



# Aberrant Expression of Pseudogene-Derived lncRNAs as an Alternative Mechanism of Cancer Gene Regulation in Lung Adenocarcinoma

Greg L. Stewart<sup>1\*</sup>, Katey S. S. Enfield<sup>1,2</sup>, Adam P. Sage<sup>1</sup>, Victor D. Martinez<sup>1</sup>, Brenda C. Minatel<sup>1</sup>, Michelle E. Pewarchuk<sup>1</sup>, Erin A. Marshall<sup>1</sup> and Wan L. Lam<sup>1</sup>

<sup>1</sup> BC Cancer Research Centre, Vancouver, BC, Canada, <sup>2</sup> The Francis Crick Institute, London, United Kingdom

## OPEN ACCESS

### Edited by:

Yujing Li,  
Emory University, United States

### Reviewed by:

Sergio Verjovski-Almeida,  
University of São Paulo, Brazil  
Nithyananda Thorenoor,  
Pennsylvania State University,  
United States

### \*Correspondence:

Greg L. Stewart  
gstewart@bccrc.ca

### Specialty section:

This article was submitted to  
RNA,  
a section of the journal  
Frontiers in Genetics

**Received:** 16 November 2018

**Accepted:** 11 February 2019

**Published:** 06 March 2019

### Citation:

Stewart GL, Enfield KSS, Sage AP, Martinez VD, Minatel BC, Pewarchuk ME, Marshall EA and Lam WL (2019) Aberrant Expression of Pseudogene-Derived lncRNAs as an Alternative Mechanism of Cancer Gene Regulation in Lung Adenocarcinoma. *Front. Genet.* 10:138. doi: 10.3389/fgene.2019.00138

Transcriptome sequencing has led to the widespread identification of long non-coding RNAs (lncRNAs). Subsequently, these genes have been shown to hold functional importance in human cellular biology, which can be exploited by tumors to drive the hallmarks of cancer. Due to the complex tertiary structure and unknown binding motifs of lncRNAs, there is a growing disparity between the number of lncRNAs identified and those that have been functionally characterized. As such, lncRNAs deregulated in cancer may represent critical components of cancer pathways that could serve as novel therapeutic intervention points. Pseudogenes are non-coding DNA sequences that are defunct relatives of their protein-coding parent genes but retain high sequence similarity. Interestingly, certain lncRNAs expressed from pseudogene loci have been shown to regulate the protein-coding parent genes of these pseudogenes in *trans* particularly because of this sequence complementarity. We hypothesize that this phenomenon occurs more broadly than previously realized, and that aberrant expression of lncRNAs overlapping pseudogene loci provides an alternative mechanism of cancer gene deregulation. Using RNA-sequencing data from two cohorts of lung adenocarcinoma, each paired with patient-matched non-malignant lung samples, we discovered 104 deregulated pseudogene-derived lncRNAs. Remarkably, many of these deregulated lncRNAs (i) were expressed from the loci of pseudogenes related to known cancer genes, (ii) had expression that significantly correlated with protein-coding parent gene expression, and (iii) had lncRNA protein-coding parent gene expression that was significantly associated with survival. Here, we uncover evidence to suggest the lncRNA-pseudogene-protein-coding gene axis as a prominent mechanism of cancer gene regulation in lung adenocarcinoma, and highlights the clinical utility of exploring the non-coding regions of the cancer transcriptome.

**Keywords:** long non-coding RNAs, pseudogenes, lung cancer, gene regulation, non-coding RNAs

## INTRODUCTION

Lung cancer is an enormous health burden, representing the most common cause of cancer death worldwide. While advances in imaging technology have improved early detection, the outcome for lung cancer patients remains poor. This is largely attributed to limited therapeutic options and short-term response due to tumor heterogeneity. Despite extensive sequencing efforts to characterize exomes and protein-coding transcriptomes of lung tumors, less than half of lung adenocarcinoma (LUAD) patients harbor clinically targetable mutated driver genes, highlighting the need to explore alternative mechanisms of cancer gene deregulation (Chan and Hughes, 2015).

Long non-coding RNAs (lncRNAs, > 200 nt) have emerged as important players in cell biology. lncRNAs have been observed to function through a wide variety of regulatory mechanisms, targeting DNA, proteins, and other RNA species; specifically with defined roles in RNA degradation or stabilization, protein translocation and complex formation, and recruitment of complexes to transcriptional loci (Lee et al., 1999; Wang et al., 2011; Kung et al., 2013; Simon et al., 2013). This broad functional repertoire has been shown to be exploited by many cancer types, including LUAD, to drive various hallmarks of cancer (Bhan et al., 2017; Peng et al., 2017). One of the first cancer-associated lncRNA transcripts, *MALAT1* (Metastasis Associated Lung Adenocarcinoma Transcript 1) was discovered in lung cancer. *MALAT1* is overexpressed in metastatic lung tumors and functions *in trans* through transcriptional regulation of multiple genes involved in cell motility (Tano et al., 2010; Gutschner et al., 2013). Since the discovery of *MALAT1*, many other lncRNAs have been shown to play a direct role in nearly every major cancer type. Tumor suppressive lncRNA transcripts are commonly observed to be downregulated, for instance the lncRNA *TARID*, which promotes the expression of the tumor suppressor gene *TCF21* through active promoter demethylation (Arab et al., 2014). Conversely oncogenic lncRNAs can be overexpressed, such as *NEAT1*, an architectural lncRNA associated with metastasis in many cancer types (Hirose et al., 2014; Blume et al., 2015; Yamazaki et al., 2018; Zhang et al., 2018). As such, lncRNAs may represent critical regulators of oncogenic-driver pathways that could serve as undiscovered clinical intervention points in LUAD.

While lncRNAs have been observed to be important in cancer biology, functional prediction of newly-discovered lncRNAs remains a major challenge. However, many lncRNAs expressed from pseudogene loci have been shown to regulate the specific genes with which they have sequence homology. Pseudogenes are DNA sequences that are defunct relatives of functional protein-coding genes (herein referred to as parent genes) and arise during either gene duplication events, or the reverse transcription of an mRNA transcript into a new genomic location. Through evolution these duplicated genes have acquired mutations such as premature stop codons and frameshifts, which results in the loss of protein coding ability, while still retaining a high degree of sequence homology with the original parent gene (Khachane and Harrison, 2009). Recently, pseudogene-derived lncRNAs have been shown to regulate their parent genes and this novel mechanism has been observed in many tumor types, including

lung cancer (Sun et al., 2017; Hu et al., 2018). A prominent example is the tumor suppressor gene *PTEN* (chromosome 10), regulated both positively by *PTENP1* (chromosome 9), a lncRNA transcribed from the sense strand of the pseudogene locus, and negatively by the lncRNA *PTENP1-AS1*, which is transcribed from the strand antisense to the parent gene (Johnsson et al., 2013).

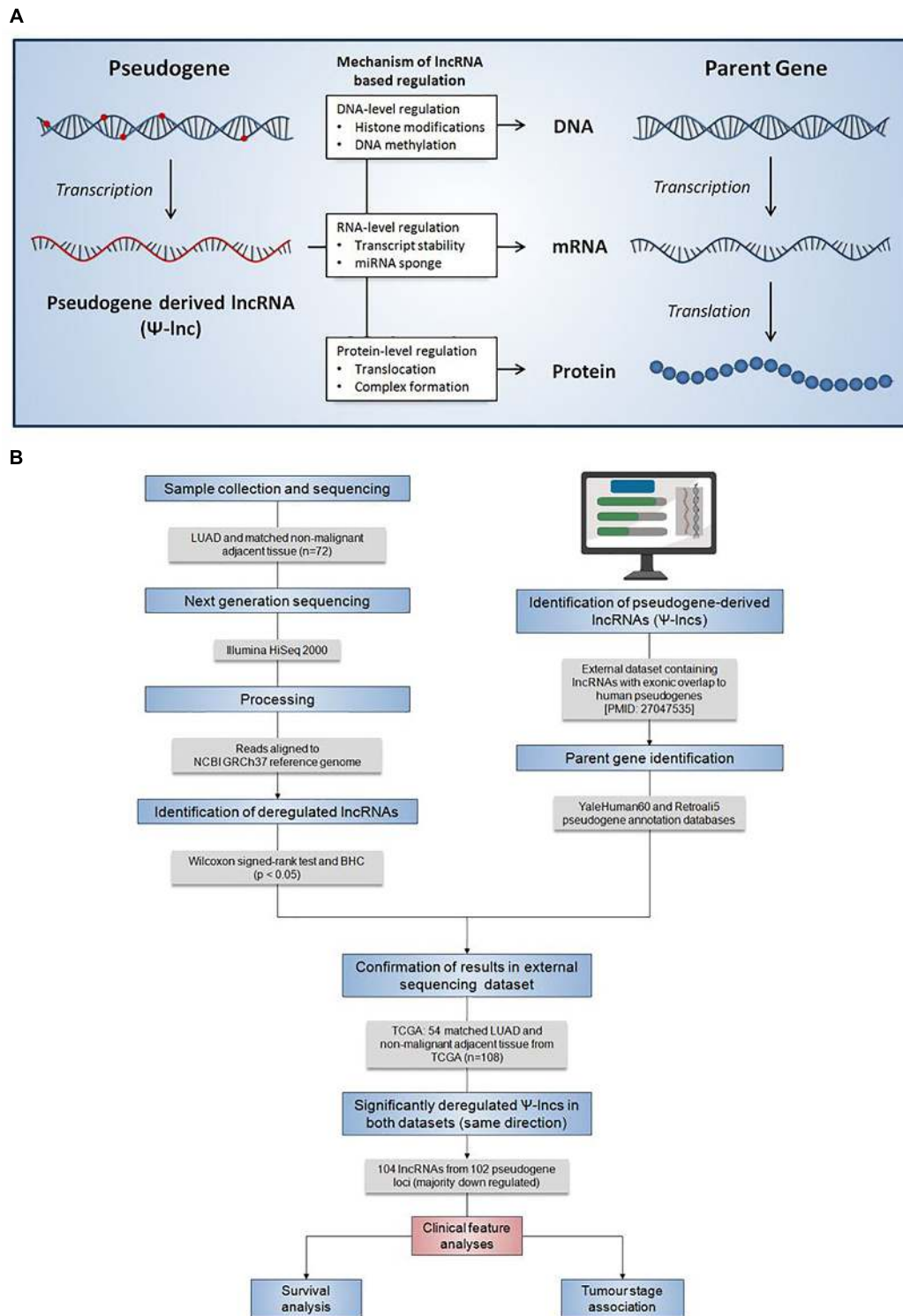
Pseudogenes have been continually omitted from large RNA-sequencing datasets due to the complexity of separating highly similar pseudogene sequences from parent genes. However, Milligan et al. (2016) recently generated an atlas of lncRNAs overlapping pseudogenes, which has provided a foundation for their analysis in RNA sequencing datasets. We hypothesize that the functions of pseudogene-derived lncRNAs are an under-explored mechanism of gene regulation that occurs more broadly than previously realized, and that these events contribute to the tumorigenesis of LUAD. We performed next generation RNA-sequencing on microdissected LUAD tumors and matched non-malignant tissue to identify deregulated lncRNAs expressed from pseudogene loci (herein referred to as  $\Psi$ -lncs). We then explored the relationships of these  $\Psi$ -lncs with their parent genes, and explored their significance in relation to patient clinical features in our discovery dataset as well as a validation dataset.

## RESULTS

### $\Psi$ -lnc Expression Is Deregulated in Lung Adenocarcinoma

Pseudogenes vary widely in terms of length, gene fraction, and identity to parent genes, and can be expressed as lncRNAs that are sense, antisense, partial overlapping, or internal to the parent. In light of this variation,  $\Psi$ -lncs are observed to have vastly different regulatory effects on downstream target genes (Figure 1A). In our curation of  $\Psi$ -lncs in LUAD, we have included those that have exonic overlap with a pseudogene (partial or full length) and considered both sense and antisense transcripts (Supplementary Table S1).  $\Psi$ -lncs were analyzed in an in-house discovery (BCCA,  $n = 72$ ) and external validation (TCGA,  $n = 108$ ) cohort of LUAD and paired non-malignant lung tissues (Table 1). We identified aberrantly expressed  $\Psi$ -lncs that are significantly deregulated in both the discovery and validation datasets with the same direction of expression alteration ( $\Psi$ -lncs upregulated or downregulated in tumors compared to matched non-malignant tissue).

We found 104 lncRNAs expressed from 102 pseudogene loci to be significantly deregulated in LUAD (Supplementary Table S2). To our surprise, we found that the majority of these deregulated  $\Psi$ -lncs were downregulated in tumors (Figures 2A,B). Most of these were unannotated lncRNAs, such as *RP11-1007O24.3*, which was downregulated in tumors, with only 24 of the total deregulated  $\Psi$ -lncs having been previously described in scientific literature annotated in PubMed, albeit none in the field of pseudogene-mediated deregulation (Supplementary Table S3 and Figure 2B). Twenty of these 24 have been described in the context of cancer, with only four in lung cancer. This includes *DGCR5*, a lncRNA we found to be overexpressed in



**FIGURE 1** | Summary of the regulatory mechanisms of  $\Psi$ -lncs and the analysis pipeline for the identification of their deregulation in lung adenocarcinoma.

**(A)** Summary of the regulatory mechanisms of pseudogene-derived lncRNAs ( $\Psi$ -lncs) that retain sequence homology with the parent gene. Overall, lncRNAs have been shown to function through a variety of regulatory mechanisms, acting on the DNA, RNA, and protein levels. **(B)** Flow diagram description of the analysis pipeline applied for the identification of deregulated  $\Psi$ -lncs. Patient LUAD samples were collected and subjected to next generation sequencing to quantify RNA expression. Gene expression was then compared between tumors and matched non-malignant tissue to identify significantly deregulated transcripts. lncRNAs with exonic overlap to known pseudogenes were then identified and confirmed in a 2nd set of LUAD and matched non-malignant tissue. Deregulated  $\Psi$ -lncs were then assayed to determine associations with clinical features.

**TABLE 1** | Patient clinical characteristics of the discovery (BCCA), validation (TCGA), and survival analysis (KmPlotter) datasets.

Characteristic	BCCA	TCGA	KmPlotter
Samples (pairs)	72 (36)	108 (54)	866 (n/a)
Sex			
Male	10	24	344
Female	26	30	318
Average age	70	67	–
Stage			
I	20 (56%)	28 (52%)	370 (69%)
II	11 (31%)	14 (30%)	136 (25%)
III	3 (8%)	10 (19%)	24 (4%)
IV	1 (3%)	2 (4%)	4 (1%)
Ethnicity			
Caucasian	11 (31%)	51 (94%)	–
Asian	14 (39%)	–	–
Unknown	11 (31%)	–	–
Black	–	3 (6%)	–
Smoking			
Current	5 (14%)	7 (13%)	246 (Ever) <sup>a</sup>
Former	6 (17%)	36 (67%)	
Never	25 (69%)	5 (9%)	246

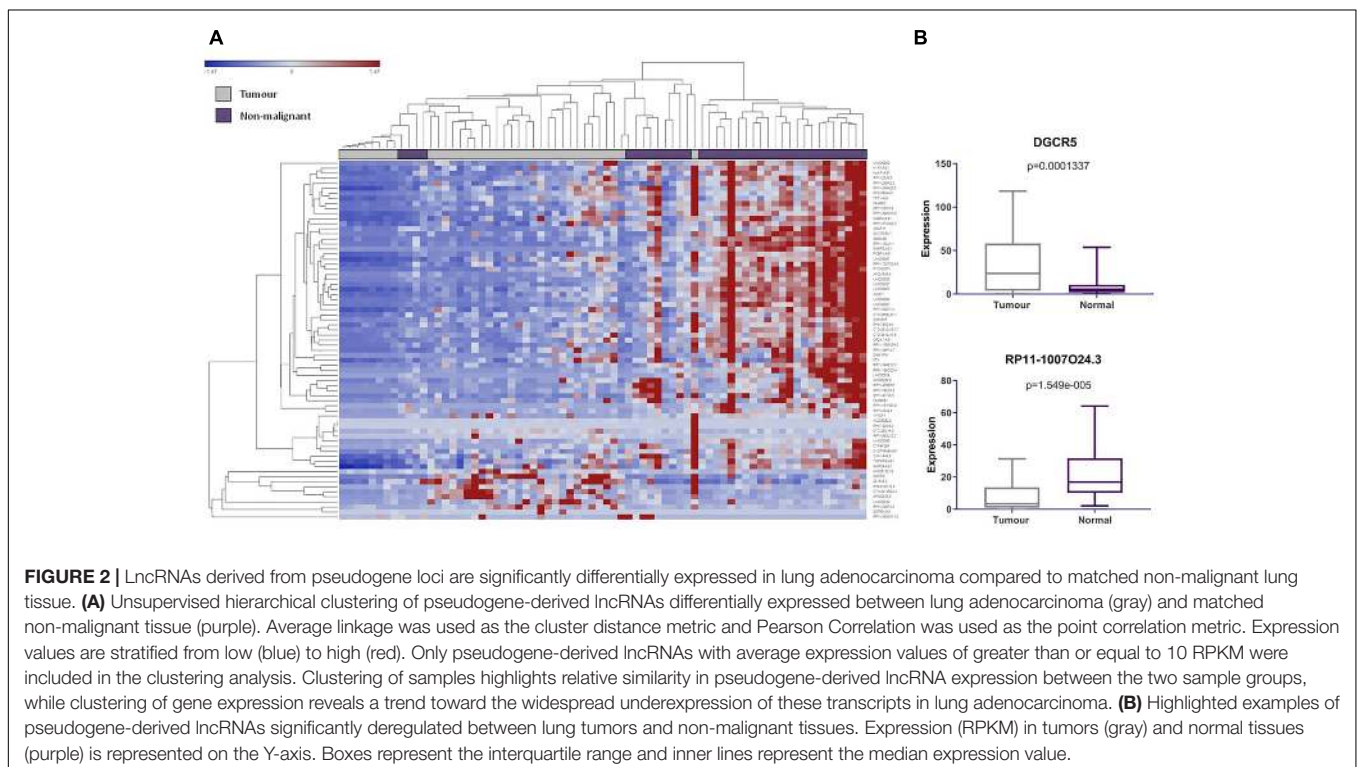
<sup>a</sup> Ever smokers is a term that includes both current and former smokers.

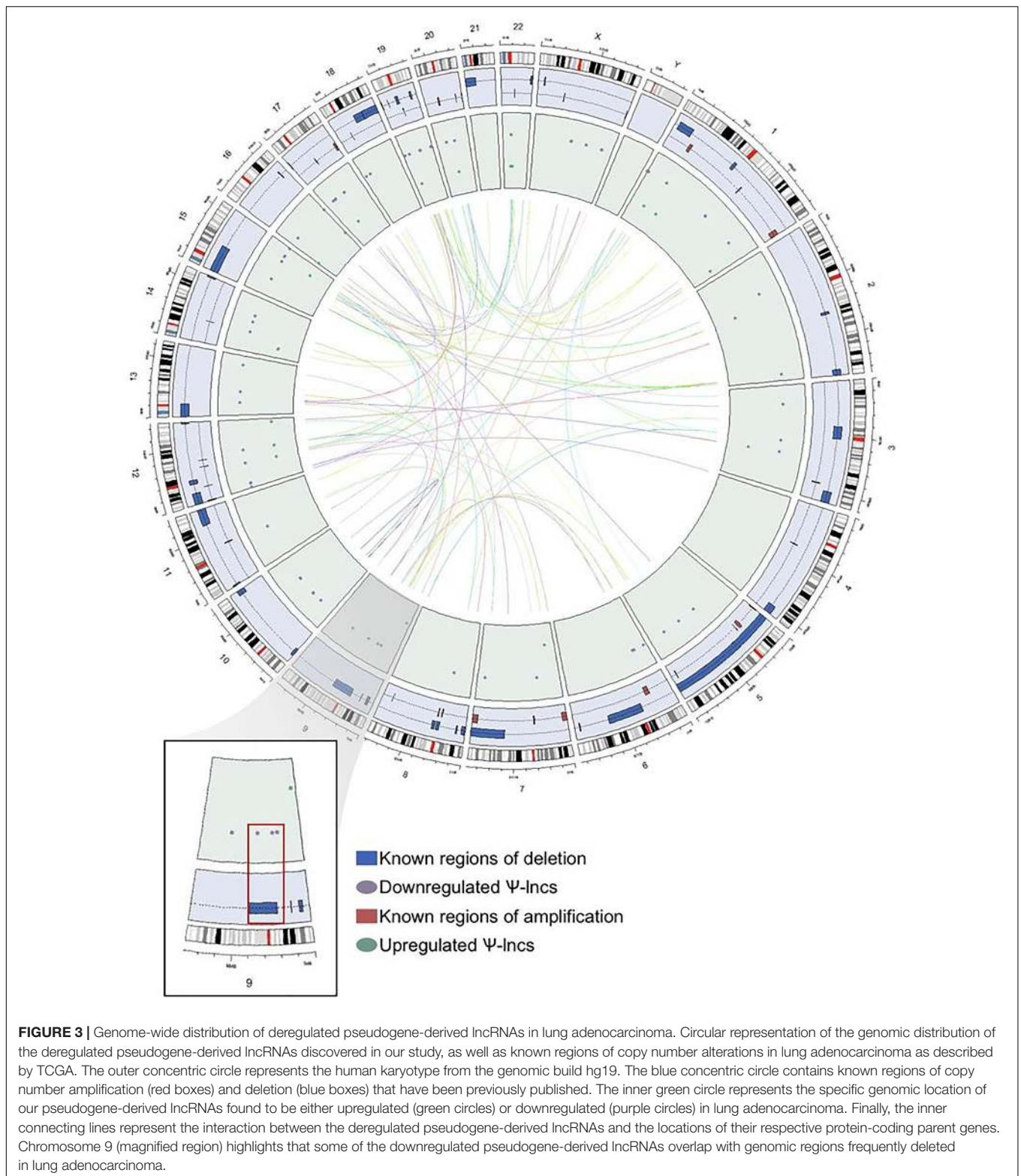
tumors. *DGCR5* has been reported to promote LUAD progression by sequestering a variety of miRNAs involved in cell cycle regulation, although it has not been investigated with regard to its pseudogene-derived nature (Chen et al., 2017; Dong et al., 2018; Luo et al., 2018).

We were interested in examining the genetic events that could impact pseudogene loci, and thus affect  $\Psi$ -lnc expression. We mapped the chromosomal distribution of the deregulated  $\Psi$ -lncs, finding them to be distributed throughout the genome and detected on most chromosomes, except for chromosomes 4 and Y (Figure 3). The locations of each of the parent genes of deregulated  $\Psi$ -lncs are similarly distributed through the genome (Supplementary Table S4). We then determined the overlap of these genes with regions of recurrent chromosomal amplification and deletion as determined by The Cancer Genome Atlas (TCGA) for LUAD (Zack et al., 2013). While some  $\Psi$ -lncs overlap with regions of recurrent deletion, the majority do not, indicating that they may be regulated by mechanisms other than copy number alteration (Figure 3).

## Global Patterns of $\Psi$ -lnc and Parental Gene Expression

As a first step to identify deregulated  $\Psi$ -lncs that may function through regulation of their respective parent genes, we explored whether  $\Psi$ -lncs with significantly deregulated expression were associated with altered parent gene transcript levels (Poliseno et al., 2010; Johnsson et al., 2013; Feng et al., 2017; Huang et al., 2018). We obtained parent gene information for the 95 deregulated  $\Psi$ -lncs and determined that they shared 104 parent genes. Some pseudogenes contained multiple lncRNAs, and some lncRNAs overlapped multiple pseudogenes, constituting a total number of 116  $\Psi$ -lnc-parent gene pairings. For each  $\Psi$ -lnc-parent pair we compared groups of tumors with high levels of  $\Psi$ -lnc expression to those with low levels of  $\Psi$ -lnc expression.





We found that 33  $\Psi$ -lncs have a significant expression relationship with their parent gene in at least 1 dataset (**Supplementary Table S5**). This included 21 sense  $\Psi$ -lnc-parent-gene pairs, and 13 antisense  $\Psi$ -lnc-parent gene pairs. Having identified

$\Psi$ -lncs with expression associated with parent gene expression, we investigated whether parent genes had known oncogenic or tumor suppressive roles. We performed a literature search to determine if any had been previously described in the context

**TABLE 2** | Parent genes of deregulated  $\Psi$ -lncs previously described in cancer literature.

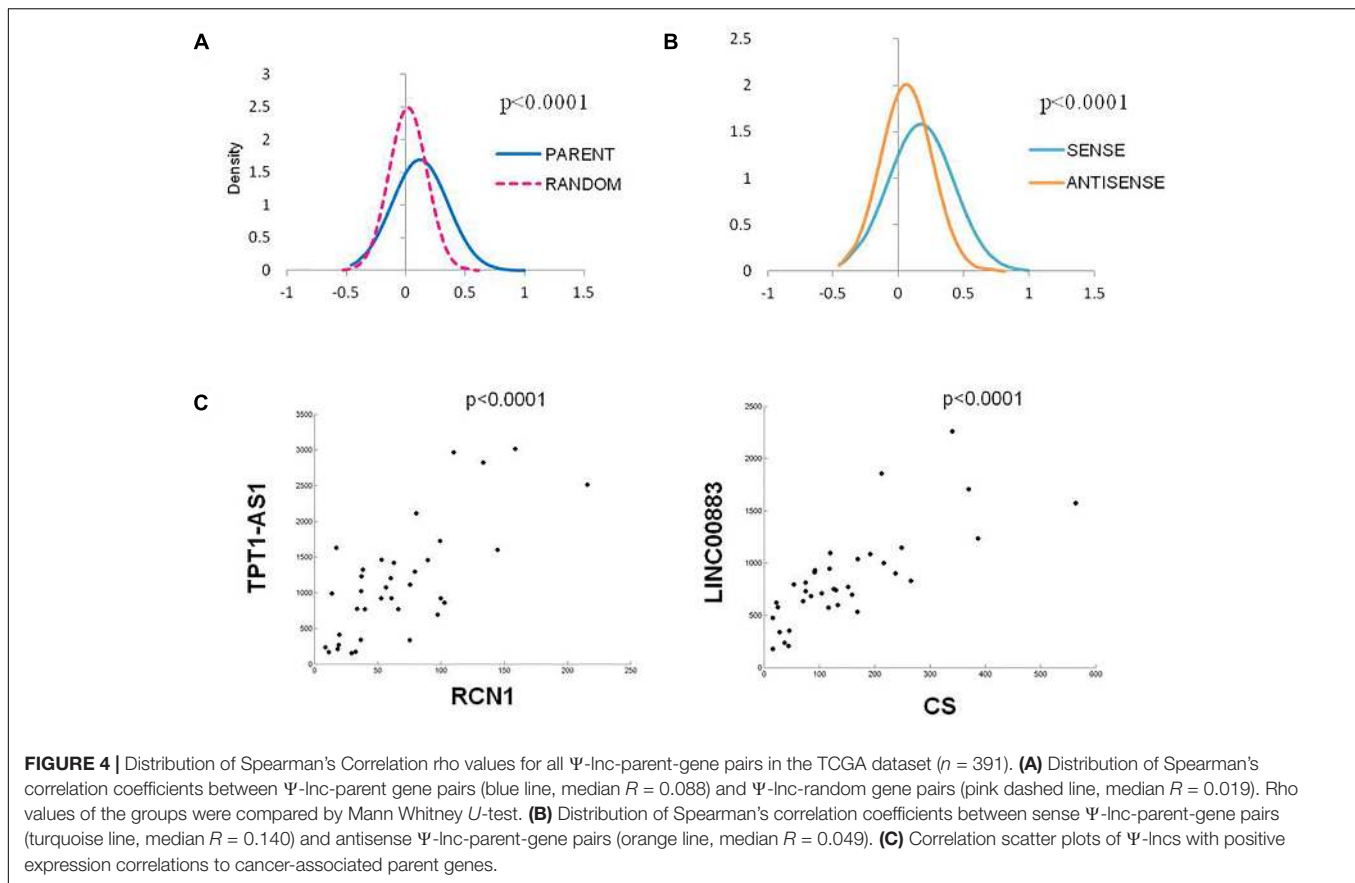
Parent Gene	Cancer <sup>a</sup>	Lung <sup>a</sup> Cancer	$\Psi$ -lnc	Parent Gene	Cancer <sup>a</sup>	Lung <sup>a</sup> Cancer	$\Psi$ -lncRNA
ADC	8	–	RP11-439L8.3	MT1	>1000	95	MMP24-AS1
AGAP1	12	–	AGAP11	NEBL	17	–	LINC00342
ARIH1	4	–	RP11-1007O24.3	NUP210	9	2	H1FX-AS1
ATXN7L3	5	–	RP11-56G10.2	PCBP2	26	–	PCBP1-AS1
BCR	>1000	>1000	AC008132.13	PKD1	238	5	RP11-1186N24.5
BMS1	4	–	AGAP11	POTEF	4	–	RP11-193H5.1
CDC42	1657	129	RP11-390F4.3	PPARGC1B	37	1	RP11-527N22.1
CECR7	3	–	AP000525.9;	PPY	91	5	CTD-2008P7.8
			CTD-2314B22.3	PRKX	12	2	RP11-526I2.1
CELSR1	11	2	DGCR5	PTCHD3	2	–	PTCHD3P1
CHRN1	8	–	RP11-650L12.2	PTMA	34	–	LINC00987
CIC	–	96	RP11-34P13.7	PZP	20	–	LINC00987
CS	–	10	LINC00883;	RAB11FIP1	16	1	RP3-368A4.5
			RP11-446H18.5	RAB40B	5	–	LL0XNCO1-250H12.3
CSPG4	62	5	DNM1P51	RCN1	17	1	TPT1-AS1
CTAGE1	49	–	RP1-122P22.2	RNASEH1	7	–	RP11-344E13.3
CYP4F2	29	–	CYP4F35P	RPL21	5	–	AC005062.2
CYP4F3	10	1	FAM95B1	RPL23A	8	–	HLA-F-AS1
CYP4F31P	10	1	CYP4F35P;	RPSA	51	3	FTX
			FAM95B1	RPSAP58	–	3	LINC00466
DFFB	33	1	TOPORES-AS1	SEMA3A	163	25	SEMA3B
DRD2	121	10	AP000438.2	SHQ1	10	1	RP11-50E11.3
EGLN1	158	5	RP11-182J1.1	SNAPC5	4	–	LY86-AS1
FAM103A1	1	–	RP11-324H6.5	SNX18	2	–	RP11-435B5.5
GPR39	8	–	RP11-399O19.9	SRSF9	5	–	RP11-752G15.3
HMGB1	>1000	111	RP11-349A22.5;	TACC3	122	21	LINC00667
			ZBED3-AS1	TOMM40	10	2	CTD-2314B22.3
HMG2P46	81	1	RAMP2-AS1	TPTE	69	7	TPTEP1
KRT8	94	7	RP5-1198O20.4	TUBB4B	5	–	RP11-386G11.10
LINC00657	2	–	CTA-14H9.5;	TULP3	5	–	LINC00359
			HCG11	VENTX	8	1	RP11-81H3.2
LMNB2	3	1	RP11-161H23.9	VWF	733	52	TPTEP1
MARK4	25	3	CTD-2201G3.1	ZNF14	2	–	CTD-2666L21.1;
MICE	1	–	HLA-F-AS1				ZNF833P
MIPEPP3	4	1	C1QTNF9B-AS1	ZNF44	1	–	CTD-2666L21.1
				ZNF584	1	–	CTD-2619J13.17

<sup>a</sup> denotes number of entries in PubMed.

of cancer. Interestingly, we found that 65 of these parent genes had been previously described in cancer, and of those, 33 had been described in lung cancer (Table 2). Of the 34 significantly differentially expressed  $\Psi$ -lnc-parent gene pairs, 25 parent genes were described in cancer. This includes lung cancer associated genes like CS, which affects tumor drug response, as well as RCN1, which is associated with poor prognosis and tumor progression in lung cancer (Chen et al., 2014, 2018).

As the vast majority of the deregulated  $\Psi$ -lncs that were correlated with their parent gene had positive associations, we were interested whether this was a global phenomenon or exclusive to deregulated genes. We performed a Spearman's correlation analysis on every  $\Psi$ -lnc-parent-gene pair with expression data in our dataset irrespective of deregulation status ( $n = 390$  gene pairs). We plotted the distribution of Spearman's rho ( $\rho$ ) values for the  $\Psi$ -lnc-parent-gene pairs

and compared them to the rho values for  $\Psi$ -lncs paired to randomly selected genes. We found that the  $\Psi$ -lnc-parent-gene pairs have significantly more positive relationships than the random gene pairs in both the BCCA (Mann Whitney  $U$ -test,  $p < 0.0001$ ) and TCGA datasets (Mann Whitney  $U$ -test,  $p < 0.0001$ ; Figure 4A). Studies have shown that lncRNAs transcribed from opposite strands can have different regulatory effects on target genes (Johnsson et al., 2013; Balbin et al., 2015). To determine if transcriptional orientation has an effect on  $\Psi$ -lnc-parent relationships we compared the Spearman's rho values of sense  $\Psi$ -lnc-parent pairs to antisense  $\Psi$ -lnc-parent pairs. In both datasets we observed that the sense  $\Psi$ -lnc-parent pairs to have significantly more positive relationships than the antisense  $\Psi$ -lnc-parent pairs (Mann Whitney  $U$ -test, TCGA set ( $p < 0.0001$ ), and BCCA set ( $p < 0.0025$ ) (Figure 4B and Supplementary Table S6). Strongly positively



correlated  $\Psi$ -lnc-parent pairs include *TPT1-AS1/RCN1* and *LINC00887/CS* (Figure 4C).

### $\Psi$ -lncs and Their Parent Genes Are Associated With Patient Survival

If the aberrant expression of  $\Psi$ -lncs is biologically relevant, it follows that they may be relevant in tumor aggressiveness, stage, and patient survival. We performed a two-group analysis using a Mann Whitney  $U$ -test between Stage I tumors and Stage II-IV tumors, as the majority of our tumors fell into these categories (Table 1). Of the deregulated  $\Psi$ -lncs we found *CTC-250I14.3* to be associated with Stage 1 disease and downregulated in both the BCCA and TCGA LUAD cohorts (Supplementary Figure S1). Our discovery datasets were limited in sample size for survival association analysis; therefore, we examined a third cohort of 719 LUAD from the KM Plotter database (Kaplan–Meier Plotter)<sup>1</sup> (Table 1). This cohort was limited to genes with probe coverage on microarray platforms. A total of 19 of the deregulated  $\Psi$ -lncs were represented on this platform, yet, the majority of these  $\Psi$ -lncs (16 of 19) were significantly associated with poor overall survival (log-rank  $p < 0.05$ , Table 3).

While we were not able to investigate the survival associations of all deregulated  $\Psi$ -lncs as many were not covered by the microarray platforms, the majority (72 out of 103) of

their parent genes were represented. We discovered that 67 of these parent genes were associated with patient survival. Twenty-eight of these survival associated parent genes were also significantly associated with the expression of their paired deregulated  $\Psi$ -lnc. Furthermore, we found 11 pairs where both  $\Psi$ -lnc and parent gene are associated with patient survival. For example *RP11-1007O24.3*, a  $\Psi$ -lnc downregulated in tumors, is positively associated with expression of survival-associated parent gene *ARIH1* in both the BCCA and TCGA datasets (Table 3 and Figure 5A). Further examples also include  $\Psi$ -lnc-parent pairs such as *ZBED3-AS1* and *HMGB1*, which are positively correlated at the expression level, and both significantly associated with survival (Figure 5B). We also observe  $\Psi$ -lnc-parent pairs that are associated with survival, but do not share an expression relationship such as *LINC00667* and parent gene *TACC3* (Figure 5C). Collectively, our discovery of the broad deregulation of  $\Psi$ -lncs, many of which are survival-associated and associated with parent gene expression, may indicate that  $\Psi$ -lncs impact LUAD biology through *trans* regulation of their cancer-associated parent genes.

### DISCUSSION

Here we expand upon the work done by Milligan et al. (2016), completing the first large-scale analysis of lncRNA expression from pseudogene loci in LUAD and paired non-malignant lung

<sup>1</sup><http://kmplot.com/analysis/>

**TABLE 3** | Associations between  $\Psi$ -lncRNA, parent gene expression, and patient outcome.

$\Psi$ -lncRNA	Survival $p$ (KmPlotter)	Parent Gene	Survival $p$ (KmPlotter) <sup>a</sup>	Association with Parent ( $\rho$ , BCCA)	Association with Parent ( $\rho$ , TCGA)
AC008132.13	–	BCR	0.00043 <sup>a</sup>	0.0173	–
AGAP11	0.0099 <sup>b</sup>	AGAP1	5.80E-10 <sup>b</sup>	0.0005	–
AGAP11	–	BMS1	0.00018 <sup>a</sup>	0.0121	–
AP000525.9	–	CECR7	3.20E-06 <sup>a</sup>	–	0.0371
CTA-14H9.5	–	LINC00657	6.00E-06 <sup>b</sup>	–	–
CTD-2314B22.3	–	CECR7	3.20E-06 <sup>a</sup>	–	0.0402
CTD-2619J13.17	–	ZNF584	1.70E-08 <sup>a</sup>	–	0.0205
CTD-2666L21.1	–	ZNF44	2.00E-15 <sup>b</sup>	–	0.0013
CTD-2666L21.1	–	ZNF14	1.90E-12 <sup>a</sup>	–	–
DGCR5	4.00E-04 <sup>a</sup>	CELSR1	2.70E-02 <sup>a</sup>	–	–
FAM66C	3.00E-04 <sup>b</sup>	DEFB130	–	–	–
FAM95B1	–	CYP4F3	2.00E-01 <sup>b</sup>	–	–
HCG11	1.40E-08 <sup>b</sup>	LINC00657	6.00E-06 <sup>b</sup>	–	0.0046
LINC00342	1.80E-06 <sup>b</sup>	NEBL	2.30E-05 <sup>b</sup>	–	–
LINC00466	4.60E-06 <sup>a</sup>	RPSAP58	–	–	–
LINC00639	0.0087 <sup>b</sup>	ZFP41	0.014 <sup>b</sup>	–	–
LINC00667	4.50E-08 <sup>b</sup>	TACC3	6.20E-09 <sup>a</sup>	–	–
LINC00883	2.90E-03 <sup>b</sup>	CS	0.00041 <sup>a</sup>	0.0449	–
LINC00957	0.042 <sup>b</sup>	RASA4B	0.035 <sup>a</sup>	–	–
LINC00982	1.30E-06 <sup>b</sup>	<i>n/a</i> <sup>c</sup>	–	–	–
LINC00987	–	PZP	0.02 <sup>a</sup>	–	<0.0001
LY86-AS1	0.00025 <sup>b</sup>	SNAPC5	5e-0.6 <sup>b</sup>	–	–
MMP24-AS1	0.023 <sup>a</sup>	MT1	1.20E-12 <sup>a</sup>	–	–
PCBP1-AS1	–	PCBP2	4.30E-13 <sup>b</sup>	–	–
RAMP2-AS1	0.014 <sup>b</sup>	HMG2P46	–	–	–
RP11-1007O24.3	–	ARIH1	0.00093 <sup>b</sup>	<0.0001	0.0030
RP11-1186N24.5	–	PKD1	0.0013 <sup>b</sup>	–	<0.0001
RP11-182J1.1	–	EGLN1	1.20E-08 <sup>b</sup>	0.0011	–
RP11-344E13.3	–	RNASEH1	1.00E-07 <sup>b</sup>	0.0007	–
RP11-349A22.5	–	HMG1	7.50E-10	–	–
RP11-34P13.7	–	CIC	1.20E-08 <sup>a</sup>	–	0.0205
RP11-390F4.3	–	CDC42	1.30E-05	0.0449	–
RP11-439L8.3	–	ADC	3.30E-09	0.0205	–
RP11-446H18.5	–	CS	0.00041 <sup>a</sup>	–	–
RP11-93K22.13	–	FAM86B3P	6.30E-05 <sup>b</sup>	–	0.0129
TOPORS-AS1	–	DFFB	9.10E-08 <sup>b</sup>	–	0.0033
TPT1-AS1	0.00019 <sup>b</sup>	RCN1	0.016 <sup>a</sup>	0.0242	–
ZBED3-AS1	1.10E-10 <sup>b</sup>	HMG1	7.50E-10 <sup>b</sup>	0.0100	–
ZNF833P	–	ZNF491	0.016 <sup>b</sup>	–	–
ZNF833P	–	ZNF14	1.90E-12 <sup>a</sup>	–	0.0079

<sup>a</sup> Denotes poor overall survival associated with high gene expression.

<sup>b</sup> Denotes poor overall survival associated with low gene expression.

<sup>c</sup> No information for the parent gene.

tissue. We discovered a broadly positive association between  $\Psi$ -lncs and parent-gene expression, suggestive of an alternative mechanism of cancer gene regulation. While there have been singular examples of deregulated pseudogene-derived lncRNAs in cancer, we show that this phenomenon is widespread in LUAD. In addition to being correlated with  $\Psi$ -lnc expression, we find that many of the parent genes of these deregulated lncRNAs are annotated cancer genes and are significantly associated with patient survival, highlighting how these previously unappreciated non-coding genes may affect LUAD biology.

While the identification of lncRNAs associated with cancer phenotypes is increasing, a great challenge in the field remains the accurate downstream prediction of lncRNA function. Unlike protein-coding genes or small non-coding RNAs, features like complex folding patterns and unknown binding motifs have

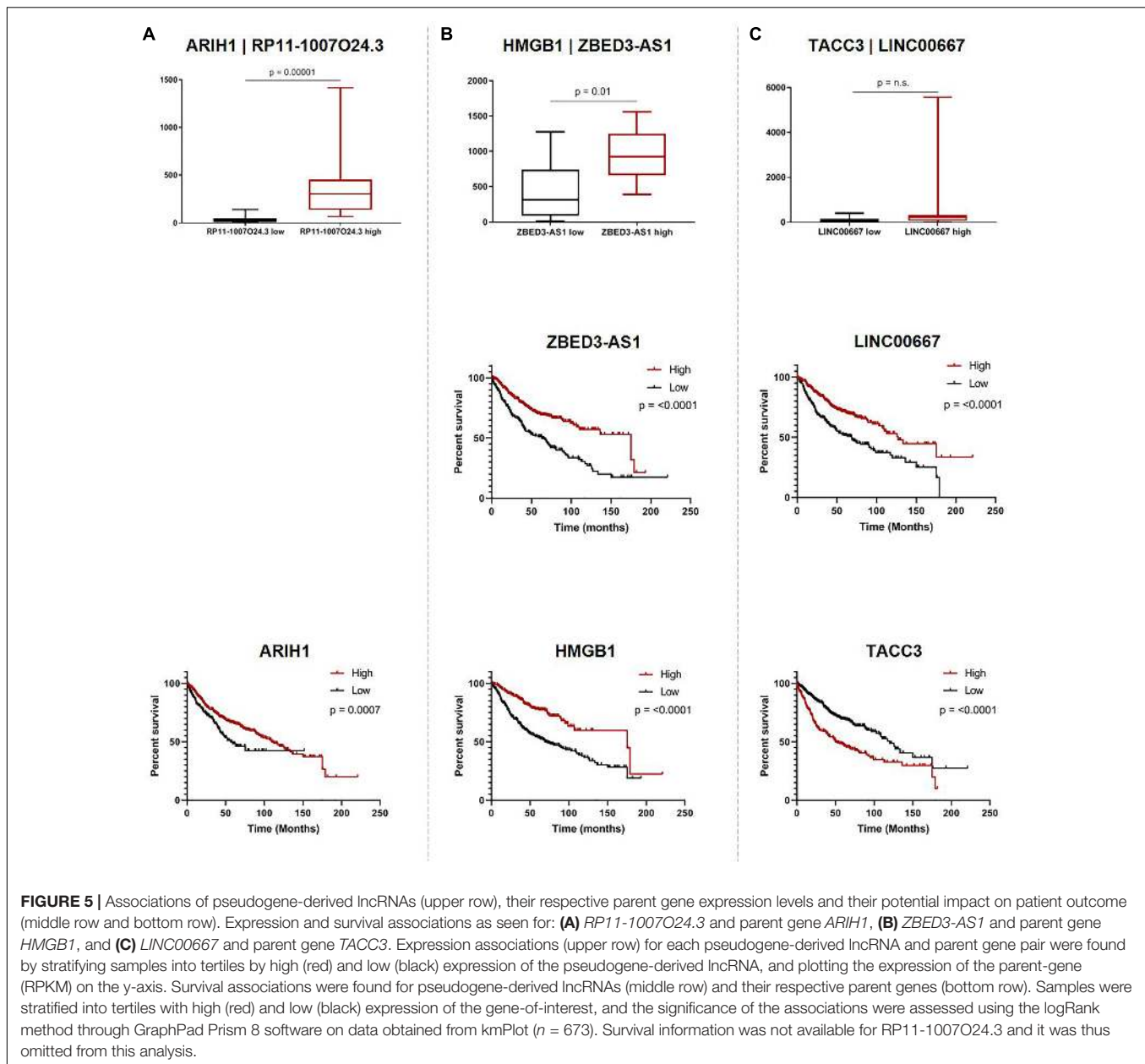
contributed to the challenging functional characterization of lncRNAs. We utilized the sequence similarity found between lncRNAs expressed from pseudogene loci, and their parent genes to predict the function of this subset of lncRNAs in LUAD. We identified a set of 104  $\Psi$ -lncs deregulated in LUAD in two independent datasets. This greatly increases the number of deregulated lncRNAs known to be expressed from pseudogene loci in LUAD. Interestingly, the majority of these  $\Psi$ -lncs were under-expressed in tumors compared to non-malignant tissues, suggesting that they may have tumor-suppressive roles, and that their downregulation is advantageous to LUAD tumorigenesis.

Under-expression was not significantly associated with regions of recurrent copy number deletion in LUAD, although a subset of deregulated  $\Psi$ -lnc loci were localized to these regions (Figure 3). These observations suggest that they may be regulated by alternative molecular mechanisms, including broad chromosomal aberrations that affect whole chromosome arms, or epigenetic mechanisms. For example, endogenous retroviruses and repetitive elements often become aberrantly expressed in cancer due to deregulated methylation patterns (Kassiotis, 2014). We did not observe enrichment of  $\Psi$ -lncs or their parent genes on any chromosomes, despite the fact that pseudogenes are known to be overabundant on the human X chromosome (Drouin, 2006).

The direction of transcription often affects lncRNA function (Balbin et al., 2015). For example, the *PTEN* pseudogene (*PTENP1*) expressed in the sense direction can function as a decoy for inhibitory miRNAs that would otherwise cause translational inhibition of the *PTEN* parent mRNA. Conversely, when the antisense lncRNA is expressed from the *PTENP1* locus, the transcript is able to localize to the *PTEN* parent locus and recruit chromatin-remodeling machinery, which leads to the silencing of *PTEN* transcription (Poliseno et al., 2010; Johnsson et al., 2013). Both mechanisms have been coopted by cancer cells for their respective tumor suppressive and oncogenic roles (Hu et al., 2018). We found sense  $\Psi$ -lnc-parent pairs, (which account for 208 out of 391  $\Psi$ -lncs examined) to be more positively correlated than antisense  $\Psi$ -lnc-parent pairs in both cohorts (Figure 4B). This may imply that  $\Psi$ -lncs are more likely to regulate their parent gene in a positive manner which may occur through mechanisms such as miRNA sponging or transcript stabilization when transcribed in the sense direction (Glenfield and McLysaght, 2018; Hu et al., 2018). The distribution of Spearman's  $\rho$ -values for antisense-parent gene pairs suggests a more even split between positive and negative regulation. A limitation of this study is that we cannot discount the possibility that sequencing reads for sense overlapping  $\Psi$ -lncs that have sequence homology with their parent gene are being mapped to the parent gene instead of the  $\Psi$ -lnc. This potential issue warrants further investigation considering both the large number of annotated pseudogenes in the genome ( $n = 13,000$ ) and the possibility of false interpretation of sequencing data for both protein-coding and non-coding genes.<sup>2</sup> While these alignment errors could affect sense  $\Psi$ -lnc-parent gene pairs,

<sup>2</sup><https://www.genenames.org/cgi-bin/statistics>





antisense  $\Psi$ -lnc-parent gene pairs are not subjected to this same technical artifact. Recently, RNA-sequencing analysis strategies have emerged that begin to address this issue, and long read RNA-sequencing could be used to reduce errors in sequence alignment.

When looking at the parent genes of these deregulated  $\Psi$ -lncs we were interested to find that many had previously described roles in cancer (Table 2). This includes *EGLN1*, a well described cancer gene involved in regulation of tumor hypoxia, and *CDC42*, an oncogene involved in cell cycle control (Chan and Giaccia, 2010; Maldonado and Dharmawardhane, 2018). Many of these deregulated  $\Psi$ -lncs were also associated with clinical parameters such as patient survival and patient stage, in addition to the correlated expression between  $\Psi$ -lncs and their parent genes (Table 3). For example, *ZBED3-AS1*

and *HMGB1* were positively correlated at the expression level, and low expression of both genes was associated with poor patient survival (Figure 5B). We also observed  $\Psi$ -lnc-parent pairs where both genes are associated with survival, but do not share an expression relationship. *LINC00667* and parent gene *TACC3*, for example, are both survival-associated, but not correlated at the expression level. *TACC3* is a component of the *TACC3*/ch-TOG/clathrin protein complex, and roles in complex assembly have been previously observed for lncRNAs (Munschauer et al., 2018; Figure 5C). Thus, it is possible that *LINC00667* is involved in a form of regulation that would not affect transcript levels, including protein-complex assembly. While we were unable to assess survival associations for many of the deregulated  $\Psi$ -lncs, they may still impact patient survival

through regulation of their parent genes. We found 28 out of 34 expression-associated  $\Psi$ -lnc-parent gene pairs to have parent genes associated with patient survival. *RP11-1007O24.3*, for example, was positively correlated with survival-associated parent gene *ARIH1* expression in both cohorts. *ARIH1* has been previously described to be a mediator of DNA-damage response and mitophagy in cancer cells (Figure 5A; Perdomo et al., 1988; von Stechow et al., 2015). The potential regulatory impact of  $\Psi$ -lnc on their clinically relevant parent genes is considerable and may represent a novel avenue for targeted therapies. While this study focused on the broad effects of  $\Psi$ -lnc deregulation, future studies utilizing *in vitro* and *in vivo* experiments will be necessary to determine the specific mechanisms of parent gene regulation.

As  $\Psi$ -lnc may represent an unexplored area of cancer-associated parent gene regulation, their therapeutic relevance should be further explored. lncRNAs make ideal targets for therapies that target RNA products such as Antisense Oligonucleotide (ASO) therapies, since RNA is their final functional state, rather than the intermediate product for protein-coding genes (Croke et al., 2018). In addition, ASOs are easier and less costly to develop than small molecule inhibitors, and are in development as aerosol sprays that may be ideal for lung cancer treatment (Kjems and Howard, 2012; Li and Chen, 2013). However, as ASOs target through complementary sequence pairing, they would have to be designed in such a way as to not interfere with the parent gene, especially in the case of  $\Psi$ -lnc expressed from the sense strand.

This strategy of identifying lncRNAs aberrantly expressed from pseudogene loci may be useful when applied to other cancer types. Indeed, we see that several of our deregulated  $\Psi$ -lnc have been described in other tumor types, such as *TPT1-ASI* in cervical cancer, and *HGC11* in both prostate cancer and hepatocellular carcinoma (Zhang et al., 2016; Xu et al., 2017; Jiang et al., 2018). Additionally, as lncRNA expression is highly tissue specific, the application of this approach to other cancer types may yield novel disease-specific  $\Psi$ -lnc-parent pairs, highlighting the clinical utility of examining these previously underappreciated transcripts. Overall,  $\Psi$ -lnc-cancer-parent-gene axes represent alternative mechanisms of cancer gene regulation, and their identification is a critical step toward the functional characterization of lncRNAs.

## CONCLUSION

There is a growing need to functionally characterize lncRNAs. Pseudogene-derived lncRNAs have been shown to be involved in cancer and regulate the expression of their parent genes. We show here how pervasive this gene regulatory mechanism is in LUAD samples. We identify a large set of deregulated  $\Psi$ -lncs, with aberrant expression observed in RNA-sequencing data from two LUAD cohorts of paired tumor and non-malignant lung tissue samples. We show that these deregulated  $\Psi$ -lncs have clinical value and that the parent genes, many of which are correlated with  $\Psi$ -lnc expression, have been implicated in cancer phenotypes and are associated with clinical outcome. Together,

our results highlight the important roles of the non-coding transcriptome in cancer cellular biology.

## MATERIALS AND METHODS

### Next Generation Sequencing of Lung Adenocarcinoma Patient Samples Discovery Cohort

We performed Next Generation Illumina HiSeq RNA sequencing on a set of 36 micro-dissected LUAD tumors and matched adjacent non-malignant tissue ( $n = 72$ ). Our British Columbia Cancer Agency cohort (BCCA) was composed of fresh-frozen LUAD tumors and matched non-malignant lung parenchymal tissue collected from 36 patients at the Vancouver General Hospital with approval from the University of British Columbia-BCCA Research Ethics Board. Consent obtained from the tissue donors of this study was both informed and written. To avoid field effects non-malignant samples were collected from areas  $> 2$  cm away from the tumor. In order to reduce contaminating sequences derived from alternative cell types, tissue microdissection was guided by a pathologist. Samples used in this study contained  $> 80\%$  tumor cell or  $> 80\%$  non-malignant cell content. Total RNA was extracted using Trizol reagent and standard procedures.

### Processing of RNA-Sequencing Data

Total RNA was used for library construction at the Genome Sciences Center (GSC, Vancouver, BC, Canada). Briefly, samples were first analyzed using Agilent Bioanalyzer RNA nanochip, and samples that passed quality check were arrayed into a 96-well plate. PolyA<sup>+</sup> RNA was purified using the 96-well MultiMACS mRNA isolation kit on the MultiMACS 96 separator (Miltenyi Biotec, Germany) from 2  $\mu$ g total RNA with on-column DNase I-treatment as per the manufacturer's instructions. Double-stranded cDNA was synthesized from the purified polyA<sup>+</sup>-RNA using the Superscript Double-Stranded cDNA Synthesis kit (Life Technologies, United States) and random hexamer primers at a concentration of 5  $\mu$ M. The paired-end sequencing library was prepared following the GSC paired-end library preparation protocol, which is strand specific. Sequencing was performed using the Illumina HiSeq 2000 platform. Raw sequencing reads were subject to a quality control process. Reads with a length  $< 50$  nt (under two thirds of maximum read length of 75 nt) and quality level (Phred)  $< 20$  were discarded. High quality reads (.fastq files) were aligned to the NCBI GRCh37 reference human genome build using the STAR aligner (v 2.4.1d) under default parameters (Dobin et al., 2013). Aligned reads (.bam files) were quantified using Ensembl Transcripts (Release 75) reference annotations (Flicek et al., 2014). Raw RNA sequencing reads from each patient (tumor and corresponding non-malignant tissue) were deposited at BioProject.<sup>3</sup> Quantification was performed using the Partek Flow platform as reads per kilobase per million (RPKM). RNA sequencing (.bam files) and clinical

<sup>3</sup><http://www.ncbi.nlm.nih.gov/bioproject/516232>

data for a secondary set of LUAD tumors and matched non-malignant tissue ( $n = 108$ ) were downloaded from The Cancer Genome Atlas (TCGA) Data Portal for validation purposes<sup>4</sup> (Table 1). Expression profiles from TCGA were processed as described above.

## Identification of lncRNAs Expressed From Pseudogene Loci and Corresponding Parent Genes

### $\Psi$ -lnc Annotation

Milligan et al. (2016) recently published a global atlas of lncRNAs that have exonic overlap with positionally non-redundant (unique) pseudogenes from 3 major pseudogene databases. Using this resource we obtained a list of lncRNAs overlapping pseudogene loci (Supplementary Table S1) that we used as a foundation for our expression analysis. As the degree of sequence overlap required for a pseudogene-derived lncRNA to regulate its parent gene is unknown, we did not restrict our analysis to full-length, expressed pseudogenes, and included lncRNAs with any exonic overlap, including sense and antisense transcripts in order to annotate the most comprehensive list of  $\Psi$ -lncs (lncRNAs overlapping pseudogene loci).

### Parent Gene Annotation

The parent gene information was also extracted for all pseudogenes overlapping our list of  $\Psi$ -lncs that have parent-gene annotations in the YaleHuman60 and Retroali5 databases (Supplementary Table S7). Manual literature search was performed for parent genes of deregulated  $\Psi$ -lncs that were not contained in these databases.

## Statistical Analysis

### Identification of Significantly Deregulated $\Psi$ -lncs in Paired LUAD and Non-Malignant Lung Tissue

Gene expression for protein-coding and non-coding genes was compared between tumors and non-malignant tissue and significantly deregulated genes were identified using a Wilcoxon signed-rank test ( $p < 0.05$ ) and subjected to a Benjamini-Hochberg (BH) FDR correction.  $\Psi$ -lncs as identified previously were extracted, and those that were significantly deregulated between tumors and non-malignant tissue, in both our discovery (BCCA) and validation (TCGA) were selected for further analysis (Figure 1B).

### $\Psi$ -lncs and Parent Gene Expression

Tumors were sorted by  $\Psi$ -lnc expression for each  $\Psi$ -lnc-parent gene pair, and grouped into top and bottom  $\Psi$ -lnc expressing tertiles. Parent gene expression was then compared between the two groups using the Mann Whitney  $U$ -test ( $p$ -value  $\leq 0.05$ ). We performed a global expression analysis to determine whether  $\Psi$ -lnc-parent pairs were more positive or negatively correlated than random chance. For all  $\Psi$ -lncs with expression data in both datasets, Spearman's correlation rho

values were calculated for  $\Psi$ -lnc-parent gene pairs ( $n = 390$ ) and compared to  $\Psi$ -lnc-random gene pairs. Random genes in our expression matrices were selected to pair with each  $\Psi$ -lnc. Each gene was assigned a number and pairs were chosen by using a random number generator.<sup>5</sup> Spearman's rho values were then plotted, smoothed, and compared (Mann Whitney  $U$ -test,  $p$ -value  $\leq 0.05$ ). Rho value distribution was also compared between sense lnc-parent gene pairs ( $n = 208$ ) and antisense lnc-parent pairs ( $n = 182$ ).

### Literature Searches

To determine if each gene of interest ( $\Psi$ -lnc or parent gene) had been previously described in the context of tumors we searched Pubmed using the terms "gene + cancer" or "gene + lung cancer."

### Hierarchical Clustering and Data Visualization

Unsupervised hierarchical clustering was performed in order to visualize and examine the expression of the  $\Psi$ -lncs in individual samples (Figure 2A). Average Linkage was used as a cluster distance metric, while Pearson Correlation was used as a point distance metric. To visualize the expression patterns of the most highly expressed  $\Psi$ -lncs, those with an average expression value of  $\geq 10$  RPKM in either tumor or non-malignant samples were included in the analysis.

### Distribution of Deregulated $\Psi$ -lncs Across Genome

Locations of  $\Psi$ -lncs and parent genes were compared to identify their genomic position (Supplementary Table S4). Circular plot visualization was performed using the R-package Circlize (Figure 3; Gu et al., 2014). LUAD-specific regions of significant recurrent somatic copy number alterations had been previously identified by TCGA, and were used in this study to determine if the deregulated  $\Psi$ -lncs overlapped with frequently altered regions (Zack et al., 2013). All genomic coordinates correspond to the NCBI GRCh37 reference human genome build.

## Clinical Features

### Survival Analysis

A large public clinical database (Kaplan-Meier Plotter)<sup>6</sup> comprised of 719 LUAD samples was used to determine the association between both protein-coding and non-coding gene expression with patient outcomes. Similar to the BCCA and TCGA cohorts, the patient samples in this 3rd dataset were mostly comprised of Stage I and Stage II tumors (Table 1). Of the 104 deregulated  $\Psi$ -lncs, 19 were represented in this database, while 70% of parent genes (72 out of 103) were present. Default settings were used and a log-rank (Mantel-Cox) test was applied to compare survival between groups of tumors with high and low expression of each gene tested, where  $p < 0.05$  was considered statistically significant. The optimal expression cut-off was selected for each gene.

<sup>4</sup><https://portal.gdc.cancer.gov/>

<sup>5</sup><https://www.random.org/>

<sup>6</sup><http://kmplot.com/analysis/>

## Association With Tumor Stage

The majority of tumors for BCCA and TCGA fell into the categories of Stage I and Stage II (Table 1). We compared expression of deregulated  $\Psi$ -lncs between Stage I tumors and tumors classified as Stage II and above using a Mann Whitney  $U$ -test ( $p$ -value  $\leq 0.05$ ).

## AUTHOR CONTRIBUTIONS

GS was responsible for the project design. GS, KE, AS, VM, BM, MP, EM, and WL contributed to data acquisition, data analysis, interpretation of results, and manuscript preparation. WL was the principal investigator of this project. All authors have read, edited, and approved the final manuscript.

## FUNDING

This study was supported by grants from the Canadian Institutes for Health Research (CIHR FDN-143345) and Scholarships from CIHR. EM is a Vanier Canada Graduate Scholar, and is additionally supported by UBC Faculty of Medicine.

## REFERENCES

- Arab, K., Park, Y. J., Lindroth, A. M., Schafer, A., Oakes, C., Weichenhan, D., et al. (2014). Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A. *Mol. Cell* 55, 604–614. doi: 10.1016/j.molcel.2014.06.031
- Balbin, O. A., Malik, R., Dhanasekaran, S. M., Prensner, J. R., Cao, X., Wu, Y. M., et al. (2015). The landscape of antisense gene expression in human cancers. *Genome Res.* 25, 1068–1079. doi: 10.1101/gr.180596.114
- Bhan, A., Soleimani, M., and Mandal, S. S. (2017). Long noncoding RNA and cancer: a new paradigm. *Cancer Res.* 77, 3965–3981. doi: 10.1158/0008-5472.CAN-16-2634
- Blume, C. J., Hotz-Wagenblatt, A., Hullein, J., Sellner, L., Jethwa, A., Stolz, T., et al. (2015). p53-dependent non-coding RNA networks in chronic lymphocytic leukemia. *Leukemia* 29, 2015–2023. doi: 10.1038/leu.2015.119
- Chan, B. A., and Hughes, B. G. (2015). Targeted therapy for non-small cell lung cancer: current standards and the promise of the future. *Transl. Lung. Cancer Res.* 4, 36–54. doi: 10.3978/j.issn.2218-6751.2014.05.01
- Chan, D. A., and Giaccia, A. J. (2010). PHD2 in tumour angiogenesis. *Br. J. Cancer* 103, 1–5. doi: 10.1038/sj.bjc.6605682
- Chen, E. G., Zhang, J. S., Xu, S., Zhu, X. J., and Hu, H. H. (2017). Long non-coding RNA DGCR5 is involved in the regulation of proliferation, migration and invasion of lung cancer by targeting miR-1180. *Am. J. Cancer Res.* 7, 1463–1475.
- Chen, L., Liu, T., Zhou, J., Wang, Y., Wang, X., Di, W., et al. (2014). Citrate synthase expression affects tumor phenotype and drug resistance in human ovarian carcinoma. *PLoS One* 9:e115708. doi: 10.1371/journal.pone.0115708
- Chen, X., Shao, W., Huang, H., Feng, X., Yao, S., and Ke, H. (2018). Overexpression of RCN1 correlates with poor prognosis and progression in non-small cell lung cancer. *Hum. Pathol.* 83, 140–148. doi: 10.1016/j.humpath.2018.08.014
- Crooke, S. T., Witztum, J. L., Bennett, C. F., and Baker, B. F. (2018). RNA-targeted therapeutics. *Cell Metab.* 27, 714–739. doi: 10.1016/j.cmet.2018.03.004
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Dong, H. X., Wang, R., Jin, X. Y., Zeng, J., and Pan, J. (2018). LncRNA DGCR5 promotes lung adenocarcinoma (LUAD) progression via inhibiting hsa-mir-22-3p. *J. Cell Physiol.* 233, 4126–4136. doi: 10.1002/jcp.26215

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00138/full#supplementary-material>

**FIGURE S1** | Expression of lncRNA CTC-25014.3 is significantly associated with Stage I disease in both the BCCA and TCGA LUAD cohorts (Mann Whitney  $U$ -test).

**TABLE S1** | Detailed information of all pseudogene-derived lncRNAs.

**TABLE S2** | Statistical analysis of the deregulated lncRNAs in the BCCA and TCGA cohorts (BHC-corrected  $p$ -values).

**TABLE S3** | Number of PUBMED entries for lncRNA cancer-association.

**TABLE S4** | Genomic locations of significant lncRNAs and their respective parent genes.

**TABLE S5** | Expression correlations between the lncRNAs and their respective parent genes in the BCCA and TCGA cohorts.

**TABLE S6** | Spearman's correlation coefficients for global expression analysis of  $\Psi$ -lnc-parent pairs.

**TABLE S7** | Parent gene information for pseudogenes contained in Retrolali5 and Yale60 databases.

- Drouin, G. (2006). Processed pseudogenes are more abundant in human and mouse X chromosomes than in autosomes. *Mol. Biol. Evol.* 23, 1652–1655. doi: 10.1093/molbev/msl048
- Feng, F., Qiu, B., Zang, R., Song, P., and Gao, S. (2017). Pseudogene PHBP1 promotes esophageal squamous cell carcinoma proliferation by increasing its cognate gene PHB expression. *Oncotarget* 8, 29091–29100. doi: 10.18632/oncotarget.16196
- Flicek, P., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., et al. (2014). Ensembl 2014. *Nucleic Acids Res.* 42, D749–D755. doi: 10.1093/nar/gkt1196
- Glenfield, C., and McLysaght, A. (2018). Pseudogenes provide evolutionary evidence for the competitive endogenous RNA hypothesis. *Mol. Biol. Evol.* 35, 2886–2899. doi: 10.1093/molbev/msy183
- Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize Implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812. doi: 10.1093/bioinformatics/btu393
- Gutschner, T., Hammerle, M., Eissmann, M., Hsu, J., Kim, Y., Hung, G., et al. (2013). The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.* 73, 1180–1189. doi: 10.1158/0008-5472.CAN-12-2850
- Hirose, T., Virnicchi, G., Tanigawa, A., Naganuma, T., Li, R., Kimura, H., et al. (2014). NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol. Biol. Cell* 25, 169–183. doi: 10.1091/mbc.E13-09-0558
- Hu, X., Yang, L., and Mo, Y. Y. (2018). Role of pseudogenes in tumorigenesis. *Cancers* 10:256. doi: 10.3390/cancers10080256
- Huang, J. L., Cao, S. W., Ou, Q. S., Yang, B., Zheng, S. H., Tang, J., et al. (2018). The long non-coding RNA PTTG3P promotes cell growth and metastasis via up-regulating PTTG1 and activating PI3K/AKT signaling in hepatocellular carcinoma. *Mol. Cancer* 17:93. doi: 10.1186/s12943-018-0841-x
- Jiang, H., Huang, G., Zhao, N., Zhang, T., Jiang, M., He, Y., et al. (2018). Long non-coding RNA TPT1-AS1 promotes cell growth and metastasis in cervical cancer via acting AS a sponge for miR-324-5p. *J. Exp. Clin. Cancer Res.* 37:169. doi: 10.1186/s13046-018-0846-8
- Johnsson, P., Ackley, A., Vidarsdottir, L., Lui, W. O., Corcoran, M., Grander, D., et al. (2013). A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. *Nat. Struct. Mol. Biol.* 20, 440–446. doi: 10.1038/nsmb.2516
- Kassiotis, G. (2014). Endogenous retroviruses and the development of cancer. *J. Immunol.* 192, 1343–1349. doi: 10.4049/jimmunol.1302972

- Khachane, A. N., and Harrison, P. M. (2009). Assessing the genomic evidence for conserved transcribed pseudogenes under selection. *BMC Genomics* 10:435. doi: 10.1186/1471-2164-10-435
- Kjems, J., and Howard, K. A. (2012). Oligonucleotide delivery to the lung: waiting to inhale. *Mol. Ther. Nucleic Acids* 1:e1. doi: 10.1038/mtna.2011.1
- Kung, J. T., Colognori, D., and Lee, J. T. (2013). Long noncoding RNAs: past, present, and future. *Genetics* 193, 651–669. doi: 10.1534/genetics.112.146704
- Lee, J. T., Davidow, L. S., and Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat. Genet.* 21, 400–404. doi: 10.1038/7734
- Li, C. H., and Chen, Y. (2013). Targeting long non-coding RNAs in cancers: progress and prospects. *Int. J. Biochem. Cell Biol.* 45, 1895–1910. doi: 10.1016/j.biocel.2013.05.030
- Luo, J., Zhu, H., Jiang, H., Cui, Y., Wang, M., Ni, X., et al. (2018). The effects of aberrant expression of lncRNA DGCR5/miR-873-5p/TUSC3 in lung cancer cell progression. *Cancer Med.* doi: 10.1002/cam4.1566 [Epub ahead of print].
- Maldonado, M. D. M., and Dharmawardhane, S. (2018). Targeting Rac and Cdc42 GTPases in cancer. *Cancer Res.* 78, 3101–3111. doi: 10.1158/0008-5472.CAN-18-0619
- Milligan, M. J., Harvey, E., Yu, A., Morgan, A. L., Smith, D. L., Zhang, E., et al. (2016). Global intersection of long non-coding RNAs with processed and unprocessed pseudogenes in the human genome. *Front. Genet.* 7:26. doi: 10.3389/fgene.2016.00026
- Munschauer, M., Nguyen, C. T., Sirokman, K., Hartigan, C. R., Hogstrom, L., Engreitz, J. M., et al. (2018). The NORAD lncRNA assembles a topoisomerase complex critical for genome stability. *Nature* 561, 132–136. doi: 10.1038/s41586-018-0453-z
- Peng, W. X., Koirala, P., and Mo, Y. Y. (2017). lncRNA-mediated regulation of cell signaling in cancer. *Oncogene* 36, 5661–5667. doi: 10.1038/onc.2017.184
- Perdomo, R., Alvarez, C., Geninazzi, H., Ferreira, C., Monti, J., Parada, R., et al. (1988). Early diagnosis of hydatidosis by ultrasonography. *Lancet* 1:244. doi: 10.1016/S0140-6736(88)91097-5
- Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W. J., and Pandolfi, P. P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 465, 1033–1038. doi: 10.1038/nature09144
- Simon, M. D., Pinter, S. F., Fang, R., Sarma, K., Rutenberg-Schoenberg, M., Bowman, S. K., et al. (2013). High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* 504, 465–469. doi: 10.1038/nature12719
- Sun, M., Nie, F. Q., Zang, C., Wang, Y., Hou, J., Wei, C., et al. (2017). The pseudogene DUXAP8 promotes non-small-cell lung cancer cell proliferation and invasion by epigenetically silencing EGR1 and RHOB. *Mol. Ther.* 25, 739–751. doi: 10.1016/j.ymthe.2016.12.018
- Tano, K., Mizuno, R., Okada, T., Rakwal, R., Shibato, J., Masuo, Y., et al. (2010). MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes. *FEBS Lett.* 584, 4575–4580. doi: 10.1016/j.febslet.2010.10.008
- von Stechow, L., Typas, D., Carreras Puigvert, J., Oort, L., Siddappa, R., Pines, A., et al. (2015). The E3 ubiquitin ligase ARIH1 protects against genotoxic stress by initiating a 4EHP-mediated mRNA translation arrest. *Mol. Cell. Biol.* 35, 1254–1268. doi: 10.1128/MCB.01152-14
- Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472, 120–124. doi: 10.1038/nature09819
- Xu, Y., Zheng, Y., Liu, H., and Li, T. (2017). Modulation of IGF2BP1 by long non-coding RNA HCG11 suppresses apoptosis of hepatocellular carcinoma cells via MAPK signaling transduction. *Int. J. Oncol.* 51, 791–800. doi: 10.3892/ijo.2017.4066
- Yamazaki, T., Fujikawa, C., Kubota, A., Takahashi, A., and Hirose, T. (2018). CRISPRa-mediated NEAT1 lncRNA upregulation induces formation of intact paraspeckles. *Biochem. Biophys. Res. Commun.* 504, 218–224. doi: 10.1016/j.bbrc.2018.08.158
- Zack, T. I., Schumacher, S. E., Carter, S. L., Cherniack, A. D., Saksena, G., Tabak, B., et al. (2013). Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* 45, 1134–1140. doi: 10.1038/ng.2760
- Zhang, M., Weng, W., Zhang, Q., Wu, Y., Ni, S., Tan, C., et al. (2018). The lncRNA NEAT1 activates Wnt/beta-catenin signaling and promotes colorectal cancer progression via interacting with DDX5. *J. Hematol. Oncol.* 11:113. doi: 10.1186/s13045-018-0656-7
- Zhang, Y., Zhang, P., Wan, X., Su, X., Kong, Z., Zhai, Q., et al. (2016). Downregulation of long non-coding RNA HCG11 predicts a poor prognosis in prostate cancer. *Biomed. Pharmacother.* 83, 936–941. doi: 10.1016/j.biopha.2016.08.013

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Stewart, Enfield, Sage, Martinez, Minatel, Pewarchuk, Marshall and Lam. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.