detected. It should be noted that TWIST1 and SNAI2 were not substantially expressed in Panc-1 cells (data not shown). However, intersection of upregulated and downregulated target genes in both cell systems highlighted genes important for EMT and cytoskeletal organization, such as CDH1, formin 1 (FMN1) and myosin light chain kinase (MYLK), to be regulated by LINC00261. Specifically, CDH1 expression was robustly decreased both at the RNA (Figure 21A) and at the protein level (Figure 21B) in both CRISPR systems. To test the idea of a potential transcriptional regulation of CDH1 by LINC00261, luciferase reporter assays were performed using a CDH1 promoter construct that was cloned in front of the luciferase gene. The CDH1 promoter construct was transfected into Panc-1 wild-type or LINC00261-depleted cells, and luciferase activity was measured 48 h later. Intriguingly, the reduced expression level of LINC00261 in KO clones resulted in a significantly lower CDH1 promoter activity as compared to WT clones (Figure 21C). CDH1 was also significantly downregulated in LINC00261-depleted cells established by the CRISPRi system (Figure 21D). The mesenchymal counterpart of CDH1, CDH2 (N-cadherin), was significantly upregulated in LINC00261<sup>low</sup> cells established by the CRISPRi system, but not in LINC00261 promoter KO cells (Supplementary Figure S1).



**Figure 21: LINC00261 regulated E-cadherin expression.**

(A, B) Expression of E-cadherin mRNA (A) and protein levels (B) in three wild-type and three promoter knockout clones (\*\*\*\*  $p < 0.0001$ , two-way ANOVA). Quantification of protein expression was carried out by using the Image Studio™ Acquisition Software. RPL7 served as loading control; (C) Relative luciferase activity of the CDH1 gene promoter constructs normalized to the pGL3-minCMV vector. The average of the rel. luciferase activity in wild-type clones was set to 1.0 (\*\*\*\*  $p < 0.0001$ , two-way ANOVA); (D) Differences in E-cadherin mRNA expression in LINC00261<sup>low</sup> cells established by the CRISPRi system (\*\*  $p < 0.01$ , one-way ANOVA). The figure was modified from Dorn et al.<sup>58</sup>.

These findings support the idea of LINC00261 being involved in maintaining a pro-epithelial cell identity whereas loss of LINC00261 induces transcriptional and morphological changes potentially via regulating CDH1 expression. A well-characterized EMT pathway that regulates CDH1 expression is the TGF $\beta$  signaling pathway<sup>86</sup>. To analyze this connection, the responsiveness of Panc-1 cells to TGF $\beta$  was initially determined. The stimulation of Panc-1 cells for 24 h, 48 h and 72 h caused EMT-like morphological changes (Figure 22A) and expression changes of EMT-associated genes, both on RNA and protein level (Figure 22B and C). Furthermore, the phosphorylation of SMAD2 and SMAD3 proteins as mediators of the canonical TGF $\beta$  signaling inside the cell were induced (Figure 22C). Intriguingly, Panc-1 cells significantly downregulated LINC00261 expression upon exposure to TGF $\beta$  for 24 h, 48 h and 72 h via the TGFBR1 receptor (Figure 22D), suggesting that the activation of the TGF $\beta$  signaling may be responsible for the mesenchymal features of the LINC00261<sup>low</sup> cancer cells. However, the analysis of the phosphorylation of the proteins SMAD2 and SMAD3 in LINC00261-depleted cells did not show an elevated activation of the canonical SMAD-TGF $\beta$  signaling (Figure 22E and F). These results demonstrated that LINC00261 is regulated by TGF $\beta$ , but they refute the idea that the classical SMAD-TGF $\beta$  pathway is responsible for the observed EMT phenotype in LINC00261<sup>low</sup> cells.

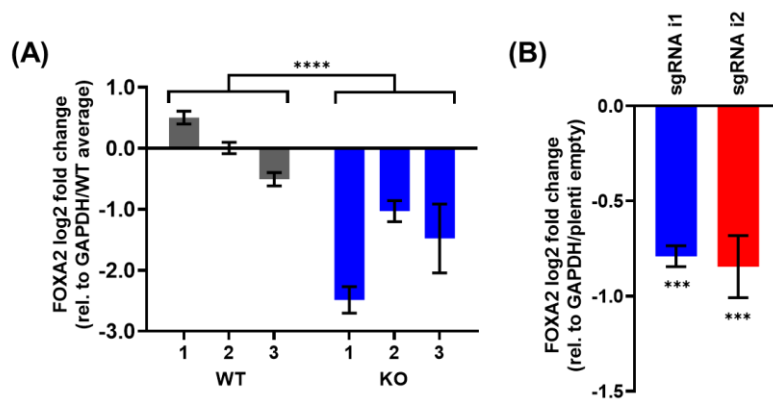


**Figure 22: Canonical TGF $\beta$  signaling is not responsible for CDH1 downregulation in LINC00261<sup>low</sup> cells.**

(A) Brightfield images of untreated and TGF $\beta$  treated Panc-1 cells (scale bar = 100  $\mu$ m); (B, C) Analysis of mRNA (B) and protein levels (C) of genes associated with EMT by qRT-PCR and western Blot, respectively (\*  $p < 0.05$ , \*\*  $p < 0.01$ , two-way ANOVA, Šídák's multiple comparison test); (D) LINC00261 regulation in Panc-1 cells treated with TGF $\beta$ , TGFBR1 inhibitor (RepSox) or both after 72 h measured by qRT-PCR (\*\*  $p < 0.01$ , unpaired t-test); (E, F) Analysis of phosphorylation of SMAD2 and SMAD3 in LINC00261<sup>low</sup> cells established by the CRISPRi (E) or CRISPR/Cas9 system (F). Vinculin (E) or GAPDH (F) were used as loading controls. Quantification of protein expression was carried out by using the Image Studio™ Acquisition Software ( $\pm$  standard deviation). Parts of the figure were published by Dorn et al.<sup>58</sup>

### 5.3.2 Analysis of FOXA2 gene expression

In addition to the EMT transcription factors and negative regulators of CDH1 gene expression mentioned earlier (SNAI1, SNAI2, TWIST, and ZEB1<sup>129–133</sup>), several transcription factors such as p300, hepatocyte nuclear factor 1 homeobox A (HNF1 $\alpha$ ), and FOXA2 have been shown to positively regulate CDH1 expression<sup>134</sup>. Intriguingly, upon analysis of the RNA-seq data, downregulation of the LINC00261-adjacent transcription factor FOXA2 was observed. The qRT-PCR analysis confirmed a log<sub>2</sub> fold change of 1–2.5 in LINC00261-depleted cells established with the CRISPR/Cas9 system (Figure 23A) and a log<sub>2</sub> fold change of around 0.75 in cells established with the CRISPRi system (Figure 23B). It is noteworthy that FOXA2, like LINC00261, was downregulated following TGF $\beta$  treatment in Panc-1 cells (Supplementary Figure S2).



**Figure 23: FOXA2 expression is significantly downregulated in LINC00261<sup>low</sup> cells.**

(A) Expression of FOXA2 mRNA in three wild-type and three promoter knockout clones (\*\*\*\*  $p < 0.0001$ , two-way ANOVA); (B) Expression of FOXA2 mRNA in LINC00261<sup>low</sup> cells established by the CRISPRi system (\*\* $p < 0.001$ , one-way ANOVA).

Previously reported findings indicate that CDH1 is regulated by FOXA2<sup>69,71,134</sup>. Thus, CDH1 expression could be indirectly influenced by LINC00261 through its impact on FOXA2. Liu et al.<sup>134</sup> and Bow et al.<sup>71</sup> demonstrated that FOXA2 can directly bind to the CDH1 promoter, indicating that regulation of CDH1 can occur without the involvement of EMT TF factors. Consequently, the next experiments aimed to investigate the regulation of FOXA2 by LINC00261 and the contribution of FOXA2 to the observed changes in PDAC cells.

## 6. Analysis of the LINC00261-FOXA2 regulatory circuit

First, the regulatory network between LINC00261 and FOXA2 was further investigated.

### 6.1 Correlation of LINC00261 and FOXA2 gene expression

Importantly, both the TCGA (Figure 24A,  $R^2=0.65$ ) and the Bailey dataset (Figure 24B,  $R^2=0.69$ ) confirmed that the transcription factor FOXA2 was positively correlated with LINC00261 expression. In line with its strong correlation to LINC00261, FOXA2 showed a similar expression pattern as LINC00261 across various PDAC subtypes in the Bailey dataset (Supplementary Figure S3). Furthermore, an analysis of LINC00261 and FOXA2 expression in multiple PDAC cell lines demonstrated a positive correlation between these two genes (Figure 24C,  $R^2=0.83$ ).

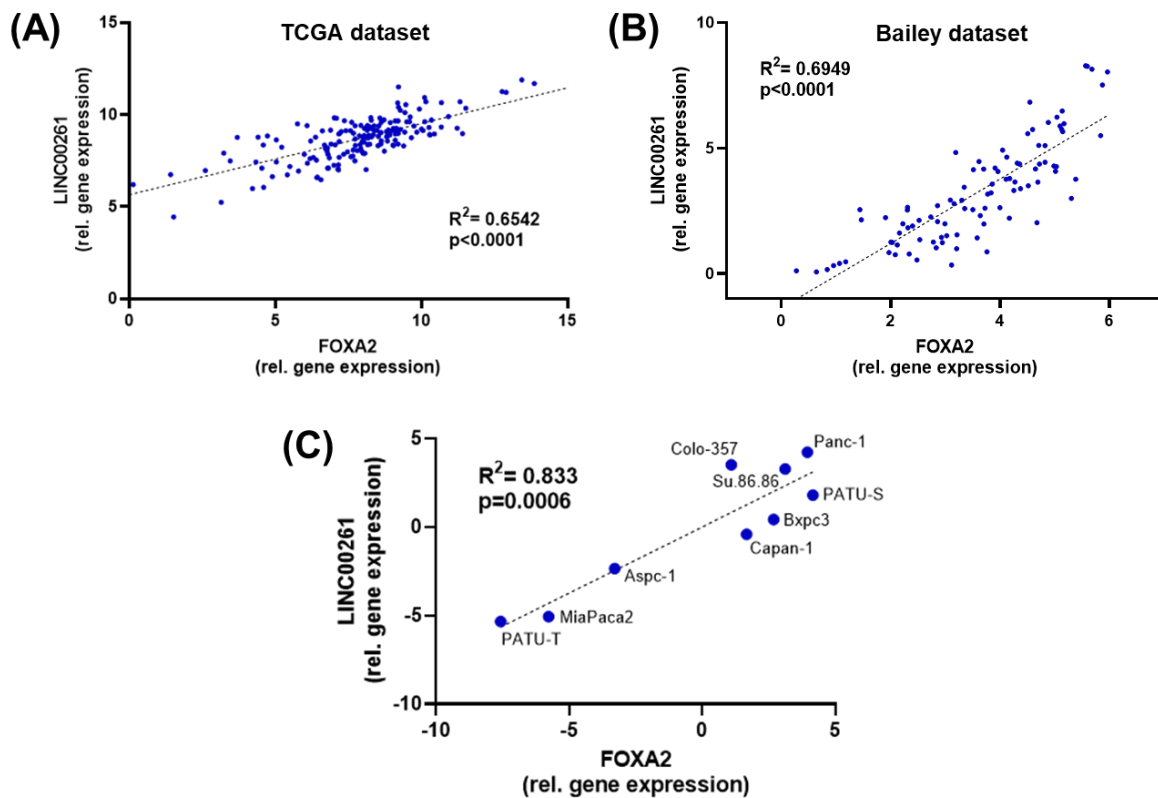
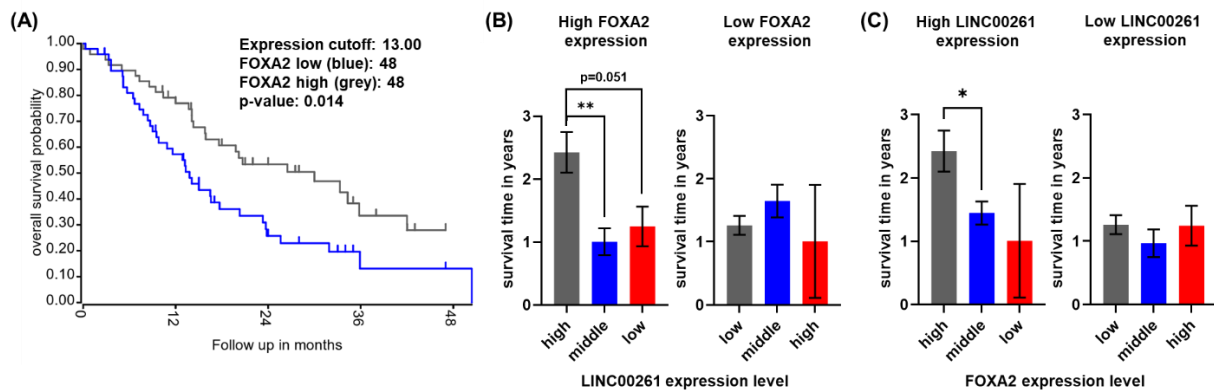


Figure 24: LINC00261 and FOXA2 are positively correlated in PDAC samples and cell lines.

(A, B) LINC00261 and FOXA2 gene expressions are positively correlated in both the TCGA dataset (A,  $R^2 = 0.65$ ) and Bailey dataset (B,  $R^2 = 0.69$ ); (C) LINC00261 and FOXA2 gene expressions are positively correlated in several PDAC cell lines ( $R^2 = 0.83$ ).

## 6.2 Analysis of the patient survival time considering FOXA2 expression

Based on the observed co-regulation of LINC00261 and FOXA2, the impact of FOXA2 expression on patient survival was analyzed (Figure 25A). Similar to LINC00261, high FOXA2 expression was associated with significantly better overall survival. In addition, the survival of PDAC patients was investigated considering both the expression of FOXA2 and LINC00261. To this end, all TCGA samples ( $n = 177$ ) with available survival data were categorized into nine groups based on low, middle, and high LINC00261 or FOXA2 expression, and the average survival time was calculated. Although correlation analysis confirmed the co-regulation of FOXA2 and LINC00261 (Figure 24A), some samples showed substantially different expression levels of LINC00261 and FOXA2. Interestingly, the survival benefit was observed only in patients who exhibited high expression levels of both FOXA2 and LINC00261. In other words, a decrease in the expression of either FOXA2 or LINC00261 was found to reduce the survival time of PDAC patients. The results indicate that LINC00261 may modulate the progression of PDAC through pathways that are not solely dependent on the regulation of FOXA2.



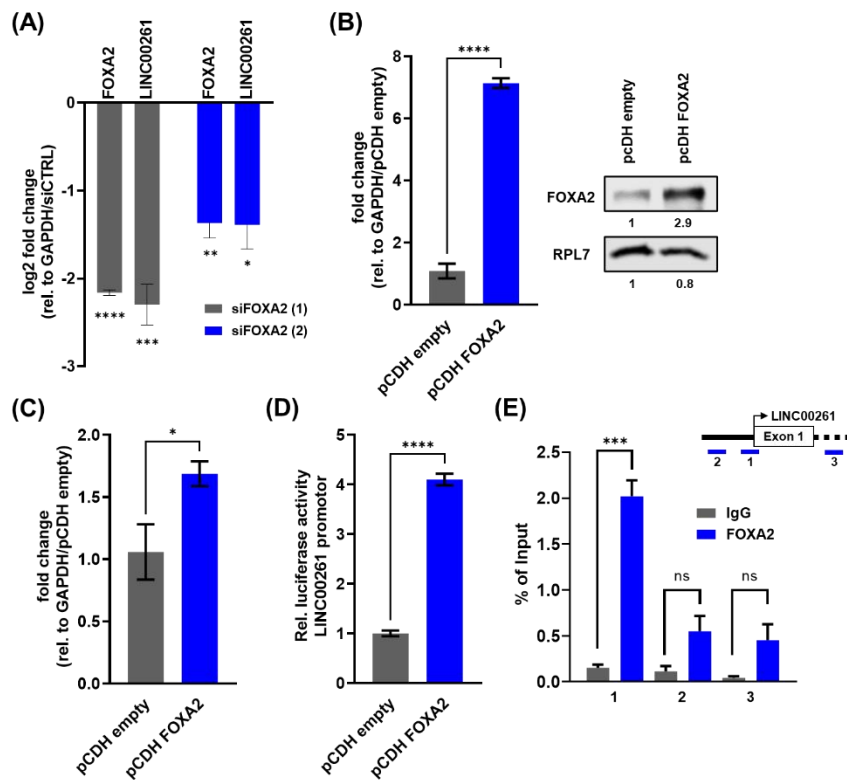
**Figure 25: Impact of FOXA2 and LINC00261 expression on the survival of PDAC patients.**

(A) Survival analysis of PDAC patients with low FOXA2 (blue line, n=48) versus high FOXA2 (grey line, n=48) expression (Bailey dataset, <http://r2.amc.nl>, Log rank test); (B) Survival of patients with either high or low FOXA2 expression, along with corresponding high/middle/low LINC00261 expression; (C) Survival of patients with high or low LINC00261 expression and corresponding high/middle/low FOXA2 expression (\*  $p < 0.05$ , \*\*  $p < 0.01$ , Dunnett's T3 multiple comparisons test). The sample sizes of the groups are provided in Supplementary Table S2.

### 6.3 Molecular regulation of LINC00261 expression by FOXA2

To investigate the potential regulatory role of FOXA2 on LINC00261 expression in pancreatic cancer cells, FOXA2 levels were manipulated in Panc-1 cells. The knockdown of FOXA2 using two different siRNAs resulted in a significant reduction of LINC00261 transcript levels (Figure 26A). Furthermore, to explore a direct transcriptional regulation of LINC00261 by FOXA2, Panc-1 cells were stably overexpressed with FOXA2 (Figure 26B), and LINC00261 expression was examined (Figure 26C). The results proved that LINC00261 expression was significantly upregulated in FOXA2 overexpressing cells. Furthermore, luciferase assays were conducted to determine LINC00261 promoter activity. The overexpression of FOXA2 in Panc-1 cells was found to result in a significant increase in LINC00261 promoter activity (Figure 26D). Finally, the physical interaction of FOXA2 with the LINC00261 promoter was examined using a ChIP-qPCR experiment. The ChIP-qPCR analysis revealed an association of FOXA2 with the promoter region of LINC00261, while no binding of FOXA2 to the upstream proximal or downstream intragenic regions of LINC00261 was detected (Figure 26E). Altogether, these findings imply a direct regulation and a close interconnection between LINC00261 and its genomic neighbor FOXA2.





**Figure 26: LINC00261 regulation by its genomic neighbor FOXA2.**

(A) siRNA-mediated knockdown of FOXA2 significantly downregulated FOXA2 and LINC00261 RNA levels (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ , one-way ANOVA); (B) Stable overexpression of FOXA2 in Panc-1 was achieved on RNA (left panel; GAPDH was used as the reference gene, \*\*\*\*  $p < 0.0001$ , unpaired t-test) and protein level (right panel; RPL7 protein was used as a loading control); (C) LINC00261 expression levels in control or FOXA2 overexpressing Panc-1 cells (\*  $p < 0.05$ , unpaired t-test); (D) Luciferase activity of a LINC00261 promoter reporter after stable FOXA2 or empty control vector overexpression in Panc-1 cells (\*\*\*\*  $p < 0.0001$ , unpaired t-test); (E) ChIP followed by qPCR analysis using primers located upstream (1, 2) and downstream (3) of the LINC00261 transcriptional start site (upper panel) confirmed binding of FOXA2 to the LINC00261 promoter region (lower panel, \*\*\*  $p < 0.001$ , unpaired t-test). The figure was modified from Dorn et al.<sup>58</sup>.

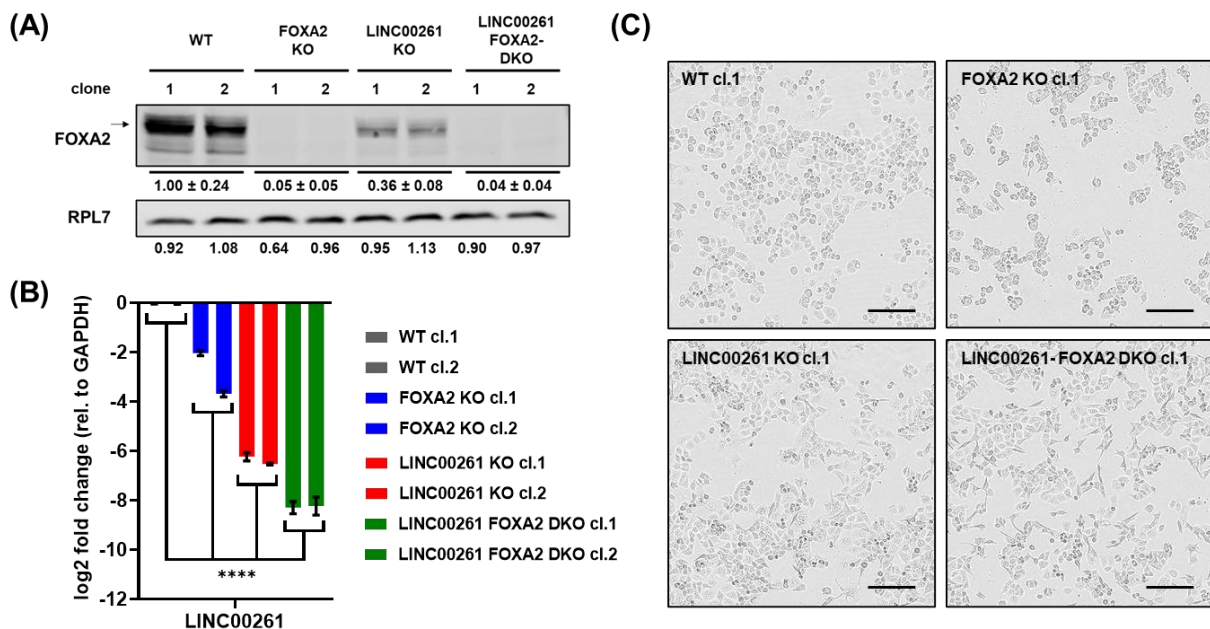
## 7. Cellular analysis of the interdependencies between LINC00261 and FOXA2

Previous experiments have confirmed a co-regulation of FOXA2 and LINC00261, suggesting a potential contribution of FOXA2 to the effects observed in LINC00261-depleted PDAC cells. However, survival analysis of PDAC patients has indicated that LINC00261 effects may occur independently of FOXA2. Hence, the objective of the subsequent chapter was to gain further insights into the regulatory mechanisms of these two genes and to characterize individual target genes of LINC00261 and FOXA2.

### 7.1 CRISPR-based knockout of FOXA2 in WT and LINC00261<sup>low</sup> cells

To discover individual target genes, FOXA2 KO cells have been established using CRISPR/Cas9. Specifically, FOXA2 was deleted in LINC00261-proficient WT Panc-1 cells

and in LINC00261 KO cells resulting in the generation of FOXA2 KO and FOXA2-LINC00261 double KO clones. The successful knockout of FOXA2 was confirmed by western blot analysis (Figure 27A). Additionally, disruption of the FOXA2 gene on both alleles was verified by PCR of genomic DNA following TOPO™ TA™ cloning and Sanger sequencing (data not shown). Expression analysis of established cell lines by western blot (Figure 27A) and qPCR (Figure 27B) demonstrated a reduction of FOXA2 protein expression in LINC00261 KO cells, as well as a downregulation of LINC00261 expression in FOXA2 KO cells, thereby providing evidence for the co-regulation of both genes. Remarkably, expression of LINC00261 in LINC00261-FOXA2 double KO cells was significantly lower when compared to single LINC00261 KO cells (Figure 27B). Brightfield images illustrated a more spindle-like morphology of LINC00261 KO and LINC00261-FOXA2 double KO cells in comparison to WT cells. In contrast, FOXA2 KO cells exhibited a rounded shape, formed clusters, and showed weak attachment to the plate (Figure 27C).



**Figure 27: Establishment of FOXA2 KO cells and LINC00261-FOXA2 double KO cells.**

(A) Western blot showing the depletion of FOXA2 protein in FOXA2 KO and LINC00261-FOXA2 double KO cells and the downregulation of FOXA2 expression in LINC00261 KO cells. RPL7 served as loading control. Quantification of protein expression was carried out by using the Image Studio™ Acquisition Software ( $\pm$  SD); (B) LINC00261 expression is downregulated in FOXA2 KO cells as compared to WT cells and is further reduced in LINC00261-FOXA2 double KO cells in comparison to sole LINC00261 KO cells (\*\*\*\*  $p < 0.0001$ , two-way ANOVA); (C) Brightfield images of established cell lines (4x objective, scale bar: 200  $\mu$ m).

The phenotype of the newly established cell lines was initially characterized *in vitro*. To assess the proliferation of the cells, proliferation assays were conducted using the live cell analysis imaging system IncuCyte. The proliferative capacity of the LINC00261<sup>low</sup> cells was not affected by the knockout of FOXA2 (LINC00261-FOXA2 DKO cells). However, a

significant reduction in proliferative capacity was observed, depicted as a higher doubling time (h), when FOXA2 was knocked out in the WT Panc-1 cells (Figure 28).

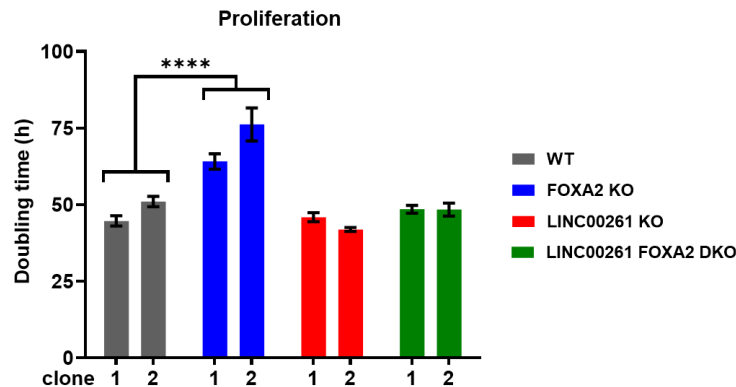


Figure 28: Proliferation of WT, FOXA2 KO, LINC00261 KO and LINC00261-FOXA2 DKO Panc-1 cells. Cell doubling times of the newly established cell lines (\*\*\*\*  $p < 0.0001$ , two-way ANOVA).

Next, clonogenic assays were performed using the same cell lines to assess the ability of single cells to survive, self-renew and form colonies. Results showed that FOXA2 KO cells formed fewer, but larger and more well-defined colonies compared to WT cells. On the other hand, LINC00261 KO and DKO cells formed extensive colonies with a larger area but less defined compared to WT and FOXA2 KO cells (Figure 29A). Quantitative analysis revealed that FOXA2 KO cell colonies occupied ~20% less area of the well compared to WT cells. Moreover, the LINC00261-FOXA2 double KO colonies occupied a slightly larger area of the well than LINC00261 KO cells. While LINC00261 KO cells displayed a more dispersed growth pattern compared to WT cells, quantitative analysis did not reveal a significant difference in percentual colony area (Figure 29B).

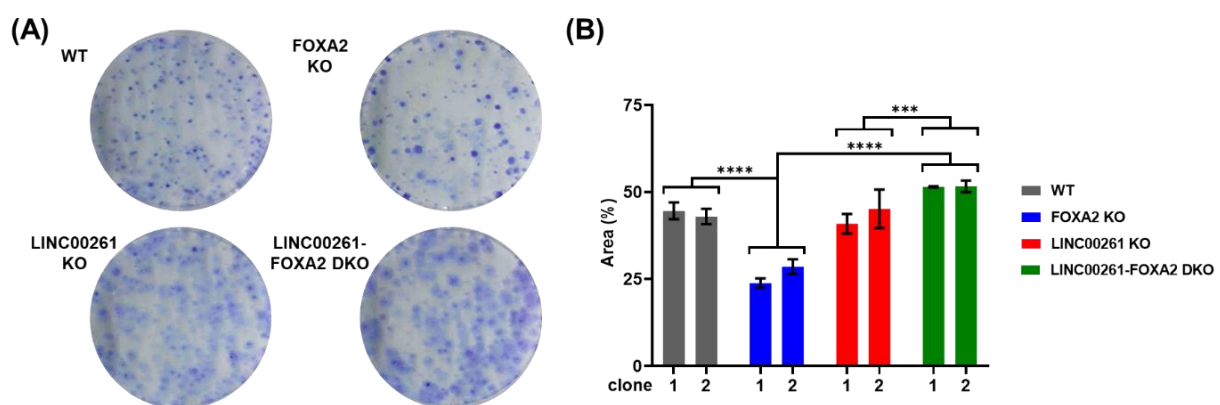
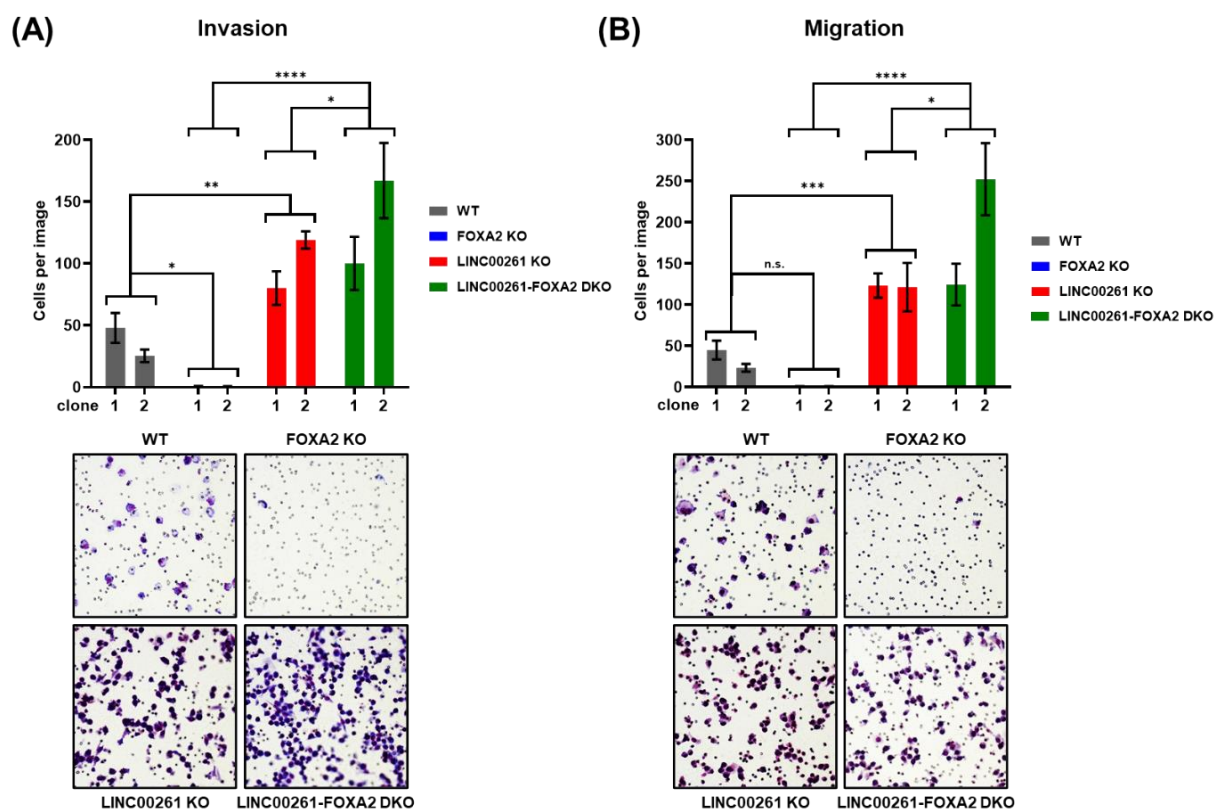


Figure 29: Clonogenic assay with WT, FOXA2 KO, LINC00261 KO and double KO Panc-1 cells.

(A) Representative images of clonogenicity studies with WT, FOXA2 KO, LINC00261 KO and double KO cells; (B) Quantification of the clonogenic growth of the established cell lines by using the ImageJ software (\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ , two-way ANOVA).

In the transwell assay, it was observed that the invasive (Figure 30A) and migratory (Figure 30B) capacity of FOXA2 KO cells were opposite to that of LINC00261 KO or double KO cells. Notably, the LINC00261-FOXA2 double KO cells displayed a significant increase in invasion and migration compared to the single LINC00261 KO cells, potentially due to the more pronounced downregulation of LINC00261 expression in the former (Figure 27B). These findings suggest that the enhanced cellular motility observed in the LINC00261 promoter KO cells was not attributed to the downregulation of FOXA2, but rather due to a LINC00261-specific effect that operates independently of FOXA2.



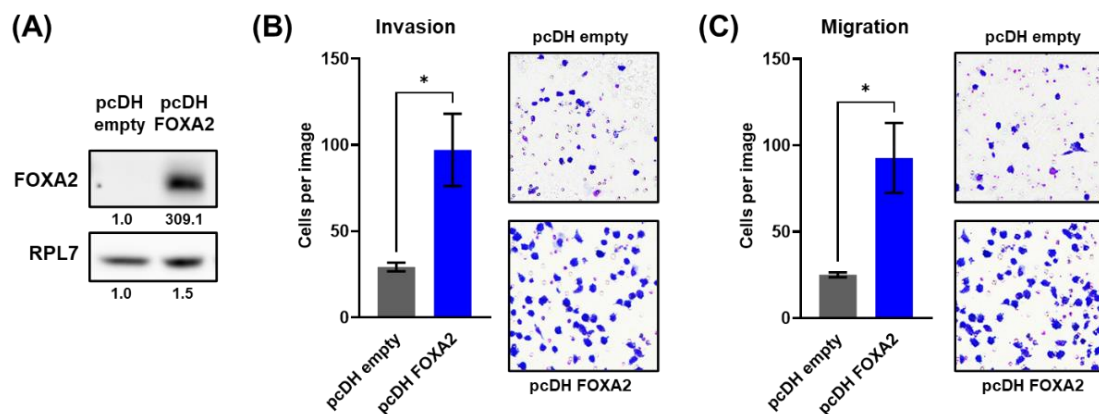
**Figure 30: Cell invasion and migration of newly established Panc-1 cell lines.**

Transwell assays demonstrate that both cell invasion (A) and migration (B) were decreased in FOXA2 KO cells but increased in LINC00261 KO and double KO cells compared to WT cells. The quantification of migrated/invaded cells was conducted from five random fields after Eosin Y/Methylene blue staining using light microscopy (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ , two-way ANOVA).

## 7.2 Stable overexpression of FOXA2

To investigate the consequence of FOXA2 overexpression in PDAC cells, PATU-T cells were stably transfected with a FOXA2-expressing vector, which led to a substantial increase in FOXA2 protein expression (Figure 31A). The overexpression of FOXA2 was found to significantly enhance both cell invasion (Figure 31B) and migration (Figure 31C), exhibiting a contrasting effect compared to LINC00261 overexpressing cells. Critically, a reduction in the migratory and invasive capacities of PATU-T cells was observed following

transfection with the empty pcDH puro vector in comparison to cells transfected with the previously utilized empty LBid vector (Figure 18). Consequently, effects of the vectors cannot be excluded in these cell assays, necessitating the inclusion of untransfected parental cells as an additional control. However, the alterations relative to the respective empty vectors can be solely attributed to the overexpressed RNA/protein. Moreover, the overexpression of FOXA2 in Panc-1 WT cells resulted in a similarly substantial increase in migration and invasion, supporting a consistent impact of FOXA2 in both PDAC cell lines (Supplementary Figure S4).



**Figure 31: Overexpression of FOXA2 in PATU-T cells enhances cell invasion and migration.**

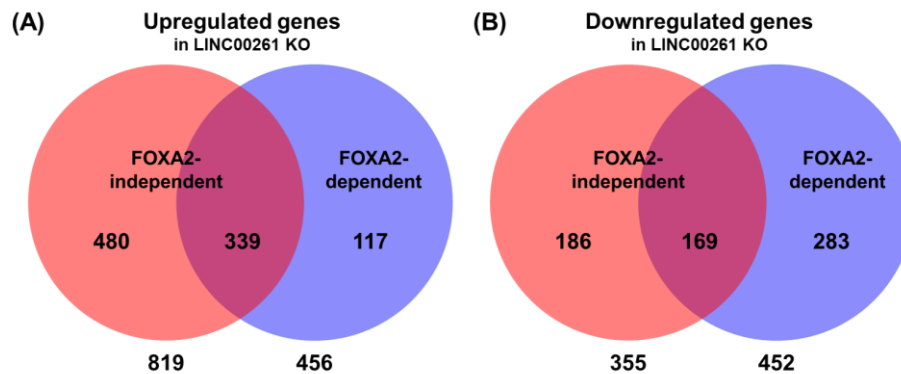
(A) Western blot displaying clearly increased FOXA2 protein expression in FOXA2 overexpressing PATU-T cells. RPL7 served as loading control; (B, C) Transwell invasion (B) and migration (C) assays in FOXA2-overexpressing PATU-T cells. Quantification of migrated/invaded cells from five random fields after Eosin Y/Methylene blue staining using light microscopy (\*  $p < 0.05$ , unpaired t-test).

## 8. Molecular analysis of the interdependencies between LINC00261 and FOXA2

### 8.1 FOXA2-dependent and -independent regulation of genes by LINC00261

In order to characterize mutual or individual LINC00261 and FOXA2 target genes, RNA-seq analysis was performed, and gene expression was compared against respective single KO cells to identify FOXA2-independent LINC00261 as well as LINC00261-independent FOXA2 target genes. The initial focus was on identifying LINC00261 target genes that were regulated dependently or independently of FOXA2. To identify FOXA2-independently regulated genes, gene expression data of FOXA2 KO cells was compared to that of LINC00261-FOXA2 double KO cells (Figure 32). The analysis revealed that 819 genes were strongly upregulated ( $\log_2 > 1$ ,  $FDR < 0.05$ ,  $FPKM \geq 1$ , Figure 32A), and 355 were strongly downregulated ( $\log_2 < -1$ ,  $FDR < 0.05$ ,  $FPKM \geq 1$ , Figure 32B). Genes dependent on FOXA2 expression were identified by determining those no longer regulated by LINC00261 in cells lacking FOXA2. This involved overlapping the deregulated genes identified in the comparison between WT and LINC00261 KO with these between FOXA2 KO and double KO. In this category, 117 genes were upregulated (Figure 32A), and 283 downregulated (Figure

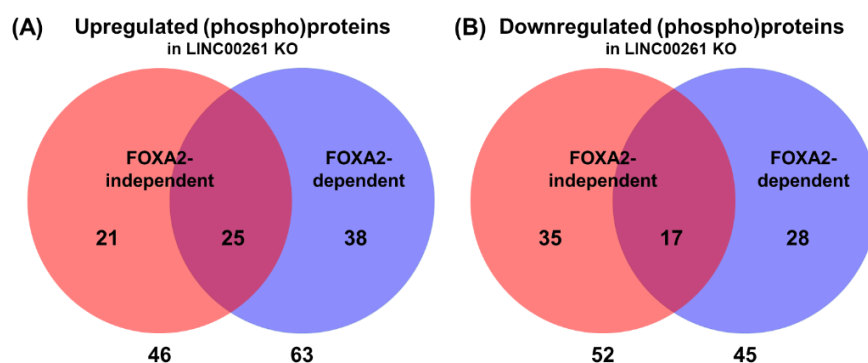
32B). The complete lists of genes can be found in Supplementary Table S5 (upregulated genes) and Supplementary Table S6 (downregulated genes), respectively.



**Figure 32: RNA-sequencing reveals FOXA2-dependent and -independent LINC00261 target genes.**

(A) 819 genes were FOXA2-independently upregulated, and 117 genes were FOXA2-dependently upregulated after LINC00261-depletion; (B) 355 genes were FOXA2-independently downregulated, and 283 genes were FOXA2-dependently downregulated after LINC00261 depletion.

Furthermore, the expression of several hundred proteins and phosphoproteins was analyzed using a high-throughput protein microarray called reverse phase protein assay (RPPA). The same approach as in RNA-seq data analysis was utilized to identify proteins regulated both dependently and independently of FOXA2. The Venn diagrams in Figure 33 illustrate the amount of proteins, including phosphoproteins, which were FOXA2-independently or -dependently up- (Figure 33A) or downregulated (Figure 33B) by LINC00261 depletion. The complete lists of significantly deregulated proteins and phosphoproteins are presented in Supplementary Table S7 (upregulated proteins) and Supplementary Table S8 (downregulated proteins).



**Figure 33: Determination of FOXA2-independently and -dependently regulated (phospho)proteins by RPPA.**

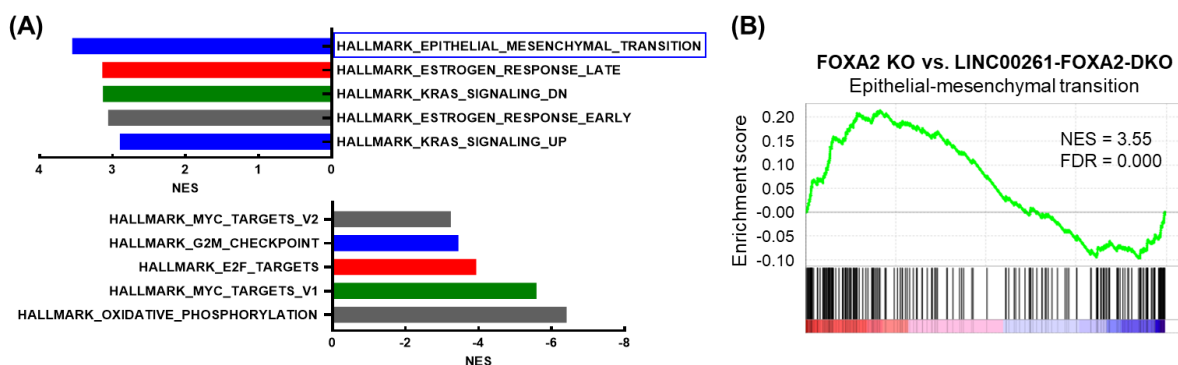
The expression of 499 unique proteins, including phosphoproteins, was evaluated. (A) LINC00261 depletion led to the upregulation of 46 (phospho)proteins independently of FOXA2 and 38 (phospho)proteins in a FOXA2-dependent manner; (B) LINC00261 depletion resulted in the downregulation of 52 (phospho)proteins independently of FOXA2 and 28 (phospho)proteins in a FOXA2-dependent manner. Proteins were considered significantly up- or downregulated if  $p < 0.05$ .

### 8.1.1 FOXA2-independently regulated genes

The focus of interest was initially on the effects of LINC00261 that occurred independently of its adjacent transcription factor FOXA2.

#### 8.1.1.1 Analysis of FOXA2-independently regulated genes and proteins

Therefore, the genes that exhibited differential expression in LINC00261-FOXA2 double KO cells compared to single FOXA2 KO cells were further analyzed. Interestingly, gene set enrichment analysis of these genes revealed a significant enrichment of the epithelial-mesenchymal transition hallmark gene set in double KO cells (Figure 34). This outcome indicated that the observed phenotype of increased cell migration/invasion and EMT signature in LINC00261 KO and LINC00261-FOXA2 double KO cells was attributed to the downregulation of LINC00261 rather than FOXA2.



**Figure 34: GSEA analysis of FOXA2-independently regulated LINC00261 target genes.**

(A) Overview of the five most positively and negatively regulated GSEA hallmark gene sets; (B) GSEA analysis of RNA-sequencing data of FOXA2-independently regulated genes revealed a positive enrichment of the EMT hallmark gene set ( $p < 0.0001$ ).

To investigate the signaling cascades that lead to the LINC00261-associated transcriptomic program with EMT signature, an unbiased enrichment analysis of the FOXA2-independently deregulated (phospho)proteins identified through RPPA (Figure 33) was conducted using the STRING database<sup>135</sup>. It should be noted that care should be taken when interpreting the enrichment analysis due to the biased selection of antibodies in the RPPA assay and the resulting unbalanced classification of the analyzed proteins into ontology categories. However, the pathway assignment of proteins helped in interpreting the datasets. Interestingly, enrichment analysis using the Reactome pathways revealed that 45 out of the 94 deregulated (phospho)proteins played a role in signal transduction (Reactome pathway HSA-162582, FDR= 5.56e-13), with 17 proteins assigned to the PI3K-Akt signaling pathway (WikiPathway WP4172, FDR= 3.63e-16) and 13 to the focal adhesions (WikiPathway WP306, FDR= 7.25e-14, shaded in Supplementary Table S7 and Supplementary Table S8, Figure 35). Additional significantly enriched signaling pathways

of interest were the EGF/EGFR signaling pathway (10 proteins, WikiPathway WP437, FDR= 1.09e-10) and the RAC1/PAK1/p38/MMP2 pathway (10 proteins, WikiPathway WP3303, FDR= 1.13e-13). To validate the results of the RPPA analysis, western blot analyses were performed on key players of these pathways (e.g., PI3K, Akt, EGFR, ERK, p38, c-Src, PXN). In several cases, such as the phosphorylation of ERK, p38, and c-Src, the observations from the RPPA analysis could not be confirmed in the western blot analyses. However, disturbances were observed in the focal adhesion signaling pathway, prompting further investigation.

Reactome Pathway	Count in Network	FDR		WikiPathway	Count in Network	FDR
Signal Transduction	45 of 2540	5.56e-13	→	Malignant pleural mesothelioma	22 of 434	1.19e-21
Disease	35 of 1702	3.77e-11		PI3K-Akt signaling pathway	17 of 336	3.63e-16
Immune System	29 of 1979	3.26e-06		VEGFA-VEGFR2 signaling	14 of 428	1.95e-11
Gene expression (Transcription)	28 of 1476	4.01e-08		Gastrin signaling pathway	13 of 113	3.63e-16
Generic Transcription Pathway	27 of 1215	3.93e-09		Focal adhesion	13 of 195	7.15e-14
Signaling by Receptor Tyrosine Kinases	21 of 521	5.93e-11		Focal adhesion: PI3K-Akt-mTOR-signaling pathway	13 of 301	5.87e-12

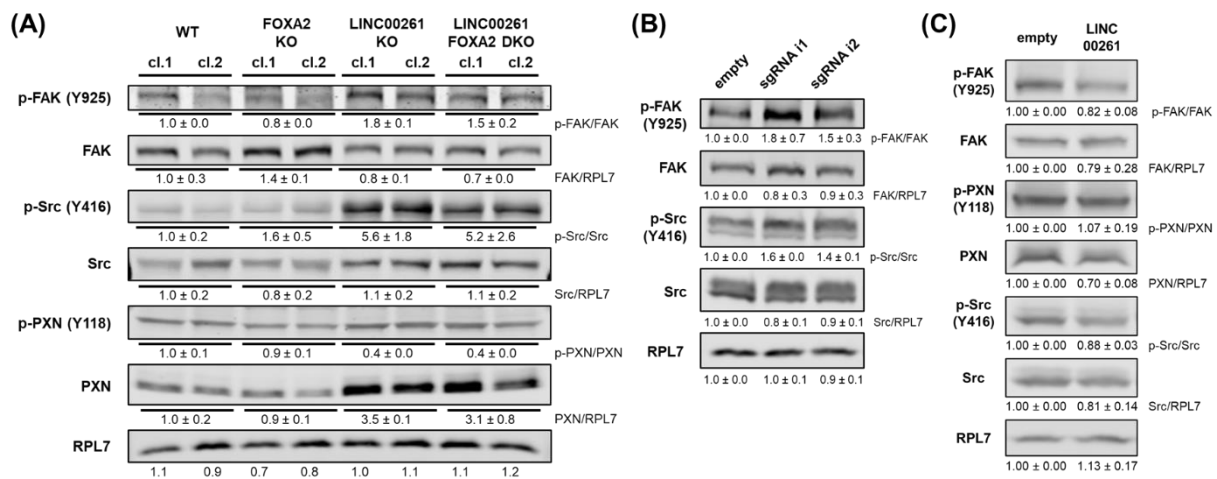
**Figure 35: STRING pathway analysis of FOXA2-independently deregulated (phospho)proteins.**

The (phospho)proteins deregulated independently of FOXA2, as identified through RPPA, underwent analysis using the STRING database<sup>135</sup>. Reactome pathway analysis revealed a significant enrichment of the Signal Transduction pathway. The proteins within this pathway were further examined using WikiPathway analysis.

### 8.1.1.2 Regulation of the FAK/c-Src/PXN axis by LINC00261

RPPA analysis suggested an involvement of the focal adhesion signaling, with particular importance of three signaling proteins: Focal adhesion kinase (FAK), proto-oncogenic tyrosine protein kinase Src (c-Src), and paxillin (PXN). The involvement of these proteins was confirmed using western blot analyses, which showed increased tyrosine phosphorylation of FAK (Y925) and c-Src (Y416), indicating activation of both signaling proteins. Additionally, higher total protein expression of the scaffold protein PXN was observed in LINC00261 KO cells compared to WT and FOXA2 KO cells (Figure 36A). Similar observations were made in LINC0061<sup>low</sup> cells generated using the CRISPRi system with regard to c-Src and FAK. An assessment of PXN protein expression in this cell system is currently pending (Figure 36B). Moreover, overexpression of LINC00261 in PATU-T cells led to a decrease in FAK and Src phosphorylation and paxillin expression (Figure 36C). Consequently, it is possible that the FAK/c-Src/PXN axis may be responsible for the observed migratory and invasive phenotype of LINC00261 KO cells.

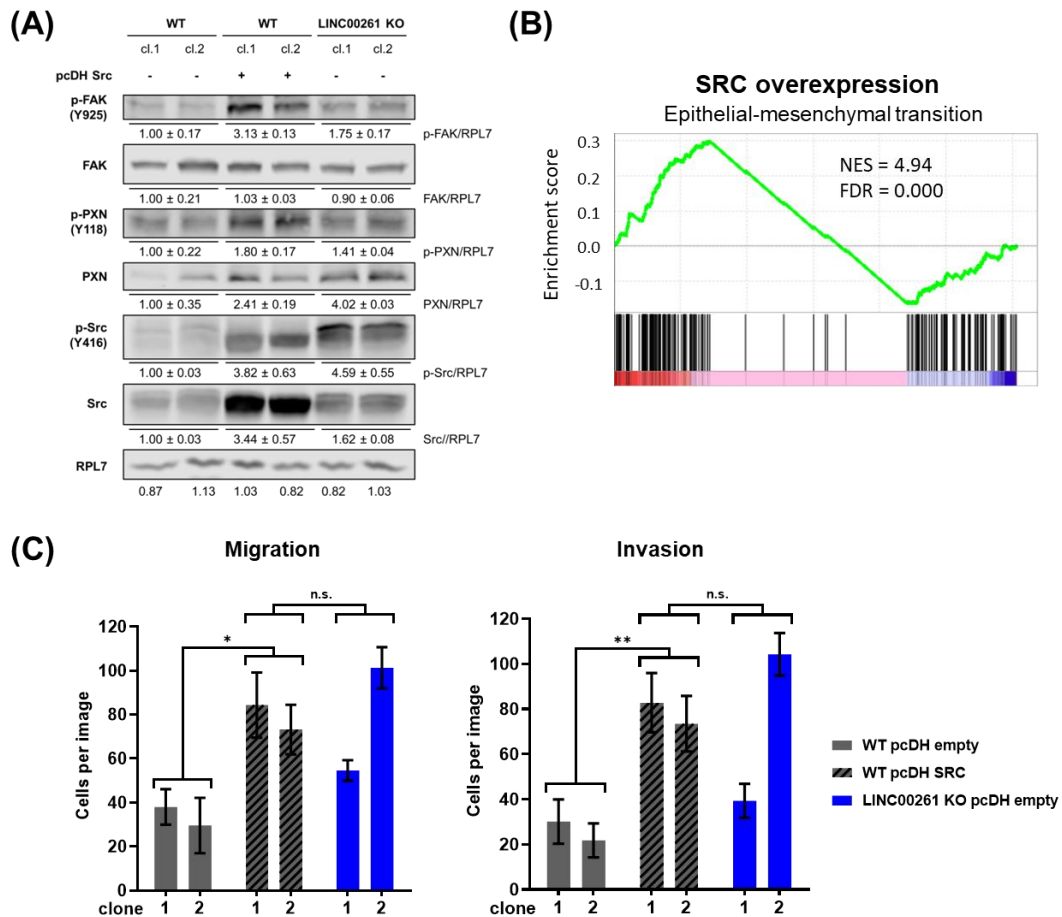




**Figure 36: Western Blots showing FOXA2-independent regulation of c-Src kinase signaling by LINC00261.** (A) FAK and c-Src phosphorylation and PXN expression were enhanced in LINC00261 KO and double KO cells compared to WT/FOXA2 KO cells; (B) Downregulation of LINC00261 using the CRISPRi system with two sgRNAs (i1, i2) resulted in enhanced FAK and c-Src phosphorylation; (C) In PATU-T cells overexpressing LINC00261, a decrease in FAK and c-Src phosphorylation, and a decrease in PXN expression, was observed. RPL7 served as loading control. Quantification of protein expression was carried out by using the Image Studio™ Acquisition Software ( $\pm$  standard deviation).

### 8.1.1.3 The role of the FAK/c-Src/PXN axis in cell migration and invasion

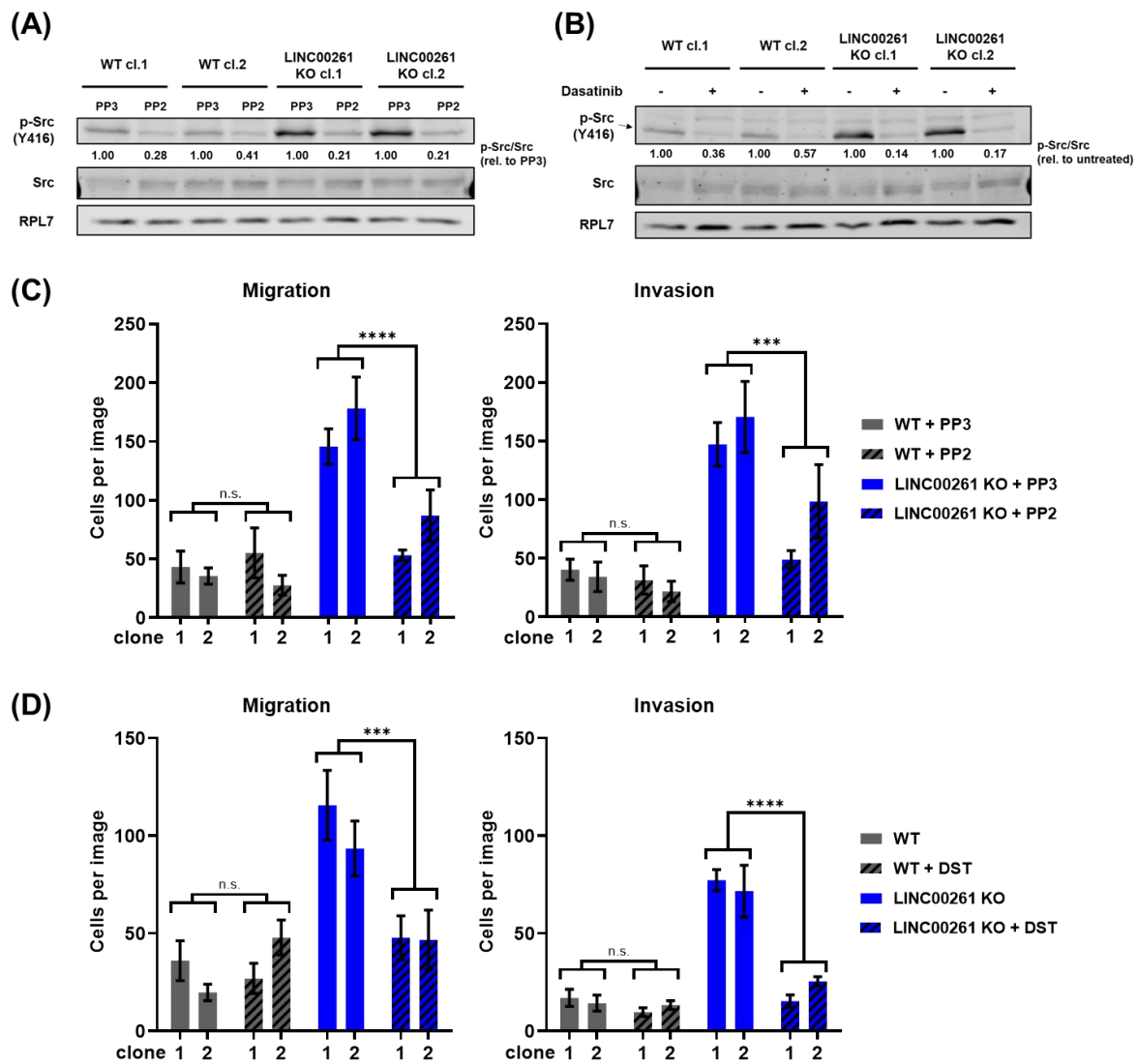
The FAK/c-Src/PXN signaling has been implicated in EMT (see 1.5.3.2). To explore the potential involvement of this signaling pathway in the observed migratory and invasive phenotype of Panc-1 cells, along with the associated EMT gene signature, c-Src was overexpressed in Panc-1 WT cells. Virus titration was employed to achieve a comparable level of active c-Src (phosphorylation of Y416) to that observed in LINC00261 KO cells. Elevated activation of FAK (phosphorylation of Y925) and enhanced PXN expression were discovered in Panc-1 cells overexpressing c-Src (Figure 37A). Moreover, RNA-seq of c-Src-overexpressing Panc-1 cells followed by GSEA analysis revealed significant positive enrichment of the EMT hallmark gene set (Figure 37B), similar to what was observed in LINC00261 KO cells (Figure 20B). In addition, migration and invasion assays showed that Panc-1 WT cells overexpressing c-Src exhibited similarly elevated migratory and invasive potential as LINC00261 KO cells (Figure 37C).



**Figure 37: c-Src overexpression activates FAK/c-Src/PXN axis and induces migration and invasion.**

(A) c-Src overexpression led to the activation of the FAK/c-Src/PXN axis. RPL7 served as loading control. Quantification of protein expression was carried out by using the Image Studio™ Acquisition Software ( $\pm$  standard deviation); (B) GSEA analysis revealed a positive enrichment of EMT-related genes in c-Src-overexpressing cells; (C) c-Src overexpression led to an enhanced migration and invasion of Panc-1 WT cells (\*  $p < 0.05$ , \*\*  $p < 0.01$ , two-way ANOVA).

The impact of the FAK/c-Src/PXN axis on the phenotype of LINC00261 KO cells was further explored by investigating the influence of two different Src inhibitors on cell migration and invasion. Diminished invasion and migration of LINC00261 KO cells were observed after treatment with both the Src inhibitor PP2 (Figure 38A and C) and Dasatinib (Figure 38B and D). The finding that inhibition of the Src kinase reversed the elevated migration and invasion initiated by the downregulation of LINC00261 in Panc-1 cells suggested a potential association between c-Src signaling and LINC00261 expression.

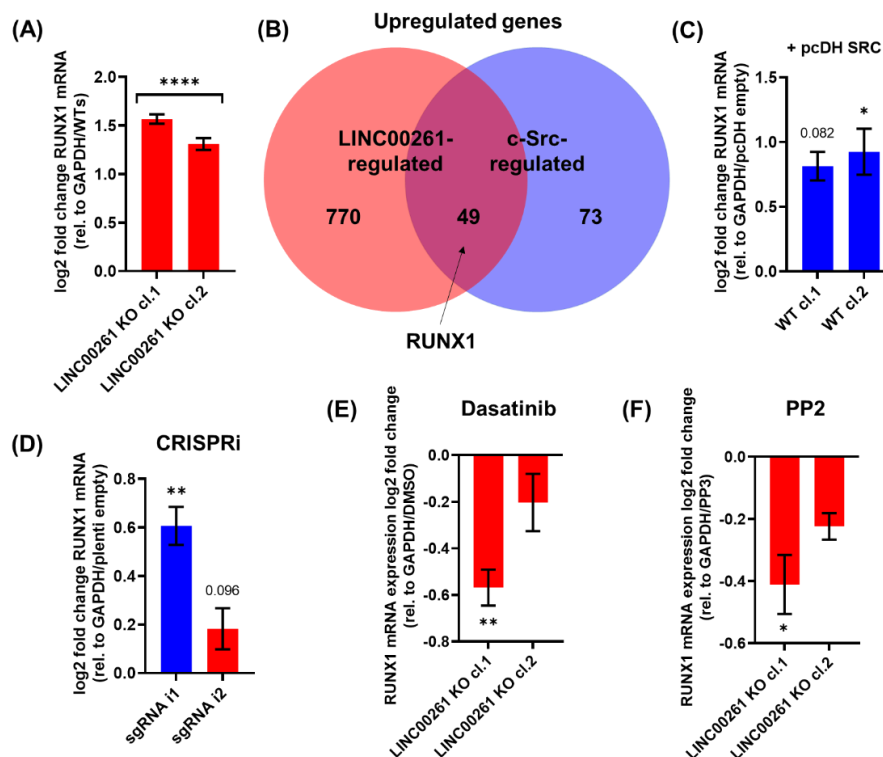


**Figure 38: The migratory and invasive effects observed in LINC00261 KO were reversed by Src inhibition.**

(A) Western Blot confirming a reduction in c-Src phosphorylation following treatment with the c-Src inhibitor PP2. PP3, an inactive analog of PP2, was used as a negative control; (B) Western Blot confirming a reduction in c-Src phosphorylation following treatment with the c-Src inhibitor Dasatinib (DST). RPL7 served as loading control. Quantification of protein expression was carried out by using the Image Studio™ Acquisition Software; (C) Treatment with PP2, but not PP3, reversed the migration and invasion effects observed in LINC00261 KO cells; (D) Treatment with Dasatinib (DST) reversed the migration and invasion effects observed in LINC00261 KO cells (\*\*\*)  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ , two-way ANOVA).

Next, the aim was to identify the transcription factor(s) responsible for the transcriptomic reprogramming towards EMT observed in LINC00261 KO cells, as FOXA2 was excluded as a candidate due to the opposite effects observed in FOXA2 KO cells. Given the involvement of c-Src signaling in the observed reprogramming, an overlap analysis was performed using RNA-seq data to identify common genes regulated by LINC00261 and c-Src. The overlap analysis of genes that were downregulated in LINC00261 KO cells and in Src-overexpressing cells did not yield any interesting candidates of EMT-repressing transcription factors (Supplementary Table S9). However, the overlap of genes upregulated

in both LINC00261 KO cells and Src-overexpressing cells identified 49 commonly regulated genes, including one transcription factor called runt-related transcription factor 1 (RUNX1), which has been associated with tumor cell metastasis and EMT<sup>100,104,136</sup> (Figure 39B, Supplementary Table S9). The upregulation of RUNX1 in LINC00261 KO cells compared to WT cells (Figure 39A) and the upregulation of RUNX1 in Src-overexpressing WT cells were confirmed by qRT-PCR analysis (Figure 39C). Relevantly, RUNX1 was also upregulated in LINC00261<sup>low</sup> cells established by the CRISPRi system (sgRNA i1:  $p < 0.01$ , sgRNA i2:  $p < 0.1$ , Figure 39D). Moreover, the inhibition of Src by treatment with Dasatinib or PP2 led to a significant decrease in RUNX1 expression in LINC00261 KO clone 1, while inhibition of Src in LINC00261 KO clone 2 resulted in a noticeable but not significant decrease in RUNX1 expression (Figure 39E, Dasatinib and Figure 39F, PP2). Taken together, these results suggest that RUNX1 may be a critical downstream effector of the elevated Src signaling in LINC00261 KO cells, and partially responsible for the observed transcriptomic reprogramming in these cells.



**Figure 39: Possible role of c-Src-induced RUNX1 expression in transcriptomic reprogramming.**

(A) qRT-PCR analysis confirming significant upregulation of RUNX1 in LINC00261 KO cells compared to WT cells (\*\*\*\*  $p < 0.0001$ , two-way ANOVA); (B) Venn diagram showing the overlap of genes upregulated in both LINC00261 KO cells and Src-overexpressing WT cells. The RUNX1 gene coding for the transcription factor RUNX1 was upregulated in both cell systems; (C) Confirmation of RUNX1 upregulation in Src-overexpressing Panc-1 WT cells by qRT-PCR (\*  $p < 0.05$ , unpaired t-test); (D) RUNX1 was also upregulated in LINC00261<sup>low</sup> cells generated by the CRISPRi system (\*\*  $p < 0.01$ , unpaired t-test); (E, F) Treatment with c-Src inhibitors Dasatinib (E) and PP2 (F) led to decreased RUNX1 expression in LINC00261 KO cells (\*  $p < 0.05$ , \*\*  $p < 0.01$ , unpaired t-test).

## 8.1.2 FOXA2-dependently regulated genes

Next, the analysis of FOXA2-dependent LINC00261 target genes was performed, revealing CDH1 as a gene regulated by FOXA2 (Supplementary Table S6). The downregulation of CDH1 in FOXA2 KO, as well as in LINC00261 KO and double KO cells, was confirmed by qRT-PCR (Figure 40A) and western blot analysis (Figure 40B). Vice versa, upregulation of CDH1 protein expression was observed in both FOXA2 (Figure 40C) and LINC00261 (Figure 40D) overexpressing PATU-T cells. Intriguingly, CHIP-qRT-PCR analysis demonstrated that FOXA2 binds to the promoter region of CDH1, thereby activating its gene expression, indicating that CDH1 may be regulated independently of the EMT TFs mentioned above. Furthermore, treatment with TGF $\beta$  decreased the binding of FOXA2 to the CDH1 promoter, leading to the downregulation of CDH1 gene expression (Figure 40E). Although the regulatory role of FOXA2 in CDH1 expression was demonstrated by direct binding to the promoter region, experimental results suggest that LINC00261 may be necessary for the regulatory function of FOXA2. Particularly, overexpression of FOXA2 increased the expression of E-cadherin, but this effect was partially blunted in LINC00261 promoter KO cells, despite similar levels of FOXA2 expression in both WT and KO clones (Figure 40F). In summary, these findings suggest that while FOXA2 is capable of regulating CDH1 independently, the full regulatory capacity requires the presence of LINC00261.

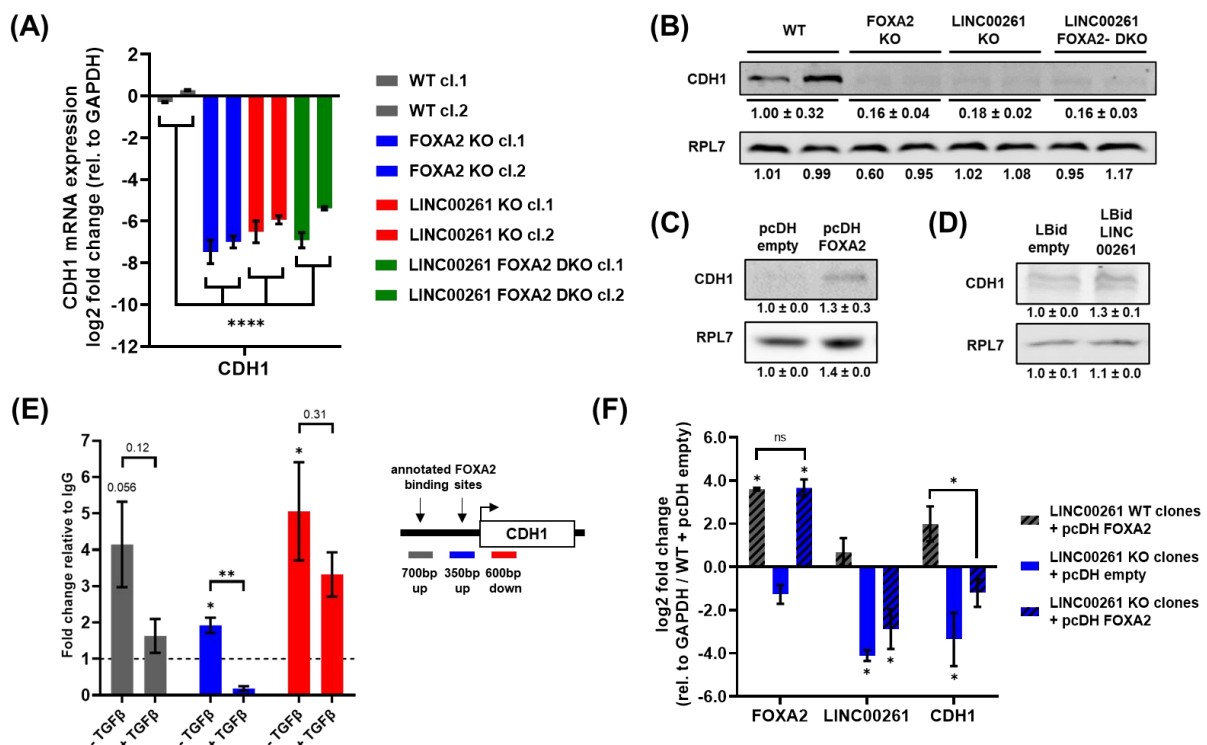


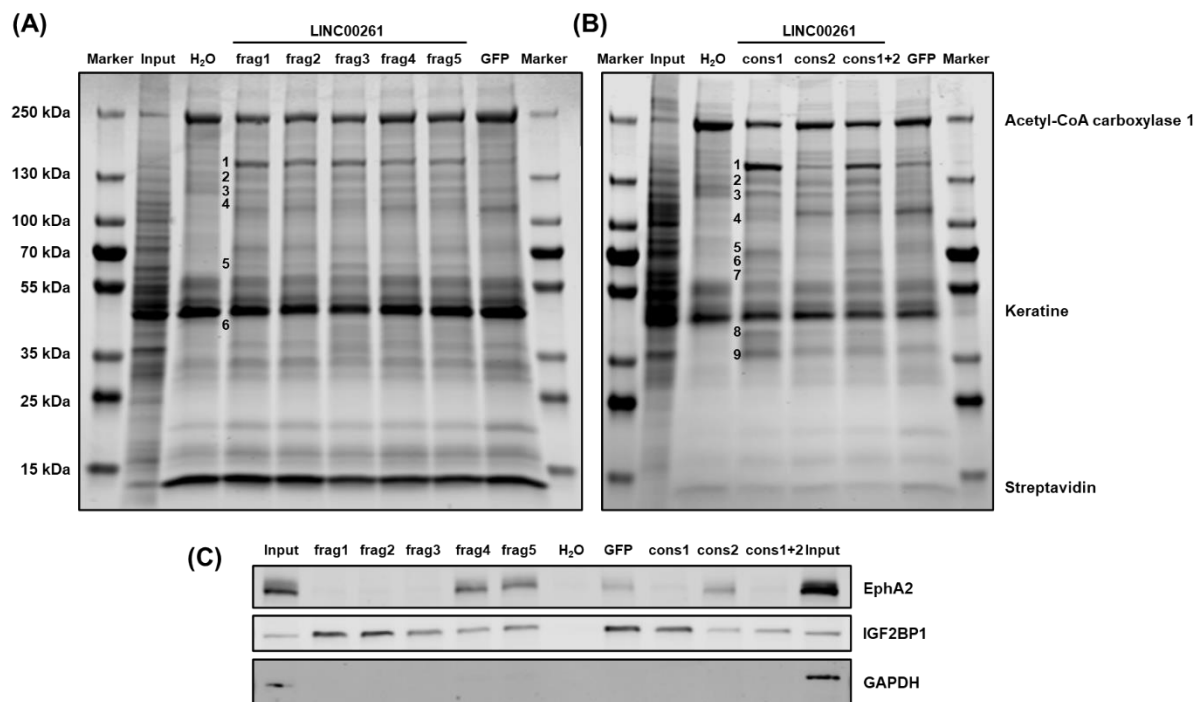
Figure 40: Regulation of CDH1 by FOXA2 and LINC00261.

(A) CDH1 RNA expression was strongly downregulated in FOXA2 KO, LINC00261 KO and double KO Panc-1 cells compared to WT cells (\*\*\*\*  $p < 0.0001$ , two-way ANOVA); (B) Western blot showing disappearance of E-cadherin protein expression in FOXA2 KO, LINC00261 KO and double KO cells; (C, D) Both FOXA2 (C) and

LINC00261 (D) overexpression enhanced E-cadherin protein expression in PATU-T cells; (E) ChIP followed by qPCR analysis using primers located upstream and downstream of the CDH1 transcriptional start site (right panel) confirmed binding of FOXA2 to the CDH1 promoter region (left panel, \*  $p < 0.05$ , \*\*  $p < 0.01$ , unpaired t-tests); (F) FOXA2, LINC00261 and CDH1 expression in LINC00261 WT and KO clones with or without FOXA2 overexpression (\*  $p < 0.05$ , two-way ANOVA). RPL7 served as loading control for western blot analyses. Quantification of protein expression was carried out by using the Image Studio™ Acquisition Software ( $\pm$  SD).

### 8.1.3 Identification of LINC00261-binding proteins

As outlined in the introduction, lncRNAs can interact with a diverse array of biomolecules, including DNA, RNA, and proteins, where they can serve as scaffolds, guides, decoys, or signaling molecules. The focus in this study was on identifying specific proteins that interact with LINC00261, thereby elucidating potential regulatory mechanisms underlying the cellular observations. To achieve this goal, protein pull-down experiments were conducted using *in vitro* transcribed and biotinylated RNA. To minimize steric hindrance and enhance the accessibility of binding sites within the pull-down experiment, five fragments of the 4912 nts long LINC00261, each ~1000 nts in length, were synthesized. Additionally, the pull-down assay was performed using two conserved regions from LINC00261's exon 4 (444 and 736 nts, Figure 12, red boxes B and C), due to their potential physiological significance. Following the pull-down experiment, proteins were separated by SDS-PAGE, and differential protein bands were analyzed using mass spectrometry (Figure 41A, B). To date, the analysis has focused exclusively on the examination of the nine bands as depicted in Figure 41B. Proteins demonstrating high abundance in mass spectrometry analysis were subjected to western blot analyses. Intriguingly, the binding of FOXA2 could not be demonstrated in the pull-down experiment. Beyond FOXA2, the focus was also on proteins involved in EMT signaling cascades, such as the TGF $\beta$  and FAK/c-Src/PXN pathways. Interestingly, Ephrin type-A receptor 2 (EphA2), a known participant in c-Src signaling with implications for tumor progression<sup>137-139</sup>, was identified in gel bands 2, 3, and 4. Insulin-like growth factor 2 mRNA-binding protein 1 (IGF2BP1), recently demonstrated to interact with LINC00261<sup>140</sup> and implicated in EMT<sup>141</sup>, exhibited substantial presence in gel band 7. In the western blot analyses, robust binding interactions between EphA2 and specific regions of LINC00261 were observed. Specifically, strong binding was detected between EphA2 and LINC00261 fragments 4 and 5, along with the conserved region 2 of LINC00261. In the lanes corresponding to conserved regions 1 and 1+2, faint bands were noted, suggesting a potential, albeit weaker, interaction with these regions. Notably, a distinct band was also observed in the lane corresponding to GFP mRNA, which served as a control for nonspecific binding, indicating an interaction between EphA2 and GFP mRNA. IGF2BP1 demonstrated an affinity for all LINC00261 fragments, as well as GFP mRNA, reflecting its versatile RNA-binding capabilities (Figure 41C).



**Figure 41: Pulldown of RNA-binding proteins with biotinylated LINC00261 fragments.**

(A, B) Lysates of Panc-1 cells were incubated with five *in vitro* transcribed, biotinylated LINC00261 RNA fragments, each ~1000 nucleotides in length (A), or two highly conserved regions of LINC00261's exon 4, as depicted in Figure 12 (B). Streptavidin beads incubated with unrelated GFP mRNA (GFP) or water (H<sub>2</sub>O) served as controls for nonspecific protein binding. RNA-binding proteins were separated by SDS-PAGE and stained with Coomassie Blue. Inserted numbers indicate prominent differential bands. Bands 1-9 in gel B were analyzed by mass spectrometry; (C) Western blot analyses demonstrate the binding of EphA2 and IGF2BP1 to LINC00261 and GFP mRNA, two proteins of interest identified in the mass spectrometry analysis.

## 8.2 LINC00261-independent regulation of genes by FOXA2

Lastly, the effects of FOXA2 occurring independently of LINC00261 were examined. However, due to the significant downregulation of FOXA2 in LINC00261 KO cells, a comparison between LINC00261 KO and double KO cells would result in a substantial loss of information. As a result, the differences in gene expression between WT and FOXA2 KO cells were initially assessed, and subsequently, the targets of interest were validated by overexpressing FOXA2 in the same cells.

Initially, GSEA was conducted on the genes that were deregulated in FOXA2 KO cells. Interestingly, the hallmark of EMT was highly negatively enriched, which contrasts with the KO of LINC00261 (Figure 42A, B). This finding is consistent with the opposite behavior observed in the migration and invasion assays (Figure 30).

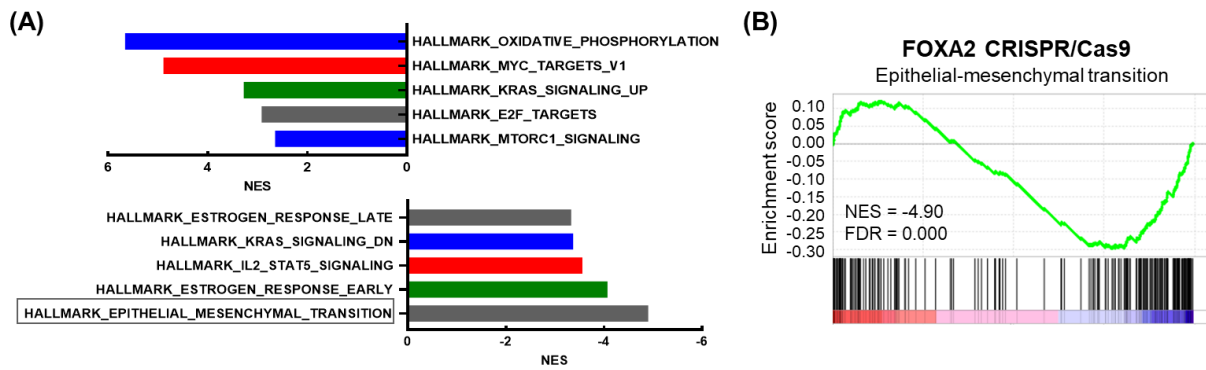


Figure 42: GSEA analysis of FOXA2-regulated genes.

(A) Overview of the five most positively and negatively regulated GSEA hallmark gene sets in FOXA2 KO cells compared to WT cells; (B) GSEA analysis of RNA-sequencing data of FOXA2-regulated genes revealed a significant negative enrichment of the EMT hallmark gene set.

In previous experiments, it was demonstrated that the FAK/PXN/c-Src axis was not deregulated in FOXA2 KO cells (Figure 36), ruling out a possible role of this signaling pathway in the negative regulation of EMT-related genes. Interestingly, in addition to the EMT hallmark gene set, GSEA analysis of the RNA-seq data also showed a significant negative enrichment of the TGF $\beta$  hallmark gene set in FOXA2 KO cells (Figure 43A). Accordingly, qRT-PCR analysis confirmed deregulation of genes associated with the TGF $\beta$  signaling and/or EMT, including transforming growth factor  $\beta$  2 (TGFB2), transforming growth factor  $\beta$  receptor 2 (TGFB2) and N-cadherin (CDH2) (Figure 43B). This effect seemed to be exclusively dependent on FOXA2 since it was not observed following the depletion of LINC00261.

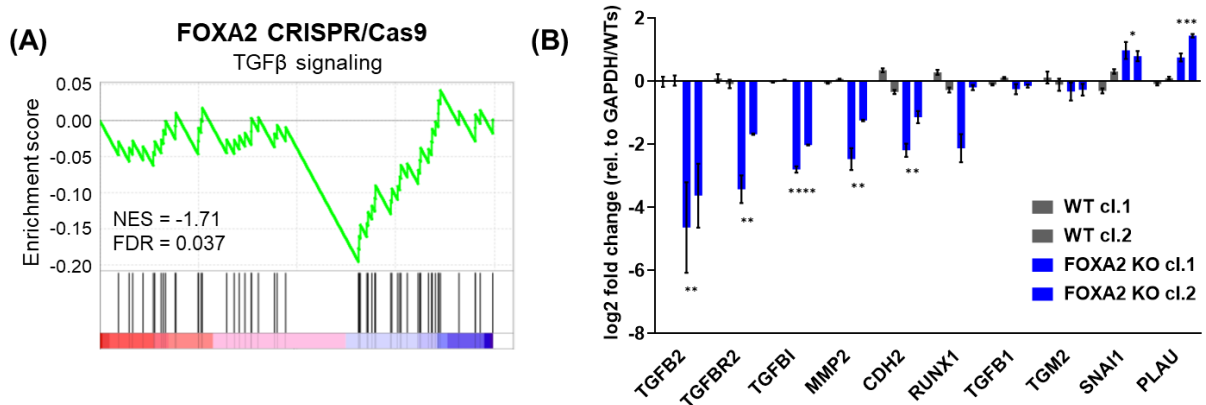
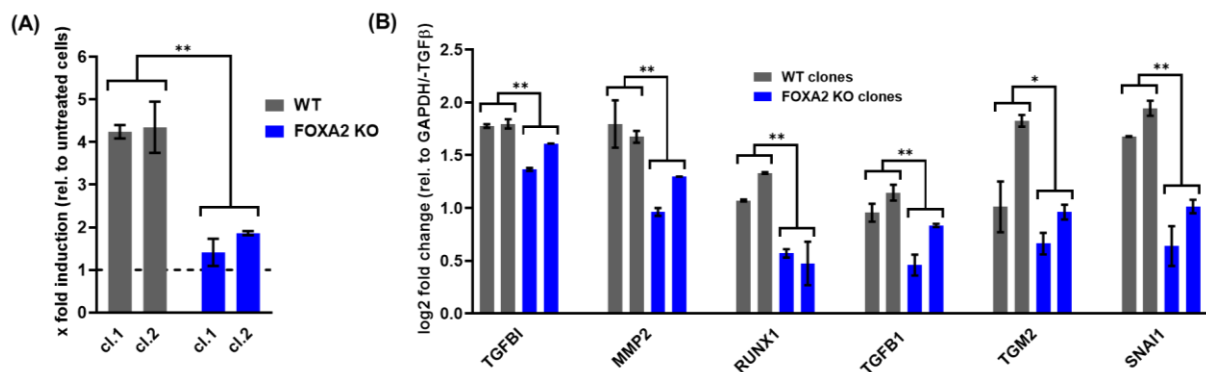


Figure 43: Loss of FOXA2 negatively regulates TGF $\beta$  hallmark gene set.

(A) GSEA analysis of RNA-sequencing data revealed a significant negative enrichment of the TGF $\beta$  hallmark gene set in FOXA2 KO cells; (B) qRT-PCR analysis confirmed deregulation of several EMT and TGF $\beta$  target genes (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ , two-way ANOVA).



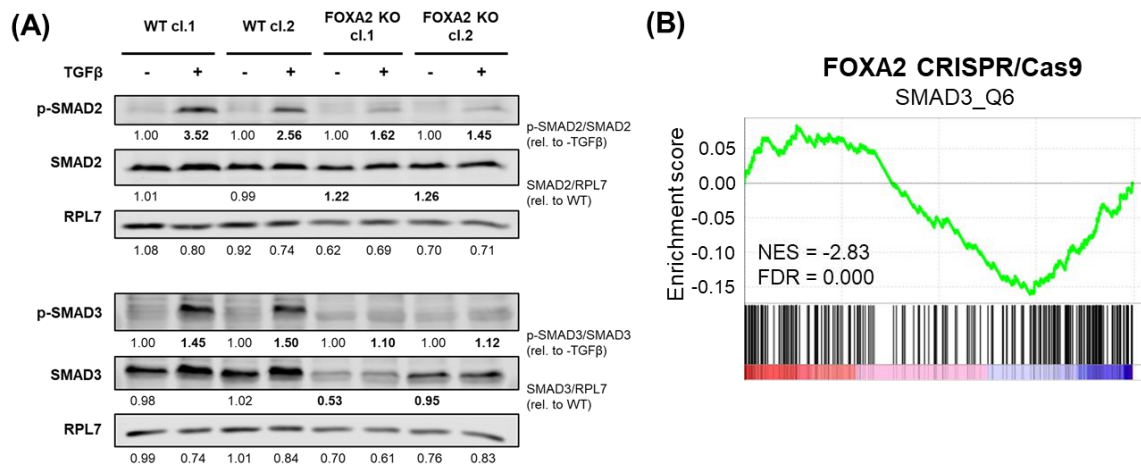
In order to examine the involvement of FOXA2 in the regulation of the canonical TGF $\beta$  signaling pathway, both WT and FOXA2 KO cells were treated with TGF $\beta$ , and the luciferase activity of the SMAD-binding element (SBE) reporter was measured. The SBE reporter is a specific DNA sequence that is activated by SMAD proteins to initiate gene expression. The induction of the SBE reporter was critically reduced in FOXA2 KO cells. Specifically, treatment with TGF $\beta$  resulted in a 4-fold induction of luciferase activity in WT cells, while FOXA2 KO cells only displayed a 1.5-fold induction (Figure 44A). Quantitative PCR analyses confirmed the significantly reduced induction of the expression of several TGF $\beta$ -inducible genes (Figure 44B).



**Figure 44: Attenuation of canonical TGF $\beta$  signaling in FOXA2 KO cells.**

(A) Induction of the luciferase activity of a SMAD-binding element reporter in Panc-1 WT cells and FOXA2 KO cells after treatment with 10 ng/ml TGF $\beta$  for 24h (\*\*  $p < 0.01$ , two-way ANOVA); (B) After treatment with 10 ng/ml TGF $\beta$  for 24h, several TGF $\beta$ -target genes were found to be less upregulated in FOXA2 KO cells compared to WT cells (\*  $p < 0.05$ , \*\*  $p < 0.01$ , two-way ANOVA).

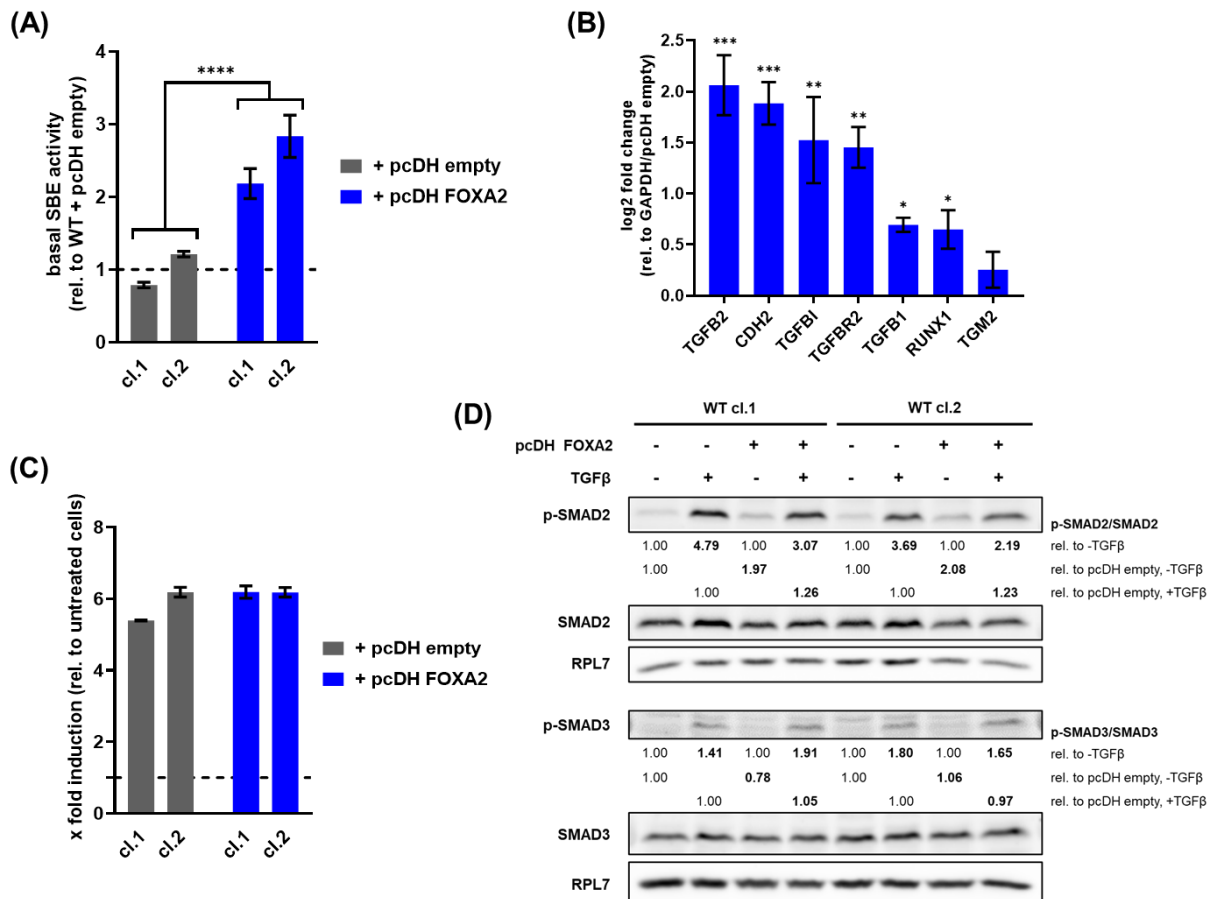
These results suggested that FOXA2 is required for a thorough activation of the canonical TGF $\beta$  signaling, which involves the SMAD proteins (see 1.5.3.1). RPPA analysis showed that the protein expression of SMAD3, an indispensable component of the canonical TGF $\beta$  pathway, was considerably downregulated after FOXA2 depletion (Supplementary Table S10). The Western blot analysis of the two major players in canonical TGF $\beta$  signaling, SMAD2 and SMAD3, not only confirmed the downregulation of SMAD3, but also revealed a significant reduction in the activation of both SMAD proteins upon TGF $\beta$  treatment in FOXA2 KO cells (Figure 45A). Specifically, SMAD2 phosphorylation by TGF $\beta$  was weaker in FOXA2 KO cells than in WT cells, whereas SMAD3 phosphorylation was almost completely absent. GSEA analysis with respect to transcription factor motifs revealed negative enrichment of the SMAD3\_Q6 motif in FOXA2 KO cells, indicating reduced expression of genes regulated specifically by SMAD3 (Figure 45B).



**Figure 45: Depletion of FOXA2 leads to significant reduction in SMAD signaling.**

(A) Western blot analysis of expression and activation of SMAD2 and SMAD3 in TGF $\beta$ -treated WT and FOXA2 KO cells. RPL7 served as loading control. Quantification of protein expression was carried out by using the Image Studio™ Acquisition Software; (B) GSEA analysis of transcription factor binding sites reveals a negative enrichment of genes containing the SMAD3 binding motif TGTCTGTCT in the regions spanning 4 kb centered on their transcription starting sites [-2kb, +2kb].

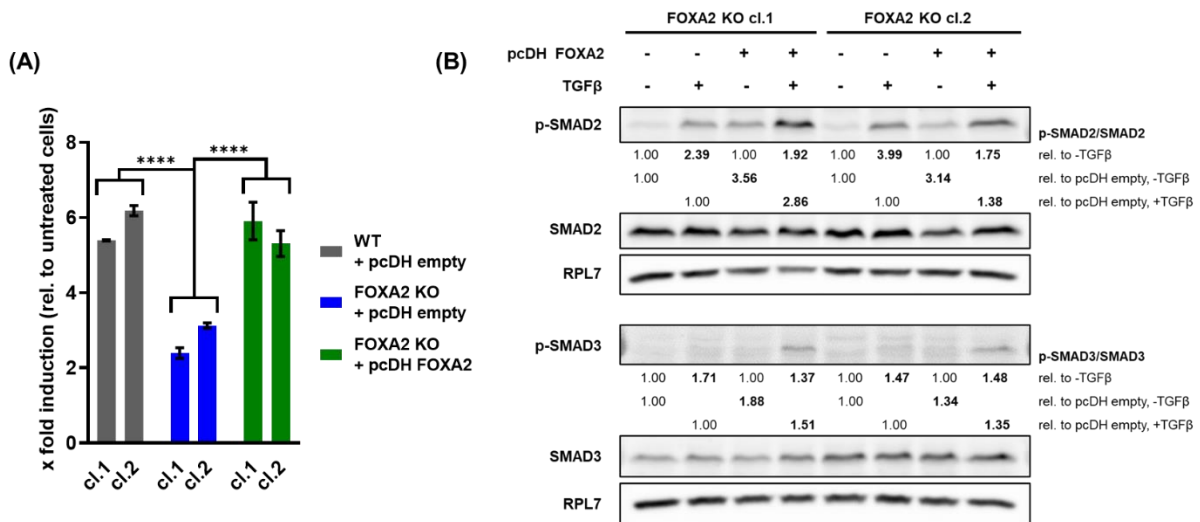
To confirm the influence of FOXA2 on TGF $\beta$  signaling, FOXA2 was overexpressed in WT cells. The basal SBE reporter activity was evaluated and found to be 2–3 times higher in the FOXA2 overexpressing cells compared to WT cells, confirming regulation by FOXA2 (Figure 46A). Moreover, many of the genes downregulated in FOXA2 KO cells were upregulated in FOXA2 overexpressing cells (Figure 46B). However, elevated expression of FOXA2 did not result in an increased induction of the SBE reporter by TGF $\beta$  (Figure 46C). Consistent with this, Western blot analysis of SMAD2 and SMAD3 phosphorylation showed no increased activation of either protein after TGF $\beta$  treatment in FOXA2-overexpressing cells. On the other hand, in line with the elevated basal SBE activity, the basal phosphorylation level of SMAD2 was increased in FOXA2-overexpressing cells compared to WT cells. Conversely, neither the basal phosphorylation level nor the total protein expression of SMAD3 was enhanced in FOXA2-overexpressing cells (Figure 46D).



**Figure 46: FOXA2 overexpression modulates TGFβ signaling.**

(A) FOXA2 overexpression resulted in a higher basal SMAD signaling activity (\*\*\*\* p < 0.0001, two-way ANOVA); (B) Several TGFβ target genes are upregulated by FOXA2 overexpression (\*\*\* p < 0.001, \*\* p < 0.01, \* p < 0.05, unpaired t-test, Holm-Šídák correction); (C) FOXA2 overexpression did not lead to an elevated induction of the signaling by TGFβ; (D) Western blot analysis of expression and activation of SMAD2 and SMAD3 in TGFβ-treated WT cells stably transfected with pcDH empty or pcDH FOXA2 plasmids and treated with 10 ng/ml TGFβ for 24h. RPL7 served as loading control. Quantification of protein expression was carried out by using the Image Studio™ Acquisition Software.

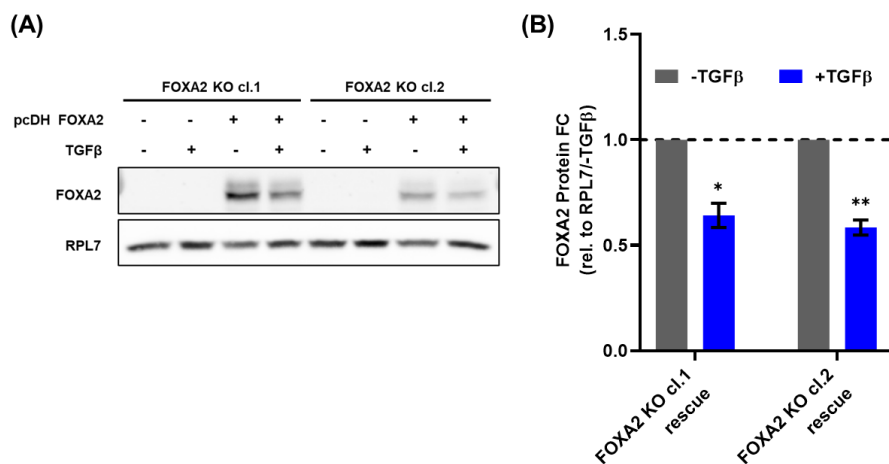
Next, the expression of FOXA2 was rescued in both FOXA2 KO clones to validate the observed effect of FOXA2 on the canonical TGFβ signaling. The rescue of FOXA2 led to the recovery of the inducibility of the SBE reporter activity by TGFβ (Figure 47A). Furthermore, the phosphorylation of the SMAD proteins by TGFβ treatment was recovered (SMAD3) or improved (SMAD2), respectively (Figure 47B).



**Figure 47: Rescue of FOXA2 recovers TGFβ signaling in FOXA2-deficient cells.**

(A) Induction of luciferase activity of a SMAD-binding element reporter in Panc-1 WT, FOXA2 KO and FOXA2 rescue cells after treatment with 10 ng/ml TGFβ for 24h (\*\* p<0.01, two-way ANOVA); (B) Western blot analysis of expression and activation of SMAD2 and SMAD3 in TGFβ-treated WT cells stably transfected with pcDH empty or pcDH FOXA2 plasmids and treated with 10 ng/ml TGFβ for 24h. RPL7 served as loading control. Quantification of protein expression was carried out by using the Image Studio™ Acquisition Software.

Interestingly, a reduction of FOXA2 protein expression was observed in the rescued cells after TGFβ treatment (Figure 48A). Particularly, the protein level decreased by around 40% in both clones following treatment (Figure 48B). Since the protein was exogenously introduced, these findings suggest a post-transcriptional and/or -translational downregulation of FOXA2.



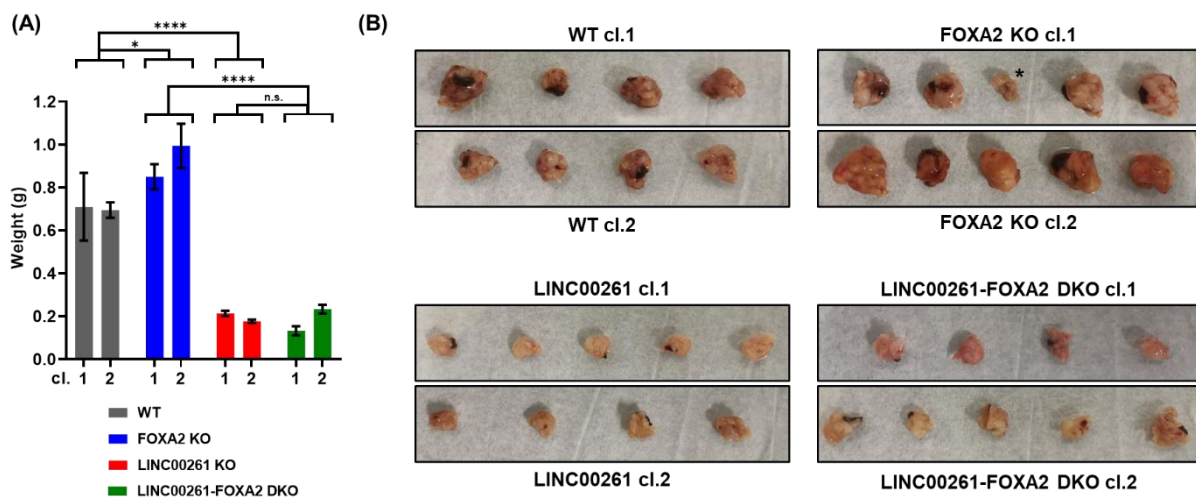
**Figure 48: TGFβ treatment downregulates FOXA2 post-transcriptionally and/or -translationally.**

(A) Western blot demonstrating downregulation of exogenous FOXA2 by TGFβ; (B) Quantification of the western blot analysis confirms a significant downregulation of exogenous FOXA2 by 10 ng/ml TGFβ after 24h (\* p<0.05, \*\* p<0.01, unpaired t-test). RPL7 served as loading control. Quantification of protein expression was carried out by using the Image Studio™ Acquisition Software.

Altogether, these experiments illuminate FOXA2 as a critical positive regulator of the canonical TGF $\beta$  pathway, while also unveiling its participation in a regulatory negative feedback loop, in which FOXA2 itself experiences downregulation by TGF $\beta$  at a post-transcriptional and/or post-translational level.

### 9. Importance of LINC00261 and FOXA2 for *in vivo* tumor growth and metastasis

To investigate the roles of LINC00261 and FOXA2 within a more complex biological setting, *in vivo* models were employed. Tumor growth of respective cell lines was assessed using an orthotopic mouse model, wherein  $4 \times 10^5$  cells were orthotopically injected into five mice for each group. Unfortunately, two mice died during the surgical procedure, and two others died later for unknown reasons. The tumors within a group grew consistently with one exception in the FOXA2 KO cl.1 group, which was identified as an outlier. Interestingly, the size of the tumors after six weeks originating from the FOXA2 KO cells were significantly larger compared to the ones originating from the WT cells. On the other hand, the tumors deriving from LINC00261 KO and LINC00261-FOXA2 double KO cells were significantly smaller than the WT tumors (Figure 49).



**Figure 49: Impact of LINC00261 and FOXA2 KO on primary tumor growth in an orthotopic mouse model.** (A) Determination of the tumor weight 6 weeks after the injection (in gram (g), \*  $p < 0.05$ , \*\*\*\*  $p < 0.0001$ , two-way ANOVA); (B) Images of the tumors directly after harvesting. The weight of the tumor indicated with an asterisk (\*) was identified as an outlier by the Grubbs outlier test ( $\alpha = 0.05$ ).

The primary tumors were subjected to hematoxylin and eosin (H&E) staining to examine their morphological characteristics. The staining revealed that all examined tumors displayed a notable degree of intratumoral heterogeneity. This heterogeneity was evident through irregularly dispersed angiogenesis, as well as variations in cell size and density within the tumor tissue. However, there were no discernible disparities observed when comparing primary tumors of different cell lines (Figure 50).

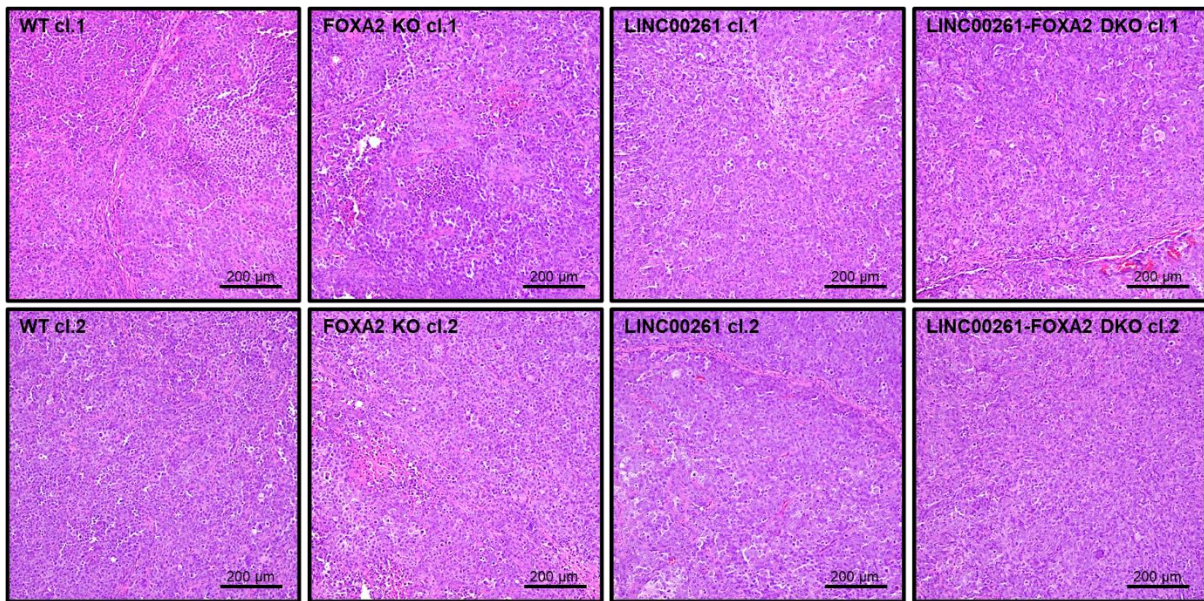


Figure 50: Microscopic images of primary tumors originating from established cell lines. Representative images of primary tumors stained by hematoxylin and eosin (H&E).

The transfected luciferase-expressing vector allowed for the analysis of tumor cell metastasis to the liver and the lung by measuring bioluminescence. The measurement of the luminescence intensities revealed a higher number of metastases in the liver than in the lungs. Notably, the luminescence signal in the livers (Figure 51A) and lungs (Figure 51B) of mice injected with the FOXA2 KO cells was significantly higher compared to all other mice. In contrast, no difference was observed between WT tumors and LINC00261 KO or LINC00261-FOXA2 double KO tumors.

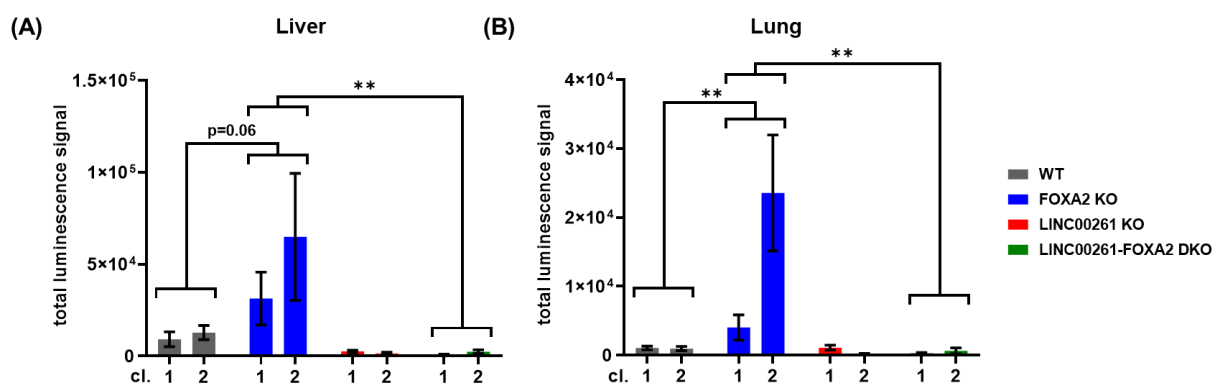


Figure 51: Metastasis of tumor cells to liver and lung analyzed by luminescence measurement.

To determine the extent of tumor cell metastasis, luminescence in the liver (A) and lung (B) was quantified using the *in vivo* imaging system (IVIS).

Figure 52 shows representative images of the liver metastases. The mice injected with FOXA2 KO cells exhibited the largest and most extensive metastases in the liver, followed by the animals injected with WT cells. In contrast, the livers of LINC00261 KO and LINC00261-FOXA2 double KO mice showed only small or no metastases. The metastasis

to the lung was minimal in most mice, thus representative images could not be obtained. Therefore, caution should be exercised in interpreting the lung data presented in Figure 51B.

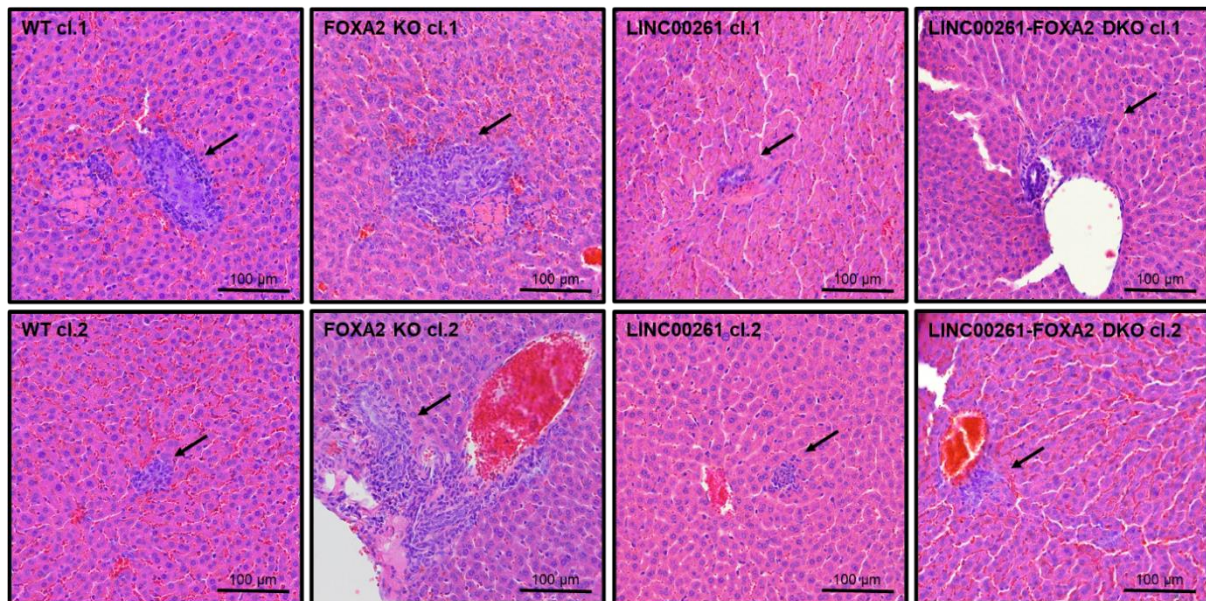


Figure 52: Microscopic images of metastases in the liver.

Representative images of PDAC metastases in the liver (black arrows) stained with H&E.

## 10. Tabular overview of the results of this study

The main findings on reduced LINC00261 and FOXA2 expression in pancreatic cancer (cells) are summarized in the following table.

Table 20: Overview of the main findings of this study.

Expression level (vs. WT*)		LINC00261 low	FOXA2 low	LINC00261 & FOXA2 low
Patient survival (TCGA)		↘	↘	↘
<i>In vitro</i> phenotype	Proliferation	-	↘	-
	Migration	↗	↘	↗↗
	Invasion	↗	↘	↗↗
	Clonogenicity	-	↘	↗↗
<i>In vivo</i> phenotype	Tumor growth	↘	↗	↘
	Metastasis	↘	↗	↘
Expression of EMT-related genes (GSEA)		↗	↘	↗
EMT-related pathways/genes	TGFβ pathway	-	↘	↘
	c-Src pathway	↗	-	↗
	CDH1	↘	↘	↘

\* For patient survival: Expression relative to all tumor samples.

## V. DISCUSSION

The discussion contains revised and adapted content from the original research article “LINC00261 Is Differentially Expressed in Pancreatic Cancer Subtypes and Regulates a Pro-Epithelial Cell Identity”, Dorn et al., *Cancers* (2020)<sup>58</sup>.

### 1. Deregulated and subtype-specific lncRNAs in PDAC

Recent studies have uncovered a broad spectrum of lncRNA functions in cancer, including their roles in tumor initiation and progression<sup>41</sup>. However, only a limited number of studies have investigated the role of lncRNAs in PDAC. A comprehensive and systematic analysis of differential lncRNA expression in pancreatic cancer has identified specific lncRNAs that may serve as potential biomarkers for disease and patient survival<sup>142–147</sup>. Overall, a diverse set of lncRNAs in pancreatic cancer has been identified; however, the functions of these are largely unknown. Recently, large-scale RNA sequencing analyses of PDAC samples have provided additional insights into pancreatic carcinogenesis. In these studies, different molecular subtypes of prognostic and biological relevance have been identified<sup>15–17</sup>, and differential expression of lncRNAs has been associated with these subtypes<sup>56</sup>. The potential of these large datasets was leveraged in the present study to conduct a comprehensive analysis of the cohort of Bailey et al.<sup>15</sup>, and the NMF algorithm<sup>118</sup> was applied to identify the four previously reported disease subtypes of PDAC. Notably, PDAC patients assigned to the squamous subtype exhibit the worst overall survival due to the highly aggressive disease histology associated with this subtype. As lncRNAs have been shown to affect gene expression on multiple levels, it was hypothesized that these transcripts could actively contribute to the disease biology of the aggressive squamous subtype of PDAC. Therefore, the lncRNA expression landscape was analyzed across PDAC subtypes in order to identify lncRNAs specifically associated with the squamous subtype (Figure 9A).

### 2. Expression of LINC00261 in PDAC

The subtype-specific analysis led to the identification of LINC00261, whose expression was found to be variable across PDAC subtypes and correlated with stage and grade, as well as favorable patient survival (Figure 9 and Figure 10). More specifically, a significant downregulation of LINC00261 expression was found in the squamous subtype of the Bailey dataset<sup>15</sup> and in the basal-like subtype of the TCGA dataset, as defined by Moffitt et al.<sup>16</sup>, including only those samples with a high tumor cell content, as described by the Cancer Genome Atlas Research Network<sup>56</sup>. These results propose that LINC00261 may be one of several important factors contributing to the establishment of a gene expression program that is characteristic for the squamous/basal-like subtype of pancreatic cancer. Compared to normal pancreatic tissue, the expression of LINC00261 was markedly reduced in



pancreatic cancer, aligning with prior reports in different tumor types that have shown dysregulated LINC00261 expression in cancer as opposed to normal tissue<sup>78,148,149</sup>. Moreover, higher LINC00261 expression was observed in low-grade and early-stage PDAC samples. Altogether, the analyses conducted in this study in pancreas carcinoma, along with additional published reports in other cancer entities, strongly suggest that LINC00261 may function as a tumor-suppressive lncRNA. However, studies on cholangiocarcinoma and neuroendocrine prostate cancer have reported that LINC00261 may also have pro-tumorigenic functions<sup>150,151</sup>.

### 3. Characterization of LINC00261

The selected lncRNA, LINC00261, was initially characterized by analyzing its sequence conservation across species, its coding probability, and its subcellular localization.

The conservation analysis revealed that the LINC00261 locus is present in all mammalian genomes, with particularly strong conservation in primates, suggesting a physiological importance of this lncRNA in this biological order. In contrast, conservation of LINC00261 outside of mammals, such as in birds, amphibians, and fish, is limited to specific regions in exon 1 and exon 4 (Figure 12). These highly conserved regions may be crucial for the secondary/tertiary structure of LINC00261, which is proposed to be more functionally relevant than the nucleotide sequence<sup>37,38</sup>. Palazzo and Koonin<sup>35</sup> postulated that most lncRNAs originate from transcriptional events of surrounding genes, particularly those involved in transcription-related processes. Indeed, numerous studies have observed that lincRNAs are frequently located in close proximity to genes encoding transcription factors<sup>37,152,153</sup>. Given this, it is plausible that LINC00261 evolved from transcriptional events originating from the nearby transcription factor FOXA2.

The protein-coding potential of an RNA is generally determined by the presence and the length of ORFs, which is also the basis of the algorithms used in this study<sup>124,125</sup>. Importantly, both algorithms classified LINC00261 as a non-coding RNA, and the coding probability value was similar to that of well-characterized lncRNAs such as MALAT1, XIST, and NEAT1 (Figure 13A). Ji et al., however, reported that up to 40% of transcripts annotated as non-coding could be translated and encode small peptides, but these appeared to be mostly unstable by-products with no function<sup>154</sup>. Indeed, Gaertner et al. detected seven microproteins produced by the LINC00261 transcript using Ribo-seq and *in vitro* translation<sup>155</sup>. However, their observations implied a function of the LINC00261 transcript independent of the produced microproteins. Also, they did not provide a protein-level evidence for the endogenous production and stability of the microproteins in their cell system<sup>155</sup>. Moreover, LINC00261 has not been previously detected by sORF analyses in other cell types, questioning the presence of these ORFs<sup>154,156–160</sup>. Taken together, both the

conducted analysis in this study and the current literature classify LINC00261 as non-coding, and the single study that detected ORFs in LINC00261 declared them as non-functional.

Interestingly, Ji et al. also found that translated lncRNAs are preferentially localized in the cytoplasm, whereas untranslated lncRNAs preferentially localize in the nucleus<sup>154</sup>. Consistent with these findings, the cell fractionation analysis of a panel of pancreas cell lines conducted in this study revealed a predominant nuclear localization of LINC00261 (Figure 13B), which is also supported by studies in mouse hepatocytes<sup>161</sup>, esophageal cancer cells<sup>162</sup>, and lung epithelial cells<sup>163</sup>. This localization pattern suggests a role for LINC00261 in the control of target gene transcription, possibly through recruitment of transcription factors or by regulating higher order chromatin folding. Notably, the nuclear enrichment of LINC00261 was not as prominent as that observed for well-known nuclear lncRNAs such as MALAT1 or NEAT1, which were used as positive controls in the experiments<sup>126</sup>. Thus, LINC00261 may shuttle between the cytoplasm and the nucleus, and its subcellular localization may dictate its various molecular functions. Moreover, the localization and function of LINC00261 may differ between normal and cancer cells or even between different cancer types, which warrants further investigation to fully understand the regulation and biological role of LINC00261 in cancer.

#### 4. Role of LINC00261 in EMT of PDAC cells

Squamous or basal-like tumors are characterized by gene expression changes related to oncogenic signaling, including EMT<sup>18</sup>. The process of EMT and the expression of EMT transcription factors have been linked to cancer progression and therapy resistance in PDAC<sup>164</sup>. Moreover, the EMT status of patient-derived tumor specimens, as determined by the expression of epithelial and mesenchymal markers, as well as EMT-associated transcription factors, is predictive of pancreatic cancer prognosis<sup>165,166</sup>. However, one study suggested that EMT may be dispensable for metastasis, but still important for chemoresistance in pancreatic cancer<sup>167</sup>. Similar conclusions were drawn for breast cancer<sup>168</sup>. The reduced expression of LINC00261 in the squamous subtype of PDAC may be causally linked to disease progression rather than only being a bystander effect. Particularly, LINC00261 may actively contribute to the disease subtype by modulating the expression of epithelial and mesenchymal genes, thereby supporting the migratory and invasive phenotypes of cancer cells. To address this question, two complementary CRISPR strategies were employed to reduce LINC00261 levels in Panc-1 cells, and migratory and invasive behavior *in vitro* was analyzed (Figure 14). A strong induction of cell migration and invasion was revealed after LINC00261 downregulation (Figure 16 and Figure 17). Furthermore, overexpression of LINC00261 in PATU-T cells led to a decreased cell migration and invasion (Figure 18). These findings are supported by studies in pancreatic

cancer<sup>140,169–171</sup>, as well as other cancer types, such as gastric cancer<sup>149</sup>, hepatocellular carcinoma<sup>148</sup>, and lung adenocarcinoma<sup>172</sup>. However, contrasting observations were made in two studies on cholangiocarcinoma and neuroendocrine prostate cancer, proposing LINC00261 as a driver of cell migration and invasion<sup>150,151</sup>. Consistent with the migratory and invasive phenotype, a significant enrichment of genes related to EMT was found after LINC00261 depletion using CRISPR interference or CRISPR/Cas9-mediated LINC00261 promoter deletion (Figure 20). Further, *in silico* analysis in LINC00261<sup>high</sup> versus LINC00261<sup>low</sup> tumors confirmed the association with EMT in a very comprehensive way (Figure 19). Intriguingly, the important epithelial marker E-cadherin was robustly downregulated in both LINC00261-depleted cell systems (Figure 21). Notably, it has been shown that a decrease of CDH1 expression can solely be responsible for pancreatic cancer metastasis<sup>173</sup>. Additionally, proteomic analyses revealed that low CDH1 correlated with poor disease outcome<sup>174</sup>.

The role of LINC00261 in EMT has also been described in multiple studies, although different mechanisms have been proposed. For example, Wang et al.<sup>175</sup> investigated the role of LINC00261 in high-grade serous ovarian cancer and demonstrated that LINC00261 interacts with and inhibits miR-552, leading to increased expression of autophagy-related protein 10 (ATG10), which may suppress EMT by regulating SNAI2, TWIST1, N-cadherin, and E-cadherin. Chen et al. also suggested a similar mechanism in the pancreatic cancer cell lines Panc-1 and MiaPaca2, proposing that LINC00261 functions as a ceRNA by sponging miR-552<sup>169</sup>. In contrast to Wang et al., the authors attributed the deregulation of E-cadherin, N-cadherin, and vimentin to the upregulation of forkhead box O3 (FOXO3) via the Wnt pathway<sup>176</sup>. Both studies partially support the findings of the present study. However, it is worth noting that neither ATG10 nor FOXO3 was identified as a commonly deregulated gene in this study, and miR-552 was not expressed in the analyzed PDAC cell line. Also, other studies have suggested that LINC00261 may function as a sponge for miRNAs, as observed in several comprehensive analyses of ceRNA networks in different types of cancer<sup>177–179</sup>. Nevertheless, there is a need for more rigorous experimental evidence to firmly establish the ceRNA function of LINC00261. Li et al.<sup>171</sup> postulated that LINC00261 might be involved in regulating the mTOR-P70S6K1-S6 signaling pathway, thereby regulating E-cadherin, vimentin and MMP2. However, no mechanistic experiments have been conducted to validate this hypothesis.

In order to unravel the LINC00261 functions in the present study, the involvement of the TGF $\beta$  signaling pathway was initially investigated, given its established role in inducing EMT and downregulating CDH1 gene expression. Previous studies have shown that TGF $\beta$  signaling pathway components can be modulated by lncRNA expression, and lncRNAs can regulate TGF $\beta$  signaling.<sup>180–184</sup> In this study, it was demonstrated that TGF $\beta$  stimulation induced a fast and strong downregulation of LINC00261 expression in TGF $\beta$ -responsive

Panc-1 cells undergoing EMT (Figure 22A-D). An additional study has demonstrated that TGF $\beta$  can regulate LINC00261 expression in TGF $\beta$ -responsive lung cancer cell lines<sup>78</sup>. Consequently, it was hypothesized that LINC00261 is involved in regulating a pro-epithelial phenotype, thereby potentially influencing pancreatic cancer differentiation and patient survival. However, differences in SMAD protein activation were not observed in LINC00261-depleted cell systems (Figure 22E, F), indicating that increased activation of the canonical TGF $\beta$  pathway was not responsible for the downregulation of CDH1 in these cells.

Therefore, the expression of other potentially involved pathways and transcription factors was analyzed. Known negative regulators of CDH1 gene expression include SNAI1, SNAI2, TWIST1, and ZEB1<sup>129-133</sup>, while p300, HNF1 $\alpha$ , and FOXA2 are believed to positively regulate CDH1 expression<sup>134</sup>. A previous study has shown that LINC00261 can bind to SNAI2 and promote its degradation in gastric cancer<sup>149</sup>. However, SNAI2 expression is low in Panc-1 cells, similar to TWIST, suggesting that these EMT transcription factors may not play a significant regulatory role in this particular cell system. The transcription factors SNAI1, ZEB1, p300, and HNF1 $\alpha$  were not deregulated in LINC00261-depleted cell systems, indicating that they are not responsible for the strong downregulation of CDH1 in these cells. However, FOXA2, the transcription factor adjacent to LINC00261, exhibited significant downregulation in LINC00261<sup>low</sup> cells, suggesting its potential role in regulating CDH1 expression and the initiation of observed EMT (Figure 23). Indeed, multiple studies have shown that FOXA2 directly regulates CDH1<sup>69,71,134</sup>. On this basis, it was postulated that LINC00261 might indirectly regulate CDH1 expression via its effects on FOXA2. In order to explore this possibility, experiments were conducted to assess the regulation of FOXA2 by LINC00261 and to determine whether this mechanism contributes to the observed changes in PDAC cells.

### 5. The LINC00261-FOXA2 regulatory circuit

Indeed, correlation analysis of the TCGA and Bailey dataset as well as various pancreatic cancer cell lines confirmed a strong positive correlation of FOXA2 and LINC00261 gene expression (Figure 24). In addition, survival analysis revealed that high FOXA2 expression was associated with significantly better overall survival, as previously shown for LINC00261 (Figure 25A). Overexpression and knockdown experiments verified a regulatory circuit between FOXA2 and LINC00261 in both directions. ChIP and luciferase analyses revealed that FOXA2 transcriptionally regulates LINC00261 expression through direct binding to the LINC00261 promoter (Figure 26). These results are supported by studies of lung cancer that have indicated a tight interconnection between these two genes<sup>78,172,185</sup>. The control of FOXA2 expression by LINC00261 has also been observed in lung cancer cells and mouse hepatocytes<sup>161,163,185</sup>. However, the survival analysis of patients with unequal expression of

FOXA2 and LINC00261 suggested that LINC00261 can affect the progression of PDAC independently of FOXA2 regulation (Figure 25B). Consequently, subsequent experiments aimed to provide further insights into the regulatory mechanisms of FOXA2 and LINC00261 and to characterize their individual target genes.

## 6. Interdependencies between LINC00261 and FOXA2

To discover individual target genes, FOXA2 KO cells and FOXA2-LINC00261 double KO cells were established using the CRISPR/Cas9 technique (Figure 27). The expression analysis of these cell lines revealed a reduced FOXA2 protein expression in LINC00261 KO cells as well as a LINC00261 downregulation in FOXA2 KO cells, emphasizing the co-regulation of both genes. The LINC00261 expression was significantly lower in LINC00261-FOXA2 double KO cells compared to single LINC00261 KO cells (Figure 27A, B). Thus, the residual gene expression activity likely enabled by alternative promoters in the LINC00261 promoter region was further diminished by the loss of FOXA2, confirming FOXA2 as a fundamental regulator of LINC00261 gene expression. Interestingly, FOXA2-LINC00261 double KO cells showed a more spindle-like morphology and higher migratory and invasive behavior than single LINC00261 KO cells. In contrast, single FOXA2 KO cells had a round shape, formed clusters, showed weak attachment to the plate, and had a low rate of migration and invasion through the transwell membrane (Figure 27C, Figure 30). These findings are consistent with a previous study by Milan et al.<sup>79</sup> that also used CRISPR/Cas9-mediated genome editing to delete FOXA2 in Panc-1 cells and observed reduced migration capacity and strongly reduced adhesion to lamin of cells lacking FOXA2.

The results of this study indicate that LINC00261, rather than FOXA2, plays a crucial role in the process of EMT in pancreatic cancer cells. This conclusion is based on the observation that downregulation of LINC00261 expression promoted cellular invasiveness and migratory ability, while FOXA2 KO cells exhibited a contrary behavior. Additionally, overexpression of FOXA2 led to a surge in migrating and invading cells, which further supports the opposing roles of LINC00261 and FOXA2 in this cellular process (Figure 31). These findings imply individual functions of both LINC00261 and FOXA2 independent of the respective neighboring gene. Notably, FOXA2 has been implicated in pancreas development<sup>63</sup> and potential tumor suppression<sup>69,70,76</sup>, but controversy exists regarding its potential role as an oncogene in certain cancer types<sup>72,186</sup>.

To identify both individual and mutual functions of the two genes, molecular analysis of WT, LINC00261 KO, FOXA2 KO and LINC00261-FOXA2 double KO cells were performed. Particularly, RNA-seq and RPPA were utilized to determine deregulated genes and pathways.

## 6.1 FOXA2-independent LINC00261 functions

The RNA-seq and GSEA analyses of the newly established cell lines led to the conclusion that the observed EMT signature in LINC00261<sup>low</sup> cells in the first part of the study was exclusively attributed to the loss of LINC00261 (Figure 34). However, the reduced expression of E-cadherin was not found to be responsible for the change in cell morphology and behavior. Pathway analysis indicated that the changes provoked by the loss of LINC00261 may be regulated by the FAK/PXN/c-Src axis (Figure 35 and Figure 36). These signaling molecules, which are crucial for the assembling and remodeling of focal adhesions, have been shown to impact tumor progression, invasion and metastasis of various cancer types<sup>92,96</sup>. This is likely due to the activation of these signaling molecules leading to the initiation of intracellular signaling cascades, resulting in significant changes in gene expression, including genes involved in EMT and cell migration and invasion<sup>97</sup>. The role of activated c-Src signaling in EMT was confirmed in the study by overexpressing c-Src following RNA-seq and GSEA (Figure 37B). c-Src and associated proteins such as FAK and paxillin were repeatedly described as key factors for EMT in various cancer types affecting numerous downstream cascades<sup>92</sup>. The transcriptome analysis revealed several specific target genes of c-Src in Panc-1 cells. To make a direct comparison between the observations in the LINC00261 KO cells and the c-Src overexpression cells, the amount of activated c-Src was titrated to be equivalent in both cell lines (Figure 37A). Interestingly, the impact on cell migration and invasion was similar in both the c-Src overexpression cells and the LINC00261 KO cells, suggesting that the c-Src pathway and associated molecules are responsible for the observed phenotype (Figure 37C). Also, experiments with c-Src inhibitors Dasatinib and PP2 support this hypothesis, as they showed reduced invasion and migration of LINC00261 KO cells upon treatment (Figure 38). c-Src has no intrinsic transcriptional activity, but it has been shown to phosphorylate and activate the transcriptional activity of several transcription factors<sup>95,98-103,108</sup>, which have been demonstrated to activate expression of genes involved in EMT, cell migration, and invasion<sup>100,104-107</sup>. The RNA-seq data was utilized to identify genes involved in EMT that were regulated by both LINC00261 and c-Src, in order to determine transcription factor(s) responsible for the transcriptomic reprogramming observed in LINC00261 KO cells. In this analysis, RUNX1 emerged as a candidate of interest, given its known links to tumor cell metastasis and EMT<sup>100,104,136</sup> (Figure 39A-C). RUNX1 is a master regulator that is overexpressed in various human malignancies and has been associated with a poor prognosis in PDAC<sup>187,188</sup>. Additionally, it has been implicated in several oncogenic processes and signaling pathways, including enhanced cell invasion, migration, and EMT<sup>100,104,105,189</sup>. Furthermore, studies propose that the RUNX1 protein regulates its own gene transcription<sup>190,191</sup>. Interestingly, c-Src has also been shown to phosphorylate several tyrosine residues on RUNX1, leading to increased activity and stability of the transcription factor<sup>101,102</sup>. The involvement of RUNX1 was supported by validation experiments showing

an upregulation of RUNX1 in LINC00261<sup>low</sup> cells generated by the CRISPRi system (Figure 39D). However, treatment with c-Src inhibitors only partially abolished the increased gene expression of RUNX1 in LINC00261 KO cells, indicating that c-Src is not solely responsible for LINC00261-associated RUNX1 regulation (Figure 39E, F). Nevertheless, the elevated FAK/PXN/c-Src signaling in LINC00261 KO cells may have a critical downstream effect on RUNX1 and partly account for the observed transcriptomic reprogramming in these cells. To test this hypothesis, additional experiments should be performed to examine the effects of a downregulation of RUNX1 in LINC00261 KO cells, for example by using small hairpin RNAs (shRNAs) or inhibitors. Furthermore, the role of RUNX1 in EMT in PDAC cells should be investigated by either overexpressing or downregulating RUNX1.

Moreover, to further strengthen the evidence supporting the FOXA2-independent role of LINC00261 and to completely exclude the possible impact of both the FOXA2 protein and the consequences of the intensive manipulation of the genomic region, future experiments should include LINC00261 overexpression and knockdown studies in FOXA2 KO cells. Given the low efficiency observed in initial experiments using siRNA-induced knockdown of the primarily nuclear LINC00261, the utilization of locked nucleic acids<sup>192</sup> may be contemplated to enhance knockdown efficiency.

Mechanistically, the influence of LINC00261 on the FAK/PXN/c-Src signaling pathway, leading to the observed phenotypic changes, remains elusive. To elucidate the molecular interactions, protein pulldown experiments were conducted using five LINC00261 fragments covering the entire RNA sequence and two conserved regions from exon 4 of LINC00261 (Figure 41A, B). To date, the analysis by mass spectrometry used to identify LINC00261-binding proteins has been limited to the prominent bands within conserved region 1 of LINC00261 (see Figure 12, red box B). Subsequent western blot analyses, comprising both LINC00261 fragments and control samples, have been conducted to evaluate the candidates of interest. Notably, the initial findings for the two chosen candidates, EphA2 and IGF2BP1, have demonstrated their binding affinity not only to the LINC00261 fragments but also to unrelated GFP mRNA (Figure 41C). This observation implies a broad-spectrum RNA-binding capability of EphA2 and IGF2BP1, raising questions about the relevance of the discovered interactions with LINC00261. Unlike IGF2BP1, EphA2, according to the comprehensive RNA-binding protein database RBP2GO<sup>193</sup>, has not yet been documented as an RNA-binding protein, necessitating further investigations in this context. At a mechanistic level, LINC00261 may bind to EphA2, thereby inhibiting the activation of c-Src, which is proposed to be activated by interacting with the phosphorylated juxtamembrane region of Eph receptors via its SH2 domain<sup>138,139</sup>. In the context of IGF2BP1, a plausible mechanism includes the binding of LINC00261 to IGF2BP1, preventing the subsequent activation of the c-Src signaling cascade by IGF2BP1, which was demonstrated in a recent publication by Bley et al.<sup>141</sup>. To prove these hypotheses, it is

recommended to conduct IP experiments using EphA2/IGF2BP1 antibodies, followed by qPCR analysis. To gain a more comprehensive understanding of LINC00261-binding proteins, all proteins captured in the pulldown experiment should undergo mass spectrometry in future investigations. Additionally, the inclusion of the control samples in the mass spectrometric assessment is necessary to identify and exclude nonspecific binding events. To comprehensively encompass the entire LINC00261 molecule, all five fragments of LINC00261 should undergo a thorough analysis. This approach will help to elucidate whether LINC00261 influences the FAK/c-Src/PXN signaling pathway by binding to one or several key players within this pathway.

## 6.2 Mutual LINC00261 and FOXA2 functions

The next analyzed category, "FOXA2-dependent LINC00261 target genes", encompasses both genes regulated exclusively by FOXA2 through indirect regulation and genes regulated by both LINC00261 and FOXA2. Notably, CDH1 was classified under FOXA2-dependent gene regulation as it was strongly downregulated in all FOXA2 KO, LINC00261 KO, and double KO cells (Figure 40A, B). Vice versa, CDH1 was upregulated by both FOXA2 and LINC00261 in PATU-T cells (Figure 40C, D). Liu et al.<sup>134</sup> and Bow et al.<sup>71</sup> demonstrated that FOXA2 can directly bind to the CDH1 promoter, providing an explanation for CDH1 regulation without the involvement of the classical EMT transcription factors. Indeed, CHIP-qRT-PCR and luciferase assay confirmed the binding of FOXA2 to the promoter and the resulting activation of CDH1 gene expression. Moreover, TGF $\beta$  treatment decreased the binding of FOXA2 to the CDH1 promoter, leading to a reduction in gene expression, possibly as a consequence of FOXA2 downregulation by TGF $\beta$ , as demonstrated herein and by others<sup>194,195</sup> (Figure 40E). Although CDH1 regulation by FOXA2 was confirmed, overexpression experiments in LINC00261 KO cells suggest that LINC00261 must be present in the cells for FOXA2 to exert its full regulatory capacity (Figure 40F). Previous studies have indicated that CDH1 expression is highly dependent on the methylation status of its promoter<sup>196,197</sup>, suggesting a potential epigenetic cooperation between LINC00261 and FOXA2. However, FOXA2 was not detected in the mass spectrometry analysis of LINC00261-binding proteins conducted in this study, and the western blot analysis of FOXA2 in this experiment is still pending. Additionally, preliminary IP experiments using FOXA2 antibodies have not provided significant evidence of LINC00261 binding to FOXA2 (data not shown), raising questions about a direct interaction between the two. Additional experiments are essential to comprehensively assess the potential interaction.



### 6.3 LINC00261-independent FOXA2 functions

The effects of FOXA2, independent of LINC00261, were further examined by evaluating the differences in gene expression between WT and FOXA2 KO cells and subsequently validating the targets of interest through overexpression of FOXA2 in the same cells. Curiously, GSEA of the genes deregulated in FOXA2 KO cells showed that the hallmark of EMT was highly negatively enriched (Figure 42), which was surprising given the fact that FOXA2 has been proposed as a tumor suppressor by inhibiting EMT in several cancer types<sup>69,70,76,198</sup>. Another study, however, stated that FOXA2 promotes esophageal squamous cell carcinoma progression by activation of the EMT-inducer ZEB2<sup>73</sup>. In these studies, a variety of markers including upregulation of N-cadherin, vimentin, SNAI1, MMP2 or ZEB2 and downregulation of E-cadherin or tight junction protein 1 (TJP1/ZO-1) have been used to assess EMT. Especially, the reduction of E-cadherin (CDH1), which was also demonstrated in this study, was considered as evidence for EMT. However, the observed contrary migration and invasion behavior and EMT signature of LINC00261 KO and FOXA2 KO cells in the present study, despite the loss of CDH1 in both cell lines, suggest that the reduction of CDH1 is not a reliable indication of EMT. The RNA-seq dataset showed that N-cadherin (CDH2) and vimentin were also significantly deregulated after deletion of FOXA2, but both were downregulated, indicating negative regulation of EMT upon FOXA2 depletion. This observation once again contrasts with the positive regulation of EMT in LINC00261 KO cells. It aligns with the opposite behavior observed in migration and invasion assays, suggesting that the cells' capability to invade or migrate across the transwell membrane in the *in vitro* assay is linked to their EMT signature.

In addition to the EMT gene set, the TGF $\beta$  hallmark gene set was found to be negatively enriched, and several deregulated genes related to the TGF $\beta$ /SMAD signaling and EMT were identified (Figure 43). This effect was solely dependent on FOXA2, as it was not observed after depletion of LINC00261. In a previous study by Milan et al.<sup>79</sup>, ChIP-seq of FOXA2 was performed in Panc-1 cells, and gene ontology (GO) terms associated with the ChIP-peaks were analyzed using the genomic regions enrichment of annotations tool (GREAT), which calculates the enrichment of GO terms in a set of genomic regions based on the weighted distance between peaks and genes with annotated functions<sup>199</sup>. Interestingly, the most enriched category linked to the ChIP-peaks was the TGF $\beta$  receptor signaling pathway (GO:0007179)<sup>79</sup>. Moreover, FOXA2 depletion in Panc-1 cells was also carried out by Milan et al. using CRISPR/Cas9-mediated genome editing, and RNA-seq analysis was performed. The expression of TGF $\beta$  signaling-related genes, including TGFB2, TGFBR2, TGFBI, and CDH2, which were strongly downregulated in the FOXA2 KO cells of the present study, was also downregulated in their dataset, albeit not to the same extent<sup>79</sup>. Another study by Lee et al. also demonstrated a decrease in TGFBI expression upon FOXA2 siRNA treatment in lung adenocarcinoma<sup>200</sup>. Notably, the luciferase assays, which evaluated the activity of the

SMAD binding element, and qRT-PCR analysis of TGF $\beta$ -inducible genes, showed that the induction of the canonical TGF $\beta$  signaling by TGF $\beta$  was significantly impaired in FOXA2-deficient cells (Figure 44). The evaluation of the expression and phosphorylation of SMAD2 and SMAD3 proteins also confirmed the reduced activity of the signaling in FOXA2 KO cells (Figure 45A). The GSEA transcription factor analysis indicated that specifically genes regulated by SMAD3 were affected (Figure 45B). This is a plausible observation given that the activation of SMAD3 in response to TGF $\beta$  treatment was entirely abolished in FOXA2 KO cells. Interestingly, SMAD3 itself is a regulator of FOXA2, as it directly binds to the FOXA2 promoter and controls its transcriptional activity<sup>195</sup>. Overexpression of FOXA2 in Panc-1 WT cells confirmed its positive regulation of canonical TGF $\beta$  signaling at basal levels (Figure 46). However, the lack of increased induction upon TGF $\beta$  treatment in cells overexpressing FOXA2 implied that the cellular amount of FOXA2 was already sufficient for full capacity TGF $\beta$ /SMAD signaling. Subsequent rescue experiments provided evidence of the influence of FOXA2 on canonical TGF $\beta$  signaling, as the inducibility of SBE reporter activity and SMAD protein phosphorylation upon TGF $\beta$  treatment were restored (Figure 47). This strong impact of FOXA2 on TGF $\beta$  signaling may be attributed to its robust regulation of TGFBR2 (Figure 43B and Figure 46B), which is crucial for transmitting the signal from the plasma membrane to the cell interior<sup>93</sup>. This possibility could be explored in the future by either overexpressing TGFBR2 in FOXA2 KO cells or inhibiting TGFBR2 in FOXA2-overexpressing or rescuing cells. However, TGFBR2 downregulation alone cannot fully account for these findings since SMAD2 activation was still possible in FOXA2 KO cells, whereas SMAD3 activation was completely abolished. Furthermore, GSEA analysis revealed explicit downregulation of SMAD3 targets, indicating that additional factors are involved. Intriguingly, Minoo et al.<sup>201</sup> demonstrated that SMAD3 binds to the winged helix DNA-binding domain (DBD) of FOXA1, which shares 93% amino acid homology with the DBD of FOXA2 and binds to the same DNA consensus sequences<sup>202</sup>. They proposed a mechanism in which TGF $\beta$  activates SMAD3, which then binds to FOXA1 and prevents it from binding to FOXA1-controlled promoters<sup>201</sup>. A regulatory mechanism involving the interaction between FOXA2 and SMAD3, leading to the activation of SMAD3 target gene expression, may account for the significant impact of FOXA2 loss on the expression of SMAD3 target genes. This, however, necessitates further investigation.

In this study, a novel finding was the downregulation of exogenous FOXA2 protein expression by TGF $\beta$ , indicating a post-transcriptional or -translational mechanism of regulation (Figure 48). However, to rule out any potential impacts of TGF $\beta$  on the CMV promoter responsible for driving the expression of exogenous FOXA2, an additional control should be incorporated in future experiments. One suitable control option could involve using the same CMV-driven vector encoding an easily detectable, externally derived protein that remains unaffected by TGF $\beta$  regulation. The downregulation of FOXA2 may be part of a negative feedback mechanism, as FOXA2 was found to upregulate TGF $\beta$ 1 (and TGF $\beta$ 2)

expression in the present study. TGF $\beta$ -regulated miRNAs targeting FOXA2 mRNA may be responsible for the post-transcriptional regulation. TGF $\beta$  regulates several miRNAs, and conversely, most members of the TGF $\beta$  pathway are targeted by one or more miRNAs, suggesting the possibility of their involvement in the regulation of FOXA2 mRNA expression<sup>203,204</sup>. However, only few miRNAs have been shown to regulate the expression of FOXA2 mRNA<sup>205–207</sup>. Interestingly, Chen et al. have demonstrated that miR200a, a miRNA known to be regulated by TGF $\beta$ , regulates FOXA2 expression<sup>206</sup>. Certainly, additional research is required to elucidate the role of miRNAs in the TGF $\beta$ -mediated regulation of FOXA2 expression. Moreover, FOXA2 has been shown to be post-translational modified by phosphorylation, sumoylation, acetylation and ubiquitination, which may influence its protein stability and thereby regulate FOXA2 protein expression<sup>208–210</sup>. However, the relationship between TGF $\beta$  signaling and post-translational modification of FOXA2 has not been explored thus far. The regulation of FOXA2 might also involve RNA binding. Despite FOXA2 not being traditionally categorized as an RNA-binding protein, a proteome-wide quantitative analysis of RNA-dependent protein complexes, identified through density gradient ultracentrifugation followed by mass spectrometry<sup>211</sup>, suggests a potential RNA-dependent regulatory role for FOXA2. The analysis revealed an RNA-dependent shift in FOXA2, resembling the behavior of well-established RNA-binding proteins such as IGF2BP1 and other chromatin-associated factors like CCCTC-Binding Factor (CTCF). Notably, this behavior appears to be specific to FOXA2, as other cancer-relevant transcription factors, like SMAD3, STAT3, and Krüppel-like factor 4 (KLF4), did not exhibit an RNA-dependent shift<sup>211</sup>. In the future, further experiments, such as IP experiments with FOXA2 followed by RNA sequencing, should be conducted to identify specific RNAs binding to FOXA2.

## 7. The role of LINC00261 and FOXA2 for *in vivo* tumor growth and metastasis

In order to further investigate the biological and therapeutic significance of LINC00261 and FOXA2 for PDAC growth and metastasis, *in vivo* xenograft mouse models were utilized for the eight established cell lines. Tumors originating from FOXA2 KO cells were found to grow faster, while those originating from LINC00261 KO and LINC00261-FOXA2 double KO cells grew more slowly than WT tumors (Figure 49), contradicting the *in vitro* results that showed reduced proliferation of FOXA2 KO cells and no alterations of LINC00261 KO and LINC00261-FOXA2 DKO cells. This discrepancy between the *in vitro* and *in vivo* outcomes implies that factors such as cell-cell interactions within the tumor, as well as interactions with the tumor microenvironment, have a crucial impact on tumor growth. In accordance with tumor growth, according to our IVIS evaluation and HE staining, the strongest metastasis occurred in tumor models originating from FOXA2 KO cells, while the lowest metastasis occurred in those deriving from LINC00261 KO and DKO cells (Figure 51 and Figure 52). To improve the quantification of the metastases, additional staining of the

affected tissues should be conducted. Specifically, anti-GFP staining would be valuable for visualizing metastases in the lung and liver, as tumor cells have been labeled with GFP in addition to luciferase. Furthermore, staining with pan-cytokeratin (pan-CK) enables precise visualization and identification of tumor cells, further supporting a comprehensive assessment of metastasis<sup>212</sup>. Interestingly, colony formation/clonogenic assays revealed that FOXA2 KO cells formed fewer but larger and more well-defined colonies compared to WT cells, whereas LINC00261 KO and DKO cells formed extensive colonies with a larger area but less defined than WT cells and FOXA2 KO cells (Figure 29). These observations indicate that FOXA2 KO cells have a higher ability to form spheres than LINC00261 KO and LINC00261-FOXA2 DKO cells. These findings were further supported by 3D sphere experiments, where neither Panc-1 WT cells nor LINC00261 KO and DKO cells were able to form spheres in 3D culture, whereas FOXA2 KO cells were capable of doing so (data not shown). Thus, the observed differences in the xenograft model may arise from distinct abilities to form tumors and metastases, rather than from differences in cell proliferation and invasion. Notably, clonogenic and sphere-forming assays are commonly employed to identify cancer stem cells (CSCs), which are subpopulations of cancer cells that exhibit stem-like properties, such as self-renewal and the ability to differentiate into multiple developmental lineages. These cells play a critical role in promoting tumor growth and heterogeneity<sup>213-216</sup>. Although the observation that only FOXA2 KO cells exhibit enhanced tumor growth may appear inconsistent with the lack of a similar effect in LINC00261-FOXA2 double KO cells, it should be acknowledged that the double KO cells were generated from LINC00261 single KO cells, which had already undergone transcriptomic reprogramming. Previous research has suggested that the function of the transcription factor FOXA2 in PDAC is dependent on the differentiation grade of the cancer cells, possibly due to its varying genomic distribution that controls distinct gene expression programs in partnership with other transcription factors<sup>79</sup>. As previously mentioned, deletion of LINC00261 resulted in notable modifications of cell features, including changes in the expression of multiple transcription factors, leading to alterations in the transcriptomic profile. Moreover, the expression of FOXA2 was affected in LINC00261 KO cells, resulting in a distinct basal expression level prior to FOXA2 depletion, which could have contributed to the varying outcome observed in the LINC00261-FOXA2 double KO cells in comparison to the FOXA2 single KO cells. To elucidate the precise impact of the consecutive deletion of LINC00261 and FOXA2, it would be helpful to generate double KO cells in the reverse order, wherein FOXA2 KO precedes LINC00261 deletion. Additionally, it's worth noting that techniques enabling the simultaneous knockout of multiple genes do exist, although they are predominantly reliant on viral CRISPR/Cas9 systems<sup>217</sup>. Interestingly, the protein expression of the pancreatic CSC marker CD44<sup>218</sup> was significantly upregulated in FOXA2 KO cells, potentially accounting for the observed heightened tumor growth and metastasis *in vivo* (Supplementary Table S10, western blot not shown). FOXA2 has been reported to

play a crucial role in pancreatic cell differentiation, and the loss of FOXA2 may prompt the re-acquisition of stem-cell-like properties, facilitating tumor growth and metastasis<sup>63</sup>. However, the expression of other pancreatic cancer cell markers, such as CD24 and EPCAM<sup>218</sup>, was downregulated in FOXA2 KO cells, indicating that this aspect requires further investigation (data not shown). Additionally, GSEA highlighted hallmark pathways, including oxidative phosphorylation, KRAS signaling, and MYC proto-oncogene targets (Figure 42A) that may contribute to the *in vivo* phenotype of FOXA2 KO cells and should be explored in future studies. Overall, the *in vivo* behavior of the cells contrasted with *in vitro* findings, emphasizing the significance of tumor- and metastasis-forming abilities for disease progression.

The *in vivo* results are in agreement with survival analyses of PDAC patients with respect to FOXA2, which revealed that decreased FOXA2 expression correlated with a poorer prognosis. Similarly, decreased LINC00261 expression was associated with an aggressive squamous subtype and poorer prognosis. In contrast, reduced tumor growth and metastasis were observed in the LINC00261 KO cells compared to the WT cells in the *in vivo* model, leading to the hypothesis that the reduced tumor formation capability is responsible for this observation. Nonetheless, it should be noted that the experiment was terminated after six weeks, and further monitoring of tumor growth and metastasis was not performed. Tumor progression may have accelerated once a certain size was reached. Furthermore, apart from the FOXA2 KO tumors, the metastatic potential of the tumors was substantially low. Thus, the time frame for assessing disparities between the WT, LINC00261 KO, or DKO tumors was suboptimal. In prospective investigations, *in vivo* passaging may be considered as a strategy to enhance tumor metastasis, a method previously employed by Metildi et al.<sup>219</sup>. In addition, it should be mentioned that the patients from whom survival data was collected underwent radiotherapy and/or chemotherapy. Consequently, the reduced survival of patients with low LINC00261 expression could be attributed to the cancer cells' resistance to therapy. In fact, several studies have found an association between EMT and therapy resistance in pancreatic cancer<sup>220-222</sup>. Hence, further investigation is needed to determine whether the EMT signature in LINC00261 KO and LINC00261-FOXA2 DKO cells could contribute to therapy resistance.

## 8. The impact of EMT for *in vivo* tumor growth and metastasis

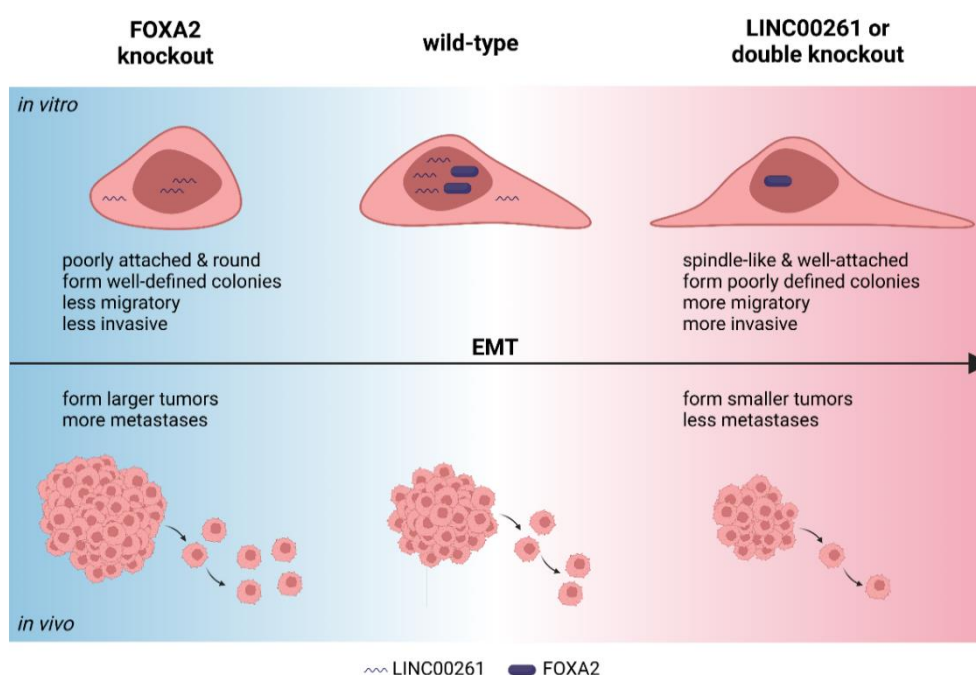
A positive regulation of EMT was observed in LINC00261 KO and LINC00261-FOXA2 double KO cells, but a negative regulation of EMT was observed in FOXA2 KO cells. As discussed above, the upregulation of c-Src signaling in cells with reduced LINC00261 expression may account for the EMT signature through transcriptomic changes implicated in EMT. This EMT phenotype was demonstrated *in vitro* by spindle-like cell morphology, and migratory

and invasive behavior. However, *in vivo* xenografts of these cells indicate that this phenotype might not be responsible for tumor metastasis. Instead, FOXA2 KO cells, which exhibited low migratory and invasive abilities *in vitro* and showed a reduced expression of EMT-related genes, developed large tumors and metastasized to the liver and lung. This observation contrasts with several studies demonstrating that EMT promotes metastasis of cancer cells<sup>223,224</sup>. However, other studies have shown that tumor cells with an epithelial phenotype are more likely to survive in the circulation and form distant metastases<sup>225-227</sup>, which is consistent with the findings in this study. Overall, the results suggest that an epithelial rather than mesenchymal cell phenotype promotes metastasis. Further studies are necessary to determine whether the epithelial signature or stemness of FOXA2 KO cells is responsible for increased tumor progression. Furthermore, it is necessary to substantiate the *in vivo* findings obtained from the orthotopic model through validation with alternative models, e.g. by using genetically modified mouse models (GEMMs), such as the KPC mouse model for the induction of PDAC<sup>228</sup>. In the context of our study, these mice should additionally have genetic alterations in LINC00261, FOXA2, and the combination of both.

## VI. SUMMARY

Pancreatic cancer has a dismal prognosis due to late diagnosis, frequent local and distant metastasis and high degree of resistance to therapies. Since less than 20% of pancreatic ductal adenocarcinomas (PDACs) are resectable at time of diagnosis, it is essential to identify molecular mechanisms of tumorigenesis in order to understand possible pathways that are responsible for progression to invasive cancer. Integrative genomic analyses in PDAC provided valuable insights into pancreatic carcinogenesis and identified different disease subtypes that have prognostic and biological relevance and are related to differences in therapy response. Bioinformatic analysis identified long non-coding RNAs (lncRNAs) to be associated with these subtypes suggesting a subtype-specific expression and hypothetical function of these lncRNAs in PDAC. Analysis of publicly available datasets and differential lncRNA expression in normal pancreas versus PDAC tissue, as well as in different molecular subtypes, has identified LINC00261 as a downregulated lncRNA in PDAC, particularly in the squamous subtype of PDAC. This subtype has the worst prognosis and is characterized by a gene signature related to epithelial-to-mesenchymal transition (EMT). Consistently, LINC00261 expression was inversely correlated with disease stage, grade and patient survival, and bioinformatic analysis of LINC00261<sup>high</sup> vs. LINC00261<sup>low</sup> PDAC samples revealed an enrichment of genes related to EMT in LINC00261<sup>low</sup> tumors. Furthermore, CRISPR-mediated knockdown and promoter knockout of LINC00261 induced an EMT-related transcription program that enhanced cancer cell invasion and migration and decreased expression of the epithelial marker CDH1. Although the EMT-inducer TGF $\beta$  downregulated LINC00261, it was discovered that the canonical TGF $\beta$  signaling pathway was not responsible for EMT in LINC00261<sup>low</sup> cells. Instead, a regulatory circuit between LINC00261 and its genomic neighbor, the transcription factor FOXA2, has been found, and in consequence, FOXA2 KO and FOXA2-LINC00261 double KO cell clones were established. Through the analysis of the newly established cell lines in both *in vitro* and *in vivo* settings, a complex regulatory network between LINC00261 and FOXA2 has been revealed. Interestingly, both the *in vitro* and the *in vivo* behavior of FOXA2 KO cells were opposite to those of LINC00261 KO or double KO cells, uncovering a partial independence of both factors in regulating cellular behavior. The depletion of LINC00261 resulted in highly migratory and invasive cells with an EMT gene expression signature, but with reduced ability to grow and metastasize in *in vivo* xenograft models. This observation contradicted the findings from survival analyses, which indicated a diminished survival rate among patients exhibiting low LINC00261 expression. The possibility of tumor resistance to therapy as a potential explanation for this inconsistency should be investigated in future studies. Pathway analyses suggested that the elevated FAK/c-Src/PXN axis may be responsible for the observed EMT phenotype in LINC00261 KO cells by leading to transcriptomic reprogramming, supposedly partially driven by the transcription factor

RUNX1. Notably, loss of CDH1, which has been shown to depend on both LINC00261 and FOXA2, has been disproved as a driver of EMT. On the other hand, the deletion of FOXA2 resulted in cells with a low ability to migrate and invade in *in vitro* assays, combined with a negatively regulated EMT signature, but an aggressive tumor growth and metastasis *in vivo*. These findings indicate that tumor growth and metastasis of pancreatic cancer cells *in vivo* are driven by an epithelial cell signature and/or other factors, such as stemness-like features, rather than a mesenchymal phenotype. Additionally, this study has revealed the significance of FOXA2 as a crucial positive regulator of the canonical TGF $\beta$  pathway, alongside its involvement in a regulatory negative feedback loop wherein FOXA2 is downregulated by TGF $\beta$  at a post-transcriptional or post-translational level. Overall, this project has further unraveled the complicated regulatory network of LINC00261 and FOXA2 and has shed light on novel, independent functions of both molecules. These findings provide valuable insights into the underlying molecular mechanisms driving the metastasis of PDAC. Nevertheless, given the complex nature of the regulatory network between LINC00261 and FOXA2, there are still unanswered questions that warrant further investigation in future studies. Addressing these will contribute to a more comprehensive understanding of the roles played by LINC00261 and FOXA2 in pancreatic cancer metastasis.



**Figure 53: Overview of effects of FOXA2 and/or LINC00261 depletion in PDAC cells.**

Depletion of LINC00261 (and depletion of both LINC00261 and FOXA2) in PDAC cells led to highly migratory and invasive cells with an EMT gene expression signature, but reduced ability to grow and metastasize *in vivo*. Conversely, deletion of FOXA2 resulted in low migration and invasion *in vitro* but aggressive tumor growth and metastasis *in vivo*. The findings of this study indicate that an epithelial cell phenotype promotes tumor growth and metastasis, while a mesenchymal cell phenotype exerts an inhibitory effect on these processes. The figure was created with BioRender.com.



---

## VII. REFERENCES

1. Karpińska, M. & Czauderna, M. Pancreas-Its Functions, Disorders, and Physiological Impact on the Mammals' Organism. *Frontiers in physiology* **13**, 807632; 10.3389/fphys.2022.807632 (2022).
2. Betts, J. G. *et al.* *Anatomy and Physiology* (Rice University, 2015).
3. Pan, F. C. & Brissova, M. Pancreas development in humans. *Current opinion in endocrinology, diabetes, and obesity* **21**, 77–82; 10.1097/MED.0000000000000047 (2014).
4. Sung, H. *et al.* Global Cancer Statistics 2020. GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a cancer journal for clinicians* **71**, 209–249; 10.3322/caac.21660 (2021).
5. Mathers, C. D., Fat, D. M., Inoue, M., Rao, C. & Lopez, A. D. Counting the dead and what they died from. An assessment of the global status of cause of death data. *Bulletin of the World Health Organization* **83**, 171–177 (2005).
6. Lynch, S. M. *et al.* Cigarette smoking and pancreatic cancer. A pooled analysis from the pancreatic cancer cohort consortium. *American journal of epidemiology* **170**, 403–413; 10.1093/aje/kwp134 (2009).
7. Larsson, S. C., Orsini, N. & Wolk, A. Body mass index and pancreatic cancer risk. A meta-analysis of prospective studies. *International journal of cancer* **120**, 1993–1998; 10.1002/ijc.22535 (2007).
8. Huxley, R., Ansary-Moghaddam, A., Berrington de González, A., Barzi, F. & Woodward, M. Type-II diabetes and pancreatic cancer. A meta-analysis of 36 studies. *British journal of cancer* **92**, 2076–2083; 10.1038/sj.bjc.6602619 (2005).
9. Park, W., Chawla, A. & O'Reilly, E. M. Pancreatic Cancer. A Review. *JAMA* **326**, 851–862; 10.1001/jama.2021.13027 (2021).
10. Kleeff, J. *et al.* Pancreatic cancer. *Nature reviews. Disease primers* **2**, 16022; 10.1038/nrdp.2016.22 (2016).
11. Hosoda, W. & Wood, L. D. Molecular Genetics of Pancreatic Neoplasms. *Surgical pathology clinics* **9**, 685–703; 10.1016/j.path.2016.05.011 (2016).
12. *AJCC cancer staging handbook. From the AJCC cancer staging manual 6.ed.* 7th ed. (Springer, New York, 2010).
13. Amin, M. B. *et al.* The Eighth Edition AJCC Cancer Staging Manual. Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA: a cancer journal for clinicians* **67**, 93–99; 10.3322/caac.21388 (2017).
14. Öberg, K., Knigge, U., Kwekkeboom, D. & Perren, A. Neuroendocrine gastro-entero-pancreatic tumors. ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of*

- oncology : official journal of the European Society for Medical Oncology* **23 Suppl 7**, vii124-30; 10.1093/annonc/mds295 (2012).
15. Bailey, P. *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47–52; 10.1038/nature16965 (2016).
  16. Moffitt, R. A. *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nature genetics* **47**, 1168–1178; 10.1038/ng.3398 (2015).
  17. Collisson, E. A. *et al.* Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nature medicine* **17**, 500–503; 10.1038/nm.2344 (2011).
  18. Collisson, E. A., Bailey, P., Chang, D. K. & Biankin, A. V. Molecular subtypes of pancreatic cancer. *Nat Rev Gastroenterol Hepatol* **16**, 207–220; 10.1038/s41575-019-0109-y (2019).
  19. Rahib, L. *et al.* Projecting cancer incidence and deaths to 2030. The unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer research* **74**, 2913–2921; 10.1158/0008-5472.CAN-14-0155 (2014).
  20. Sally, Á., McGowan, R., Finn, K. & Moran, B. M. Current and Future Therapies for Pancreatic Ductal Adenocarcinoma. *Cancers* **14**; 10.3390/cancers14102417 (2022).
  21. Hangauer, M. J., Vaughn, I. W. & McManus, M. T. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS genetics* **9**, e1003569; 10.1371/journal.pgen.1003569 (2013).
  22. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108; 10.1038/nature11233 (2012).
  23. Palazzo, A. F. & Lee, E. S. Non-coding RNA. What is functional and what is junk? *Frontiers in genetics* **6**, 2; 10.3389/fgene.2015.00002 (2015).
  24. Cobb, M. Who discovered messenger RNA? *Current biology : CB* **25**, R526-32; 10.1016/j.cub.2015.05.032 (2015).
  25. Lodish, H. *Molecular cell biology*. 4th ed. (Freeman, New York, NY, 2002).
  26. RajBhandary, U. L. & Köhrer, C. Early days of tRNA research. Discovery, function, purification and sequence analysis. *Journal of biosciences* **31**, 439–451; 10.1007/BF02705183 (2006).
  27. Weinberg, R. A. & Penman, S. Small molecular weight monodisperse nuclear RNA. *Journal of Molecular Biology* **38**, 289–304; 10.1016/0022-2836(68)90387-2 (1968).
  28. Maxwell, E. S. & Fournier, M. J. The small nucleolar RNAs. *Annual review of biochemistry* **64**, 897–934; 10.1146/annurev.bi.64.070195.004341 (1995).
  29. Jonas, S. & Izaurralde, E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nature reviews. Genetics* **16**, 421–433; 10.1038/nrg3965 (2015).

- 
30. Weick, E.-M. & Miska, E. A. piRNAs. From biogenesis to function. *Development (Cambridge, England)* **141**, 3458–3471; 10.1242/dev.094037 (2014).
31. Kristensen, L. S. *et al.* The biogenesis, biology and characterization of circular RNAs. *Nature reviews. Genetics* **20**, 675–691; 10.1038/s41576-019-0158-7 (2019).
32. Frankish, A. *et al.* GENCODE 2021. *Nucleic acids research* **49**, D916–D923; 10.1093/nar/gkaa1087 (2021).
33. Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F. & Mattick, J. S. Specific expression of long noncoding RNAs in the mouse brain. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 716–721; 10.1073/pnas.0706729105 (2008).
34. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & development* **25**, 1915–1927; 10.1101/gad.17446611 (2011).
35. Palazzo, A. F. & Koonin, E. V. Functional Long Non-coding RNAs Evolve from Junk Transcripts. *Cell* **183**, 1151–1161; 10.1016/j.cell.2020.09.047 (2020).
36. Ponting, C. P. Biological function in the twilight zone of sequence conservation. *BMC Biol* **15**, 85; 10.1186/s12915-017-0411-5 (2017).
37. Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H. & Bartel, D. P. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**, 1537–1550; 10.1016/j.cell.2011.11.055 (2011).
38. Diederichs, S. The four dimensions of noncoding RNA conservation. *Trends in genetics : TIG* **30**, 121–123; 10.1016/j.tig.2014.01.004 (2014).
39. Ma, L., Bajic, V. B. & Zhang, Z. On the classification of long non-coding RNAs. *RNA biology* **10**, 925–933; 10.4161/rna.24604 (2013).
40. Dahariya, S. *et al.* Long non-coding RNA. Classification, biogenesis and functions in blood cells. *Molecular immunology* **112**, 82–92; 10.1016/j.molimm.2019.04.011 (2019).
41. Gutschner, T. & Diederichs, S. The hallmarks of cancer. *RNA biology* **9**, 703–719; 10.4161/rna.20481 (2014).
42. Wang, K. C. & Chang, H. Y. Molecular mechanisms of long noncoding RNAs. *Molecular cell* **43**, 904–914; 10.1016/j.molcel.2011.08.018 (2011).
43. Gupta, R. A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076; 10.1038/nature08975 (2010).
44. Jain, A. K. *et al.* LncPRESS1 Is a p53-Regulated LncRNA that Safeguards Pluripotency by Disrupting SIRT6-Mediated De-acetylation of Histone H3K56. *Molecular cell* **64**, 967–981; 10.1016/j.molcel.2016.10.039 (2016).

- 
45. Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. & Akoulitchev, A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* **445**, 666–670; 10.1038/nature05519 (2007).
  46. Li, S. *et al.* Long non-coding RNA metastasis-associated lung adenocarcinoma transcript 1 promotes lung adenocarcinoma by directly interacting with specificity protein 1. *Cancer science* **109**, 1346–1356; 10.1111/cas.13587 (2018).
  47. Beltran, M. *et al.* A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes & development* **22**, 756–769; 10.1101/gad.455708 (2008).
  48. Wang, J. *et al.* CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. *Nucleic acids research* **38**, 5366–5383; 10.1093/nar/gkq285 (2010).
  49. Yuan, J. *et al.* A long noncoding RNA activated by TGF- $\beta$  promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer Cell* **25**, 666–681; 10.1016/j.ccr.2014.03.010 (2014).
  50. Qi, W. *et al.* LncRNA GABPB1-AS1 and GABPB1 regulate oxidative stress during erastin-induced ferroptosis in HepG2 hepatocellular carcinoma cells. *Scientific reports* **9**, 16185; 10.1038/s41598-019-52837-8 (2019).
  51. Willingham, A. T. *et al.* A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science (New York, N.Y.)* **309**, 1570–1573; 10.1126/science.1115901 (2005).
  52. Clemson, C. M. *et al.* An architectural role for a nuclear noncoding RNA. NEAT1 RNA is essential for the structure of paraspeckles. *Molecular cell* **33**, 717–726; 10.1016/j.molcel.2009.01.026 (2009).
  53. Gil, N. & Ulitsky, I. Regulation of gene expression by cis-acting long non-coding RNAs. *Nature reviews. Genetics* **21**, 102–117; 10.1038/s41576-019-0184-5 (2020).
  54. Gutschner, T., Richtig, G., Haemmerle, M. & Pichler, M. From biomarkers to therapeutic targets—the promises and perils of long non-coding RNAs in cancer. *Cancer metastasis reviews* **37**, 83–105; 10.1007/s10555-017-9718-5 (2018).
  55. Li, Y., Al Hallak, M. N., Philip, P. A., Azmi, A. S. & Mohammad, R. M. Non-Coding RNAs in Pancreatic Cancer Diagnostics and Therapy. Focus on lncRNAs, circRNAs, and piRNAs. *Cancers* **13**; 10.3390/cancers13164161 (2021).
  56. Raphael, B. J. *et al.* Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* **32**, 185–203.e13; 10.1016/j.ccell.2017.07.007 (2017).
  57. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome research* **12**, 996–1006; 10.1101/gr.229102 (2002).

- 
58. Dorn, A. *et al.* LINC00261 Is Differentially Expressed in Pancreatic Cancer Subtypes and Regulates a Pro-Epithelial Cell Identity. *Cancers* **12**; 10.3390/cancers12051227 (2020).
59. Jiang, W., Liu, Y., Liu, R., Zhang, K. & Zhang, Y. The lncRNA DEANR1 facilitates human endoderm differentiation by activating FOXA2 expression. *Cell reports* **11**, 137–148; 10.1016/j.celrep.2015.03.008 (2015).
60. Zhang, M. *et al.* LINC00261: a burgeoning long noncoding RNA related to cancer. *Cancer cell international* **21**, 274; 10.1186/s12935-021-01988-8 (2021).
61. Ang, S. L. *et al.* The formation and maintenance of the definitive endoderm lineage in the mouse. Involvement of HNF3/forkhead proteins. *Development (Cambridge, England)* **119**, 1301–1315; 10.1242/dev.119.4.1301 (1993).
62. Burtscher, I. & Lickert, H. Foxa2 regulates polarity and epithelialization in the endoderm germ layer of the mouse embryo. *Development (Cambridge, England)* **136**, 1029–1038; 10.1242/dev.028415 (2009).
63. Lee, K. *et al.* FOXA2 Is Required for Enhancer Priming during Pancreatic Differentiation. *Cell reports* **28**, 382–393.e7; 10.1016/j.celrep.2019.06.034 (2019).
64. Gao, N. *et al.* Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes & development* **22**, 3435–3448; 10.1101/gad.1752608 (2008).
65. Rojas, A., Schachterle, W., Xu, S.-M., Martín, F. & Black, B. L. Direct transcriptional regulation of Gata4 during early endoderm specification is controlled by FoxA2 binding to an intronic enhancer. *Developmental biology* **346**, 346–355; 10.1016/j.ydbio.2010.07.032 (2010).
66. Dassaye, R., Naidoo, S. & Cerf, M. E. Transcription factor regulation of pancreatic organogenesis, differentiation and maturation. *Islets* **8**, 13–34; 10.1080/19382014.2015.1075687 (2016).
67. Sahoo, S. S. *et al.* FOXA2 suppresses endometrial carcinogenesis and epithelial-mesenchymal transition by regulating enhancer activity. *The Journal of clinical investigation* **132**; 10.1172/JCI157574 (2022).
68. Ouyang, C., Fu, Q., Xie, Y. & Xie, J. Forkhead box A2 transcriptionally activates hsa-let-7 g to inhibit hypoxia-induced epithelial-mesenchymal transition by targeting c14orf28 in colorectal cancer. *Arab journal of gastroenterology : the official publication of the Pan-Arab Association of Gastroenterology*; 10.1016/j.ajg.2022.04.002 (2022).
69. Tang, Y., Shu, G., Yuan, X., Jing, N. & Song, J. FOXA2 functions as a suppressor of tumor metastasis by inhibition of epithelial-to-mesenchymal transition in human lung cancers. *Cell research* **21**, 316–326; 10.1038/cr.2010.126 (2011).
70. Wang, J. *et al.* FOXA2 suppresses the metastasis of hepatocellular carcinoma partially through matrix metalloproteinase-9 inhibition. *Carcinogenesis* **35**, 2576–2583; 10.1093/carcin/bgu180 (2014).

- 
71. Bow, Y.-D. *et al.* Silencing of FOXA2 decreases E-cadherin expression and is associated with lymph node metastasis in oral cancer. *Oral diseases* **26**, 756–765; 10.1111/odi.13282 (2020).
72. Wang, B., Liu, G., Ding, L., Zhao, J. & Lu, Y. FOXA2 promotes the proliferation, migration and invasion, and epithelial mesenchymal transition in colon cancer. *Experimental and therapeutic medicine* **16**, 133–140; 10.3892/etm.2018.6157 (2018).
73. Gao, H., Yan, Z., Sun, H. & Chen, Y. FoXA2 promotes esophageal squamous cell carcinoma progression by ZEB2 activation. *World journal of surgical oncology* **19**, 286; 10.1186/s12957-021-02358-4 (2021).
74. Wruck, W. *et al.* The pioneer and differentiation factor FOXA2 is a key driver of yolk-sac tumour formation and a new biomarker for paediatric and adult yolk-sac tumours. *Journal of cellular and molecular medicine* **25**, 1394–1405; 10.1111/jcmm.16222 (2021).
75. Liu, Q. *et al.* Histone demethylase PHF8 drives neuroendocrine prostate cancer progression by epigenetically upregulating FOXA2. *The Journal of pathology* **253**, 106–118; 10.1002/path.5557 (2021).
76. Song, Y., Washington, M. K. & Crawford, H. C. Loss of FOXA1/2 is essential for the epithelial-to-mesenchymal transition in pancreatic cancer. *Cancer research* **70**, 2115–2125; 10.1158/0008-5472.CAN-09-2979 (2010).
77. Kaushik, G. *et al.* Selective inhibition of stemness through EGFR/FOXA2/SOX9 axis reduces pancreatic cancer metastasis. *Oncogene* **40**, 848–862; 10.1038/s41388-020-01564-w (2021).
78. Dhamija, S. *et al.* LINC00261 and the Adjacent Gene FOXA2 Are Epithelial Markers and Are Suppressed during Lung Cancer Tumorigenesis and Progression. *Non-coding RNA* **5**; 10.3390/ncrna5010002 (2018).
79. Milan, M. *et al.* FOXA2 controls the cis-regulatory networks of pancreatic cancer cells in a differentiation grade-specific manner. *The EMBO journal* **38**, e102161; 10.15252/embj.2019102161 (2019).
80. Li, Y. *et al.* Mutant Kras co-opts a proto-oncogenic enhancer network in inflammation-induced metaplastic progenitor cells to initiate pancreatic cancer. *Nature cancer* **2**, 49–65; 10.1093/nar/gkx247 (2021).
81. Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *The Journal of clinical investigation* **119**, 1420–1428; 10.1172/JCI39104 (2009).
82. Dongre, A. & Weinberg, R. A. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nature reviews. Molecular cell biology* **20**, 69–84; 10.1038/s41580-018-0080-4 (2019).
83. Yang, J. *et al.* Guidelines and definitions for research on epithelial-mesenchymal transition. *Nature reviews. Molecular cell biology* **21**, 341–352; 10.1038/s41580-020-0237-9 (2020).

- 
84. Hollier, B. G. *et al.* FOXC2 expression links epithelial-mesenchymal transition and stem cell properties in breast cancer. *Cancer research* **73**, 1981–1992; 10.1158/0008-5472.CAN-12-2962 (2013).
85. Wang, X. *et al.* Krüppel-like factor 8 induces epithelial to mesenchymal transition and epithelial cell invasion. *Cancer research* **67**, 7184–7193; 10.1158/0008-5472.CAN-06-4729 (2007).
86. Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nature reviews. Molecular cell biology* **15**, 178–196; 10.1038/nrm3758 (2014).
87. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell systems* **1**, 417–425; 10.1016/j.cels.2015.12.004 (2015).
88. Krantz, S. B., Shields, M. A., Dangi-Garimella, S., Bentrem, D. J. & Munshi, H. G. Contribution of epithelial-mesenchymal transition to pancreatic cancer progression. *Cancers* **2**, 2084–2097; 10.3390/cancers2042084 (2010).
89. Aiello, N. M. *et al.* EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. *Developmental cell* **45**, 681–695.e4; 10.1016/j.devcel.2018.05.027 (2018).
90. Cano, C. E., Motoo, Y. & Iovanna, J. L. Epithelial-to-mesenchymal transition in pancreatic adenocarcinoma. *TheScientificWorldJournal* **10**, 1947–1957; 10.1100/tsw.2010.183 (2010).
91. Ahmed, S., Bradshaw, A.-D., Gera, S., Dewan, M. Z. & Xu, R. The TGF- $\beta$ /Smad4 Signaling Pathway in Pancreatic Carcinogenesis and Its Clinical Significance. *Journal of clinical medicine* **6**; 10.3390/jcm6010005 (2017).
92. Ortiz, M. A. *et al.* Src family kinases, adaptor proteins and the actin cytoskeleton in epithelial-to-mesenchymal transition. *Cell communication and signaling : CCS* **19**, 67; 10.1186/s12964-021-00750-x (2021).
93. Massagué, J. TGF $\beta$  signalling in context. *Nature reviews. Molecular cell biology* **13**, 616–630; 10.1038/nrm3434 (2012).
94. Lu, W. & Kang, Y. Epithelial-Mesenchymal Plasticity in Cancer Progression and Metastasis. *Developmental cell* **49**, 361–374; 10.1016/j.devcel.2019.04.010 (2019).
95. Mahajan, N. P. *et al.* Activated Cdc42-associated kinase Ack1 promotes prostate cancer progression via androgen receptor tyrosine phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 8438–8443; 10.1073/pnas.0700420104 (2007).
96. Mitra, S. K. & Schlaepfer, D. D. Integrin-regulated FAK–Src signaling in normal and cancer cells. *Current Opinion in Cell Biology* **18**, 516–523; 10.1016/j.ceb.2006.08.011 (2006).
97. Playford, M. P. & Schaller, M. D. The interplay between Src and integrins in normal and tumor biology. *Oncogene* **23**, 7928–7946; 10.1038/sj.onc.1208080 (2004).

- 
98. Guo, Z. *et al.* Regulation of androgen receptor activity by tyrosine phosphorylation. *Cancer Cell* **10**, 309–319; 10.1016/j.ccr.2006.08.021 (2006).
99. Nath, D. *et al.* Abi1 loss drives prostate tumorigenesis through activation of EMT and non-canonical WNT signaling. *Cell communication and signaling : CCS* **17**, 120; 10.1186/s12964-019-0410-y (2019).
100. Lu, C., Yang, Z., Yu, D., Lin, J. & Cai, W. RUNX1 regulates TGF- $\beta$  induced migration and EMT in colorectal cancer. *Pathology, research and practice* **216**, 153142; 10.1016/j.prp.2020.153142 (2020).
101. Friedman, A. D. *et al.* Src Kinase Can Activate RUNX1 Activity Via Phosphorylation Of C-Terminal Tyrosines and Activated RUNX1 Stimulates Granulopoiesis. *Blood* **122**, 1210; 10.1182/blood.V122.21.1210.1210 (2013).
102. Leong, W. Y. *et al.* Runx1 Phosphorylation by Src Increases Trans-activation via Augmented Stability, Reduced Histone Deacetylase (HDAC) Binding, and Increased DNA Affinity, and Activated Runx1 Favors Granulopoiesis. *The Journal of biological chemistry* **291**, 826–836; 10.1074/jbc.M115.674234 (2016).
103. Cao, X., Tay, A., Guy, G. R. & Tan, Y. H. Activation and association of Stat3 with Src in v-Src-transformed cell lines. *Molecular and cellular biology* **16**, 1595–1603; 10.1128/MCB.16.4.1595 (1996).
104. Li, Q. *et al.* RUNX1 promotes tumour metastasis by activating the Wnt/ $\beta$ -catenin signalling pathway and EMT in colorectal cancer. *Journal of experimental & clinical cancer research : CR* **38**, 334; 10.1186/s13046-019-1330-9 (2019).
105. Liu, S. *et al.* The lncRNA RUNX1-IT1 regulates C-FOS transcription by interacting with RUNX1 in the process of pancreatic cancer proliferation, migration and invasion. *Cell death & disease* **11**, 412; 10.1038/s41419-020-2617-7 (2020).
106. Zhu, M.-L. & Kyprianou, N. Role of androgens and the androgen receptor in epithelial-mesenchymal transition and invasion of prostate cancer cells. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **24**, 769–777; 10.1096/fj.09-136994 (2010).
107. Wendt, M. K., Balanis, N., Carlin, C. R. & Schiemann, W. P. STAT3 and epithelial-mesenchymal transitions in carcinomas. *JAK-STAT* **3**, e28975; 10.4161/jkst.28975 (2014).
108. Byers, L. A. *et al.* Reciprocal regulation of c-Src and STAT3 in non-small cell lung cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **15**, 6852–6861; 10.1158/1078-0432.CCR-09-0767 (2009).
109. Glaß, M., Dorn, A., Hüttelmaier, S., Haemmerle, M. & Gutschner, T. Comprehensive Analysis of LincRNAs in Classical and Basal-Like Subtypes of Pancreatic Cancer. *Cancers* **12**; 10.3390/cancers12082077 (2020).



- 
110. Emmrich, S. *et al.* miR-99a/100~125b tricistrons regulate hematopoietic stem and progenitor cell homeostasis by shifting the balance between TGF $\beta$  and Wnt signaling. *Genes & development* **28**, 858–874; 10.1101/gad.233791.113 (2014).
111. Deer, E. L. *et al.* Phenotype and genotype of pancreatic cancer cell lines. *Pancreas* **39**, 425–435; 10.1097/MPA.0b013e3181c15963 (2010).
112. Loukopoulos, P. *et al.* Orthotopic transplantation models of pancreatic adenocarcinoma derived from cell lines and primary tumors and displaying varying metastatic activity. *Pancreas* **29**, 193–203; 10.1097/00006676-200410000-00004 (2004).
113. Belo, A. I., van der Sar, A. M., Tefsen, B. & van Die, I. Galectin-4 Reduces Migration and Metastasis Formation of Pancreatic Cancer Cells. *PLoS one* **8**, e65957; 10.1371/journal.pone.0065957 (2013).
114. Bartsch, D. *et al.* Higher frequency of DPC4/Smad4 alterations in pancreatic cancer cell lines than in primary pancreatic adenocarcinomas. *Cancer Letters* **139**, 43–49; 10.1016/S0304-3835(98)00380-2 (1999).
115. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)* **25**, 402–408; 10.1006/meth.2001.1262 (2001).
116. Tibes, R. *et al.* Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. *Molecular cancer therapeutics* **5**, 2512–2521; 10.1158/1535-7163.MCT-06-0334 (2006).
117. Grund, S. E., Polycarpou-Schwarz, M., Luo, C., Eichmüller, S. B. & Diederichs, S. Rare Drosha splice variants are deficient in microRNA processing but do not affect general microRNA expression in cancer cells. *Neoplasia (New York, N.Y.)* **14**, 238–248; 10.1593/neo.111586 (2012).
118. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367; 10.1186/1471-2105-11-367 (2010).
119. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* **26**, 139–140; 10.1093/bioinformatics/btp616 (2010).
120. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic acids research* **46**, W537–W544; 10.1093/nar/gky379 (2018).
121. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29**, 15–21; 10.1093/bioinformatics/bts635 (2013).
122. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)* **30**, 923–930; 10.1093/bioinformatics/btt656 (2014).

- 
123. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550; 10.1186/s13059-014-0550-8 (2014).
124. Camargo, A. P., Sourkov, V., Pereira, G. A. G. & Carazzolle, M. F. RNAsamba: neural network-based assessment of the protein-coding potential of RNA sequences. *NAR genomics and bioinformatics* **2**, lqz024; 10.1093/nargab/lqz024 (2020).
125. Kang, Y.-J. *et al.* CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic acids research* **45**, W12-W16; 10.1093/nar/gkx428 (2017).
126. Hutchinson, J. N. *et al.* A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC genomics* **8**, 39; 10.1186/1471-2164-8-39 (2007).
127. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183; 10.1016/j.cell.2013.02.022 (2013).
128. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature protocols* **8**, 2281–2308; 10.1038/nprot.2013.143 (2013).
129. Batlle, E. *et al.* The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nature cell biology* **2**, 84–89; 10.1038/35000034. (2000).
130. Cano, A. *et al.* The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nature cell biology* **2**, 76–83; 10.1038/35000025. (2000).
131. Craene, B. de *et al.* The transcription factor snail induces tumor cell invasion through modulation of the epithelial cell differentiation program. *Cancer research* **65**, 6237–6244; 10.1158/0008-5472.CAN-04-3545. (2005).
132. Wang, F., Sloss, C., Zhang, X., Lee, S. W. & Cusack, J. C. Membrane-bound heparin-binding epidermal growth factor like growth factor regulates E-cadherin expression in pancreatic carcinoma cells. *Cancer research* **67**, 8486–8493; 10.1158/0008-5472.CAN-07-0498. (2007).
133. Yang, J. *et al.* Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117**, 927–939; 10.1016/j.cell.2004.06.006. (2004).
134. Liu, Y.-N. *et al.* Regulatory mechanisms controlling human E-cadherin gene expression. *Oncogene* **24**, 8277–8290; 10.1038/sj.onc.1208991 (2005).
135. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic acids research* **51**, D638–D646; 10.1093/nar/gkac1000 (2023).
136. Wang, T. *et al.* COL4A1 promotes the growth and metastasis of hepatocellular carcinoma cells by activating FAK-Src signaling. *Journal of experimental & clinical cancer research : CR* **39**, 148; 10.1186/s13046-020-01650-7 (2020).

- 
137. Pasquale, E. B. Eph receptors and ephrins in cancer: bidirectional signalling and beyond. *Nature reviews. Cancer* **10**, 165–180; 10.1038/nrc2806 (2010).
138. Kullander, K. & Klein, R. Mechanisms and functions of Eph and ephrin signalling. *Nature reviews. Molecular cell biology* **3**, 475–486; 10.1038/nrm856 (2002).
139. Zisch, A. H., Kalo, M. S., Chong, L. D. & Pasquale, E. B. Complex formation between EphB2 and Src requires phosphorylation of tyrosine 611 in the EphB2 juxtamembrane region. *Oncogene* **16**, 2657–2670; 10.1038/sj.onc.1201823 (1998).
140. Zhai, S. *et al.* Epigenetic silencing of LncRNA LINC00261 promotes c-myc-mediated aerobic glycolysis by regulating miR-222-3p/HIPK2/ERK axis and sequestering IGF2BP1. *Oncogene* **40**, 277–291; 10.1038/s41388-020-01525-3 (2021).
141. Bley, N. *et al.* IGF2BP1 is a targetable SRC/MAPK-dependent driver of invasive growth in ovarian cancer. *RNA biology* **18**, 391–403; 10.1080/15476286.2020.1812894 (2021).
142. Fu, X.-L. *et al.* Analysis of long non-coding RNA expression profiles in pancreatic ductal adenocarcinoma. *Scientific reports* **6**, 33535; 10.1038/srep33535. (2016).
143. Giulietti, M., Righetti, A., Principato, G. & Piva, F. LncRNA co-expression network analysis reveals novel biomarkers for pancreatic cancer. *Carcinogenesis* **39**, 1016–1025; 10.1093/carcin/bgy069. (2018).
144. Hao, S. *et al.* Genome-Wide Analysis Identified a Number of Dysregulated Long Noncoding RNA (lncRNA) in Human Pancreatic Ductal Adenocarcinoma. *Technology in cancer research & treatment* **17**, 1533034617748429; 10.1177/1533034617748429. (2018).
145. Yu, X., Lin, Y., Sui, W., Zou, Y. & Lv, Z. Analysis of distinct long noncoding RNA transcriptional fingerprints in pancreatic ductal adenocarcinoma. *Cancer medicine* **6**, 673–680; 10.1002/cam4.1027. (2017).
146. Zhou, C. *et al.* A Long Non-coding RNA Signature to Improve Prognostic Prediction of Pancreatic Ductal Adenocarcinoma. *Frontiers in oncology* **9**, 1160; 10.3389/fonc.2019.01160. (2019).
147. Zhou, M. *et al.* Construction and analysis of dysregulated lncRNA-associated ceRNA network identified novel lncRNA biomarkers for early diagnosis of human pancreatic cancer. *Oncotarget* **7**, 56383–56394; 10.18632/oncotarget.10891. (2016).
148. Zhang, H.-F., Li, W. & Han, Y.-D. LINC00261 suppresses cell proliferation, invasion and Notch signaling pathway in hepatocellular carcinoma. *Cancer biomarkers : section A of Disease markers* **21**, 575–582; 10.3233/CBM-170471. (2018).
149. Yu, Y. *et al.* Long non-coding RNA linc00261 suppresses gastric cancer progression via promoting Slug degradation. *Journal of cellular and molecular medicine* **21**, 955–967; 10.1111/jcmm.13035. (2017).

- 
150. Gao, J. *et al.* Up-regulated LINC00261 predicts a poor prognosis and promotes a metastasis by EMT process in cholangiocarcinoma. *Pathology, research and practice* **216**, 152733; 10.1016/j.prp.2019.152733. (2020).
151. Mather, R. L. *et al.* The evolutionarily conserved long non-coding RNA LINC00261 drives neuroendocrine prostate cancer proliferation and metastasis via distinct nuclear and cytoplasmic mechanisms. *Molecular oncology* **15**, 1921–1941; 10.1002/1878-0261.12954 (2021).
152. Ponjavic, J., Oliver, P. L., Lunter, G. & Ponting, C. P. Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. *PLoS genetics* **5**, e1000617; 10.1371/journal.pgen.1000617 (2009).
153. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227; 10.1038/nature07672 (2009).
154. Ji, Z., Song, R., Regev, A. & Struhl, K. Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife* **4**, e08890; 10.7554/eLife.08890 (2015).
155. Gaertner, B. *et al.* A human ESC-based screen identifies a role for the translated lncRNA LINC00261 in pancreatic endocrine differentiation. *eLife* **9**; 10.7554/eLife.58659 (2020).
156. Bazzini, A. A. *et al.* Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *The EMBO journal* **33**, 981–993; 10.1002/embj.201488411 (2014).
157. Chen, J. *et al.* Pervasive functional translation of noncanonical human open reading frames. *Science (New York, N.Y.)* **367**, 1140–1146; 10.1126/science.aay0262 (2020).
158. Martinez, T. F. *et al.* Accurate annotation of human protein-coding small open reading frames. *Nature chemical biology* **16**, 458–468; 10.1038/s41589-019-0425-0 (2020).
159. Raj, A. *et al.* Thousands of novel translated open reading frames in humans inferred by ribosome footprint profiling. *eLife* **5**; 10.7554/eLife.13328 (2016).
160. van Heesch, S. *et al.* Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome biology* **15**, R6; 10.1186/gb-2014-15-1-r6 (2014).
161. Sergeeva, O. V., Korinfskaya, S. A., Kurochkin, I. I. & Zatsepin, T. S. Long Noncoding RNA LL35/Falcor Regulates Expression of Transcription Factor Foxa2 in Hepatocytes in Normal and Fibrotic Mouse Liver. *Acta naturae* **11**, 66–74; 10.32607/20758251-2019-11-3-66-74. (2019).
162. Lin, K. *et al.* Long noncoding RNA LINC00261 induces chemosensitization to 5-fluorouracil by mediating methylation-dependent repression of DPYD in human esophageal cancer. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **33**, 1972–1988; 10.1096/fj.201800759R. (2019).

163. Swarr, D. T. *et al.* The long noncoding RNA Falcor regulates Foxa2 expression to maintain lung epithelial homeostasis and promote regeneration. *Genes & development* **33**, 656–668; 10.1101/gad.320523.118. (2019).
164. Nieto, M. A., Huang, R. Y.-J., Jackson, R. A. & Thiery, J. P. EMT: 2016. *Cell* **166**, 21–45; 10.1016/j.cell.2016.06.028. (2016).
165. Hotz, B. *et al.* Epithelial to mesenchymal transition: expression of the regulators snail, slug, and twist in pancreatic cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **13**, 4769–4776; 10.1158/1078-0432.CCR-06-2926. (2007).
166. Yamada, S. *et al.* Epithelial-to-mesenchymal transition predicts prognosis of pancreatic cancer. *Surgery* **154**, 946–954; 10.1016/j.surg.2013.05.004. (2013).
167. Zheng, X. *et al.* Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **527**, 525–530; 10.1038/nature16064. (2015).
168. Fischer, K. R. *et al.* Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**, 472–476; 10.1038/nature15748. (2015).
169. Chen, T. *et al.* Linc00261 inhibits metastasis and the WNT signaling pathway of pancreatic cancer by regulating a miR-552-5p/FOXO3 axis. *Oncology reports* **43**, 930–942; 10.3892/or.2020.7480. (2020).
170. Liu, S. *et al.* Methylation-mediated LINC00261 suppresses pancreatic cancer progression by epigenetically inhibiting c-Myc transcription. *Theranostics* **10**, 10634–10651; 10.7150/thno.44278 (2020).
171. Li, B. *et al.* The inhibitory effect of LINC00261 upregulation on the pancreatic cancer EMT process is mediated by KLF13 via the mTOR signaling pathway. *Clinical & translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico* **24**, 1059–1072; 10.1007/s12094-021-02747-x (2022).
172. Dang, H. X. *et al.* Long non-coding RNA LCAL62 / LINC00261 is associated with lung adenocarcinoma prognosis. *Heliyon* **6**, e03521; 10.1016/j.heliyon.2020.e03521. (2020).
173. Burstin, J. von *et al.* E-cadherin regulates metastasis of pancreatic cancer in vivo and is suppressed by a SNAIL/HDAC1/HDAC2 repressor complex. *Gastroenterology* **137**, 361–71, 371.e1–5; 10.1053/j.gastro.2009.04.004. (2009).
174. Hu, D. *et al.* Proteomic analyses identify prognostic biomarkers for pancreatic ductal adenocarcinoma. *Oncotarget* **9**, 9789–9807; 10.18632/oncotarget.23929. (2018).
175. Wang, L., Wang, H. & Chen, J. Linc00261 Inhibited High-Grade Serous Ovarian Cancer Progression through miR-552-ATG10-EMT Axis. *Computational and Mathematical Methods in Medicine* **2022**, 9450353; 10.1155/2022/9450353 (2022).
176. Iyer, S. *et al.* FOXOs attenuate bone formation by suppressing Wnt signaling. *The Journal of clinical investigation* **123**, 3409–3419; 10.1172/JCI68049. (2013).

177. Ouyang, D., Li, R., Li, Y. & Zhu, X. Construction of a Competitive Endogenous RNA Network in Uterine Corpus Endometrial Carcinoma. *Medical science monitor : international medical journal of experimental and clinical research* **25**, 7998–8010; 10.12659/MSM.915798. (2019).
178. Qi, L. *et al.* Identification of lncRNAs associated with lung squamous cell carcinoma prognosis in the competitive endogenous RNA network. *PeerJ* **7**, e7727; 10.7717/peerj.7727. (2019).
179. Zhao, D. *et al.* Identifying prognostic biomarkers in endometrial carcinoma based on ceRNA network. *Journal of cellular biochemistry* **121**, 2437–2446; 10.1002/jcb.29466. (2020).
180. Fan, Y. *et al.* TGF- $\beta$ -induced upregulation of malat1 promotes bladder cancer metastasis by associating with suz12. *Clinical cancer research : an official journal of the American Association for Cancer Research* **20**, 1531–1541; 10.1158/1078-0432.CCR-13-1455 (2014).
181. Kawasaki, N. *et al.* Long noncoding RNA NORAD regulates transforming growth factor- $\beta$  signaling and epithelial-to-mesenchymal transition-like phenotype. *Cancer science* **109**, 2211–2220; 10.1111/cas.13626 (2018).
182. Papoutsoglou, P. *et al.* The TGFB2-AS1 lncRNA Regulates TGF- $\beta$  Signaling by Modulating Corepressor Activity. *Cell reports* **28**, 3182–3198.e11; 10.1016/j.celrep.2019.08.028 (2019).
183. Richards, E. J. *et al.* Long non-coding RNAs (lncRNA) regulated by transforming growth factor (TGF)  $\beta$ : lncRNA-hit-mediated TGF $\beta$ -induced epithelial to mesenchymal transition in mammary epithelia. *The Journal of biological chemistry* **290**, 6857–6867; 10.1074/jbc.M114.610915 (2015).
184. Tang, R., Zhang, G., Wang, Y.-C., Mei, X. & Chen, S.-Y. The long non-coding RNA GAS5 regulates transforming growth factor  $\beta$  (TGF- $\beta$ )-induced smooth muscle cell differentiation via RNA Smad-binding elements. *The Journal of biological chemistry* **292**, 14270–14278; 10.1074/jbc.M117.790030 (2017).
185. Shahabi, S. *et al.* LINC00261 Is an Epigenetically Regulated Tumor Suppressor Essential for Activation of the DNA Damage Response. *Cancer research* **79**, 3050–3062; 10.1158/0008-5472.CAN-18-2034. (2019).
186. Villacorte, M. *et al.*  $\beta$ -Catenin signaling regulates Foxa2 expression during endometrial hyperplasia formation. *Oncogene* **32**, 3477–3482; 10.1038/onc.2012.376. (2013).
187. Tuo, Z. *et al.* RUNX1 is a promising prognostic biomarker and related to immune infiltrates of cancer-associated fibroblasts in human cancers. *BMC cancer* **22**, 523; 10.1186/s12885-022-09632-y (2022).
188. Birnbaum, D. J. *et al.* Expression of Genes with Copy Number Alterations and Survival of Patients with Pancreatic Adenocarcinoma. *Cancer genomics & proteomics* **13**, 191–200 (2016).
189. Cheng, Y. *et al.* RUNX1 promote invasiveness in pancreatic ductal adenocarcinoma through regulating miR-93. *Oncotarget* **8**, 99567–99579; 10.18632/oncotarget.20433 (2017).

- 
190. Martinez, M. *et al.* Transcriptional Auto-Regulation of RUNX1 P1 Promoter. *PLoS one* **11**, e0149119; 10.1371/journal.pone.0149119 (2016).
191. Schütte, J. *et al.* An experimentally validated network of nine haematopoietic transcription factors reveals mechanisms of cell state stability. *eLife* **5**, e11469; 10.7554/eLife.11469 (2016).
192. Wahlestedt, C. *et al.* Potent and nontoxic antisense oligonucleotides containing locked nucleic acids. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5633–5638; 10.1073/pnas.97.10.5633 (2000).
193. Caudron-Herger, M., Jansen, R. E., Wassmer, E. & Diederichs, S. RBP2GO: a comprehensive pan-species database on RNA-binding proteins, their interactions and functions. *Nucleic acids research* **49**, D425–D436; 10.1093/nar/gkaa1040 (2021).
194. Kondratyeva, L. G. *et al.* Downregulation of expression of mater genes SOX9, FOXA2, and GATA4 in pancreatic cancer cells stimulated with TGF $\beta$ 1 epithelial-mesenchymal transition. *Doklady. Biochemistry and biophysics* **469**, 257–259; 10.1134/S1607672916040062 (2016).
195. Zhang, Y. *et al.* High throughput determination of TGF $\beta$ 1/SMAD3 targets in A549 lung epithelial cells. *PLoS one* **6**, e20319; 10.1371/journal.pone.0020319 (2011).
196. Lombaerts, M. *et al.* E-cadherin transcriptional downregulation by promoter methylation but not mutation is related to epithelial-to-mesenchymal transition in breast cancer cell lines. *British journal of cancer* **94**, 661–671; 10.1038/sj.bjc.6602996. (2006).
197. Wang, Y.-Q., Yuan, Y., Jiang, S. & Jiang, H. Promoter methylation and expression of CDH1 and susceptibility and prognosis of eyelid squamous cell carcinoma. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* **37**, 9521–9526; 10.1007/s13277-016-4851-2. (2016).
198. Zhang, Z. *et al.* FOXA2 attenuates the epithelial to mesenchymal transition by regulating the transcription of E-cadherin and ZEB2 in human breast cancer. *Cancer Letters* **361**, 240–250; 10.1016/j.canlet.2015.03.008 (2015).
199. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology* **28**, 495–501; 10.1038/nbt.1630 (2010).
200. Lee, M.-R. *et al.* FOXA2 and STAT5A regulate oncogenic activity of KIF5B-RET fusion. *American journal of cancer research* **13**, 638–653 (2023).
201. Minoo, P. *et al.* SMAD3 prevents binding of NKX2.1 and FOXA1 to the SpB promoter through its MH1 and MH2 domains. *Nucleic acids research* **36**, 179–188; 10.1093/nar/gkm871 (2008).
202. Overdier, D. G., Porcella, A. & Costa, R. H. The DNA-binding specificity of the hepatocyte nuclear factor 3/forkhead domain is influenced by amino-acid residues adjacent to the recognition helix. *Molecular and cellular biology* **14**, 2755–2766; 10.1128/MCB.14.4.2755 (1994).

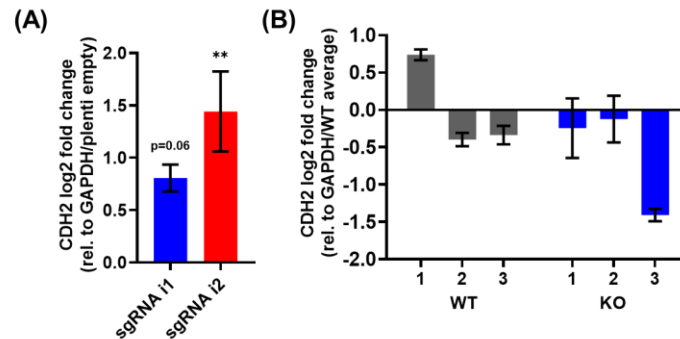
203. Janakiraman, H. *et al.* The Long (lncRNA) and Short (miRNA) of It: TGF $\beta$ -Mediated Control of RNA-Binding Proteins and Noncoding RNAs. *Molecular cancer research : MCR* **16**, 567–579; 10.1158/1541-7786.MCR-17-0547 (2018).
204. Butz, H., Rácz, K., Hunyady, L. & Patócs, A. Crosstalk between TGF- $\beta$  signaling and the microRNA machinery. *Trends in pharmacological sciences* **33**, 382–393; 10.1016/j.tips.2012.04.003 (2012).
205. Baroukh, N. *et al.* MicroRNA-124a regulates Foxa2 expression and intracellular signaling in pancreatic beta-cell lines. *Journal of Biological Chemistry* **282**, 19575–19588; 10.1074/jbc.M611841200 (2007).
206. Chen, S.-Y. *et al.* MicroRNA-200a inhibits cell growth and metastasis by targeting Foxa2 in hepatocellular carcinoma. *Journal of Cancer* **8**, 617–625; 10.7150/jca.17394 (2017).
207. Yu, Y., Yu, F. & Sun, P. MicroRNA-1246 Promotes Melanoma Progression Through Targeting FOXA2. *OncoTargets and therapy* **13**, 1245–1253; 10.2147/OTT.S234276 (2020).
208. Belaguli, N. S., Zhang, M., Brunicardi, F. C. & Berger, D. H. Forkhead box protein A2 (FOXA2) protein stability and activity are regulated by sumoylation. *PloS one* **7**, e48019; 10.1371/journal.pone.0048019 (2012).
209. van Gent, R. *et al.* SIRT1 mediates FOXA2 breakdown by deacetylation in a nutrient-dependent manner. *PloS one* **9**, e98438; 10.1371/journal.pone.0098438 (2014).
210. Wolfrum, C., Asilmaz, E., Luca, E., Friedman, J. M. & Stoffel, M. Foxa2 regulates lipid metabolism and ketogenesis in the liver during fasting and in diabetes. *Nature* **432**, 1027–1032; 10.1038/nature03047 (2004).
211. Caudron-Herger, M. *et al.* R-Deep: Proteome-wide and Quantitative Identification of RNA-Dependent Proteins by Density Gradient Ultracentrifugation. *Molecular cell* **75**, 184–199.e10; 10.1016/j.molcel.2019.04.018 (2019).
212. Karamitopoulou, E. *et al.* Loss of Raf-1 kinase inhibitor protein (RKIP) is strongly associated with high-grade tumor budding and correlates with an aggressive phenotype in pancreatic ductal adenocarcinoma (PDAC). *Journal of translational medicine* **11**, 311; 10.1186/1479-5876-11-311 (2013).
213. Fu, Z. *et al.* Endogenous miRNA Sponge LincRNA-ROR promotes proliferation, invasion and stem cell-like phenotype of pancreatic cancer cells. *Cell death discovery* **3**, 17004; 10.1038/cddiscovery.2017.4 (2017).
214. Chung, W.-M. *et al.* MicroRNA-21 promotes the ovarian teratocarcinoma PA1 cell line by sustaining cancer stem/progenitor populations in vitro. *Stem cell research & therapy* **4**, 88; 10.1186/scrt247 (2013).
215. Kang, H.-Y. MicroRNA-21 regulates stemness in cancer cells. *Stem cell research & therapy* **4**, 110; 10.1186/scrt321 (2013).



216. Rajendran, V. & Jain, M. V. In Vitro Tumorigenic Assay: Colony Forming Assay for Cancer Stem Cells. *Methods in molecular biology (Clifton, N.J.)* **1692**, 89–95; 10.1007/978-1-4939-7401-6\_8 (2018).
217. Kabadi, A. M., Ousterout, D. G., Hilton, I. B. & Gersbach, C. A. Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic acids research* **42**, e147; 10.1093/nar/gku749 (2014).
218. Li, C. *et al.* Identification of pancreatic cancer stem cells. *Cancer research* **67**, 1030–1037; 10.1158/0008-5472.CAN-06-2030 (2007).
219. Metildi, C. A., Kaushal, S., Hoffman, R. M. & Bouvet, M. In vivo serial selection of human pancreatic cancer cells in orthotopic mouse models produces high metastatic variants irrespective of Kras status. *The Journal of surgical research* **184**, 290–298; 10.1016/j.jss.2013.03.049 (2013).
220. Jin, H. *et al.* Hyperthermia inhibits the motility of gemcitabine-resistant pancreatic cancer PANC-1 cells through the inhibition of epithelial-mesenchymal transition. *Molecular medicine reports* **17**, 7274–7280; 10.3892/mmr.2018.8763 (2018).
221. Hiramoto, H. *et al.* miR-509-5p and miR-1243 increase the sensitivity to gemcitabine by inhibiting epithelial-mesenchymal transition in pancreatic cancer. *Scientific reports* **7**, 4002; 10.1038/s41598-017-04191-w (2017).
222. Gaianigo, N., Melisi, D. & Carbone, C. EMT and Treatment Resistance in Pancreatic Cancer. *Cancers* **9**; 10.3390/cancers9090122 (2017).
223. Craene, B. de & Berx, G. Regulatory networks defining EMT during cancer initiation and progression. *Nature reviews. Cancer* **13**, 97–110; 10.1038/nrc3447 (2013).
224. Lambert, A. W., Pattabiraman, D. R. & Weinberg, R. A. Emerging Biological Principles of Metastasis. *Cell* **168**, 670–691; 10.1016/j.cell.2016.11.037 (2017).
225. Tsuji, T. *et al.* Epithelial-mesenchymal transition induced by growth suppressor p12CDK2-AP1 promotes tumor cell local invasion but suppresses distant colony growth. *Cancer research* **68**, 10377–10386; 10.1158/0008-5472.CAN-08-1444 (2008).
226. Celià-Terrassa, T. *et al.* Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. *The Journal of clinical investigation* **122**, 1849–1868; 10.1172/JCI59218 (2012).
227. Chui, M. H. Insights into cancer metastasis from a clinicopathologic perspective: Epithelial-Mesenchymal Transition is not a necessary step. *International journal of cancer* **132**, 1487–1495; 10.1002/ijc.27745 (2013).
228. Hingorani, S. R. *et al.* Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**, 469–483; 10.1016/j.ccr.2005.04.023 (2005).

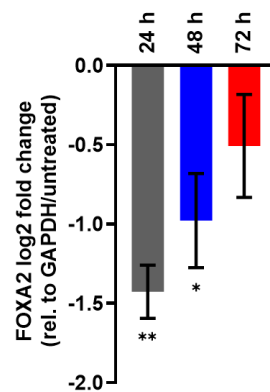
## VIII. APPENDIX

## 1. Supplementary Figures

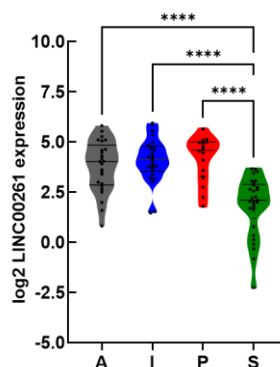


Supplementary Figure S1: Expression of CDH2 in LINC00261<sup>low</sup> cells.

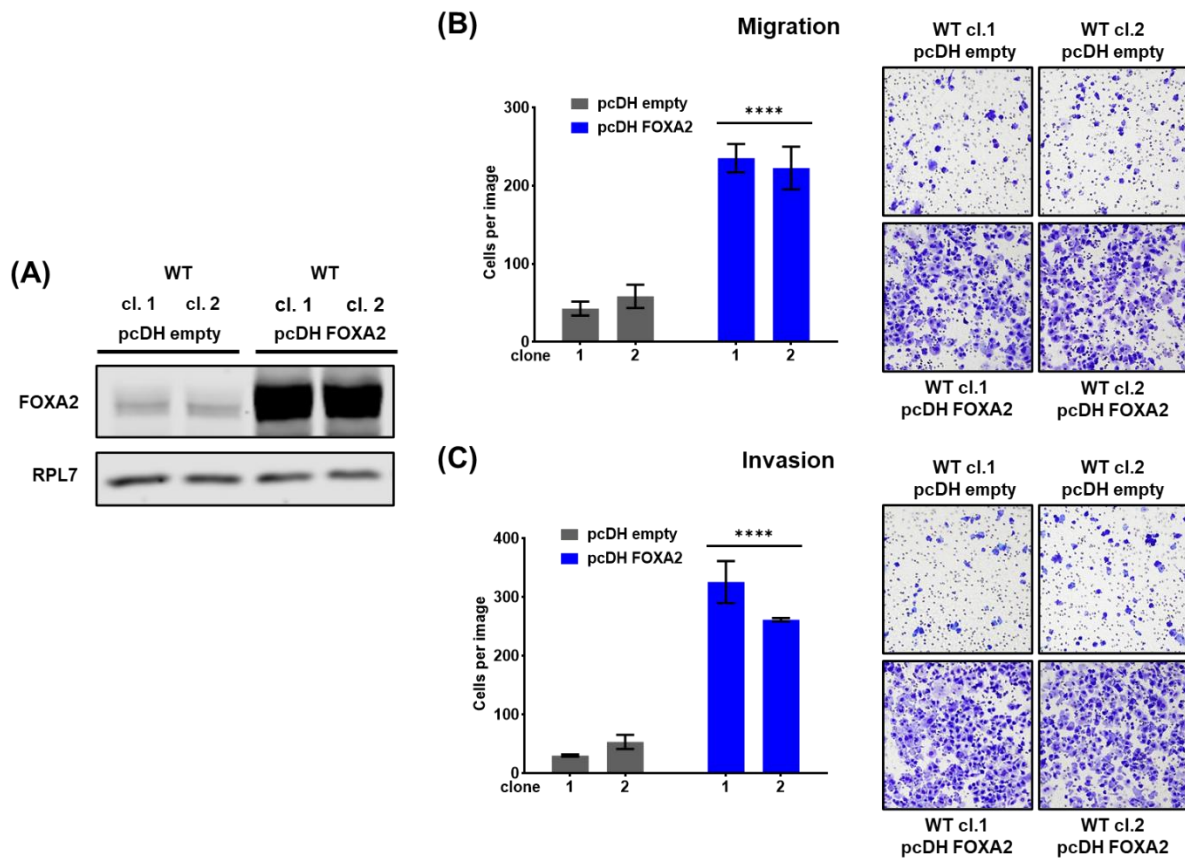
(A) CDH2 (N-cadherin) expression is significantly upregulated in LINC00261-depleted cells established by the CRISPRi system (\*\*  $p < 0.01$ , one-way ANOVA); (B) Expression of CDH2 in three wild-type and three promoter knockout clones.



Supplementary Figure S2: Expression of FOXA2 after treatment with 10 ng/ml TGF $\beta$  in Panc-1 cells. FOXA2 expression is significantly downregulated by TGF $\beta$  after 24 h and 48 h.



Supplementary Figure S3: Analysis of the Bailey PDAC dataset revealed a significant downregulation of FOXA2 expression in the squamous (S) compared to the pancreatic progenitor (P), immunogenic (I) and ADEX (A) subtype (\*\*\*\*  $p < 0.0001$ ; one-way ANOVA).



Supplementary Figure S4: Overexpression of FOXA2 in Panc-1 WT cells enhances cell invasion and migration. (A) Western blot displaying clearly increased FOXA2 protein expression in FOXA2 overexpressing Panc-1 WT cells. RPL7 served as loading control; (B, C) Transwell migration (B) and invasion (C) assays in FOXA2-overexpressing Panc-1 WT cells. Quantification of migrated/invaded cells from five random fields after Eosin Y/Methylene blue staining using light microscopy (\*\*\*\*  $p < 0.0001$ ; two-way ANOVA).

## 2. Supplementary Tables

Supplementary Table S1: Patient characteristics of FFPE tissue samples.

Age	Median (range)	68 (43-82)
Gender	male	19
	female	23
Tumor size	Median (range)	3.85 (1.5-6.0)
Grading	G1	2
	G2	15
	G3	25
Staging	T1	3
	T2	20
	T3	18
	T4	1
Lymph node status	N0	9
	N1	21
	N2	12
Distant metastasis	M0	39
	M1	3
Vascular invasion	L0	18
	L1	24
	V0	34
	V1	8
Perineural invasion	Pn0	7
	Pn1	35
UICC stage	IB	6
	IIA	3
	IIB	18
	III	12
	IV	3

Supplementary Table S2: Group sizes of TCGA samples categorized into low, middle and high LINC00261 and FOXA2 expression.

LINC00261 expression level	FOXA2 expression level	Group size
low	low	40
low	middle	13
low	high	6
middle	low	17
middle	middle	24
middle	high	17
high	low	2
high	middle	21
high	high	36

Supplementary Table S3: Differently expressed genes between 3 WT clones vs. 3 LINC00261 KO clones (p-adj.&lt;0.05).

Downregulated genes		Upregulated genes	
Gene Symbol	log2 fold change	Gene Symbol	log2 fold change
LINC00261	-4.09	SGK1	3.20
INHBB	-3.62	FHL1	2.32
PTCHD1	-3.23	COL23A1	2.32
ALPPL2	-3.01	CPPED1	2.26
CDH1	-2.85	NTN1	2.24
FMN1	-2.83	SCNN1G	2.20
F2RL2	-2.71	IGFBP7	2.16
GCNT1	-2.48	TRIM9	2.07
SPN	-2.41	ARHGEF4	2.04
RCSD1	-2.36	SCNN1B	2.01
COL6A3	-2.32	TMEM229B	1.97
CHST15	-2.30	RNF157	1.96
GALNT16	-2.25	ADAP1	1.96
RHOU	-2.14	THSD4	1.94
RP11-798K3.3	-2.13	VWDE	1.92
MUC5AC	-2.13	RAMP1	1.88
EDN1	-2.04	CYP4F22	1.87
CABLES1	-2.01	CARMIL2	1.84
TRNP1	-1.97	KCNB1	1.82
ARMCX2	-1.96	NMUR1	1.82
RP11-167H9.3	-1.96	FNDC4	1.81
RFLNA	-1.93	ENTPD2	1.79
KDF1	-1.86	MAATS1	1.79
NLRP3P1	-1.84	RGS17	1.78
LINC01447	-1.78	NOS2	1.78
PROC	-1.78	NSG1	1.77
PHYHD1	-1.77	MYLK	1.77
CAPN8	-1.76	SMO	1.77
TRPM8	-1.75	GNG4	1.77
KCNQ10T1	-1.74	SIPA1L2	1.76
MIR503HG	-1.63	TCF4	1.76
HMGB3	-1.62	TMEM30B	1.74
AKAP12	-1.53	TMEM198	1.74
RBBP8	-1.52	ETV1	1.73
CHML	-1.46	KSR2	1.73
MAGI2-AS3	-1.44	CHPF	1.72
MXRA8	-1.44	KISS1R	1.70
SLC4A11	-1.43	COL6A2	1.70
TWIST2	-1.37	COL13A1	1.69
CENPI	-1.37	ITGB4	1.68
ERCC6L	-1.35	CCDC184	1.68
MICALCL	-1.34	PDE2A	1.68
FAM111B	-1.31	ETNK2	1.68

Downregulated genes		Upregulated genes	
Gene Symbol	log2 fold change	Gene Symbol	log2 fold change
AHCYL1	-1.30	ABAT	1.67
SMC1A	-1.22	ARHGAP44	1.67
C18orf8	-1.20	PLCD1	1.67
TAF4B	-1.14	TLE2	1.66
MID2	-1.09	PIK3AP1	1.66
SETD7	-1.08	SPRY4	1.64
CABYR	-1.05	HSD3B7	1.64
NDC80	-1.05	CACNG7	1.63
CD99L2	-1.00	TUBB4A	1.61
CEP192	-0.99	TLL2	1.59
MMS22L	-0.90	PQLC2L	1.58
EIF3M	-0.82	ZDHHHC8P1	1.58
		TSPAN33	1.56
		HK2	1.55
		PLAUR	1.54
		SLC46A3	1.54
		SOX8	1.52
		ABTB1	1.46
		PCDHGB2	1.45
		SLC47A1	1.44
		CALHM2	1.44
		DUSP15	1.40
		PELI1	1.38
		APOBEC3G	1.34
		LARGE2	1.33
		MED12L	1.29
		COL9A2	1.21
		REEP2	1.21
		ZBTB46	1.10
		PPP2R5B	1.09
		MYO1E	1.06
		C3orf18	0.98

Supplementary Table S4: Differently expressed genes of CRISPRi-mediated knockdown of LINC00261 (p-adj.<0.05).

Downregulated genes		Upregulated genes	
Gene Symbol	log2FC	Gene Symbol	log2FC
LINC00261	-3.72	RCSD1	1.72
PKDCC	-2.39	C15orf48	1.58
CCDC80	-2.29	FHOD3	1.56
TMEM255A	-2.22	RBPM5	1.56
CXCR4	-2.21	TNFAIP2	1.52
TENM3	-2.16	ADRA1B	1.48
NEDD9	-2.15	ABCG2	1.47

Downregulated genes		Upregulated genes	
Gene Symbol	log2FC	Gene Symbol	log2FC
CD34	-1.96	AC005355.2	1.47
GPC4	-1.90	KCNIP3	1.43
GALNT16	-1.77	TRIB2	1.42
ANK2	-1.75	CSGALNACT1	1.41
KCNS3	-1.69	ARHGAP26	1.33
EYA2	-1.66	IL32	1.33
ANKRD1	-1.61	LAMA4	1.31
SUSD4	-1.59	RASD2	1.28
GJA1	-1.56	MYLK	1.28
GATA6	-1.55	CHST15	1.26
RASD1	-1.53	GAS6	1.25
VGLL3	-1.53	ASIC2	1.24
IGFBP7	-1.52	ROR1	1.23
AC108142.1	-1.49	KLHL29	1.23
CLU	-1.37	GDA	1.22
RP11-757G1.6	-1.37	WLS	1.22
HOOK1	-1.33	CDS1	1.22
IGFBP7-AS1	-1.30	CADPS2	1.21
SYNE1	-1.28	TRPA1	1.20
SLC7A2	-1.27	KRT7	1.20
ADARB2	-1.26	CPA4	1.19
WNT11	-1.23	CD82	1.18
FAM150A	-1.21	NR2F1-AS1	1.18
PROC	-1.21	S100A16	1.17
RASGRP3	-1.19	NTN4	1.16
SLC12A8	-1.16	TGM2	1.16
DEPTOR	-1.13	B3GNT5	1.16
PPP1R16B	-1.13	TBC1D4	1.14
MT1M	-1.11	PRSS22	1.14
HS3ST6	-1.09	CAMK2N1	1.14
PTGIS	-1.05	GRIP2	1.14
DACT2	-1.04	SEMA7A	1.13
ARFGEF3	-1.01	LAMC2	1.13
TCAF2	-1.00	RAB27B	1.13
IKZF2	-1.00	TBXAS1	1.11
CPNE4	-0.98	PPARG	1.10
MOXD1	-0.97	EGFLAM	1.10
PTX3	-0.97	MATN2	1.10
EMP1	-0.95	MB	1.09
NID2	-0.95	ITGA3	1.08
INPP4B	-0.94	ADAMTS15	1.08
DSC3	-0.92	SH3RF3	1.08
NDRG1	-0.91	THBD	1.07
CDH1	-0.91	TNFRSF1B	1.06
SCD5	-0.89	PLCL2	1.06

Downregulated genes		Upregulated genes	
Gene Symbol	log2FC	Gene Symbol	log2FC
MED12L	-0.89	CDH2	1.06
MGARP	-0.89	HRH1	1.06
BRINP3	-0.89	MLPH	1.05
F2RL2	-0.88	ADIRF	1.05
TMPRSS2	-0.88	EVA1C	1.03
TUBB2B	-0.87	STMN3	1.03
SUSD5	-0.87	HIPK2	1.02
COL1A1	-0.86	OSBPL3	1.02
ACO68282.3	-0.85	LPAR1	1.02
PALLD	-0.84	GNA15	1.01
CLVS2	-0.84	IRX3	1.01
RAB17	-0.83	BCAR3	1.01
FMN1	-0.83	RRAD	1.01
TFAP2C	-0.83	PPP3CA	1.00
ADAM22	-0.82	TRHDE-AS1	1.00
LIN28B	-0.81	GJB2	0.99
SYNE3	-0.78	COL11A1	0.98
TDRD5	-0.76	C10orf54	0.98
REEP5	-0.74	DRD1	0.98
DSC2	-0.74	GPRC5C	0.98
BCL2A1	-0.74	GRB10	0.97
CFAP57	-0.72	DRD2	0.97
CLDN10	-0.72	RFTN1	0.97
RP11-566K19.6	-0.70	L3MBTL4	0.96
RGCC	-0.70	HPCAL1	0.94
ADGRL2	-0.70	PRSS3	0.94
S1PR3	-0.70	SLC8A1	0.94
NTS	-0.69	ATP8B1	0.94
CNNM1	-0.68	FAM129B	0.94
KRT18	-0.68	IRF1	0.94
PDGFRA	-0.68	SSTR5-AS1	0.93
CBFB	-0.67	NUAK2	0.93
ADCY9	-0.64	SPARC	0.93
PRELP	-0.63	FBLIM1	0.92
PIM1	-0.63	TNFRSF21	0.92
PPM1H	-0.62	RP5-875H18.9	0.92
EBNA1BP2	-0.60	KLF4	0.92
ATP9A	-0.59	SGK1	0.92
		FRMD3	0.91
		APOL1	0.91
		SEMA3A	0.90
		CYGB	0.90
		UNC13D	0.90
		RNF43	0.90
		BACE2	0.89



Downregulated genes		Upregulated genes	
Gene Symbol	log2FC	Gene Symbol	log2FC
		IL18	0.89
		PSMB9	0.89
		ABCA13	0.89
		RUNX2	0.89
		C11orf86	0.89
		FAR2	0.88
		TMEM159	0.88
		KLHL13	0.87
		PLEKHA2	0.87
		TRIM38	0.87
		EDIL3	0.86
		ANPEP	0.86
		CREG1	0.86
		FBXO32	0.86
		TRHDE	0.85
		NWD1	0.85
		IFFO2	0.85
		LXN	0.85
		LONRF3	0.85
		RAMP1	0.84
		HBEGF	0.84
		C15orf52	0.84
		KSR1	0.83
		SDHAP3	0.83
		BHLHE41	0.83
		CTD-2263F21.1	0.83
		KLF2	0.83
		L1CAM	0.81
		CTD-2562J17.6	0.81
		CCDC68	0.81
		UGCG	0.80
		LGALS3	0.78
		EHD1	0.77
		PLEK2	0.76
		CARD10	0.76
		RP11-382A20.3	0.75
		PRDM8	0.74
		FAM171A1	0.74
		FAM3C	0.74
		APOBEC3B	0.74
		COTL1	0.73
		BIN1	0.72
		CYB5R2	0.68
		CXXC5	0.67
		SH2B3	0.67

Downregulated genes		Upregulated genes	
Gene Symbol	log2FC	Gene Symbol	log2FC
		SLCO4A1	0.64
		DHRS3	0.62
		KRT19	0.60
		ABHD11	0.55

Supplementary Table S5: FOXA2-independently and -dependently upregulated genes ( $\log_2 > 1$ ,  $FDR < 0.05$ ,  $FPKM \geq 1$ ).

FOXA2-independent		FOXA2-independent (overlap)		FOXA2-dependent	
DNAJC6	TM7SF2	C16orf74	FAR2	PRPF40A	GAS1
MMP2	CYB5R2	FSTL1	MNX1.AS1	PDHA1	ANKRD12
CEP250	RAB12	B3GNT5	FAM171A1	KCNG1	STK32B
UGCG	TMEM25	MTARC2	MSI1	RGMB	EIF2S3
BTG2	NUAK1	ARHGEF10L	SRPX	PLAU	NCCRP1
AL390719.1	ZNF185	LGALSL	RASL10A	C10orf55	LYST
PTPRK	AL359091.4	NPAS2	SEC14L2	HMOX1	TAB3
C17orf97	KAZALD1	ARSI	SKIDA1	STARD8	CNRIP1
EDN1	RTL5	H2BC12	CLIP4	CDK14	SH3KBP1
MST1R	MB21D2	ADRA2C	AC002401.4	AC036176.1	AC012447.1
CYP11A1	SLC2A8	CU634019.2	CDKN1C	PADI3	NOG
FOXD2	AC099568.2	STOX1	PLEKHA2	LINC00665	CERS1
CCNO	ADAP2	LRP3	MAT1A	MARCKS	TRBV26OR9.2
MDGA1	FNDC10	HES2	SERINC2	ANKEF1	EIF2S3B
CSPG4	CCM2L	CADM4	PPM1H	AC006504.7	ZNF462
KCNH3	CCN2	GPC4	DIPK1A	HHEX	HAND2
SIGIRR	AP005329.1	SNX18P7	KIF1A	MAP3K21	AC090587.1
RHOD	LIMA1	WNK2	AC003965.2	TXLNG	SETP14
ADCY7	ABTB2	ADD3	ZNF738	FOXN3	SOX17
CU633904.2	MATN2	OBSL1	AL583856.1	UAP1L1	LINC02768
KREMEN2	CCDC157	GDF15	CYP27C1	MAP3K15	CLIP2
UBASH3B	ST6GALNAC2	GNAL	WT1	GEMIN8	UPK2
SUSD5	TGFBR2	TBL1X	ERO1B	AMN1	CPE
FOXD3.AS1	ORAI2	NID1	PRSS22	HCCS	SLC16A2
MOXD1	ERVK9.11	ANKRD18B	AL354714.2	FBLIM1	CCND1
RAB42	GRIP1	EPB41	ICA1	LRFN1	CEMIP2
FNDC4	AC069503.1	NKAIN1	EMILIN2	NUP62CL	SCNN1B
AC068533.2	H2BS1	SFRP5	COL6A2	MOSPD2	RNF157
KIRREL1	COL23A1	GSC	FZD7	AL162411.1	IL15RA
ZBTB12	WWC1	PHLDA3	OSBP2	EIF1AX	AC005081.1
KIF26A	CLDN23	AP002478.1	SLC16A9	FAM167B	ZNF703
KDM4D	CCN1	DSP	OAF	AP001033.4	ARSK
NKX2.8	GTPBP6	AL031058.1	LINC02041	AL513497.1	AP003068.4
EFNA5	CEBPA.DT	NIBAN2	PTGFRN	HK2	AC012146.1
LURAP1	CACNA1H	LPAR3	CPVL	SCML2	TRPS1
SLC29A4	BEX2	PCDH1	TNFAIP2	DHRXS	ACOX2
IFITM3	VWCE	PIR	FYN	STARD4	PRKAG2

FOXA2-independent		FOXA2-independent (overlap)		FOXA2-dependent	
MRAS	MIR631	IGSF3	AMOTL1	APOO	ARAP3
SOX13	CDKN1A	SASH1	FERMT1	RBBP7	PHYHD1
LINC01106	RIMS3	OCIAD2	PALD1	PLEKHG5	Z98745.2
CRIM1.DT	LINC02298	TMEM132E	GSDME	SHROOM3	AKAP17A
TSPAN12	LSR	B3GAT1	SGK1	FOXN3.AS1	FIGN
CD276	PPP1R3B	NR2F1	AL161772.1	SDC2	GAPLINC
KNDC1	KLF2	ERVMER34.1	HMGAI1P4	CNTFR	CPQ
KRT80	COL5A2	MYOM1	LGALS3BP	GLDC	CPPED1
PALM	ARHGEF17	ITGA3	ITPKA	AC020928.1	FAM155B
CNTNAP3C	SSPN	PRSS12	SET	FZD8	RBAK
LRRN2	ME3	DOLK	MMP25	SLC9A3	FOXA3
CD59	CHD5	PKN3	RHBDL3	MGST1	BOLA2B
KCNQ4	LPCAT2	GPR137B	CELSR2	KLF14	FANCB
NRP1	LNP1	TGM2	DSE	SETSIP	TNFRSF25
FIBCD1	ABCC3	CLDN4	MAP7	CR381653.1	ZNF793.AS1
SLC22A17	PELI3	PAPSS2	OSBPL3	MSX2	TTC30A
EBF4	MGAT3	MAP3K5	GPC3	AC019069.1	MCTP1
WDR34	ENPP1	SMKR1	TOR4A	MTCL1	RNLS
UTF1	LAPTM5	SMO	MAF	SH2D2A	ZNF165
KCNIP3	AVPI1	NUP188	HS6ST2	LMO4	HMSD
PLEK2	KRT8	HTR1B	C1orf115	POLA1	RIN1
TLN2	SIM2	AP1S2	SEC14L4	ZSWIM5	
SMIM3	CSRP2	PCGF5	SPHK1		
S1PR2	EPHA7	KCNJ4	TTC39C		
B3GALNT1	ABCA2	ALDOC	CTPS2		
LIMS2	SARM1	RIN3	TBC1D2		
MFG8	HMX3	PHKA2	ENO2		
DMKN	HAPLN3	AK4P3	ACBD7		
TYMSOS	NPTX1	AK4	BMP7		
DNASE2	ZNF488	ANO5	COL26A1		
YBX2	ANTXR1	AP005329.3	LMX1B		
SLC44A2	RASSF8	RTN4R	RRAGD		
LOXL2	TNRC6C.AS1	KLHL15	SAT2		
MYRF	RTKN	GALNT6	AL135978.1		
RGS3	ADM	ITGB4	MLXIPL		
DKK3	CAPG	BANF1P2	ENC1		
MIR1915HG	OOEP	TMEM159	GNG4		
STARD10	RAB15	ROR2	AL157893.1		
GAREM2	PHGDH	GPR161	TIMP2		
TMEM52	STAC	CORO2A	AK4P1		
CA11	AL359182.1	LARGE1	SYCE1L		
FBXO27	VAX1	CACHD1	PREX1		
C19orf18	AURKC	FBXL19.AS1	NEURL1B		
SYNGR3	ZDHHC22	SOCS6	GABBR2		
RNF208	MATN3	RTTN	SH3GLB2		
GPR27	FDXR	SIRPA	PPP1R14C		

FOXA2-independent		FOXA2-independent (overlap)		FOXA2-dependent	
EML1	CDCP1	FAS	COL6A1		
GNAZ	CITED4	PLEKHG4	SEMA7A		
TMEM45A	ARHGDI3	GSTA4	CFAP58.DT		
ELOVL3	LRRC75B	DHRS4.AS1	MTSS2		
AC027682.6	LHFPL6	TCAM1P	C19orf81		
ANKRD6	ADAMTSL4	NEO1	HOXA11		
LTBP4	SLC44A1	TMEM98	TUB		
TSPAN13	NKILA	STEAP3	PYGO1		
CLDN3	GLIPR2	ABCA1	MARCHF4		
RASL10B	PPP1R3C	LAMP3	SMS		
KREMEN1	CPM	DNM1	RIMS4		
SLFN12	PARD6G	PDGFA	IGFBP3		
CYGB	UBAC2.AS1	NPW	CALB2		
AP001025.1	P2RY2	SCRN2	ACTA2		
FOXE1	DLGAP1.AS1	MCOLN3	IFT57		
AP001922.5	WNT9A	TRABD2A	AC022210.2		
NAT8L	H19	ARFGF3	RND3		
CARD10	GPRC5A	TENM3	SCNN1G		
KCNAB2	ZNF362	PSD4	IGFBP4		
AC022137.3	DRAM1	ITPKB	AL360181.1		
S1PR1	TSPAN33	NTN1	SEMA6B		
SH3BP5	PPP2R2C	C2orf72	ACP7		
TIAM2	MRC2	MSX1	LKAAEAR1		
NPAS1	RGCC	ADAM19	BCAR3		
RAP1GAP2	DUSP15	OPRD1	NMT2		
AL133346.1	CD83	PRPS2	AC061992.2		
FZD5	IMPACT	BARX2	FAM89A		
ADORA1	TPM1	AL359538.3	PNMA3		
DAB2IP	LINGO1	RNF215	FAM20C		
ST6GALNAC4	BEGAIN	AGPAT3	MME		
AC244453.3	CFAP300	STMN3	AC099518.6		
IGFBP6	HHIPL1	EHD3	CAMK2N1		
AP000695.2	CABLES1	KCNJ12	GAS6		
SDC1	ARPIN	ADAM23	OSBPL6		
TGFBR3L	ACOT4	SOX8	FRMD4A		
LINC01843	CRLF1	DCST1.AS1	FAM71D		
SLC1A4	PTPRU	CDC42EP3	CYBRD1		
IFIT5	TIGD3	LRRC8A	SOCS3		
SPRN	SLC9A2	LHX6	CXCL16		
MELTF	SELENOV	NAV1	S100A16		
CHST15	METTL27	RYR1	MYO1D		
FRMD5	ARHGFE25	MOCOS	SHC3		
CREG1	VPS37D	PAQR5	EPHB2		
SLC16A5	AC072054.1	BACE2	GATM		
PLXNA1	CERCAM	LIF	WNT7B		
AC005476.2	PDZD4	CA2	JPH3		

FOXA2-independent		FOXA2-independent (overlap)		FOXA2-dependent	
LLGL2	FAM92A1P1	ZRSR2	YY2		
CD82	TESK2	PTGES	PPP4R4		
AC023043.1	PKP3	AC005393.1	AMIGO2		
VAX2	LRRC15	ARL4C	AC244669.1		
CRACD	SELENOM	MAPK13	CBLN2		
CLSTN3	CTSH	MBTPS2	ARHGAP22		
FZD2	AC020910.5	SCD5	SLC18B1		
FZD9	FZD1	AC023043.4	SOX18		
PRKAG2.AS1	PLAC8	RALGPS2	ZNF503		
IGDCC4	LCTL	COLGALT2	PGPEP1		
RGS11	DDR1	CENPVL3	ALDH5A1		
AKAP12	PMEPA1	SLC47A1	TBC1D10A		
ORAI3	MAFG.DT	SAPCD2	GPAT3		
SH2B2	ZBTB47	JAG1	HSPA12A		
FGF2	HLA.B	CORO2B	TWSG1		
C1QL1	NINJ1	SLC10A4	NPTXR		
EFHD1	KRT8P3	TERT	MACROH2A2		
COPZ2	MXRA5	LINC02643	AL353150.1		
BCAM	SEC61A2	TP73	MNX1		
RAP1GAP	TMSB4XP4	TMEM74B	AC061708.1		
PCK2	THSD4	WWC3	CCDC8		
MUC12.AS1	LRIG1	MPV17L	SLC2A1		
CHST11	YPEL1	TNNC1	PSME1		
OTULINL	ITGB8	UBE2L6	RETREG1		
VENTX	CCNJL	SPTBN2	BMP8B		
MMP15	RBP7	MAP2	EPCAM		
ASS1	HES4	ZNF853	ATP2A1.AS1		
PAOX	TMC6	TGFA	ADCY1		
SLC4A3	ODF3B	GALNT14	FHOD3		
SCARA3	ADRB2	RUNX1	MVB12B		
MICAL2	PMP22	NIBAN1	CU633906.2		
AC004130.2	CRAT	SMAGP	ZFR2		
GJB2	HIP1	MID1	CERS4		
PLXNB1	FZD4	SNX10	ADGRB1		
OSBPL1A	PIK3CD	WFDC2	BIN1		
IRS1	AIF1L	PADI2	ARNT2		
ICAM1	TMSB4X	LONRF2	ITPR3		
AL591895.1	LRRC8E	WIPF3	CBFA2T3		
ACSL1	RBPM52	ADSS1	MISP3		
MIRLET7I	ITPR2	CEBPA	ZNF618		
BRSK2	RAB36	FHL1			
PAQR8	TTC9				
CTSV	AC080112.2				
LTBP1	FOXE3				
AC016745.1	LMNTD2.AS1				
CCDC9B	PARP9				

FOXA2-independent		FOXA2-independent (overlap)		FOXA2-dependent	
IFFO2	LARP6				
IGFBP7	H1.2				
CDH2	LGR4				
ABHD8	TGFBI				
AC012360.2	FANK1				
FGFR3	OTUB2				
PRSS23	SLC4A11				
LINC01833	AC019171.1				
ADGRL2	PIANP				
CDX2	LRRC73				
TMSB4XP8	HEXA				
DSG2	BHLHE41				
DHRS4L2	GBX2				
KRT18	CHPF				
RASD1	COCH				
SOCS1	LOXL1.AS1				
MAFA	PXK				
ULBP3	TSTD1				
RBMS2	HEYL				
KCNQ5	LFNG				
CACNG4	DENND2D				
BHLHE40	CTSF				
SLC41A2	ZDBF2				
ENHO	ADPRHL1				
AC009237.3	MAFF				
ADCY6	AMOTL2				
TPBG	INSYN1				
ATP2B4	SFRP1				
OSR2	CEROX1				
RET	TMSB4XP1				
SEMA4G	P3H3				
FADS2	HEG1				
PHLDA2	C9orf135				
CSPG5	PROCR				
MGAT5B	CRACR2B				
KHK	TRIB2				
PARP14	TNS2				
TOM1L2	ADA				
CD70	REEP2				
MESP1	ZMYND10				
TOX2	SEMA3C				
TMEM200B	MGARP				
SRC	S100A13				
PLAAT1	TRIM62				
VASN	SAT1				
AQP7P4	RPS6KA3				

FOXA2-independent		FOXA2-independent (overlap)		FOXA2-dependent	
KIAA1549L	NXP2				
ZER1	TMEM270				
ELFN1	CHMP4C				
LOXL1	CAMK2N2				
TMEM121	AL132656.1				
CHST13	PPIL6				
ADORA2B	INAFM2				
PIM1	ARHGEF40				
AP001318.2	C1QL4				
FAM169A	AL132656.4				
TMEM17	N4BP3				
THBD	MAPK8IP1				
SRR	S1PR3				
MAP9	CKAP4				
CLPSL2	FAM174B				
EMP1	CACNA2D2				
TMEM171	OSBPL5				
TSPAN1	OLFM2				
HCK	WLS				

Supplementary Table S6: FOXA2-independently and -dependently downregulated genes ( $\log_2 < 1$ ,  $FDR < 0.05$ ,  $FPKM \geq 1$ ).

FOXA2-independent		FOXA2-independent (overlap)		FOXA2-dependent	
PLCE1	ITM2A	RCN3	GNG7	RFLNA	HTRA1
DECR1	LINC01394	GPAA1	IDH3G	RNA5SP202	MYO10
PRADC1	AL359834.1	ERGIC1	HOXA.AS2	AK1	ANPEP
ARHGAP31	ZNF711	MCTS1	FAM3A	L1CAM	MAP3K2.DT
AIFM1	PRPH	RTL8A	RPL10	SLCO4A1	CCDC184
RTN1	MPC1	MMGT1	CHCHD7	TACC2	LIPH
HMG5	AC083799.1	CNTD2	KRT19	COL9A3	GALNT18
RPL39P40	NR3C1	RPL10P9	CEP41	EGR4	FAM43A
AL133330.1	TWIST2	ID3	UBL4A	THEM6	SIX3
AC105285.1	AC084337.1	GZMM	AL137003.1	WBP1L	TMEM200B
NETO2	AC025171.5	FBXL6	GAB3	ZNF837	PTGER1
LDHD	SH3BGRL	IKBKGP1	RNF113A	ANXA1	KRT15
MARCKS	ZSWIM7	SOX4	MYPN	SUSD5	NROB2
DPYSL2	RPL7P23	NR4A2	C1QTNF12	EPB41L4B	ALPP
HOXA1	SNORD99	MAGED4B	FOXL1	CITED2	LRP12
NRN1	AC131392.1	RPL36A	SMPD1	MAN1A1	PLAGL1
SCFD2	BRWD3	MTMR1	JUP	RNVU1.7	AP001107.9
CDC14B	CUL4B	SYNE3	IKBKG	HMGB3	BATF3
DDIT4	AC146949.1	LINC01842	MAGED4	TSPAN15	KLHL29
DDN.AS1	MAGEL2	ATXN1.AS1	MEG3	TPSG1	GPRASP1
PCDHGB7	SNAP25	KIF25.AS1	MYMX	F2RL2	HTR1D
BTNL9	MIR3619	ENTPD6	ZNF516	MAN1B1.DT	SYBU

FOXA2-independent		FOXA2-independent (overlap)		FOXA2-dependent	
CSTF2	LIN28B	FOXC2	ATXN1	CD24	IFITM2
TGFB3	RAPGEF1	LYN	SPSB1	MED13L	CAMTA2
C7orf57	EXOSC4	AFAP1L2	SUMO3	SUSD2	GNAO1
UQCRBP1	AMMECR1	NAP1L3	TCEAL3	LGALS7	EMP1
MIER2	SNORD17	DIAPH2	MAML3	CEND1	IDS
PNMA2	SLC7A5P1	BASP1	SHH	ESAM	OLFML2A
TRIQK	BRCC3	PALLD	GUSBP16	NXPH3	TSPAN18
HIVEP2	SYNJ1	ATP9A	SV2A	AC108047.1	ZNF185
STAG2	TOMM40L	HOXA2	TSHZ1	ELFN2	SERPINE1
LINC00513	RADX	PYGB	INTS6L	KLF7	ZMYND12
SLC25A46	AL121772.1	AL645608.6	MPP1	IRAK2	CCDC88C
LYPLA1	AC036214.2	EOLA2	XIAP	MECP2	AC002480.1
ID4	AC091133.7	LINC02206	SH3BGRL2	SOX21	C2CD4C
GNG3	BNIP3L	BCORL1	AFAP1.AS1	UGDH	INKA1
LNPEP	SNHG12	LINC00261	DUSP1	KIAA0513	SVIL
SCG2	GUSBP9	TMEM185AP1	LINC01998	CYP2S1	MT1DP
SPTSSA	AC093525.6	ARMCX5	HOXA5	CNN1	CLIP3
RIPK4	CENPI	GSTT2B	GNA15	COL9A2	STOM
AC090543.3	MAPK8IP3	AFAP1	MROH6	LRRC24	ARL10
THOC2	ETFB	ZDHHHC9	ATRNL1	LAMB1	EGR2
ZNF280C	SECTM1	LINC00173	FOSB	MXRA8	RAMP1
SLC9A6	NES	FOXA2	NAA10	PGF	MYL9
VAT1L	SQLE	SLC02A1	NINJ2	PDLIM1	ALPG
AC002116.2	NEGR1	BOP1	RASL11A	CLIC3	B3GNT6
SLC25A5	NEFL	FOS	RASA3.IT1	COL1A1	LXN
AC064807.1	ZHX1	MIR1.1HG.AS1	SOBP	EPB41L2	RNVU1.29
FO393411.1	RPS26P3	DCBLD1	RTL8C	ANKH	RNU4.2
RNU7.38P	UPF3B	CSF1	GALNT7	P2RY6	LSR
NAP1L5	MAP7D3	DENND3	RPL39	KCND1	KCNH4
CLEC2B	BAG3	TSHZ3	PTP4A3	RHOA	AC234781.5
APOOL	C8orf49	SLC40A1	AC138866.1	SH3PXD2B	AC233723.1
YWHAQ	POLR3F	MACO1	SNORA12	CCN4	HSPA2
AC097059.1	SPARC	NKRF	CACNG6	MMP14	KRT8
SAP30	AGFG2	RPL39P3	PRUNE2	NFIB	C1GALT1C1
QPRT	TTC9B	DDN	UTP14A	SORBS1	RN7SKP203
STAT5B	FAM155A	RPL10P16	FUNDC2	ITGBL1	FSTL3
DUSP2	KLHL4	FOXC2.AS1	RN7SL3	CD99L2	ANTXR1
ATP11C	FOXF2	ARHGAP45	WNK1	TSPAN13	PKDCC
CAPN11	CD9	BCAP31	TMLHE	CACNB3	AC093001.1
SP140	RPL7P32	VAMP5	ITGA2	TAPBPL	SYNJ2
CXorf56	SNORA72	TNNI3	IFITM1	HOXC12	IL13RA1
NEFM	VAMP7	CCNE2	RHCG	SUSD3	MIRLET7BHG
NNAT	GAP43	TAZ	HAP1	NDRG1	PLPPR3
AC025171.1	ZSWIM3	ABCD1	TMEM187	RNU1.1	EHD2
ELOVL2	LYPLA1P3	ANK1	RTL8B	MT1L	H19
SERF1A	DKC1	ARID5B	ZNF696	AC254633.1	GPRC5A



FOXA2-independent		FOXA2-independent (overlap)		FOXA2-dependent	
ARMCX1	DNER	NDUFA1	TSPAN5	AC113174.1	ITGB3
HPRT1	EOMES	PLCH2	TXNIP	HCN2	CGN
PIP4P2	TRMT2B	GATA5	SLC10A3	NKX6.2	RGCC
SNORA3B	STMN2	AC116533.1	GATA4	PLEKHB1	DNMT3A
SMAD9	AL034346.1	RPS6KL1	ABLIM1	NR4A1	ESRP2
LINC00342	CRISPLD2	MFAP2	IRS3P	REEP5	FRAT2
SLC38A4	PPARGC1B	SMIM14	GHDC	GMPR	ABLIM3
CCDC136	ZFPM2.AS1	TNFRSF11A	RN7SL471P	F8A1	LINC00638
FLNC	RNF128	ZNF34	SSR4	SAMD11	BEGAIN
FAM27B	DEPP1	EOLA1	CIART	NPB	CABLES1
AC084125.4	GCNT1	WBP1LP2	ATG4A	F2RL1	GABARAPL1
AC004980.1	HLA.A	STAT5A	ZADH2	CDH1	CLDN10
PC	ACSL4	HDAC9	TIMM8A	PALM3	CHDH
AC120057.4	DPYSL3	AKR7A3	MOCS1	CYS1	BX640514.2
AC027031.2	AC012676.1	AHCYL1	AC109322.1	B3GNT7	RNA5SP474
OCRL	ONECUT3	HOXA3	FAM50A	JUNB	MT1A
BEX1	NRROS	LPIN3		FOXL2	RNU4.1
MRPL15	LINC01351			SERPING1	FAM107B
ZBTB10	NSDHL			CARMIL1	PTK6
AC100821.2	SLC25A43			RNA5.8SN1	SLC52A3
ID2	SMARCA1			ZNF449	KRT8P3
POP1	TRIM6			MIR4653	SYNPO
CSTF2T	AC138649.1			RNF144B	TRIB1
TNRC6C	PSD3			P4HA3	ADRB2
SLC29A1	PTDSS1			F8A2	CTHRC1
				RNU1.3	TSPY26P
				CLU	JCAD
				SEC14L1	RNA5.8SN2
				AL022162.1	SLC45A1
				LIPG	AMOT
				PHACTR2	TFPI2
				G6PD	MAGED1
				RN7SKP71	MLPH
				PDE5A	CES1P1
				SCARA3	CALHM2
				AL022341.2	DNASE1L1
				RCSD1	SLC4A11
				TMEM92	AP000759.1
				CACNG7	INHBB
				PGAP3	FP671120.4
				EFEMP2	PLOD2
				PARP10	LRP10
				F2R	DHRS3
				PID1	ALG1L13P
				AL355312.4	AC233266.2
				FRAT1	MID2

FOXA2-independent		FOXA2-independent (overlap)		FOXA2-dependent	
				GDI1	FLNA
				LTBP1	TMEM88B
				LAGE3	ADCY9
				ANKRD44	GALNT16
				APMAP	CCL2
				AC024560.3	RHBDL1
				TSC22D3	CES1
				GPRASP2	ITGB5
				RAET1G	FSD1
				AC110285.2	LRP11
				NRP2	TMPRSS2
				IGFBP7	RAB17
				RNU1.27P	TCEAL1
				NEXN	MEDAG
				SNORD113.3	RNU5E.4P
				LINC02593	BCL2A1
				HES7	PTGIS
				NACAD	NCR3LG1
				PCMTD1	HAGLR
				WDR44	EFR3B
				AL592295.4	AGR2
				SLFNL1.AS1	RPRML
				MEG8	HBA2
				TENT5C	PROC
				KRT18	RNU1.4
				F8A3	GZF1
				SYT11	PLXNA3
				CACNG4	

Supplementary Table S7: FOXA2-independently and -dependently upregulated (phospho)proteins (p-adj.<0.05).

FOXA2-independent	FOXA2-independent (overlap)	FOXA2-dependent
Vinculin	WIPI2	XPA
GSK-3B	Rab11FIP1	PAR
eEF2K	Merlin	AceCS1
GSK-3a-b	Paxillin	Jagged1
p38-a	IGFBP2	SCD
eEF2	GAPDH	PTPN12
Rb_pS807_S811	4E-BP1_pT37_T46	BRD4
Raptor	FASN	Sox17
Tau	MLKL	IRS2
Dvl3	MEK1	EphA2_pS897
CSK	B7-H3	Src_pY527
DRP1	Rad23A	VEGFR2_pY1175
Mitofusin-1	SGK1	ACLY_pS455
PKA-a	PKM2	SHP-2_pY542

FOXA2-independent	FOXA2-independent (overlap)	FOXA2-dependent
b-Catenin	VHL-EPPK1	MSH6
Akt2	PLC-gamma1_pS1248	4E-BP1_pS65
Chk2	CREB_pS133	PMS2
Rheb	PREX1	p90RSK_pT573
Ets-1	LDHA	FRS2-alpha_pY196
p44-42-MAPK	PAK1	RSK1
ALKBH5	PLK1	SIRP-alpha
	Aurora-A	LC3A-B
	Hexokinase-II	Cyclin-B1
	FGF-basic	Notch1
	PLC-gamma1	GSK-3a-b_pS21_S9
		Src
		EphA2_pY588
		DDR1
		Src_pY416
		Histone-H3_pS10
		HES1
		EphA2
		EMA
		IR-b
		LRP6_pS1490
		NDRG1_pT346
		PAK_pS474_S602_S560

Supplementary Table S8: FOXA2-independently and -dependently downregulated (phospho)proteins (p-adj.<0.05).

FOXA2-independent	FOXA2-independent (overlap)	FOXA2-dependent
TUFM	Smac	Myt1
PAR	XIAP	PKC-a-b-II_pT638_T641
CtIP	Stat5a	PI3K-p85
ATP5A	HMHA1	Annexin-I
TFAM	Hexokinase-I	HSP27_pS82
EphA2_pS897	Cyclin-E1	MIG6
Cyclophilin-F	Lyn	Mnk1
UQCRC2	Glucocorticoid-Receptor	PAX6
Synaptophysin	PD-L1	MMP14
ER-a	CD171	ASNS
Mcl-1	STING	p38-MAPK-_pT180_Y182
SOD2	Lck	PKCa
IGFRb	Complex-II-Subunit	Gli3
PAI-1	ATP5	DNA-Ligase-IV
ERRalpha	CD49b	Lasu1
Glutaminase	Snail	G6PD
MCT4	WTAP	CD5
Calnexin		Coup-TFII
CD44		E-Cadherin

FOXA2-independent	FOXA2-independent (overlap)	FOXA2-dependent
MTCO1		Caveolin-1
ATRX		YTHDF3
Grp75		PKC-b-II_pS660
XRCC1		VHL
Bim		GATA6
Rad50		PKC-delta_pS664
ERCC5		14-3-3-zeta
EphA2		VEGFR-2
p53		Ets-1
HSP60		
TRAP1		
SLC1A5		
MAPK_pT202_Y204		
SDHA		
Src_pY416		
Glutamate-D1-2		

Supplementary Table S9: Overlap of genes down- or upregulated in both LINC00261 KO cells and Src-overexpressing WT cells (p-adj.<0.05).

Genes downregulated in both LINC00261 KO cells and Src-overexpressing WT cells		
GZMM	HOXA5	ID2
Genes upregulated in both LINC00261 KO cells and Src-overexpressing WT cells		
NPAS2	CCDC9B	PKP3
ARSI	ARL4C	CALB2
UBASH3B	MAPK13	CCNJL
KRT80	JAG1	SEMA6B
CD59	RASD1	ACP7
TMEM132E	AC009237.3	CAMK2N1
NRP1	RUNX1	S100A16
PLEK2	SMAGP	TGFBI
ITGA3	TOX2	FANK1
PAPSS2	KIAA1549L	EPHB2
GALNT6	CLIP4	WNT7B
CYGB	THBD	PXK
AP000695.2	EMP1	SFRP1
CHST15	ZNF185	ARHGAP22
FRMD5	ENC1	GPAT3
CD82	PREX1	LTBP1
MARCHF4		

Supplementary Table S10: FOXA2-regulated (phospho)proteins (p-adj.&lt;0.05).

Downregulated (phospho)proteins		
ADAR1	Ets-1	PKCa
Akt2	FASN	PKCb
Annexin-I	FGF-basic	PKC-delta_pS664
Atg7	G6PD	PKCa
Axl	GAPDH	PKM2
b-Catenin	GATA6	PRAS40_pT246
Caveolin-1	HER3	PREX1
CD171	LDHA	Rheb
CD49b	MCT4	Smad3
CDK1_pT14	MEK1_pS217_S221	Stat3
CSK	MERIT40	Stat5a
E-Cadherin	Mnk1	STING
EMA	p38a	VEGFR-2
Enolase-1	p38 MAPK	
Upregulated (phospho)proteins		
4E-BP1_pS65	Glutaminase	PDL1
Bak	Grp75	SDHA
Bim	Hexokinase-II	SIRPalpha
BRD4	HSP60	Sox17
CD44	LC3A-B	Src_pY416
Cdc6	Lyn	Src_pY527
Complex-II	Mcl-1	TFAM
Cyclophilin	p53	TRAP1
EphA2	p90RSK_pT573	TUFM
EphA2_pS897	PAI-1	UQCRC2
EphA2_pY588	PAR	ZEB1

### 3. List of Figures

Figure 1: Structure and main components of the pancreas.....	1
Figure 2: Comparison of transcriptomic PDAC subtypes.....	4
Figure 3: LncRNA classification based on their genomic location.....	7
Figure 4: Genomic locus of LINC00261 and FOXA2 on chromosome 20.....	9
Figure 5: Search results in the PubMed® database.....	10
Figure 6: Process of epithelial-to-mesenchymal transition.....	12
Figure 7: Canonical TGFβ signaling. ....	14

Figure 8: c-Src signaling pathways.....	15
Figure 9: Analysis of LINC00261 expression in PDAC. ....	40
Figure 10: LINC00261 expression correlates with PDAC stage, grade and patient survival. ....	41
Figure 11: Relative gene expression of LINC00261 in various PDAC cell lines. ....	41
Figure 12: Evolutionary sequence conservation of LINC00261.....	42
Figure 13: LINC00261 is a non-coding RNA that localizes to both the nucleus and cytoplasm.....	43
Figure 14: Downregulation of LINC00261 by using two different CRISPR approaches. ....	44
Figure 15: Proliferation is not substantially altered in LINC00261 <sup>low</sup> cells. ....	45
Figure 16: Enhanced cell migration and invasion after LINC00261 downregulation using CRISPRi. .....	46
Figure 17: Enhanced migration and invasion after LINC00261-depletion using CRISPR/Cas9. ....	46
Figure 18: Reduced invasion and migration of LINC00261 overexpressing PATU-T cells.....	47
Figure 19: LINC00261 expression inversely correlates with EMT-related gene expression. ....	48
Figure 20: RNA-Seq analysis revealed an enrichment of EMT gene set in LINC00621 <sup>low</sup> Panc-1 cells. .....	48
Figure 21: LINC00261 regulated E-cadherin expression.....	50
Figure 22: Canonical TGF $\beta$ signaling is not responsible for CDH1 downregulation in LINC00261 <sup>low</sup> cells. ....	51
Figure 23: FOXA2 expression is significantly downregulated in LINC00261 <sup>low</sup> cells. ....	52
Figure 24: LINC00261 and FOXA2 are positively correlated in PDAC samples and cell lines. ....	53
Figure 25: Impact of FOXA2 and LINC00261 expression on the survival of PDAC patients.....	54
Figure 26: LINC00261 regulation by its genomic neighbor FOXA2.....	55
Figure 27: Establishment of FOXA2 KO cells and LINC00261-FOXA2 double KO cells. ....	56
Figure 28: Proliferation of WT, FOXA2 KO, LINC00261 KO and LINC00261-FOXA2 DKO Panc-1 cells. .....	57

---

Figure 29: Clonogenic assay with WT, FOXA2 KO, LINC00261 KO and double KO Panc-1 cells.....	57
Figure 30: Cell invasion and migration of newly established Panc-1 cell lines.....	58
Figure 31: Overexpression of FOXA2 in PATU-T cells enhances cell invasion and migration.....	59
Figure 32: RNA-sequencing reveals FOXA2-dependent and -independent LINC00261 target genes. .....	60
Figure 33: Determination of FOXA2-independently and -dependently regulated (phospho)proteins by RPPA.....	60
Figure 34: GSEA analysis of FOXA2-independently regulated LINC00261 target genes.....	61
Figure 35: STRING pathway analysis of FOXA2-independently deregulated (phospho)proteins. ...	62
Figure 36: Western Blots showing FOXA2-independent regulation of c-Src kinase signaling by LINC00261.....	63
Figure 37: c-Src overexpression activates FAK/c-Src/PXN axis and induces migration and invasion. .....	64
Figure 38: The migratory and invasive effects observed in LINC00261 KO were reversed by Src inhibition. ....	65
Figure 39: Possible role of c-Src-induced RUNX1 expression in transcriptomic reprogramming. .	66
Figure 40: Regulation of CDH1 by FOXA2 and LINC00261. ....	67
Figure 41: Pulldown of RNA-binding proteins with biotinylated LINC00261 fragments. ....	69
Figure 42: GSEA analysis of FOXA2-regulated genes. ....	70
Figure 43: Loss of FOXA2 negatively regulates TGF $\beta$ hallmark gene set.....	70
Figure 44: Attenuation of canonical TGF $\beta$ signaling in FOXA2 KO cells.....	71
Figure 45: Depletion of FOXA2 leads to significant reduction in SMAD signaling.....	72
Figure 46: FOXA2 overexpression modulates TGF $\beta$ signaling.....	73
Figure 47: Rescue of FOXA2 recovers TGF $\beta$ signaling in FOXA2-deficient cells. ....	74
Figure 48: TGF $\beta$ treatment downregulates FOXA2 post-transcriptionally and/or -translationally.	74

Figure 49: Impact of LINC00261 and FOXA2 KO on primary tumor growth in an orthotopic mouse model.....	75
Figure 50: Microscopic images of primary tumors originating from established cell lines. ....	76
Figure 51: Metastasis of tumor cells to liver and lung analyzed by luminescence measurement. .	76
Figure 52: Microscopic images of metastases in the liver.....	77
Figure 53: Overview of effects of FOXA2 and/or LINC00261 depletion in PDAC cells.....	94

#### 4. List of Tables

Table 1: Pancreatic cancer stages.....	3
Table 2: Classification of ncRNAs.....	6
Table 3: Instruments.....	18
Table 4: Consumables.....	20
Table 5: Chemicals and reagents.....	20
Table 6: Kits and systems.....	21
Table 7: Plasmids.....	22
Table 8: Cloning reagents.....	23
Table 9: Oligonucleotides for cloning.....	24
Table 10: Oligonucleotides for qRT-PCR.....	24
Table 11: Small interfering RNAs (siRNAs).....	25
Table 12: Single guide RNAs (sgRNAs).....	25
Table 13: Primary antibodies.....	25
Table 14: Secondary antibodies.....	26
Table 15: Cell culture media.....	26
Table 16: Lysis and wash buffers.....	26



Table 17: Western blot buffers.....	27
Table 18: Software .....	28
Table 19: Online tools and databases .....	28
Table 20: Overview of the main findings of this study. ....	77

## 5. List of Abbreviations

Abbreviation	Full Form
%	Percentage
°C	Degree celsius
APS	Ammonium persulfate
bp	Base pair(s)
BSA	Bovine serum albumine
ChIP	Chromatin immunoprecipitation
CRISPR	Clustered regularly interspaced short palindromic repeats
c-Src	Cellular sarcoma kinase
CT	Cycle threshold
DAPI	4'.6'-Diamine-2-phenylindole
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DPBS	Dulbecco's phosphate-buffered saline
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
E. coli	Escherichia coli
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EGFP	Enhanced green fluorescent Protein
EMT	Epithelial-mesenchymal transition
ENCODE	Encyclopedia of DNA elements
FACS	Fluorescence activated cell sorting
FAK	Focal adhesion kinase
FBS	Fetal bovine serum
FOXA2	Forkhead box protein A2
FPKM	Fragment per kilo base per million mapped reads
g	Gram
<i>g</i>	Gravitational force
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GPCR	G protein-coupled receptor
h	Hour(s)
H <sub>2</sub> O	Dihydrogen monoxide. water
HLB	Hypotonic lysis buffer

Abbreviation	Full Form
lncRNA	Long intergenic non-coding RNA
JAK	Janus kinase
kb	Kilobases
KCl	Potassium chloride
KD	Knockdown
kDa	Kilodalton
kg	Kilogram
KO	Knockout
LB medium	Luria Bertani medium
LINC00261	Long intergenic non-protein coding RNA 261
lincRNA	Long intergenic RNA
lncRNA	Long non-coding RNA
log	Logarithm
M	Molar
MALAT1	Metastasis Associated Lung Adenocarcinoma Transcript 1
MAPK	Mitogen-activated protein kinase
mg	Milligram
MgCl <sub>2</sub>	Magnesium Chloride
min(s)	Minute(s)
miRNA	MicroRNA
ml	Milliliter
mM	Millimolar
MMP	Matrix metalloproteases
mRNA	Messenger RNA
MS	Mass spectrometry
MW	Molecular weight
MWS	Modified Wuarin-Schibler
n	Number of replicates
NaCl	Sodium chloride
NCBI	National Center for Biotechnology Information
ncRNA	Non-coding RNA
Neg Ctrl	Negative control
NFκB	Factor κ-light-chain-enhancer of activated B Cells
ng	Nanogram
nt	Nucleotides
ORF	Open reading frame
P	p-value
P/S	Penicillin/Streptomycin
PAA	Polyacrylamide
p-adj.	Adjusted p-value
PAGE	Polyacrylamide gel electrophoresis
PAGE	polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
pH	Negative decimal logarithm of the hydrogen ion concentration
PMSF	Phenylmethylsulfonyl fluoride
PPIA	Peptidylprolyl isomerase A
qRT-PCR	Reverse transcription - quantitative PCR
R	Correlation coefficient
RAS	Rat sarcoma virus

Abbreviation	Full Form
RBP	RNA binding proteins
RIP	RNA Immunoprecipitation
RIPA	Radioimmunoprecipitation assay buffer
RNA	Ribonucleic acid
RNAi	RNA interference
rpm	Revolutions per minute
RT	Reverse transcription
RT	Room temperature
SD	Standard deviation
SDS	Sodium dodecyl sulfate
sec	Second(s)
SEM	Standard error of mean
siRNA	Small interfering RNA
SMAD	Sma- and mad-related protein
snRNA	Small nuclear RNA
SOV	Sodium orthovanadate
ss	Single-stranded
STAT	Signal transducer and activator of transcription
TBS	Tris-buffered saline
TBS-T	Tris-buffered saline and Tween 20
TCGA	The cancer genome atlas
TEMED	Tetramethylethylenediamine
TGF $\beta$	Transforming growth factor $\beta$
TP53	Tumor protein p53
UV	Ultraviolet
WHO	World health organization
$\mu$ g	Microgram
$\mu$ l	Microliter
$\mu$ M	Micromolar
TF	Transcription factor
MET	Mesenchymal-epithelial transition
SNAI1/2	Drosophila melanogaster homologues snail 1 and 2
TWIST1	Twist family bHLH transcription factor 1
ZEB1/2	Zinc-finger E-box-binding homeobox proteins 1/2
GSEA	Gene set enrichment analysis
TGFBR1/2	TGF $\beta$ receptor
SBE	SMAD-binding element
EGFR	Epidermal growth factor receptor
FFPE	Formalin-fixed paraffin-embedded
WT	Wild-type
NEAT1	Nuclear enriched abundant transcript 1
XIST	X-inactive specific transcript
KRAS	Kirsten rat sarcoma virus
CDKN2A	Cyclin-dependent kinase inhibitor 2A
IPMN	Intraductal papillary mucinous neoplasms
MCN	Mucinous cystic neoplasms
ICGC	International cancer genome consortium
HNF1 $\alpha$	Hepatocyte nuclear factor 1 homeobox A
3'	3-Prime
5'	5-Prime

## EIDESSTATTLICHE ERKLÄRUNG

Hiermit erkläre ich, dass ich meine Dissertationsschrift selbständig und ohne fremde Hilfe verfasst habe. Ich habe keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt. Die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen habe ich als solche kenntlich gemacht.

Mit der vorliegenden Arbeit bewerbe ich mich erstmals um die Erlangung des Doktorgrades.

Datum, Unterschrift \_\_\_\_\_

## DANKSAGUNG

Aus Datenschutzgründen entfernt.

# CURRICULUM VITAE

## Personal Details

Name: Agnes Dorn

Date of birth, birthplace: **Aus Datenschutzgründen entfernt.**

Nationality: **Aus Datenschutzgründen entfernt.**

## Education

1998-2002: Grundschule Herzoghöhe, Bayreuth

2002-2011: Gymnasium Christian-Ernestinum, Bayreuth – Abitur

## Scientific Education

2011-2012: Studies in Food Chemistry at Martin Luther University Halle-Wittenberg

2012-2015: B.Sc. studies in Biochemistry at Martin Luther University Halle-Wittenberg

2015: Bachelor thesis under supervision of Prof. Dr. Guido Posern at Martin Luther University Halle-Wittenberg; Title: The Regulation of the MRTF Pathway by MICAL2

2015-2018: M.Sc. studies in Biochemistry at Martin Luther University Halle-Wittenberg

2018: Master thesis under supervision of Prof. Dr. Faramarz Dehghani at Martin Luther University Halle-Wittenberg; Title: Immunohistochemical and Functional Characterization of Cell and Tissue Cultures of Primary Human Glioblastomas

2018: Erasmus internship at Karolinska Institute, Stockholm (Sweden)

2019-2024: Doctoral studies in the laboratory of Prof. Dr. Monika Hämmerle at Martin Luther University Halle-Wittenberg

2022: DAAD research visit at MD Anderson Cancer Center, Houston (USA)

## LIST OF PUBLICATIONS AND PRESENTATIONS

### Publications:

**Dorn A.**, Glaß M., Neu CT., Heydel B., Hüttelmaier S., Gutschner T., Haemmerle M. (2020). LINC00261 Is Differentially Expressed in Pancreatic Cancer Subtypes and Regulates a Pro-Epithelial Cell Identity. *Cancers (Basel)*, 12(5):1227.

Glaß M., **Dorn A.**, Hüttelmaier S., Haemmerle M., Gutschner T. (2020). Comprehensive Analysis of LincRNAs in Classical and Basal-Like Subtypes of Pancreatic Cancer. *Cancers (Basel)*, 12(8):2077.

### Presentations:

05/2022: Poster Presentation at the 7th Research Day of the Medical Faculty in Halle, Title: Deciphering the Role of LINC00261 and FOXA2 as Transcriptional Regulators Shaping the Expression Landscape of Pancreatic Cancer

05/2021: Presentation at the RNA Club Seminar Series in Halle, Title: Deciphering the Molecular Function of LINC00261 in Pancreatic Cancer

12/2019: Poster Presentation at the 6th Research Day of the Medical Faculty in Halle, Title: Long non-coding RNA LINC00261 Plays a Co-Regulatory, Tumor-Suppressive Role in Pancreatic Cancer

02/2019: Poster Presentation at the Non-Coding Genome Workshop in Paris, Title: Differential Expression and Subtype-Specific Function of LncRNAs in Pancreatic Adenocarcinoma

09/2017: Presentation at the 2nd Science Retreat of the Institute of Anatomy and Cell Biology, Title: Immunohistochemical and Functional Characterization of Cell and Tissue Cultures of Primary Human Glioblastomas