

Article

# Comprehensive Analysis of lincRNAs in Classical and Basal-Like Subtypes of Pancreatic Cancer

Markus Glaß <sup>1</sup>, Agnes Dorn <sup>2</sup>, Stefan Hüttelmaier <sup>1</sup> and Monika Haemmerle <sup>2,\*</sup>  
and Tony Gutschner <sup>3,\*</sup>

<sup>1</sup> Institute of Molecular Medicine, Section for Cell Biology, Medical Faculty, Martin-Luther University Halle-Wittenberg, 06120 Halle/Saale, Germany; markus.glass@medizin.uni-halle.de (M.G.); stefan.huettelmaier@medizin.uni-halle.de (S.H.)

<sup>2</sup> Institute of Pathology, Section for Experimental Pathology, Medical Faculty, Martin-Luther University Halle-Wittenberg, 06120 Halle/Saale, Germany; agnes.dorn@uk-halle.de

<sup>3</sup> Junior Research Group 'RNA Biology and Pathogenesis', Medical Faculty, Martin-Luther University Halle-Wittenberg, 06120 Halle/Saale, Germany

\* Correspondence: monika.haemmerle@uk-halle.de (M.H.); tony.gutschner@uk-halle.de (T.G.)

† These authors contributed equally to this paper.

Received: 19 May 2020; Accepted: 23 July 2020; Published: 27 July 2020



**Abstract:** Pancreatic ductal adenocarcinomas (PDAC) belong to the deadliest malignancies in the western world. Mutations in TP53 and KRAS genes along with some other frequent polymorphisms occur almost universally and are major drivers of tumour initiation. However, these mutations cannot explain the heterogeneity in therapeutic responses and differences in overall survival observed in PDAC patients. Thus, recent classifications of PDAC tumour samples have leveraged transcriptome-wide gene expression data to account for epigenetic, transcriptional and post-transcriptional mechanisms that may contribute to this deadly disease. Intriguingly, long intervening RNAs (lincRNAs) are a special class of long non-coding RNAs (lncRNAs) that can control gene expression programs on multiple levels thereby contributing to cancer progression. However, their subtype-specific expression and function as well as molecular interactions in PDAC are not fully understood yet. In this study, we systematically investigated the expression of lincRNAs in pancreatic cancer and its molecular subtypes using publicly available data from large-scale studies. We identified 27 deregulated lincRNAs that showed a significant different expression pattern in PDAC subtypes suggesting context-dependent roles. We further analyzed these lincRNAs regarding their common expression patterns. Moreover, we inferred clues on their functions based on correlation analyses and predicted interactions with RNA-binding proteins, microRNAs, and mRNAs. In summary, we identified several PDAC-associated lincRNAs of prognostic relevance and potential context-dependent functions and molecular interactions. Hence, our study provides a valuable resource for future investigations to decipher the role of lincRNAs in pancreatic cancer.

**Keywords:** cancer; lincRNA; lncRNA; miRNA; pancreatic ductal adenocarcinoma; RBP

## 1. Introduction

Pancreatic cancer with its most common type, pancreatic ductal adenocarcinomas (PDAC), is currently the fourth leading cause of cancer-related deaths in developed countries [1]. Dismal prognosis of disease has several reasons. First, pancreatic cancer is typically diagnosed in late stages due to non-specific symptoms or a lack thereof as well as the unavailability of sensitive and specific biomarkers in conjunction with difficulties in imaging early stage tumours. Second, PDAC is a highly aggressive disease with extensive local growth and early distant metastases, precluding

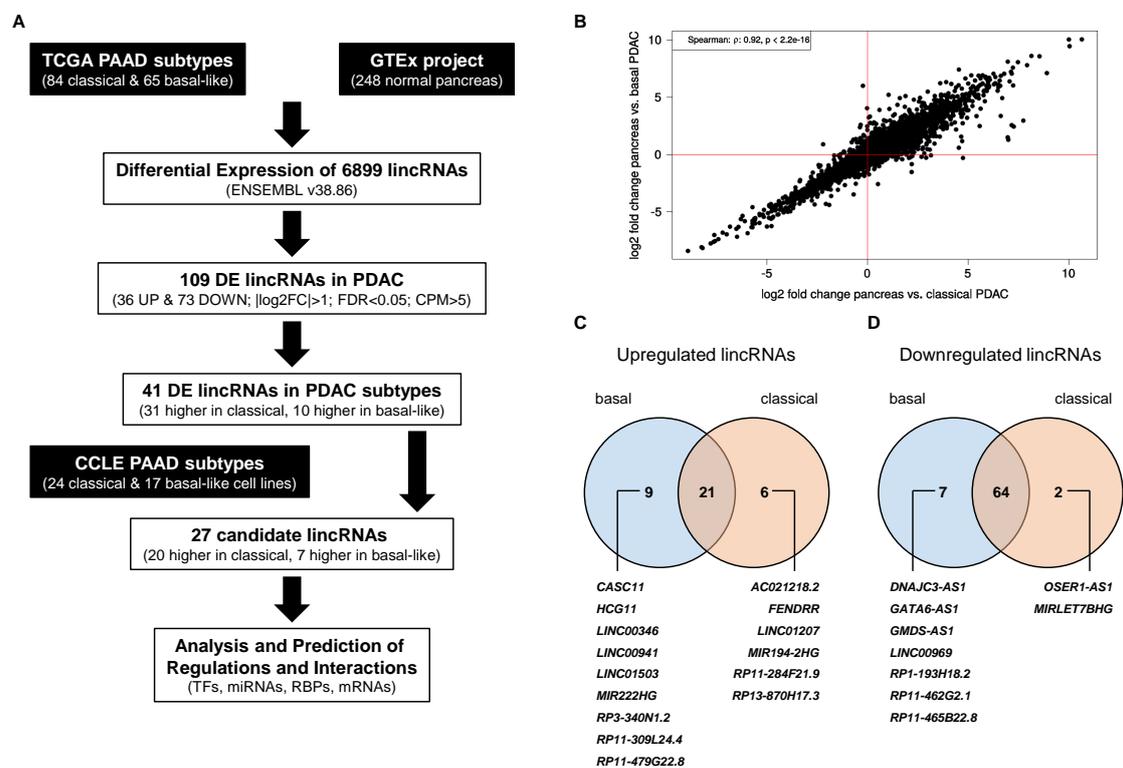
many patients from surgery. Third, pancreatic cancer is characterized by a high degree of resistance to currently available treatment options, i.e., chemotherapy, radiotherapy, or targeted therapies. All these factors result in a 5-year survival rate below 5%, if surgical resection is not possible [2]. If prognosis will not improve, pancreatic cancer is predicted to be the second leading cause of death in the next decade [1]. Molecularly, most PDACs are associated with somatic mutations, particularly affecting the KRAS, TP53, CDKN2A, and SMAD4 genes. Especially, KRAS mutations occur in more than 90% of all tumour samples [3,4]. In transgenic mouse models, activating mutations in the Kras gene were shown to be sufficient to induce PDAC [5]. Even in the few KRAS wild-type tumours, somatic mutations activating the RAS-MAPK pathway up- or downstream of KRAS have been identified [3]. However, these known and frequent mutations did not lead to clinically valuable tumour classifications accounting for different patient survival and therapy response, suggesting that origins of PDAC heterogeneity may be found at the post-genetic level [6]. Multiple transcriptomic studies provided valuable insights into pancreatic carcinogenesis and identified disease subtypes of prognostic relevance. At least three subtype classifications based on transcriptomic data have been proposed. Moffitt et al. [7] described two subtypes obtained via non-negative matrix factorization (NMF) of microarray data; a basal-like subtype, generally associated with worse outcome and a classical subtype. Furthermore, they derived two classes of stroma subtypes, normal and activated, yielding four classes for distinguishing primary tumour samples. Collisson et al. [8] suggested three subtypes also generated by NMF of microarray data: classical, quasi-mesenchymal and exocrine-like. Notably, the gene signature derived from the classical subtype showed high overlap with the signature derived from the classical subtype of Moffitt et al. [7]. Moreover, Bailey et al. [9] identified four subclasses by applying NMF to RNA-seq and microarray data of pancreatic cancer samples: squamous, pancreatic progenitor, immunogenic and aberrantly differentiated endocrine exocrine (ADEX). At least part of the differences of classifications are a consequence of different input material, either using bulk tumour tissue containing cells of the tumour microenvironment [7,9] or microdissected tumour epithelium [8] resulting in differences in tumour purity. In line with this, The Cancer Genome Atlas (TCGA) research network applied all three classification systems to their own RNA expression data and concluded that high purity tumours can be consistently classified into a basal-like/squamous group and a classical/progenitor group. In contrast, the strong association of immunogenic or ADEX as well as exocrine-like or quasimesenchymal subtypes with low purity samples within the TCGA cohort suggested that these subtypes may reflect gene expression from non-neoplastic cells [3]. Furthermore, the TCGA analysis found that the basal-like and classical subtype are characterized by expression differences of non-coding RNAs (ncRNAs), including microRNAs (miRNAs) and long non-coding RNAs (lncRNAs). LncRNAs can function as regulators of gene expression acting on the epigenetic, transcriptional as well as on the post-transcriptional level [10,11]. Several lncRNAs have been implicated in pancreatic cancer acting, e.g., as competitive endogenous RNAs (ceRNAs) that interfere with miRNA-dependent regulation [12–15] or as part of lncRNA signatures predicting PDAC patient survival [16–19]. However, an in-depth characterization of the lncRNA landscape of different PDAC subtypes, including a description of their potential interactions as well as their correlation with other protein-coding and non-coding genes, has not been systematically performed yet. Here, we aim to provide a comprehensive overview of a class of lncRNAs called long intervening RNAs (lincRNAs), which are lncRNAs that do not overlap exons of either protein-coding or other non-lincRNA types of genes [20]. LincRNAs are often transcribed by RNA Polymerase II and thus are usually capped and polyadenylated allowing their detection in large-scale RNA-sequencing-based cancer transcriptome data sets commonly generated using a polyA-selection step. Importantly, due to their non-overlapping genomic localization lincRNA expression patterns are easier to interpret than those of transcripts from loci overlapping other gene classes. Furthermore, they are amenable to specific genetic perturbations using genome engineering tools to study their cellular and organismal functions [21,22]. Hence, our analysis highlights disease-associated lincRNAs as well as their putative

molecular interactions and thereby provides a valuable resource and an ideal starting point for future investigations to expand our knowledge about the onset and progression of pancreatic cancer.

## 2. Results

### 2.1. LincRNA Expression Is Altered in PDAC Compared to Normal Pancreas

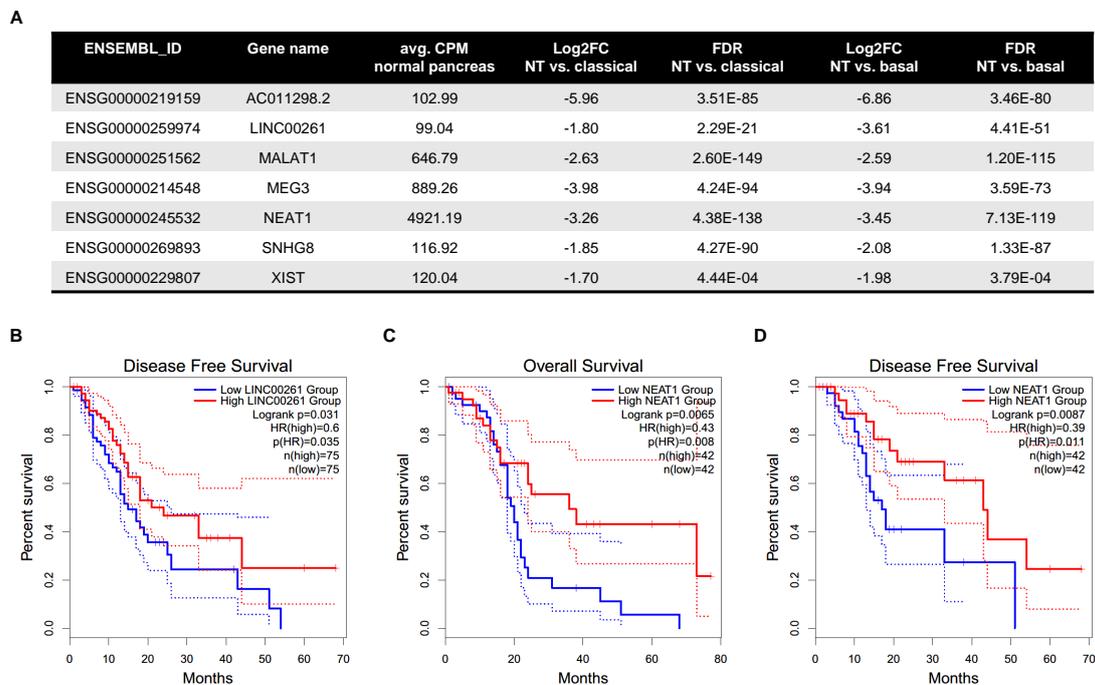
To gain insights into the expression landscape of lincRNAs in pancreatic cancer, we obtained RNA sequencing data from The Cancer Genome Atlas Pancreatic adenocarcinoma (TCGA PAAD) project [3]. However, since these data contained only four normal tissue samples, we compared the primary tumour samples with pancreas RNA-seq data from the Genotype-Tissue Expression (GTEx) project [23]. We used raw count data from both projects and processed them together to avoid RNA composition biases (see Methods). Altogether, we compared 248 normal pancreas samples against 84 classical subtype and 65 basal-like subtype tumour samples (Figure 1A). Gene set enrichment analyses (GSEA, [24]) using the fold changes obtained from the differential expression analyses showed pancreatic cancer related gene sets among the most significantly enriched sets supporting the cogency of this comparison (Figure S1A).



**Figure 1.** Analysis of lincRNA expression in PDAC (pancreatic ductal adenocarcinomas). (A) Flow-chart outlining the analysis pipeline to identify differentially expressed lincRNAs in pancreatic cancer using TCGA (The Cancer Genome Atlas) tumour samples; (B)  $\log_2$  fold changes of 6899 annotated lincRNAs in the comparison between normal pancreas tissue and classical PDAC (x-axis) and between pancreas tissue and PDAC samples categorized as basal-like subtype (y-axis). (C) Venn diagram depicting the overlap of lincRNAs upregulated in basal-like and classical tumour subtypes, respectively. (D) Venn diagram showing the overlap of lincRNAs downregulated in basal-like and classical PDAC subtypes. In total, 109 lincRNAs fulfilled the selection criteria ( $|\log_2 FC| > 1$ ; False Discovery Rate (FDR)  $< 0.05$ ; avg. Counts per million (CPM)  $> 5$  in normal pancreas or the respective tumour subtype). PAAD: Pancreatic adenocarcinoma; GTEx: Genotype-Tissue Expression; CCLC: Cancer Cell Line Encyclopedia; TFs: Transcription factors; RBPs: RNA-binding proteins.

Next, we gathered expression data of 6899 individual lincRNA genes (Table S1) and compared their respective tumour subtype-specific expression to normal pancreatic tissue. In general, the  $\log_2$  fold changes observed in both expression analyses showed a strong and highly significant correlation Spearman  $\rho = 0.92$ ,  $FDR < 2 \times 10^{-16}$ ; Figure 1B) indicating similar expression trends of lincRNAs in both subtypes. However, application of filters for fold change ( $|\log_2FC| > 1$ ), significance (False Discovery Rate (FDR)  $< 0.05$ ) and absolute expression level (average counts per million (CPM)  $\geq 5$  in either pancreas or the respective tumour samples of each subtype) revealed subtype-specific differences and identified a total of 109 differentially expressed lincRNAs (Figure S1B,C). Of note, the number of downregulated lincRNAs was more than twice the number of upregulated lincRNAs in both comparisons. Specifically, in tumours assigned to the classical subtype 93 lincRNAs showed a significantly altered expression level with 27 being up- and 66 being downregulated. In the basal-subtype group 101 lincRNAs were changed and 30 lincRNAs had a higher whereas 71 lincRNAs had a lower expression in the tumour tissue. Comparison of both lists of differentially expressed lincRNAs identified commonly deregulated transcripts as well as some candidates altered in a subtype-specific manner (Figure 1C,D). However, a closer examination of the latter cases revealed similar expression trends in the large majority of these lincRNAs yet their absolute expression level or their fold change prevented their inclusion in the list of differentially expressed lincRNAs in one subtype, but not the other. For example, FENDRR is highlighted as a lincRNA that seemed to be specifically upregulated in the classical subtype (Figure 1C). However, its expression was also significantly increased in the basal-like subtype ( $\log_2FC = 1.49$ ,  $FDR = 5.56 \times 10^{-23}$ ), but due to its low average expression level, which exceeded 5 CPM only in the classical subtype, FENDRR was not included in the list of deregulated lincRNAs in basal-like tumours besides showing the same overall expression trend. Nevertheless, further analysis identified three lincRNAs (CASC11, RP11-465B22.8, LINC01207) out of the initial 109 which showed a significant expression change in one subtype compared to normal pancreas that was not observed in the other subtype. In detail, CASC11 was strongly upregulated in the basal-like subtype ( $\log_2FC = 6.01$ ;  $FDR = 2.60 \times 10^{-129}$ ), but slightly, yet not significantly downregulated in the classical subtype ( $\log_2FC = -0.2$ ;  $FDR = 0.54$ ) whereas RP11-465B22.8 showed a decreased expression ( $\log_2FC = -1.14$ ;  $FDR = 1.84 \times 10^{-14}$ ) specific to the basal-like subtype only. In contrast, the upregulation of LINC01207 ( $\log_2FC = 2.24$ ;  $FDR = 6.35 \times 10^{-33}$ ) was specific to the classical subtype whereas its expression virtually did not change in the basal-like subtype. These specific expression patterns suggest that these lincRNAs might have context-dependent functions. In contrast, our analysis identified 85 lincRNAs that were commonly deregulated in both subtypes (21 up-/64 downregulated) suggesting a rather general role for the biology of PDAC. Hence, these lincRNAs would be interesting candidates to follow-up on. In order to further narrow down the list of 85 lincRNAs, we applied a more stringent filter to retrieve candidates with a robust expression level (average CPM  $> 99$ ). This left us with seven differentially regulated lincRNAs, which are all commonly downregulated in tumours compared to normal pancreas (Figure 2A). The strongest decrease was observed for AC011298.2 ( $\log_2FC = -5.96/-6.86$ ), a lincRNA of unknown function that was previously included in a six-lincRNA signature to predict survival of patients with bladder urothelial carcinoma [25]. Moreover, we observed a decrease in the expression of SNHG8, a host gene for a small nucleolar RNA (snoRNA), which was recently shown to be increased in PDAC, associated with an adverse prognosis [26]. In addition, we confirmed the reduced expression of LINC00261 and MEG3, which are thought to function as tumour-suppressive lincRNAs in pancreatic cancer, and especially in neuroendocrine neoplasms in the case of MEG3 [27–31]. We also detected a strong reduction of XIST, NEAT1 and MALAT1, three conserved nuclear lincRNAs with well-known links to a broad range of cancer types [32–35]. To further evaluate the clinical and prognostic relevance of these seven highly expressed and deregulated lincRNAs we performed survival analyses using the GEPIA2 data portal ([gepia2.cancer-pku.cn](http://gepia2.cancer-pku.cn) [36]). This analysis revealed striking associations between the expression of LINC00261 and NEAT1 with overall (OS) and/or disease-free survival (DFS) in TCGA PDAC patient samples. In detail, low expression of LINC00261 was predictive of

poor DFS (Figure 2B) across all 150 samples, whereas low levels of NEAT1 were associated with worse OS (Figure 2C) and DFS (Figure 2D) in classical subtype samples. Hence, based on their abundance, significant downregulation and their prognostic relevance, LINC00261 and NEAT1 might be interesting candidates for functional follow-up studies to dissect their cellular and molecular functions in pancreatic cancer.



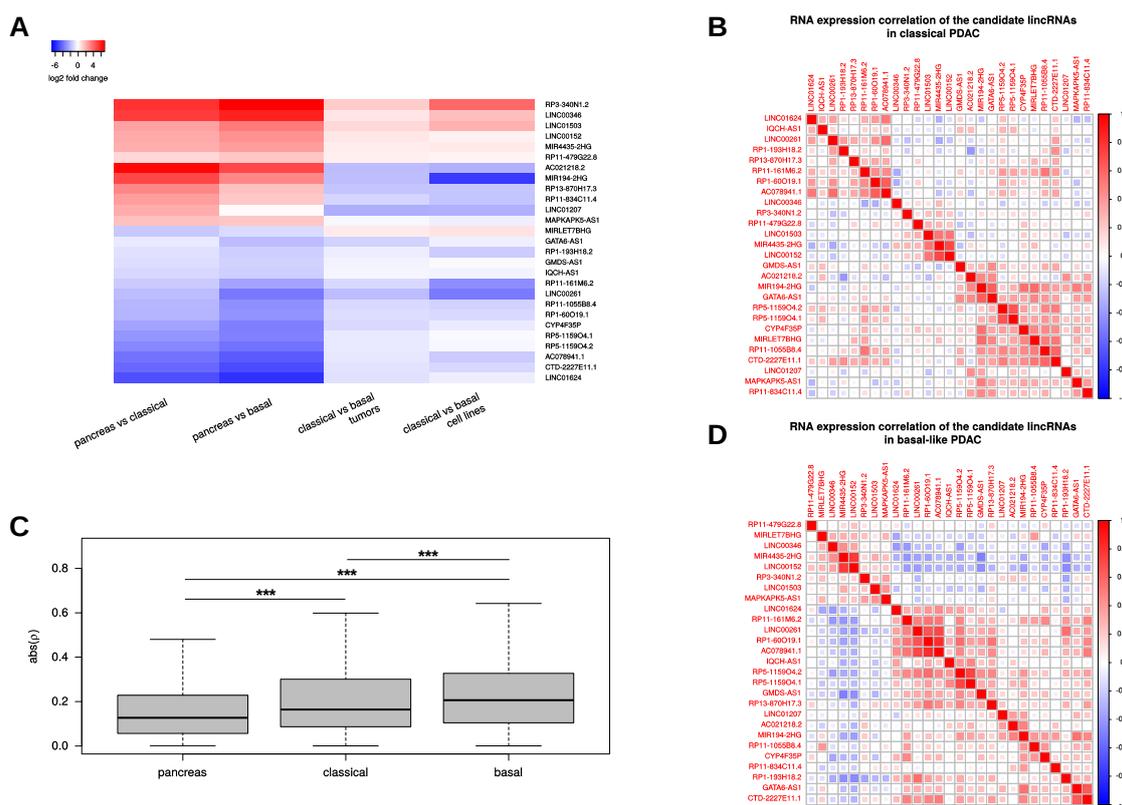
**Figure 2.** Expression and prognostic relevance of highly abundant lincRNAs. (A) Table shows the expression and regulation ( $|\log_2FC| > 1$ ;  $FDR < 0.05$ ) of highly expressed lincRNAs (avg. CPM  $> 99$  in normal pancreas). (B,D) Overall and disease-free survival analysis (OS and DFS, respectively) based on lincRNA expression was performed on either both, basal and classical TCGA PAAD samples together or separately using GEPIA2 ([36]). Log-rank test (Mantel-Cox) was used for hypothesis testing. The cox proportional hazard ratio (HR), logrank  $p$ -value and the 95% confidence interval information is included in the graphs. (B) In general, low expression of LINC00261 is associated with worse DFS of PDAC patients. (C,D) Low expression of NEAT1 is specifically associated with a poor OS (C) as well as DFS (D) in the classical subtype.

## 2.2. LincRNAs Are Differentially Expressed between PDAC Subtypes

As discussed above, our subtype-resolved gene expression analysis identified commonly deregulated lincRNAs as well as subtype-specific expression differences that might contribute to the intrinsic molecular and prognostic differences between the two subtypes. We analyzed the differential expression of lincRNAs in more detail in order to identify candidate lincRNAs whose function or regulation may depend on or contribute to the respective disease context. To this end, we considered the aforementioned 109 PDAC-associated lincRNAs (see Figure S1C) and compared their expression in the classical subtype with their respective level in the basal-like subtype. This analysis identified a total of 41 lincRNAs significantly differentially expressed ( $FDR < 0.05$ ) between both pancreatic cancer subtypes. Next, we integrated RNA-seq data from the Cancer Cell Line Encyclopedia (CCLE, ref. [37]) and examined which of the 41 lincRNAs found to be deregulated in situ showed the same trend of deregulation in cellulo when comparing PDAC cell lines classified as “classical” or “basal-like”

(Table S2) [38]. This analysis resulted in a list of 27 PDAC-associated lincRNAs that were consistently higher/lower expressed in one of the two PDAC subtypes (Figure 3A, Table S3). Of those, 7 lincRNAs were generally higher expressed in cell lines and tissues assigned to the basal-like subtype while the remaining 20 lincRNAs were higher expressed in the classical subtype samples. For example, our analysis identified LINC01207 to be upregulated in PDAC which is in line with an earlier study that reported an increased expression of LINC01207 in pancreatic adenocarcinoma as well as a putative role of this lincRNA in regulating apoptosis and autophagy [39]. However, our disease subtype comparison implicated a stronger biological relevance of LINC01207 in tumours of the classical subtype. On the other side, the highly expressed lincRNA LINC00261 was found to be generally downregulated in both cancer subtypes and this decrease was more pronounced in the basal-like subtype tissues and cell lines. LINC00261 is suggested to act as a tumour-suppressive lincRNA in lung [40], liver [41] and gastric cancer [42]. In pancreatic cancer, LINC00261 was implicated in the epithelial-to-mesenchymal transition (EMT) pathway, which is crucial for metastasis and disease progression [27,43]. Intriguingly, LINC00261 might also be important for human endoderm differentiation [44] and its reduced expression is a predictor of DFS (see Figure 2B). Additional lincRNAs whose expression is associated with OS and DFS either across all TCGA PDAC samples or only in a certain subtype are shown in Figures S2 and S3. For example, another lincRNAs of prognostic relevance that we identified with our analysis pipeline was GATA6-AS1. This lincRNA was significantly lower expressed in the basal-like subtype samples and its low expression in PDAC was associated with worse OS (Figure S2C,F) and DFS (Figure S3G,L) in the complete TCGA PDAC cohort as well as in the basal-like group. Next to LINC00261 and GATA6-AS1, which were both downregulated in the basal-like subtype, our analysis also unveiled lincRNAs with a significant upregulation in the basal-like subtype. For example, RP3-340N1.2, a novel uncharacterized and spliced lincRNA, was strongly increased in both tumour subtypes with a more than 2-fold higher expression in basal-like tumours compared to classical ones. Even more pronounced expression changes were observed for LINC00152, also referred to as CYTOR, whose herein described upregulation in pancreatic cancer confirmed a recent finding using a small cohort of six PDAC and five control tissues [45]. While functional analyses of LINC00152 in the context of pancreatic cancer are currently lacking, this lincRNA was previously shown to contribute to cancer progression acting as an oncogenic lincRNA in diverse tumour types [46]. Our survival analysis revealed that high expression of LINC00152 in PDAC is associated with a poor DFS (Figure S3H,M) suggesting a putative tumour-promoting role of LINC00152 also in pancreatic cancer, especially in the basal-like subtype.

To further characterize the expression and to identify common expression patterns among the 27 selected lincRNAs, we performed correlation analyses (Figure 3B,D). Consideration of the magnitudes of the Spearman correlation coefficients revealed a steady increase from the normal pancreas to the basal-like subtype indicating strongest associations between our candidate lincRNAs in this subtype, regardless of the direction (Figure 3C). The strongest correlation in the classical subtype was found between LINC00152 and MIR4435-2HG (Spearman  $\rho = 0.72$ ; Figure 3B). A similar strong correlation was also found in the basal-like subtype ( $\rho = 0.78$ ) as well as in normal pancreas ( $\rho = 0.76$ ) (Table S4). Intriguingly, both lincRNA genes are located on chromosome 2 and MIR4435-2HG was recently found to be a paralog of LINC00152 differing only by 13 exonic single nucleotide exchanges [47].



**Figure 3.** Expression of 27 differentially expressed and subtype-related candidate lincRNAs. (A) Heatmap displays log<sub>2</sub> fold changes of 27 lincRNAs that show expression differences between tumour subtypes and normal pancreas as well as between both cancer subtypes in situ and in cellulo. (B,D) Spearman correlation coefficients ( $\rho$ ) of the selected lincRNAs in classical (B) and basal-like (D) PDAC samples. (C) Distribution of magnitudes of Spearman correlation coefficients ( $|\rho|$ ) obtained from correlation tests between the selected lincRNAs in pancreas and PDAC subtype samples.

Since both lincRNA loci are highly similar, including their up- and downstream genomic sequences, the high correlation seen in all tissue samples might be due to shared transcriptional or similar post-transcriptional mechanisms regulating transcript production and/or decay. Another strong positive correlation seen in all tissues yet most pronounced in normal pancreas ( $\rho = 0.87$ ) existed between RP5-1159O4.1 (AC004982.2) and RP5-1159O4.2 (AC004982.1), which are both located in close proximity on chromosome 7, again indicating a common (transcriptional) regulation. On the other side, the strongest positive correlation in the basal-like subtype was found between AC078941.1 and RP1-60O19.1 ( $\rho = 0.83$ ) that do not share a common genomic locus (Figure 3D). Interestingly, the correlation between these lincRNAs was considerably lower in the classical subtype samples ( $\rho = 0.63$ ) and could not be detected in normal pancreas indicating tumour- and partially subtype-specific regulatory mechanisms responsible for the observed downregulation of both lincRNAs in neoplastic lesions (Table S4). On the opposite side, the strongest negative correlation between two lincRNAs in the basal-like subtype was observed between GMDS-AS1 and MIR4435-2HG ( $\rho = -0.49$ ). Not surprisingly, a similar strong negative correlation was observed between GMDS-AS1 and LINC00152 ( $\rho = -0.43$ ). Interestingly, both negative correlations were considerably attenuated in the classical subtype ( $\rho = -0.24$  and  $-0.12$ , respectively) and normal pancreas ( $\rho = -0.12$  and  $-0.19$ , respectively). Another example for negative correlation with a stronger emphasis in the basal-like subtype was present between LINC00261 and LINC00152 (classical:  $-0.22$ , basal:  $-0.39$ ), lending further support to the idea of a putative opposing (i.e., oncogenic) role of LINC00152 in pancreatic cancer. Of note, MIR4435-2HG showed a very similar negative correlation to LINC00261 and might

also function as a tumour-promoting lincRNA, an idea that is supported by its prognostic relevance for DFS in TCGA PDAC samples (Figure S3I). Last but not least, another strong negative correlation existed between RP1-193H18.2 and AC021218.2 ( $\rho = -0.43$ ) which was highly specific to the classical subtype (normal:  $\rho = -0.07$ ; basal-like:  $\rho = -0.03$ ). Of note, high expression of AC021218.2 in classical subtype tumours was associated with worse OS (Figure S2D). While the underlying mechanisms and the disease relevance await further investigations, this example of a strong negative correlation, which would have gone unnoticed in a global PDAC expression analysis, underscores the value of performing subtype-resolved transcriptome analyses to gain a deeper understanding of the biological processes contributing to carcinogenesis.

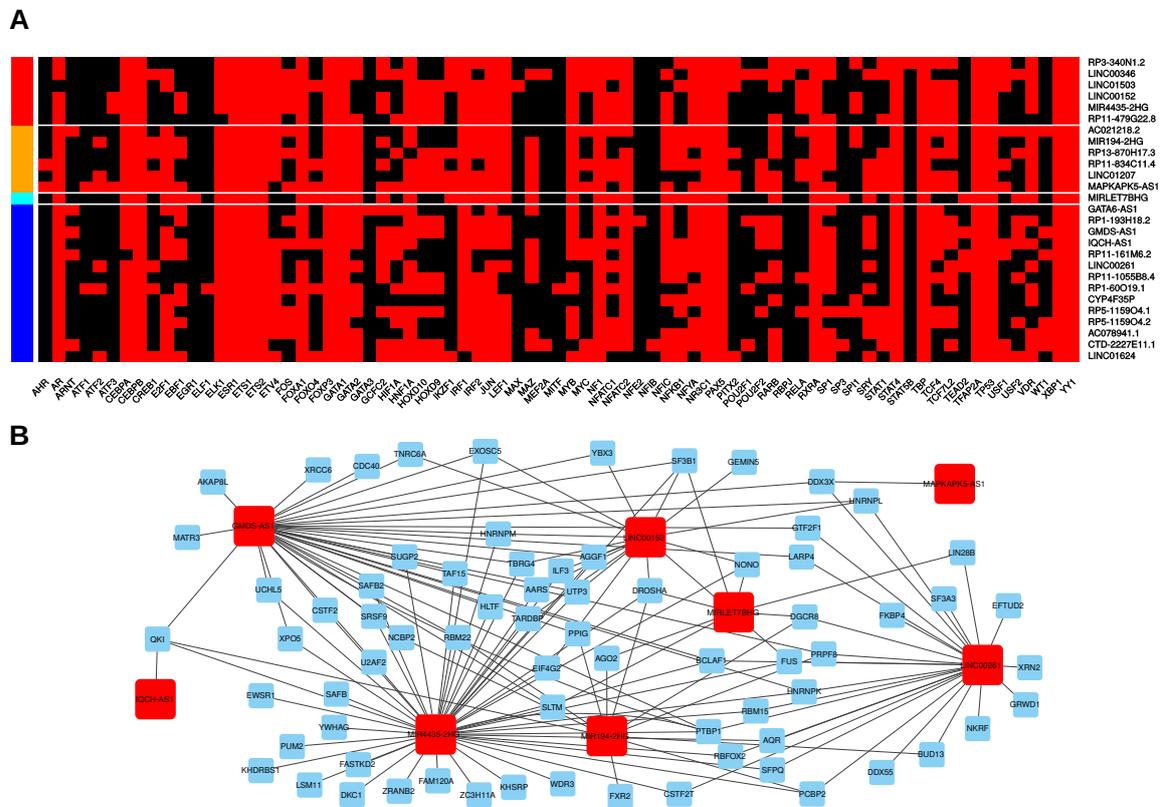
### 2.3. Putative Transcriptional and Post-Transcriptional Regulators of LincRNA Expression

So far, our analyses identified general as well as subtype-specific lincRNA expression trends in PDAC that are eventually the net result of gene transcription, RNA processing and decay. These processes are controlled by different classes of proteins, namely transcription factors (TF) and RNA-binding proteins (RBPs). Moreover, small regulatory RNAs, so called microRNAs (miRNAs), are also involved in fine-tuning the expression landscape of cells [48]. In order to gain further insights into the regulatory mechanisms that act in concert to establish specific lincRNA expression patterns, we performed *in silico* analyses to predict physical interactions. In addition, we retrieved experimental information on RBP-lincRNA binding. First, we focused on putative transcriptional regulators and used the PROMO website ([http://algggen.lsi.upc.es/cgi-bin/promo\\_v3/promo/promoinit.cgi?dirDB=TF\\_8.3](http://algggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3); ref. [49]) to search for human TFs that could bind to the promoter regions of lincRNAs (−2000+1000 bp up-/downstream of transcription start sites; Figure 4A, Table S5). In total, we found 77 TFs that might control the expression of the 27 selected lincRNAs and in each of those lincRNAs loci we found binding motifs of 35 to 56 different TFs.

Binding motifs of 16 TFs were present in all 27 candidates independent of their expression trend in tumours (up or down) compared to normal tissues. For example, the general transcription factor TBP was predicted to bind to all promoter regions consistent with its role in transcription initiation by all three human RNA polymerases. Other TFs like TP53 and YY1 have been shown to possess dual activity, i.e., these TFs are able to activate and repress transcription and their predicted up- or downregulated lincRNA targets might contribute to the context-dependent tumour-suppressive or oncogenic function of these TFs [50,51]. On the other side, this analysis also revealed some very specific putative interactions. For example, STAT5B motifs were only found in the promoter region of RP3-340N1.2 and it is tempting to speculate about an oncogenic role of this upregulated lincRNA as a downstream target of STAT5B signalling [52].

Furthermore, TEAD2, a TF that plays a key role in the Hippo signalling pathway by interacting with YAP1, was predicted to bind and regulate four lincRNAs (LINC00261, RP11-161M6.2, GMDS-AS1, MIRLET7BHG), three of them showing the strongest decrease in basal-like tumours (Figure 3A). Intriguingly, YAP1 was recently identified as a major driver of the squamous (basal-like) subtype of PDAC and its activation was associated with poor prognosis [53]. However, it remains unclear, if this context-dependent role of YAP1 is, at least partially, dependent on or mediated by lincRNAs and the three candidates identified herein should be further analyzed in the context of Hippo signalling. In addition to these examples, combination of the TF prediction with co-expression analysis (Figure 3B,D) provided additional insights into lincRNA-specific expression control. For example, analysis of the TF binding profiles of AC021218.2 and RP1-193H18.2, whose expression was found to be negatively correlated in the classical subtype, but not in the basal-like, revealed several TFs that might specifically regulate each of the lincRNAs. In detail, AC021218.2 upregulation, particularly found in the classical subtype, might be caused by WT1, USF1, RELA, MYC, ETV4, E2F1, and ARNT whereas downregulation of RP1-193H18.2, more pronounced in the basal-like subtype, might be due

to binding of SRY, NFE2, HOXD9, HOXD10, and EBF1 given the mutually exclusive predicted binding sites of these TFs to either one of the lincRNA promoter regions.



**Figure 4.** Putative transcriptional and post-transcriptional regulators of lincRNAs. **(A)** Transcription factors (TFs) predicted to bind to the promoter region ranging from 2000 bp upstream to 1000 bp downstream of the putative transcriptional start site (TSS) of the respective lincRNA. Red colour denotes a predicted TF binding. Row colour codes on the left side denote lincRNAs upregulated in both tumour subtypes compared to normal pancreas with significantly higher expression in the basal-like subtype (red); upregulated in both subtypes and significantly lower in the basal-like subtype (orange); downregulated in both subtypes and significantly higher in the basal-like subtype (cyan); downregulated in both subtypes and significantly lower in the basal-like subtype (blue). **(B)** Network representation of potential interactions between RNA-binding proteins (RBPs) and the selected candidate lincRNAs based on CLIP data. Red nodes represent lincRNAs, blue nodes RBPs. Edges are drawn, if significant CLIP sites were found for the respective lincRNA-RBP pair.

In addition to these putative transcriptional regulatory connections, we next considered post-transcriptional mechanisms that might affect lincRNA expression in PDAC. First, we analyzed enhanced crosslinking and immunoprecipitation (eCLIP) data of 153 individual RBPs that were previously generated in K562 and/or HepG2 cells [54,55] and can be downloaded from the ENCODE project website ([www.encodeproject.org](http://www.encodeproject.org), ref. [56]). We further included a Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP) dataset consisting of two biological replicates containing binding information of the Argonaute protein AGO2 performed in HEK293 cells [57]. In total, we were able to detect physical interactions between eight of the 27 lincRNA candidates and a total of 73 RBPs in these datasets (Figure 4B, Table S6). The overall numbers of lincRNA-RBP interactions ranged from one (IQCH-AS1-QKI, and MAPKAPK5-AS1-DDX3X) to 47 for MIR4435-2HG. The most interactive lincRNAs (MIR4435-2HG, GMDS-AS1, LINC00261) were predicted to associate with more than 20 different RBPs, but only three RBPs had eCLIP peaks in all of them. All three RBPs (PTBP1, HNRNPK, BCLAF1) are known

to regulate RNA processing, especially splicing [58–61]. In accordance with this, gene annotation enrichment analyses (GAEA) using Gene Ontology biological processes and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotations revealed significant enrichments of RBPs associated with splicing for all these three multi-exonic lincRNAs. Furthermore, the eCLIP dataset also included binding information for all three miRNA host genes (MIR4435-2HG, MIRLET7BHG, MIR194-2HG) as well as the MIR4435-2HG paralog LINC00152. As expected, MIRLET7BHG and MIR194-2HG contained significant eCLIP peaks of the ribonuclease DROSHA and its partner DGCR8 that jointly form the nuclear microprocessor complex, which is responsible for cleaving the stem loop of primary miRNA transcripts [62]. Furthermore, these two lincRNAs were associated with AGO2, a core protein of the mammalian RNA-induced silencing complex (RISC). However, the AGO2 CLIP peaks were all found in regions representing mature miRNAs, thus underlining AGO2's function of binding miRNAs for incorporating them into the RISC [63]. Moreover, eCLIP peaks of LIN28B mapping to MIRLET7BHG was in line with the inhibitory effect of LIN28B on let-7b biogenesis [64]. Interestingly, DGCR8 CLIP peaks were also present in LINC00261 whereas DROSHA CLIP peaks were detectable in LINC00152 suggesting putative interactions between those factors which could be caused by miRNA-like stem loop structures present within these lincRNAs. Indeed, a recent study identified a 121-bp 3'-hairpin structure (M8, see Figure S4A) within LINC00152 and overexpression of this stem loop was sufficient to increase invasion of glioblastoma cells [65]. Sequence alignments revealed that MIR4435-2HG contained the same stem loop sequence. Importantly, analysis of ribosome profiling data indicated that the stem loop interacts with one or more currently unknown RBPs. Thus, we performed a closer examination of the eCLIP data and identified six RBPs that might be able to directly bind to the stem loop sequence or within its close proximity in LINC00152 and MIR4435-2HG, respectively. Out of those six candidate RBPs only eCLIP peaks of YBX3 and PABPC4 consistently mapped to the stem loop region in both lincRNAs (Figure S4B,C). Thus, both RBPs represent compelling candidates for further analysis. For example, YBX3 has been shown to stabilize its mRNA targets [66–68] and its putative binding to LINC00152 could contribute to the increased expression of this lincRNA in PDAC, especially in the basal-like subtype.

Next to these protein-based regulatory mechanisms that might control lincRNA expression in pancreatic cancers, we also considered small RNA-mediated regulatory pathways. Hence, we performed miRNA targeting analyses for all 27 lincRNA candidates which yielded a total of 178 miRNAs predicted to bind to six lincRNAs (Figure S5A,B; Table S7). However, when considering only those interactions supported by statistically significant negative correlations in RNA expression ( $FDR < 0.05$ ), only a few putative bindings remained. In the classical subtype, MIRLET7BHG was associated with miR-4736 ( $\rho = -0.47$ ,  $FDR = 0.01$ ) and MIR4435-2HG was negatively correlated to miR-206 ( $\rho = -0.38$ ,  $FDR = 0.03$ ). For lincRNA AC021218.2, we observed significant negative correlation with the predicted miRNA binding partners miR-217 ( $\rho = -0.37$ ,  $FDR = 0.03$ ), miR-202 ( $\rho = -0.41$ ,  $FDR = 0.01$ ) and miR-20b ( $\rho = -0.35$ ,  $FDR = 0.04$ ). None of the predicted lincRNA-miRNA interactions showed significant negative correlation in the basal-like subtype samples. While this miRNA target prediction does not exclude the possibility of a regulatory effect of certain miRNAs on our lincRNA candidates in selected cell types or under certain conditions, the limited amount of putative functional miRNA-lincRNA interactions is in line with the RBP interaction analysis which identified robust and reproducible Argonaute eCLIP peaks only for the miRNA host genes MIRLET7BHG and MIR194-2HG. However, eCLIP peaks of TNRC6A, a component of the RNA-induced silencing complex (RISC) and interaction partner of Argonaute proteins, were present in GMDS-AS1 and LINC00152 (see Table S6) hinting towards a loading of both lincRNAs into the RISC and subsequent expression control by selected miRNAs. Interestingly, among all miRNAs predicted to target LINC00152 the strongest negative correlation was observed for miR-206 (see Table S7) which has been shown to regulate the expression of the closely related paralog of LINC00152, namely MIR4435-2HG, in colorectal cancer cells [69]. Moreover, GMDS-AS1 regulation by miR-96-5p might contribute to lung cancer development [70]. Hence, both lincRNAs might be subjected to

miRNA-mediated expression control in pancreatic cancer cells too and their regulation by miR-206 or miR-96-5p, respectively, should be analyzed in more detail.

#### 2.4. Putative Signalling Pathways and Direct mRNA Targets Associated with lincRNAs

Having identified subtype-specific lincRNA expression patterns as well as putative transcriptional and post-transcriptional regulators or interactors of a selected group of 27 lincRNAs leaves us with the final question of what functions these lincRNAs might have. In order to address this question, we performed a guilty-by-association analysis and calculated the expression correlation between each of the 27 lincRNAs and all protein-coding genes contained in the datasets (18,474) in normal pancreas as well as in the PDAC subtype samples. Subsequently, we used the obtained correlation coefficients to perform GSEA to identify associated signalling pathways (Figure 5A,B) using the MSigDB hallmark gene sets. These 50 gene sets contain genes serving as markers of well-defined biological states or processes [71]. Similar co-expression analyses had been performed in the past to generate hypothesis and to predict diverse roles of lincRNAs, ranging from stem cell pluripotency to cancer [72,73].

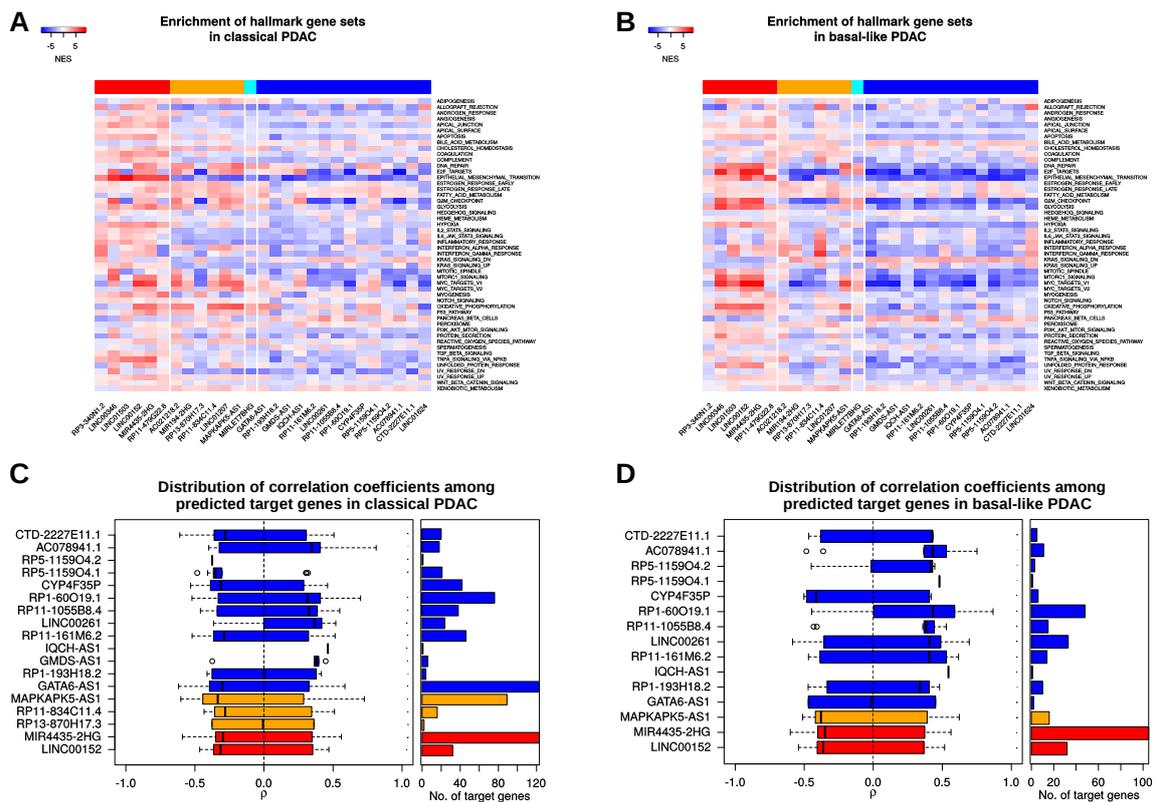
Here, our co-expression analysis revealed that all lincRNAs upregulated in PDAC and in addition also significantly upregulated in the basal-like versus the classical subtype showed a positive enrichment of genes associated with epithelial-to-mesenchymal transition (EMT) in both subtypes, whereas a negative enrichment of EMT genes was observed for most of the remaining lincRNAs (Figure 5A,B; Tables S8 and S9). Intriguingly, expression of LINC00152, LINC00346, and LINC01503 was positively correlated with genes that play a role in EMT and all three lincRNAs have been shown to regulate tumour cell migration and invasion, partially via induction of an EMT program in cholangiocarcinoma, gallbladder, gastric, or bladder cancer, respectively [74–77]. While their relevance for pancreatic cancer is largely unknown, recent studies provided evidence of a function of LINC00346 in pancreatic cancer cell migration and invasion as well as proliferation [78,79].

Hence, it might be worth to investigate the function of other lincRNAs that might positively as well as negatively influence EMT and cell motility. In addition, all upregulated lincRNAs with a higher expression in basal-like PDAC tumours showed a positive enrichment of the apoptosis gene set suggesting additional roles of these lincRNAs in cell death control, which should be further investigated as well. Interestingly, when considering co-expression in classical subtype samples, gene sets related to Interleukin (IL)2 and IL6 signalling showed a rather positive enrichment with the group of lincRNAs that were generally upregulated in PDAC, but higher in basal-like tumour whereas all other lincRNAs returned negative enrichment scores for these two gene sets. Surprisingly, several lincRNAs show a reversed relationship to IL2/IL6 gene sets in the basal-like samples suggesting subtype-specific functions of certain lincRNAs potentially in the context of IL2/IL6 signalling. Several other inversed, i.e., subtype-specific enrichment trends could be found. For example, LINC01207-related correlations showed a significant positive enrichment of genes included in the apoptosis gene set in the classical subtype (NES = 1.84, FDR = 0.02). However, in the analysis of the basal-like subtype, a negative, yet not significant enrichment (NES = −0.84, FDR = 0.68) could be detected lending further support to the idea of subtype-dependent lincRNA functions in PDAC.

Importantly, the correlations and pathway associations observed for these 27 lincRNAs could be both, the cause but also the consequence of several complex molecular interactions. For example, the expression of lincRNAs could be correlated with EMT or apoptosis genes due to overlapping transcriptional or post-transcriptional regulators, e.g., certain TFs, RBPs or miRNAs that jointly regulate the expression of certain lincRNAs and protein-coding genes. On the other side, lincRNAs could sequester or guide TFs as well as RBPs to affect transcription or RNA processing which would result in positive or negative expression correlations. Moreover, lincRNAs could sponge miRNAs to control certain expression programs, again resulting in the observed interdependencies. Importantly, another post-transcriptional mechanism that might be employed by lincRNAs to regulate transcript abundance and therefore might explain positive and negative expression correlations observed herein exists and relies on direct binding of the lincRNA to its respective target mRNA.

A prominent example of such a regulatory interaction is found between lincRNA-ATB and IL11 mRNA leading to the stabilization of the latter which results in an activated IL11/STAT3 signalling and enhanced metastatic potential of liver cancer cells [80]. Hence, we wanted to investigate whether such RNA-RNA interactions could also be present among our candidate lincRNAs and their co-regulated protein-coding transcripts in PDAC subtypes. Therefore, we obtained lincRNA-RNA target predictions from a recently developed database [81]. We only considered putative interactions that would result in expression changes of the target transcripts and that were associated with significant ( $FDR < 0.05$ ) Spearman expression correlations. In total, we were able to identify putative mRNA targets of 18 lincRNAs in the classical subtype and of 15 lincRNAs in the basal-like subtype (Figure 5C,D; Tables S10 and S11). In the classical subtype, MIR4435-2HG and GATA6-AS1 both showed significant correlations to 123 predicted target genes, with the majority of these genes having negative correlation coefficients (Figure 5C, Figure S6A). In the basal-like subtype, MIR4435-2HG also had the most predicted bindings (105), again with a trend towards negative expression correlation (Figure 5D). All but two of the predicted interacting mRNAs of GATA6-AS1 showed a significant correlation in the classical subtype, but did not do so in the basal-like subtype (Figure S6B). Also, predicted mRNA interactions of lincRNAs GMD5-AS1, RP11-834C11.4 and RP13-870H17.3 were only significant in the classical subtype. LINC00152, MAPKAPK5-AS1 and CYP4F35P consistently showed trends for negative correlation to predicted targets in both subtypes. Furthermore, AC078941.1, RP1-60O19.1, RP11-1055B8.4 and LINC00261 consistently correlated negatively to their predicted targets. Moreover, individual mRNAs were predicted to interact with two or more lincRNAs suggesting cooperativity or competition among lincRNAs. For example, MUC6 mRNA might be targeted by AC078941.1, CTD-2227E11.1 and RP1-60O19.1 in both subtypes (Figure S6A,B). Of note, all three lincRNAs were downregulated in PDAC, especially in the basal-like subtype and the same expression trend was seen for MUC6 mRNA making it tempting to speculate about a stabilizing interaction between these transcripts. On the other side, SFRP5 mRNA, which encodes a soluble modulator of Wnt signalling and was strongly downregulated in the basal-like subtype might be bound by two lincRNAs with opposite functions. While RP1-60O19.1 showed a strong positive expression correlation, LINC00152 was strongly negatively correlated, suggesting stabilizing or destabilizing effects of their interaction with SFRP5 mRNA, respectively. In general, we could observe that lincRNAs upregulated in the PDAC samples tended to be negatively correlated to their predicted target transcripts in both subtypes suggesting rather destabilizing functions among our candidate lincRNA which should be further investigated in the future.

Last but not least, we performed a preliminary analysis to investigate the impact of genetic activation of oncogenic signalling pathways on lincRNA expression. To this end, we focused on KRAS and compared the differential expression of lincRNAs between TCGA PDAC samples with wildtype KRAS (wt,  $n = 10$ ) against those samples with activated, mutant KRAS (mut,  $n = 139$ ). Applying the same criteria for the determination of differential expression (i.e.,  $FDR < 0.05$ ;  $|\log_2FC| > 1$ ; minimum average CPM  $> 5$  in either KRAS wt or mut samples), we identified only seven lincRNAs differentially expressed between these two conditions (Table S12). Interestingly, three of these were also found in our list of 27 lincRNAs significantly deregulated between the PDAC subtypes, namely RP1-193H18.2 ( $\log_2FC$  mut/wt =  $-1.37$ ), RP3-340N1.2 ( $\log_2FC$  mut/wt =  $3.76$ ), and LINC00261 ( $\log_2FC$  mut/wt =  $-3.56$ ). Interestingly, co-expressed genes of these three lincRNAs showed significant positive enrichment of the KRAS\_SIGNALING\_DN gene set as well as a significant negative enrichment of the KRAS\_SIGNALING\_UP gene set, which was also dependent on the PDAC subtype (see Figure 5A,B). Of note, although the number of KRAS wt samples was very low, most of them (8/10) had been assigned to the classical subtype, whereas only two of them were classified as basal-like, which might partially explain the subtype-dependent enrichment of co-expressed KRAS target genes. Importantly, this preliminary analysis revealed that certain lincRNA expression patterns might be caused by individual tumour-associated mutations, which warrants further investigations.



**Figure 5.** Guilty-by-association studies and potential mRNA-lincRNA-interactions. (A,B) Normalized enrichment scores (NES) obtained from gene set enrichment analyses (GSEA) using sorted Spearman correlation coefficients ( $\rho$ ) of protein-coding genes related to the respective 27 lincRNA in classical (A) and basal-like (B) PDAC samples. Considered were the 50 MSigDB hallmark gene sets. Column color codes denote lincRNAs upregulated in both tumour subtypes compared to normal pancreas with significantly higher expression in the basal-like subtype compared to the classical subtype (red); upregulated in both subtypes and significantly lower in the basal-like subtype (orange); downregulated in both subtypes and significantly higher expressed in the basal-like subtype (cyan); downregulated in both subtypes and significantly lower in the basal-like subtype (blue). (C,D) Number of predicted mRNA interactors and distribution of the RNA expression correlation coefficients ( $\rho$ ) of these putative targets and the respective lincRNAs in classical (C) and basal-like subtype (D) PDAC.

### 3. Discussion

Long non-coding RNAs are a large and heterogeneous group of transcripts that play a role in diverse biological systems. Although their general contributions to multiple human diseases, e.g., cancer, have been described in recent years, many questions remain. For example, gene expression analyses performed on thousands of tumour samples from a broad range of tumour entities have identified cancer subtypes that are defined by certain gene expression pattern [82–85]. Also lincRNAs have been assigned to these molecular subtypes, however, it is unclear, if they play an active role in shaping the expression landscape of these subtype and/or whether these lincRNAs have subtype-specific functions as well as interaction partners. Here, we provide a comprehensive analysis of lincRNAs in pancreatic cancer and we highlighted several interesting findings and generated hypotheses about the regulation and function of selected lincRNAs that require further validations as well as experimental interrogation. For example, our differential expression analysis revealed a strong reduction of NEAT1 and MALAT1, two highly conserved nuclear lincRNAs [34,35]. Interestingly, the role of NEAT1 in pancreatic cancer seems to be controversial. Loss of Neat1 in a KrasG12D-driven mouse model of pancreatic cancer has been shown to promote the development of premalignant

pancreatic intraepithelial neoplasia (PanIN) and cystic lesions [86]. On the contrary, several studies claim that NEAT1 acts through direct interaction with specific mRNAs (e.g., ELF3) or via sponging certain miRNAs thereby promoting cancer cell growth, invasion, and migration [87–90]. Based on the expression changes observed in our analysis, it is tempting to speculate that NEAT1 might indeed function as a tumour suppressor in PDAC. In addition, studies on MALAT1 in pancreatic cancer largely suggest an oncogenic function of this lincRNA. For example, MALAT1 was found to be upregulated in tumours compared to adjacent normal tissue and its downregulation inhibited cell proliferation, migration, invasion, and promoted apoptosis in pancreatic cancer cell lines [91]. Yet, no effect on tumour progression upon Malat1 depletion was observed in a highly aggressive mouse model of pancreatic cancer [92]. Importantly, the general role of MALAT1 acting as a common oncogene and driver of metastasis across multiple cancer types is currently debated and genetic studies in breast cancer indicated that it could also act as a tumour suppressor [93–96]. Our finding of a significantly reduced expression of MALAT1 in PDAC is partially contradictory to a previous study that analyzed the clinical significance of MALAT1 in pancreatic cancer using multiple datasets from different public databases (GEO, Oncomine, TCGA) [97]. However, the same study also found heterogeneity among the data with at least three other datasets (GSE3654, GSE16515, GSE1542) showing no significant expression differences [98–100]. Therefore, more sophisticated genetic tools and targeting strategies as well as more advanced cell-based model systems (e.g. patient-derived xenografts, organoids, refs. [101,102]) should be used in the future to dissect the molecular functions of MALAT1 and NEAT1 in pancreatic cancer.

In addition, our study found multiple other lincRNAs to be deregulated in PDAC subtypes and with currently unknown functions. However, a limitation of our analysis pipeline is the use of RNA-seq data from different sources (normal pancreas from the GTEx and PDAC from the TCGA project) to identify putative candidate lincRNAs for further investigations. Although we used raw count data from these two sources and normalized the samples together to avoid RNA composition biases, we cannot exclude batch effects that might arise from, e.g., differences in sample handling and processing. Wang et al. [103] proposed a pipeline for uniform processing and batch effect removal for comparing GTEx and TCGA data. The validation of this empirical correction method, however, relies on comparing adjusted expression values of normal tissue samples provided by the TCGA and the GTEx samples. Since the TCGA PDAC project offers only four normal pancreas samples, Wang et al. considered the TCGA PDAC cohort to be not suited to be processed with their pipeline. Furthermore, this method requires the TCGA normal tissues to represent adequate controls. However, these samples were collected from sites neighbouring cancerous tissue and thus might also be affected by deregulated gene expression. Thus, we did not adjust the data to account for batch effects. Nevertheless, our results are in line with previously reported findings. For example, GSEA performed on differential gene expression results confirmed that numerous protein coding genes known to be deregulated in pancreatic cancer were determined as such (Figure S1A). Further, we could confirm several reports about deregulation of specific lincRNAs. For example, we found MEG3 to be downregulated as reported in previous pancreatic cancer studies [28–31]. Upregulation of LINC01207 and LINC00152/CYTOR in PDAC was also reported previously [39,45]. Recently, we validated the herein proposed downregulation of LINC00261 by comparing RNA expression of 34 normal pancreatic tissue samples with 42 PDAC samples [43]. Further functional analyses revealed a connection of this lincRNA to the EMT programme, which we also propose herein based on the guilty-by association studies performed with GSEA. Altogether, we believe that our analyses provided a reasonable starting point for the identification of lincRNAs of particular interest for further PDAC research. However, the validation of the deregulation of each candidate lincRNA remains imperative.

Another area that requires much more attention in the future is the identification of genetic, epigenetic and post-transcriptional mechanisms that control lincRNA expression and regulation in PDAC subtypes. While we presented a broad analysis of putative connections between lincRNAs and TFs, as well as RBPs and miRNAs, which need to be validated experimentally, we did not extensively consider genetic and epigenetic factors that could explain some of the observed expression

differences. Yet, our preliminary analysis of the impact of KRAS mutation status revealed that certain lincRNA expression patterns might be caused by individual tumour-associated mutations. However, additional chromosomal alterations (amplifications, deletions, rearrangements) might as well contribute to the differential expression of individual lincRNAs. Another factor that could cause general or subtype-specific expression differences is DNA methylation, which could lead to epigenetic gene silencing. In fact, GATA6-AS1 (and GATA6) expression was shown to be controlled by two distinct mechanisms: lower expression in basal-like tumours might be caused by higher DNA methylation near the GATA6 gene, whereas classical tumours showed copy number gains of the GATA6 neighbourhood in conjunction with higher expression of GATA6 mRNA as well as GATA6-AS1 in the classical subtype [3]. Intriguingly, there seems to be a positive or negative, subtype-specific selection pressure on the GATA6/GATA6-AS1 gene locus confirming an important and complex role for both genes in PDAC and other cancers [104–106].

In summary, our study provides a comprehensive list of subtype-associated lincRNAs that might directly or indirectly contribute to subtype-specific transcription programs and cellular phenotypes. To understand the regulatory roles of these lincRNAs in more detail, comprehensive gain or loss of function experiments in a broad selection of in vitro and in vivo model systems need to be performed in the future.

## 4. Materials And Methods

### 4.1. RNA-Seq Data Processing, Differential Expression and Survival Analysis

We obtained gene-level RNA-seq read counts of TCGA primary tumour PDAC samples and GTEx V7 normal pancreas tissue via the GDC data portal ([portal.gdc.cancer.gov](https://portal.gdc.cancer.gov)) and the GTEx portal ([gtexportal.org](https://gtexportal.org)), respectively. By combining these data, we got read count information of 53,045 genes for 177 primary tumour samples and 248 normal pancreas tissue samples. Differential gene expression was assessed using R/edgeR v3.28.0 after normalizing all samples together applying the trimmed mean of M-values (TMM) normalization method [107,108]. CPM transformation was utilized to obtain normalized expression values. The classification of the TCGA tumour samples into PDAC subtypes was taken from [3]. LincRNA information was gained by extracting genes termed as lincRNAs according to ENSEMBL v38.86 annotation [109]. MiRNA expression data were also obtained by downloading read count data of TCGA PDAC samples via the GDC data portal. CPM values were generated after applying TMM normalization using edgeR. Normalized expression data (RPKM) from cell lines were obtained from the CCLE (<https://portals.broadinstitute.org/ccle/data>; version 20180929; ref. [37]). Classification of PDAC cell lines into basal-like or classical subtype origin were obtained from [38]. Only cell lines associated to a subtype class with a false discovery rate (FDR) value less than 0.05 were considered.

### 4.2. LincRNA Localization Predictions

Subcellular localizations were predicted using DeepLncRNA [110]. For each lincRNA localizations of all ENSEMBL v38.86 transcripts were predicted. Terms “Cytosol” and “Nuclear” indicate that all transcript isoforms of a particular lincRNA gene were predicted to be localized in the cytosol or nucleus, respectively. Term “both” indicates that transcript variants of the same lincRNA gene were predicted to be localized in either the cytosol or the nucleus.

### 4.3. Enrichment Analyses

Gene set enrichment analyses (GSEA) were performed using the GSEA v3.0 software [24] and MSigDB v7.0 gene sets [71], applying the pre-ranked test, 1000 permutations and the classical scoring scheme.

#### 4.4. RBP Binding Predictions

ECLIP data of 153 RNA binding proteins were obtained from the ENCODE project website ([www.encodeproject.org](http://www.encodeproject.org) [56]). These eCLIP experiments were performed in Hep-G2 or K-562 cells and always comprised two biological replicates. Statistically significant peaks (score = 1000) were extracted from the downloaded narrow peak bed files (GRCh38) and intersections of these peaks with lincRNAs were determined using bedtools v.2.25.0 [111]. Only if significant peaks for a particular lincRNA were found in both replicates of an experiment, an RBP was considered a putative binding partner of the respective lincRNA.

Peak files (hg19) of two replicates of an AGO2 PAR-CLIP experiment (GEO accession GSM714644 and GSM714645 [57]) were obtained from the CLIP-DB website ([lulab.life.tsinghua.edu.cn/clipdb](http://lulab.life.tsinghua.edu.cn/clipdb) [112]). Putative binding to lincRNAs was determined as described for the eCLIP data.

#### 4.5. MiRNA Binding Prediction

To infer putative lincRNA-miRNA bindings, we first queried for predicted miRNA bindings utilizing the R-package multiMiR v1.8.0 [113] using all eight prediction databases and the default prediction cutoff of 20%. We chose a particular lincRNA-miRNA interaction, if it was reported by at least one of the databases.

#### 4.6. LincRNA-MRNA Target Predictions

LincRNA target predictions were obtained from the database accessible at <http://rtools.cbrc.jp/cgi-bin/RNARNA/> [81]. For each transcript isoform of a certain lincRNA, we gathered the top 100 prediction results ranked by minimum energy and extracted the unique gene identifiers of the predictions of all isoforms as putative targets.

#### 4.7. TF Binding Prediction

Putative transcription factor binding sites were determined using the PROMO Website ([algggen.lsi.upc.es/cgi-bin/promo\\_v3/promo](http://algggen.lsi.upc.es/cgi-bin/promo_v3/promo) [49]) searching for human transcription factors at a maximum matrix dissimilarity of 5%. Input genomic sequences starting 2000 bp upstream of the transcription start and ranging 1000 bp into the gene body were obtained from ENSEMBL v38.86 via the biomaRt v2.42.0 R/Bioconductor-package [114,115].

#### 4.8. Survival Analysis

Prognostic relevance of lincRNA expression to predict overall and disease-free survival (OS and DFS, respectively) of PDAC patients was assessed using the GEPIA2 data portal ([gepia2.cancer-pku.cn](http://gepia2.cancer-pku.cn) [36]). Basal and/or classical subtype samples from the TCGA PAAD dataset were selected to run the analysis for each lincRNA separately. Log-rank test (Mantel-Cox) was applied by GEPIA2 for hypothesis testing. The cox proportional hazard ratio (HR), logrank *p*-value and the 95% confidence interval information was included in the respective Kaplan-Meier graphs.

## 5. Conclusions

We systematically investigated the expression of lincRNAs in pancreatic cancer and its molecular subtypes using publicly available data from large-scale studies. We identified several deregulated lincRNAs which also showed significant different expression patterns across PDAC subtypes. Subsequent expression correlation analyses implied that subtype-associated lincRNAs might directly or indirectly contribute to subtype-specific transcription programs and regulate cellular processes like migration or apoptosis. Mechanistically, we propose that lincRNAs might be regulated by transcriptional as well as post-transcriptional regulators. We predicted putative TFs as well as interacting RBPs, miRNAs and mRNAs that might be targeted by lincRNAs or might affect their

expression. In summary, we identified several PDAC-associated lincRNAs of prognostic relevance and potential context-dependent functions and molecular interactions. Hence, our study provides a valuable resource for future investigations to decipher the role of lincRNAs in pancreatic cancer.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2072-6694/12/8/2077/s1>, Figure S1: Differential expression of lincRNAs in PDAC subtypes, Figure S2: LincRNAs with prognostic relevance across all TCGA PAAD samples, Figure S3: LincRNAs with subtype-specific prognostic relevance in TCGA PAAD samples, Figure S4: M8 stem loop and its putative interaction partners, Figure S5: Potential targeting of selected lincRNAs by microRNAs, Figure S6: Putative interactions between selected lincRNAs and protein-coding transcripts, Table S1: Differential expression of lincRNAs in PDAC, Table S2: LincRNA expression in PDAC-derived cell lines, Table S3: 27 selected lincRNA candidates, Table S4: Correlation coefficients of the 27 selected lincRNA candidates, Table S5: Predicted TF bindings, Table S6: Predicted RNA RBP interactions, Table S7: Predicted miRNA interactions, Table S8: GSEA results of correlation between selected lincRNAs and protein coding genes in classical PDAC, Table S9: GSEA results of correlation between selected lincRNAs and protein coding genes in basal-like PDAC, Table S10: Predicted mRNA interactions in classical PDAC, Table S11: Predicted mRNA interactions in basal-like PDAC, Table S12: Differential expression of lincRNAs in KRAS wild type vs KRAS mut PDAC samples

**Author Contributions:** The study was conceptualized by M.G., M.H. and T.G. Data analyses were performed by M.G. and S.H. and A.D. The manuscript was written by M.G., M.H. and T.G. Figures were prepared by M.G. and A.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study is supported by the Roux program of the Medical Faculty, Martin-Luther University Halle-Wittenberg (FKZ 31/18).

**Acknowledgments:** The authors would like to thank all members of the Hüttelmaier, Hämmerle and Gutschner labs for helpful discussions. We thank the ENCODE Consortium and the Gene Yeo Lab for making eCLIP data available.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Rahib, L.; Smith, B.D.; Aizenberg, R.; Rosenzweig, A.B.; Fleshman, J.M.; Matrisian, L.M. Projecting cancer incidence and deaths to 2030: The unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.* **2014**, *74*, 2913–2921. [[CrossRef](#)] [[PubMed](#)]
- Kleeff, J.; Korc, M.; Apte, M.; La Vecchia, C.; Johnson, C.D.; Biankin, A.V.; Neale, R.E.; Tempero, M.; Tuveson, D.A.; Hruban, R.H.; et al. Pancreatic cancer. *Nat. Rev. Dis. Prim.* **2016**, *2*, 16022. [[CrossRef](#)] [[PubMed](#)]
- Cancer Genome Atlas Research Network. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* **2017**, *32*, 185–203. [[CrossRef](#)]
- Chen, C.C.; Chen, L.L.; Li, C.P.; Hsu, Y.T.; Jiang, S.S.; Fan, C.S.; Chua, K.V.; Huang, S.X.; Shyr, Y.M.; Chen, L.T.; et al. Myeloid-derived macrophages and secreted HSP90 $\alpha$  induce pancreatic ductal adenocarcinoma development. *Oncoimmunology* **2018**, *7*, e1424612. [[CrossRef](#)] [[PubMed](#)]
- Hingorani, S.R.; Petricoin, E.F.; Maitra, A.; Rajapakse, V.; King, C.; Jacobetz, M.A.; Ross, S.; Conrads, T.P.; Veenstra, T.D.; Hitt, B.A.; et al. Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* **2003**, *4*, 437–450. [[CrossRef](#)]
- Juiz, N.A.; Iovanna, J.; Dusetti, N. Pancreatic Cancer Heterogeneity Can Be Explained Beyond the Genome. *Front. Oncol.* **2019**, *9*, 246. [[CrossRef](#)]
- Moffitt, R.A.; Marayati, R.; Flate, E.L.; Volmar, K.E.; Loeza, S.G.; Hoadley, K.A.; Rashid, N.U.; Williams, L.A.; Eaton, S.C.; Chung, A.H.; et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat. Genet.* **2015**, *47*, 1168–1178. [[CrossRef](#)]
- Collisson, E.A.; Sadanandam, A.; Olson, P.; Gibb, W.J.; Truitt, M.; Gu, S.; Cooc, J.; Weinkle, J.; Kim, G.E.; Jakkula, L.; et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat. Med.* **2011**, *17*, 500–503. [[CrossRef](#)]
- Bailey, P.; Chang, D.K.; Nones, K.; Johns, A.L.; Patch, A.M.; Gingras, M.C.; Miller, D.K.; Christ, A.N.; Bruxner, T.J.; Quinn, M.C.; et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **2016**, *531*, 47–52. [[CrossRef](#)]
- 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 Rev.* **2018**, *37*, 83–105. [[CrossRef](#)]

11. Gutschner, T.; Diederichs, S. The hallmarks of cancer: A long non-coding RNA point of view. *RNA Biol.* **2012**, *9*, 703–719. [[CrossRef](#)] [[PubMed](#)]
12. Bai, X.; Lu, D.; Lin, Y.; Lv, Y.; He, L. A seven-miRNA expression-based prognostic signature and its corresponding potential competing endogenous RNA network in early pancreatic cancer. *Exp. Ther. Med.* **2019**, *18*, 1601–1608. [[CrossRef](#)] [[PubMed](#)]
13. Wang, W.; Lou, W.; Ding, B.; Yang, B.; Lu, H.; Kong, Q.; Fan, W. A novel mRNA-miRNA-lncRNA competing endogenous RNA triple sub-network associated with prognosis of pancreatic cancer. *Aging* **2019**, *11*, 2610–2627. [[CrossRef](#)] [[PubMed](#)]
14. Yao, K.; Wang, Q.; Jia, J.; Zhao, H. A competing endogenous RNA network identifies novel mRNA, miRNA and lncRNA markers for the prognosis of diabetic pancreatic cancer. *Tumour Biol.* **2017**, *39*, 1010428317707882. [[CrossRef](#)] [[PubMed](#)]
15. Zhou, M.; Diao, Z.; Yue, X.; Chen, Y.; Zhao, H.; Cheng, L.; Sun, J. Construction and analysis of dysregulated lncRNA-associated ceRNA network identified novel lncRNA biomarkers for early diagnosis of human pancreatic cancer. *Oncotarget* **2016**, *7*, 56383–56394. [[CrossRef](#)] [[PubMed](#)]
16. Shi, X.; Zhao, Y.; He, R.; Zhou, M.; Pan, S.; Yu, S.; Xie, Y.; Li, X.; Wang, M.; Guo, X.; et al. Three-lncRNA signature is a potential prognostic biomarker for pancreatic adenocarcinoma. *Oncotarget* **2018**, *9*, 24248–24259. [[CrossRef](#)]
17. Song, J.; Xu, Q.; Zhang, H.; Yin, X.; Zhu, C.; Zhao, K.; Zhu, J. Five key lncRNAs considered as prognostic targets for predicting pancreatic ductal adenocarcinoma. *J. Cell. Biochem.* **2018**, *119*, 4559–4569. [[CrossRef](#)]
18. Tahira, A.C.; Kubrusly, M.S.; Faria, M.F.; Dazzani, B.; Fonseca, R.S.; Maracaja-Coutinho, V.; Verjovski-Almeida, S.; Machado, M.C.; Reis, E.M. Long noncoding intronic RNAs are differentially expressed in primary and metastatic pancreatic cancer. *Mol. Cancer* **2011**, *10*, 141. [[CrossRef](#)]
19. Zhou, C.; Wang, S.; Zhou, Q.; Zhao, J.; Xia, X.; Chen, W.; Zheng, Y.; Xue, M.; Yang, F.; Fu, D.; et al. A Long Non-coding RNA Signature to Improve Prognostic Prediction of Pancreatic Ductal Adenocarcinoma. *Front. Oncol.* **2019**, *9*, 1160. [[CrossRef](#)]
20. Ulitsky, I.; Bartel, D.P. lincRNAs: Genomics, evolution, and mechanisms. *Cell* **2013**, *154*, 26–46. [[CrossRef](#)]
21. Haemmerle, M.; Gutschner, T. Long non-coding RNAs in cancer and development: Where do we go from here? *Int. J. Mol. Sci.* **2015**, *16*, 1395–1405. [[CrossRef](#)] [[PubMed](#)]
22. Gutschner, T.; Baas, M.; Diederichs, S. Noncoding RNA gene silencing through genomic integration of RNA destabilizing elements using zinc finger nucleases. *Genome Res.* **2011**, *21*, 1944–1954. [[CrossRef](#)] [[PubMed](#)]
23. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **2013**, *45*, 580–585. [[CrossRef](#)] [[PubMed](#)]
24. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [[CrossRef](#)]
25. Gao, X.; Zhang, S.; Chen, Y.; Wen, X.; Chen, M.; Wang, S.; Zhang, Y. Development of a novel six-long noncoding RNA signature predicting survival of patients with bladder urothelial carcinoma. *J. Cell. Biochem.* **2019**, *120*, 19796–19809. [[CrossRef](#)]
26. Song, Y.; Zou, L.; Li, J.; Shen, Z.P.; Cai, Y.L.; Wu, X.D. LncRNA SNHG8 promotes the development and chemo-resistance of pancreatic adenocarcinoma. *Eur. Rev. Med. Pharmacol. Sci.* **2018**, *22*, 8161–8168.
27. Chen, T.; Lei, S.; Zeng, Z.; Zhang, J.; Xue, Y.; Sun, Y.; Lan, J.; Xu, S.; Mao, D.; Guo, B. Linc00261 inhibits metastasis and the WNT signaling pathway of pancreatic cancer by regulating a miR-552-5p/FOXO3 axis. *Oncol. Rep.* **2020**, *43*, 930–942. [[CrossRef](#)]
28. Ma, L.; Wang, F.; Du, C.; Zhang, Z.; Guo, H.; Xie, X.; Gao, H.; Zhuang, Y.; Kornmann, M.; Gao, H.; et al. Long non-coding RNA MEG3 functions as a tumour suppressor and has prognostic predictive value in human pancreatic cancer. *Oncol. Rep.* **2018**, *39*, 1132–1140. [[CrossRef](#)]
29. Zhang, Y.Y.; Feng, H.M. MEG3 Suppresses Human Pancreatic Neuroendocrine Tumor Cells Growth and Metastasis by Down-Regulation of Mir-183. *Cell. Physiol. Biochem.* **2017**, *44*, 345–356. [[CrossRef](#)]
30. Iyer, S.; Modali, S.D.; Agarwal, S.K. Long Noncoding RNA MEG3 Is an Epigenetic Determinant of Oncogenic Signaling in Functional Pancreatic Neuroendocrine Tumor Cells. *Mol. Cell. Biol.* **2017**, *37*, e00278-17. [[CrossRef](#)]
31. Modali, S.D.; Parekh, V.I.; Kebebew, E.; Agarwal, S.K. Epigenetic regulation of the lncRNA MEG3 and its target c-MET in pancreatic neuroendocrine tumors. *Mol. Endocrinol.* **2015**, *29*, 224–237. [[CrossRef](#)] [[PubMed](#)]

32. Yang, Z.; Jiang, X.; Jiang, X.; Zhao, H. X-inactive-specific transcript: A long noncoding RNA with complex roles in human cancers. *Gene* **2018**, *679*, 28–35. [[CrossRef](#)] [[PubMed](#)]
33. Klec, C.; Prinz, F.; Pichler, M. Involvement of the long noncoding RNA NEAT1 in carcinogenesis. *Mol. Oncol.* **2019**, *13*, 46–60. [[CrossRef](#)]
34. Gutschner, T.; Haemmerle, M.; Diederichs, S. MALAT1—A paradigm for long noncoding RNA function in cancer. *J. Mol. Med.* **2013**, *91*, 791–801. [[CrossRef](#)] [[PubMed](#)]
35. Hutchinson, J.N.; Ensminger, A.W.; Clemson, C.M.; Lynch, C.R.; Lawrence, J.B.; Chess, A. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genom.* **2007**, *8*, 39. [[CrossRef](#)] [[PubMed](#)]
36. Tang, Z.; Kang, B.; Li, C.; Chen, T.; Zhang, Z. GEPIA2: An enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* **2019**, *47*, W556–W560. [[CrossRef](#)]
37. Ghandi, M.; Huang, F.W.; Jané-Valbuena, J.; Kryukov, G.V.; Lo, C.C.; McDonald, E.R.; Barretina, J.; Gelfand, E.T.; Bielski, C.M.; Li, H.; et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **2019**, *569*, 503–508. [[CrossRef](#)]
38. Yu, K.; Chen, B.; Aran, D.; Charalel, J.; Yau, C.; Wolf, D.M.; van 't Veer, L.J.; Butte, A.J.; Goldstein, T.; Sirota, M. Comprehensive transcriptomic analysis of cell lines as models of primary tumors across 22 tumor types. *Nat. Commun.* **2019**, *10*, 3574. [[CrossRef](#)]
39. Liu, C.; Wang, J.O.; Zhou, W.Y.; Chang, X.Y.; Zhang, M.M.; Zhang, Y.; Yang, X.H. Long non-coding RNA LINC01207 silencing suppresses AGR2 expression to facilitate autophagy and apoptosis of pancreatic cancer cells by sponging miR-143-5p. *Mol. Cell. Endocrinol.* **2019**, *493*, 110424. [[CrossRef](#)]
40. Dhamija, S.; Becker, A.C.; Sharma, Y.; Myacheva, K.; Seiler, J.; Diederichs, S. LINC00261 and the Adjacent Gene FOXA2 Are Epithelial Markers and Are Suppressed during Lung Cancer Tumorigenesis and Progression. *Noncod. RNA* **2018**, *5*, 2. [[CrossRef](#)]
41. Zhang, H.F.; Li, W.; Han, Y.D. LINC00261 suppresses cell proliferation, invasion and Notch signaling pathway in hepatocellular carcinoma. *Cancer Biomark* **2018**, *21*, 575–582. [[CrossRef](#)] [[PubMed](#)]
42. Yu, Y.; Li, L.; Zheng, Z.; Chen, S.; Chen, E.; Hu, Y. Long non-coding RNA linc00261 suppresses gastric cancer progression via promoting Slug degradation. *J. Cell. Mol. Med.* **2017**, *21*, 955–967. [[CrossRef](#)] [[PubMed](#)]
43. Dorn, A.; Glaß, M.; Neu, C.T.; Heydel, B.; Hüttelmaier, S.; Gutschner, T.; Haemmerle, M. LINC00261 Is Differentially Expressed in Pancreatic Cancer Subtypes and Regulates a Pro-Epithelial Cell Identity. *Cancers* **2020**, *12*, 1227. [[CrossRef](#)] [[PubMed](#)]
44. Jiang, W.; Liu, Y.; Liu, R.; Zhang, K.; Zhang, Y. The lncRNA DEANR1 facilitates human endoderm differentiation by activating FOXA2 expression. *Cell Rep.* **2015**, *11*, 137–148. [[CrossRef](#)]
45. Müller, S.; Raulefs, S.; Bruns, P.; Afonso-Grunz, F.; Plötner, A.; Thermann, R.; Jäger, C.; Schlitter, A.M.; Kong, B.; Regel, I.; et al. Next-generation sequencing reveals novel differentially regulated mRNAs, lncRNAs, miRNAs, sdRNAs and a piRNA in pancreatic cancer. *Mol. Cancer* **2015**, *14*, 94. [[CrossRef](#)] [[PubMed](#)]
46. Yu, Y.; Yang, J.; Li, Q.; Xu, B.; Lian, Y.; Miao, L. LINC00152: A pivotal oncogenic long non-coding RNA in human cancers. *Cell Prolif.* **2017**, *50*, e12349. [[CrossRef](#)]
47. Nötzold, L.; Frank, L.; Gandhi, M.; Polycarpou-Schwarz, M.; Groß, M.; Gunkel, M.; Beil, N.; Erfle, H.; Harder, N.; Rohr, K.; et al. The long non-coding RNA LINC00152 is essential for cell cycle progression through mitosis in HeLa cells. *Sci. Rep.* **2017**, *7*, 2265. [[CrossRef](#)]
48. Hausser, J.; Zavolan, M. Identification and consequences of miRNA-target interactions—beyond repression of gene expression. *Nat. Rev. Genet.* **2014**, *15*, 599–612. [[CrossRef](#)]
49. Messeguer, X.; Escudero, R.; Farré, D.; Nuñez, O.; Martínez, J.; Albà, M.M. PROMO: Detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* **2002**, *18*, 333–334. [[CrossRef](#)]
50. Sarvagalla, S.; Kolapalli, S.P.; Vallabhapurapu, S. The Two Sides of YY1 in Cancer: A Friend and a Foe. *Front. Oncol.* **2019**, *9*, 1230. [[CrossRef](#)]
51. Stein, Y.; Rotter, V.; Aloni-Grinstein, R. Gain-of-Function Mutant p53: All the Roads Lead to Tumorigenesis. *Int. J. Mol. Sci.* **2019**, *20*, 6197. [[CrossRef](#)] [[PubMed](#)]
52. Moser, C.; Ruummele, P.; Gehmert, S.; Schenk, H.; Kreutz, M.P.; Mycielska, M.E.; Hackl, C.; Kroemer, A.; Schnitzbauer, A.A.; Stoeltzing, O.; et al. STAT5b as molecular target in pancreatic cancer—inhibition of tumor growth, angiogenesis, and metastases. *Neoplasia* **2012**, *14*, 915–925. [[CrossRef](#)] [[PubMed](#)]

53. Tu, B.; Yao, J.; Ferri-Borgogno, S.; Zhao, J.; Chen, S.; Wang, Q.; Yan, L.; Zhou, X.; Zhu, C.; Bang, S.; et al. YAP1 oncogene is a context-specific driver for pancreatic ductal adenocarcinoma. *JCI Insight* **2019**, *4*, e130811. [[CrossRef](#)]
54. Van Nostrand, E.L.; Pratt, G.A.; Yee, B.A.; Wheeler, E.C.; Blue, S.M.; Mueller, J.; Park, S.S.; Garcia, K.E.; Gelboin-Burkhart, C.; Nguyen, T.B.; et al. Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biol.* **2020**, *21*, 90. [[CrossRef](#)]
55. Van Nostrand, E.L.; Pratt, G.A.; Shishkin, A.A.; Gelboin-Burkhart, C.; Fang, M.Y.; Sundararaman, B.; Blue, S.M.; Nguyen, T.B.; Surka, C.; Elkins, K.; et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **2016**, *13*, 508–514. [[CrossRef](#)] [[PubMed](#)]
56. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74. [[CrossRef](#)] [[PubMed](#)]
57. Hafner, M.; Landthaler, M.; Burger, L.; Khorshid, M.; Hausser, J.; Berninger, P.; Rothballer, A.; Ascano, M.; Jungkamp, A.C.; Munschauer, M.; et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **2010**, *141*, 129–141. [[CrossRef](#)]
58. Calabretta, S.; Bielli, P.; Passacantilli, I.; Piloizzi, E.; Fendrich, V.; Capurso, G.; Fave, G.D.; Sette, C. Modulation of PKM alternative splicing by PTBP1 promotes gemcitabine resistance in pancreatic cancer cells. *Oncogene* **2016**, *35*, 2031–2039. [[CrossRef](#)]
59. Liu, L.; Luo, C.; Luo, Y.; Chen, L.; Liu, Y.; Wang, Y.; Han, J.; Zhang, Y.; Wei, N.; Xie, Z.; et al. MRPL33 and its splicing regulator hnRNPk are required for mitochondria function and implicated in tumor progression. *Oncogene* **2018**, *37*, 86–94. [[CrossRef](#)]
60. Varia, S.; Cheedu, D.; Markey, M.; Torres-Shafer, K.; Battini, V.P.; Bubulya, A.; Bubulya, P.A. Alignment of Mitotic Chromosomes in Human Cells Involves SR-Like Splicing Factors Btf and TRAP150. *Int. J. Mol. Sci.* **2017**, *18*, 1956. [[CrossRef](#)]
61. Savage, K.I.; Gorski, J.J.; Barros, E.M.; Irwin, G.W.; Manti, L.; Powell, A.J.; Pellagatti, A.; Lukashchuk, N.; McCance, D.J.; McCluggage, W.G.; et al. Identification of a BRCA1-mRNA splicing complex required for efficient DNA repair and maintenance of genomic stability. *Mol. Cell* **2014**, *54*, 445–459. [[CrossRef](#)] [[PubMed](#)]
62. Winter, J.; Jung, S.; Keller, S.; Gregory, R.I.; Diederichs, S. Many roads to maturity: MicroRNA biogenesis pathways and their regulation. *Nat. Cell Biol.* **2009**, *11*, 228–234. [[CrossRef](#)] [[PubMed](#)]
63. Casey, M.C.; Prakash, A.; Holian, E.; McGuire, A.; Kalinina, O.; Shalaby, A.; Curran, C.; Webber, M.; Callagy, G.; Bourke, E.; et al. Quantifying Argonaute 2 (Ago2) expression to stratify breast cancer. *BMC Cancer* **2019**, *19*, 712. [[CrossRef](#)] [[PubMed](#)]
64. Piskounova, E.; Polytarchou, C.; Thornton, J.E.; LaPierre, R.J.; Pothoulakis, C.; Hagan, J.P.; Iliopoulos, D.; Gregory, R.I. Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms. *Cell* **2011**, *147*, 1066–1079. [[CrossRef](#)]
65. Reon, B.J.; Takao Real Karia, B.; Kiran, M.; Dutta, A. LINC00152 Promotes Invasion through a 3'-Hairpin Structure and Associates with Prognosis in Glioblastoma. *Mol. Cancer Res.* **2018**, *16*, 1470–1482. [[CrossRef](#)]
66. Cooke, A.; Schwarzl, T.; Huppertz, I.; Kramer, G.; Mantas, P.; Alleaume, A.M.; Huber, W.; Krijgsveld, J.; Hentze, M.W. The RNA-Binding Protein YBX3 Controls Amino Acid Levels by Regulating SLC mRNA Abundance. *Cell Rep.* **2019**, *27*, 3097–3106. [[CrossRef](#)]
67. Nie, M.; Balda, M.S.; Matter, K. Stress- and Rho-activated ZO-1-associated nucleic acid binding protein binding to p21 mRNA mediates stabilization, translation, and cell survival. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 10897–10902. [[CrossRef](#)] [[PubMed](#)]
68. Coles, L.S.; Bartley, M.A.; Bert, A.; Hunter, J.; Polyak, S.; Diamond, P.; Vadas, M.A.; Goodall, G.J. A multi-protein complex containing cold shock domain (Y-box) and polypyrimidine tract binding proteins forms on the vascular endothelial growth factor mRNA. Potential role in mRNA stabilization. *Eur. J. Biochem.* **2004**, *271*, 648–660. [[CrossRef](#)]
69. Dong, X.; Yang, Z.; Yang, H.; Li, D.; Qiu, X. Long Non-coding RNA MIR4435-2HG Promotes Colorectal Cancer Proliferation and Metastasis Through miR-206/YAP1 Axis. *Front. Oncol.* **2020**, *10*, 160. [[CrossRef](#)] [[PubMed](#)]
70. Zhao, M.; Xin, X.F.; Zhang, J.Y.; Dai, W.; Lv, T.F.; Song, Y. LncRNA GMD5-AS1 inhibits lung adenocarcinoma development by regulating miR-96-5p/CYLD signaling. *Cancer Med.* **2020**, *9*, 1196–1208. [[CrossRef](#)]
71. Liberzon, A.; Birger, C.; Thorvaldsdóttir, H.; Ghandi, M.; Mesirov, J.P.; Tamayo, P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **2015**, *1*, 417–425. [[CrossRef](#)] [[PubMed](#)]

72. Guttman, M.; Amit, I.; Garber, M.; French, C.; Lin, M.F.; Feldser, D.; Huarte, M.; Zuk, O.; Carey, B.W.; Cassady, J.P.; et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **2009**, *458*, 223–227. [[CrossRef](#)]
73. Rinn, J.L.; Chang, H.Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **2012**, *81*, 145–166. [[CrossRef](#)]
74. Zhao, J.; Liu, Y.; Zhang, W.; Zhou, Z.; Wu, J.; Cui, P.; Zhang, Y.; Huang, G. Long non-coding RNA Linc00152 is involved in cell cycle arrest, apoptosis, epithelial to mesenchymal transition, cell migration and invasion in gastric cancer. *Cell Cycle* **2015**, *14*, 3112–3123. [[CrossRef](#)] [[PubMed](#)]
75. Cai, Q.; Wang, Z.; Wang, S.; Weng, M.; Zhou, D.; Li, C.; Wang, J.; Chen, E.; Quan, Z. Long non-coding RNA LINC00152 promotes gallbladder cancer metastasis and epithelial-mesenchymal transition by regulating HIF-1 $\alpha$  via miR-138. *Open Biol.* **2017**, *7*, 160247. [[CrossRef](#)] [[PubMed](#)]
76. Ye, T.; Ding, W.; Wang, N.; Huang, H.; Pan, Y.; Wei, A. Long noncoding RNA linc00346 promotes the malignant phenotypes of bladder cancer. *Biochem. Biophys. Res. Commun.* **2017**, *491*, 79–84. [[CrossRef](#)]
77. Qu, Y.K.; Qu, X.S.; Chen, G.; Feng, Y.; Teng, X.L.; Liu, W.X.; Cheng, Z.X.; Xu, J.; Guo, L.Q. LINC01503 promotes cell proliferation, invasion and EMT process in cholangio-carcinoma. *Eur. Rev. Med. Pharmacol. Sci.* **2019**, *23*, 6445–6452. [[PubMed](#)]
78. Peng, W.X.; He, R.Z.; Zhang, Z.; Yang, L.; Mo, Y.Y. LINC00346 promotes pancreatic cancer progression through the CTCF-mediated Myc transcription. *Oncogene* **2019**, *38*, 6770–6780. [[CrossRef](#)]
79. Shi, W.; Zhang, C.; Ning, Z.; Hua, Y.; Li, Y.; Chen, L.; Liu, L.; Chen, Z.; Meng, Z. Long non-coding RNA LINC00346 promotes pancreatic cancer growth and gemcitabine resistance by sponging miR-188-3p to derepress BRD4 expression. *J. Exp. Clin. Cancer Res.* **2019**, *38*, 60. [[CrossRef](#)] [[PubMed](#)]
80. Yuan, J.H.; Yang, F.; Wang, F.; Ma, J.Z.; Guo, Y.J.; Tao, Q.F.; Liu, F.; Pan, W.; Wang, T.T.; Zhou, C.C.; et al. A long noncoding RNA activated by TGF- $\beta$  promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer Cell* **2014**, *25*, 666–681. [[CrossRef](#)]
81. Terai, G.; Iwakiri, J.; Kameda, T.; Hamada, M.; Asai, K. Comprehensive prediction of lncRNA-RNA interactions in human transcriptome. *BMC Genom.* **2016**, *17* (Suppl. 1), 12. [[CrossRef](#)] [[PubMed](#)]
82. Collisson, E.A.; Bailey, P.; Chang, D.K.; Biankin, A.V. Molecular subtypes of pancreatic cancer. *Nat. Rev. Gastroenterol. Hepatol.* **2019**, *16*, 207–220. [[CrossRef](#)] [[PubMed](#)]
83. Guinney, J.; Dienstmann, R.; Wang, X.; de Reyniès, A.; Schlicker, A.; Soneson, C.; Marisa, L.; Roepman, P.; Nyamundanda, G.; Angelino, P.; et al. The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **2015**, *21*, 1350–1356. [[CrossRef](#)] [[PubMed](#)]
84. Rudin, C.M.; Poirier, J.T.; Byers, L.A.; Dive, C.; Dowlati, A.; George, J.; Heymach, J.V.; Johnson, J.E.; Lehman, J.M.; MacPherson, D.; et al. Molecular subtypes of small cell lung cancer: A synthesis of human and mouse model data. *Nat. Rev. Cancer* **2019**, *19*, 289–297. [[CrossRef](#)] [[PubMed](#)]
85. Harbeck, N.; Gnant, M. Breast cancer. *Lancet* **2017**, *389*, 1134–1150. [[CrossRef](#)]
86. Mello, S.S.; Sinow, C.; Raj, N.; Mazur, P.K.; Biegging-Rolett, K.; Broz, D.K.; Imam, J.F.C.; Vogel, H.; Wood, L.D.; Sage, J.; et al. Neat1 is a p53-inducible lincRNA essential for transformation suppression. *Genes Dev.* **2017**, *31*, 1095–1108. [[CrossRef](#)] [[PubMed](#)]
87. Cao, J.; Zhang, Y.; Yang, J.; He, S.; Li, M.; Yan, S.; Chen, Y.; Qu, C.; Xu, L. NEAT1 regulates pancreatic cancer cell growth, invasion and migration through miR-335-5p/c-met axis. *Am. J. Cancer Res.* **2016**, *6*, 2361–2374.
88. Huang, B.; Liu, C.; Wu, Q.; Zhang, J.; Min, Q.; Sheng, T.; Wang, X.; Zou, Y. Long non-coding RNA NEAT1 facilitates pancreatic cancer progression through negative modulation of miR-506-3p. *Biochem. Biophys. Res. Commun.* **2017**, *482*, 828–834. [[CrossRef](#)]
89. Luo, Z.; Yi, Z.J.; Ou, Z.L.; Han, T.; Wan, T.; Tang, Y.C.; Wang, Z.C.; Huang, F.Z. RELA/NEAT1/miR-302a-3p/RELA feedback loop modulates pancreatic ductal adenocarcinoma cell proliferation and migration. *J. Cell. Physiol.* **2019**, *234*, 3583–3597. [[CrossRef](#)]
90. Feng, Y.; Gao, L.; Cui, G.; Cao, Y. LncRNA NEAT1 facilitates pancreatic cancer growth and metastasis through stabilizing ELF3 mRNA. *Am. J. Cancer Res.* **2020**, *10*, 237–248.
91. Jiao, F.; Hu, H.; Yuan, C.; Wang, L.; Jiang, W.; Jin, Z.; Guo, Z.; Wang, L. Elevated expression level of long noncoding RNA MALAT-1 facilitates cell growth, migration and invasion in pancreatic cancer. *Oncol. Rep.* **2014**, *32*, 2485–2492. [[CrossRef](#)]

92. Cheng, Y.; Imanirad, P.; Jutooru, I.; Hedrick, E.; Jin, U.H.; Rodrigues Hoffman, A.; Leal de Araujo, J.; Morpurgo, B.; Golovko, A.; Safe, S. Role of metastasis-associated lung adenocarcinoma transcript-1 (MALAT-1) in pancreatic cancer. *PLoS ONE* **2018**, *13*, e0192264. [[CrossRef](#)] [[PubMed](#)]
93. Kim, J.; Piao, H.L.; Kim, B.J.; Yao, F.; Han, Z.; Wang, Y.; Xiao, Z.; Siverly, A.N.; Lawhon, S.E.; Ton, B.N.; et al. Long noncoding RNA MALAT1 suppresses breast cancer metastasis. *Nat. Genet.* **2018**, *50*, 1705–1715. [[CrossRef](#)] [[PubMed](#)]
94. Sun, Y.; Ma, L. New Insights into Long Non-Coding RNA MALAT1 in Cancer and Metastasis. *Cancers* **2019**, *11*, 216. [[CrossRef](#)]
95. Arun, G.; Spector, D.L. MALAT1 long non-coding RNA and breast cancer. *RNA Biol.* **2019**, *16*, 860–863. [[CrossRef](#)] [[PubMed](#)]
96. Arun, G.; Diermeier, S.; Akerman, M.; Chang, K.C.; Wilkinson, J.E.; Hearn, S.; Kim, Y.; MacLeod, A.R.; Krainer, A.R.; Norton, L.; et al. Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. *Genes Dev.* **2016**, *30*, 34–51. [[CrossRef](#)]
97. Xie, Z.C.; Dang, Y.W.; Wei, D.M.; Chen, P.; Tang, R.X.; Huang, Q.; Liu, J.H.; Luo, D.Z. Clinical significance and prospective molecular mechanism of MALAT1 in pancreatic cancer exploration: A comprehensive study based on the GeneChip, GEO, Oncomine, and TCGA databases. *Onco Targets Ther.* **2017**, *10*, 3991–4005. [[CrossRef](#)]
98. Iacobuzio-Donahue, C.A.; Maitra, A.; Olsen, M.; Lowe, A.W.; van Heek, N.T.; Rosty, C.; Walter, K.; Sato, N.; Parker, A.; Ashfaq, R.; et al. Exploration of global gene expression patterns in pancreatic adenocarcinoma using cDNA microarrays. *Am. J. Pathol.* **2003**, *162*, 1151–1162. [[CrossRef](#)]
99. Pei, H.; Li, L.; Fridley, B.L.; Jenkins, G.D.; Kalari, K.R.; Lingle, W.; Petersen, G.; Lou, Z.; Wang, L. FKBP51 affects cancer cell response to chemotherapy by negatively regulating Akt. *Cancer Cell* **2009**, *16*, 259–266. [[CrossRef](#)]
100. Ishikawa, M.; Yoshida, K.; Yamashita, Y.; Ota, J.; Takada, S.; Kisanuki, H.; Koinuma, K.; Choi, Y.L.; Kaneda, R.; Iwao, T.; et al. Experimental trial for diagnosis of pancreatic ductal carcinoma based on gene expression profiles of pancreatic ductal cells. *Cancer Sci.* **2005**, *96*, 387–393. [[CrossRef](#)]
101. Frappart, P.O.; Walter, K.; Gout, J.; Beutel, A.K.; Morawe, M.; Arnold, F.; Breunig, M.; Barth, T.F.; Marienfeld, R.; Schulte, L.; et al. Pancreatic cancer-derived organoids—A disease modeling tool to predict drug response. *United Eur. Gastroenterol. J.* **2020**, *8*, 2050640620905183. [[CrossRef](#)] [[PubMed](#)]
102. Carugo, A.; Genovese, G.; Seth, S.; Nezi, L.; Rose, J.L.; Bossi, D.; Cicalese, A.; Shah, P.K.; Viale, A.; Pettazoni, P.F.; et al. In Vivo Functional Platform Targeting Patient-Derived Xenografts Identifies WDR5-Myc Association as a Critical Determinant of Pancreatic Cancer. *Cell Rep.* **2016**, *16*, 133–147. [[CrossRef](#)]
103. Wang, Q.; Armenia, J.; Zhang, C.; Penson, A.V.; Reznik, E.; Zhang, L.; Minet, T.; Ochoa, A.; Gross, B.E.; Iacobuzio-Donahue, C.A.; et al. Unifying cancer and normal RNA sequencing data from different sources. *Sci. Data* **2018**, *5*, 180061. [[CrossRef](#)] [[PubMed](#)]
104. Martinelli, P.; Carrillo-de Santa Pau, E.; Cox, T.; Sainz, B.; Dusetti, N.; Greenhalf, W.; Rinaldi, L.; Costello, E.; Ghaneh, P.; Malats, N.; et al. GATA6 regulates EMT and tumour dissemination, and is a marker of response to adjuvant chemotherapy in pancreatic cancer. *Gut* **2017**, *66*, 1665–1676. [[CrossRef](#)] [[PubMed](#)]
105. Li, Z.T.; Zhang, X.; Wang, D.W.; Xu, J.; Kou, K.J.; Wang, Z.W.; Yong, G.; Liang, D.S.; Sun, X.Y. Overexpressed lncRNA GATA6-AS1 Inhibits LNM and EMT via FZD4 through the Wnt/ $\beta$ -Catenin Signaling Pathway in GC. *Mol. Ther. Nucleic Acids* **2020**, *19*, 827–840. [[CrossRef](#)]
106. Liu, Z.; Dai, J.; Shen, H. Systematic analysis reveals long noncoding RNAs regulating neighboring transcription factors in human cancers. *Biochim. Biophys. Acta Mol. Basis Dis.* **2018**, *1864*, 2785–2792. [[CrossRef](#)]
107. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [[CrossRef](#)]
108. Robinson, M.D.; Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **2010**, *11*, R25. [[CrossRef](#)]
109. Cunningham, F.; Achuthan, P.; Akanni, W.; Allen, J.; Amode, M.R.; Armean, I.M.; Bennett, R.; Bhai, J.; Billis, K.; Boddu, S.; et al. Ensembl 2019. *Nucleic Acids Res.* **2019**, *47*, D745–D751. [[CrossRef](#)]
110. Gudenäs, B.L.; Wang, L. Prediction of LncRNA Subcellular Localization with Deep Learning from Sequence Features. *Sci. Rep.* **2018**, *8*, 16385. [[CrossRef](#)]

111. Quinlan, A.R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinform.* **2014**, *47*, 1–34. [[CrossRef](#)] [[PubMed](#)]
112. Yang, Y.C.; Di, C.; Hu, B.; Zhou, M.; Liu, Y.; Song, N.; Li, Y.; Umetsu, J.; Lu, Z.J. CLIPdb: A CLIP-seq database for protein-RNA interactions. *BMC Genom.* **2015**, *16*, 51. [[CrossRef](#)]
113. Ru, Y.; Kechris, K.J.; Tabakoff, B.; Hoffman, P.; Radcliffe, R.A.; Bowler, R.; Mahaffey, S.; Rossi, S.; Calin, G.A.; Bemis, L.; et al. The multiMiR R package and database: Integration of microRNA-target interactions along with their disease and drug associations. *Nucleic Acids Res.* **2014**, *42*, e133. [[CrossRef](#)]
114. Smedley, D.; Haider, S.; Durinck, S.; Pandini, L.; Provero, P.; Allen, J.; Arnaiz, O.; Awedh, M.H.; Baldock, R.; Barbiera, G.; et al. The BioMart community portal: An innovative alternative to large, centralized data repositories. *Nucleic Acids Res.* **2015**, *43*, W589–W598. [[CrossRef](#)] [[PubMed](#)]
115. Durinck, S.; Spellman, P.T.; Birney, E.; Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **2009**, *4*, 1184–1191. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).