

Deregulation of small non-coding RNAs at the *DLK1-DIO3* imprinted locus predicts lung cancer patient outcome

Katey S.S. Enfield¹, Victor D. Martinez¹, Erin A. Marshall¹, Greg L. Stewart¹, Sonia H.Y. Kung¹, Jhon R. Enterina¹, Wan L. Lam¹

¹Department of Integrative Oncology, British Columbia Cancer Research Centre, Vancouver, B.C. V5Z 1L3, Canada

Correspondence to: Katey S.S. Enfield, email: kenfield@bccrc.ca

Keywords: *DLK1-DIO3*, imprinted locus, lung cancer, microRNA, piwi-interacting RNA

Received: September 02, 2016

Accepted: October 27, 2016

Published: November 05, 2016

ABSTRACT

Deregulation of the imprinted *DLK1-DIO3* locus at chromosome 14q32.1-14q32.31 has been associated with developmental and respiratory disorders, including cancer. In lung cancer, deregulation of imprinting at *DLK1-DIO3* was recently described in smokers. Deregulated expression of a microRNA (miRNA) cluster mapping to this locus was also associated with patient outcome, suggesting the importance of this locus to lung cancer disease phenotypes. The *DLK1-DIO3* locus is complex, and encodes several protein-coding genes, in addition to long and short non-coding RNAs. While the role of miRNAs is established, the biological importance of another relevant class of small RNAs, PIWI-interacting RNAs (piRNAs), has not been investigated. When somatically expressed, piRNAs regulate gene transcription through DNA methylation. Interestingly, their expression patterns have been observed to be altered in cancer and correlated with patient outcome. Here, we characterize the somatic expression of piRNAs encoded at *DLK1-DIO3* in two independent cohorts of lung adenocarcinoma and lung squamous cell carcinoma and investigate their associations with patient outcome. We find that the expression of piRNAs encoded at *DLK1-DIO3* enhances the prognostic potential of small non-coding RNAs specific to this locus in predicting patient outcome, further emphasizing the importance of regulation at this locus in lung cancer.

INTRODUCTION

Genomic imprinting is the process by which the expression of an allele is silenced by methylation dependant on parental origin [1]. Aberrant methylation patterns at imprinted loci resulting in expression changes of encoded transcripts are common in the pathogenesis of many diseases, including cancer [2, 3]. In humans, anomalous imprinting at the *DLK1-DIO3* locus at 14q32.1-14q32.31 has been associated with respiratory insufficiency and reduced thorax development, amongst many other developmentally-related disorders [4].

The complexity of this locus is derived from the many protein-coding and non-coding RNAs it encodes. This locus encodes long non-coding RNAs (lncRNAs), and small non-coding RNAs (ncRNAs), including one of the largest microRNA (miRNA) clusters in the human genome (Figure 1A). Deregulation of small ncRNAs, mainly miRNAs, expressed from this locus has been associated with development and progression of different tumors,

including lung, in both humans and mice [5–7]. While individual genes expressed from this locus have been associated with lung cancer patient outcome, a signature of three miRNAs has been shown to better predict overall survival and recurrence-free survival [8]. This combined prediction signature suggests that the analysis of multiple genes encoded at *DLK1-DIO3* may be more biologically informative than the analysis of any single gene.

The role of other classes of small ncRNAs at this locus, such as PIWI-interacting RNAs (piRNAs), which act primarily as transcriptional regulators, has not yet been investigated in lung cancer (Figure 2). piRNAs have highly-conserved functions across species, including epigenetic silencing of transposable elements and regulation of imprinting in mice [9]. Although originally discovered in germ cells, recent evidence of their somatic expression in non-malignant human tissues and tumours suggests alternative functions and clinical importance [10–16].

In this study, we identify piRNAs expressed from *DLK1-DIO3* and determine whether their expression

patterns enhance the prognostic value of small ncRNAs encoded at this clinically important locus. We have analyzed piRNA and miRNA expression profiles from two independent cohorts of non-small cell lung cancer (NSCLC) and investigated their relationship with patient outcome.

RESULTS

The *DLK1-DIO3* locus encodes somatically expressed piRNAs

Deregulation of the *DLK1-DIO3* locus has been reported to be important to lung cancer biology, but the role of piRNAs derived from this locus has not yet been described. We analyzed expression data from two independent cohorts of lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), and non-malignant lung samples to identify somatically-expressed piRNAs encoded at this locus (Table 1). Of the 138 piRNAs encoded at the *DLK1-DIO3* locus, seven were expressed in LUAD, LUSC, and non-malignant lung samples in the discovery cohort (DQ596225, DQ596306, DQ596309,

DQ596311, DQ596354, DQ596390, DQ596863) (Figure 1B, Figure 3). Expression of all seven piRNAs was validated in the external cohort (Supplementary Figure S1). Interestingly, these somatically expressed piRNAs are encoded exclusively in the imprinted locus. In the discovery cohort of paired tumour and non-malignant lung tissues, four of seven somatically expressed piRNAs (DQ596225, DQ596306, DQ596309, DQ596354) were significantly overexpressed in LUAD and one piRNA (DQ596309) was overexpressed in LUSC (Figure 3). In the external dataset, two piRNAs (DQ596225, DQ596390) were validated to be significantly differentially expressed. Furthermore, six of seven piRNAs were significantly differentially expressed between LUAD and LUSC, with higher expression observed in LUSC (Supplementary Figure S1).

A combined miRNA+piRNA signature better predicts overall survival of lung adenocarcinoma patients

Previous work has shown that a multi-miRNA classifier (*miR-370*, *miR-376a*, and *miR-411*) was able

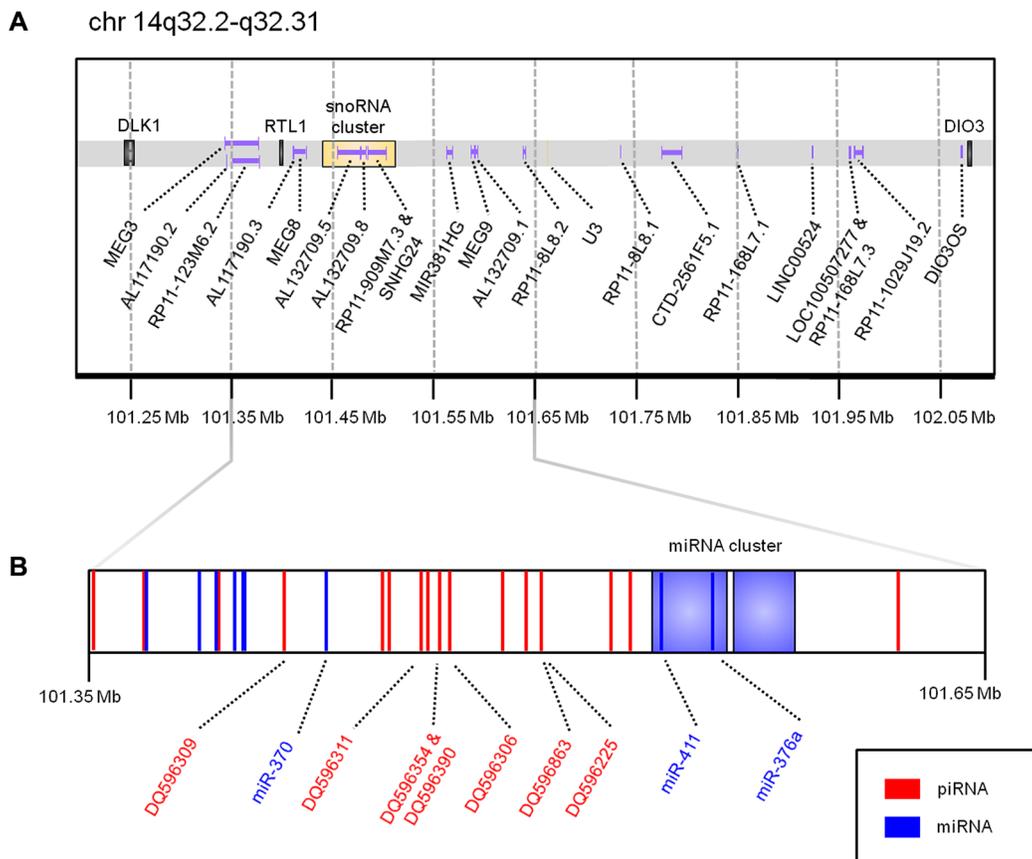


Figure 1: Schematic of the *DLK1-DIO3* imprinted locus. Genomic coordinates are derived from the UCSC Genome Browser (hg19 build). (A) illustrates protein-coding genes (black), long non-coding RNAs (purple), and snoRNAs (yellow). (B) is a zoomed view of chr14:101,350,000-101,650,000 to highlight the genomic position of miRNAs previously associated with lung cancer patient outcome (*miR-370*, *miR-411*, and *miR-376a*), as well as the seven piRNAs identified as expressed in this study. miRNAs, including two large miRNA clusters containing 41 miRNAs in total, are coloured in blue; piRNAs are coloured in red.

Table 1: Clinical features of lung cancer patient cohorts

Clinical Feature	Discovery Cohort		External Cohort	
	n (%)		n (%)	
Histological Subtype	LUAD	LUSC	LUAD	LUSC
Tumour	84	34	163	220
Non-malignant	84	34	46	45
Smoking History				
Current	35 (42)	11 (32)	44 (27)	57 (26)
Never	25 (30)	1 (3)	17 (10)	7 (3)
Former	20 (24)	22 (65)	98 (60)	147 (67)
Gender				
Male	24 (33)	10 (29)	86 (53)	165 (75)
Female	56 (67)	24 (71)	77 (47)	47 (21)
Age				
Range	45–90	58–88	40–86	39–84
Median	71	70	64	68
Stage				
IA	29 (35)	3 (9)	37 (23)	36 (16)
IB	19 (23)	11 (32)	51 (31)	61 (28)
IIA	13 (15)	4 (12)	13 (8)	30 (14)
IIB	5 (6)	7 (21)	23 (14)	40 (18)
IIIA	10 (12)	2 (6)	21 (13)	31 (14)
IIIB	0 (0)	1 (3)	6 (4)	10 (5)
IV	2 (2)	1 (3)	12 (7)	3 (1)

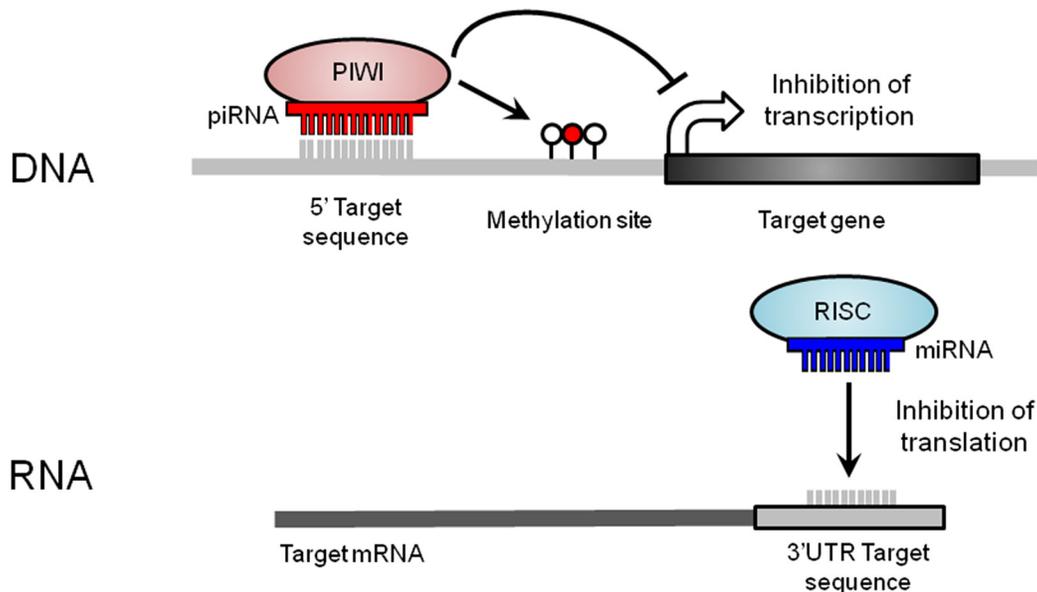


Figure 2: Small ncRNA-mediated mechanisms of gene expression regulation at the DNA and RNA levels. At the DNA level, piRNAs form a complex with PIWI (P-element-induced wimpy testis) proteins. piRNAs first target DNA sequences through base complementarity. Then, the piRNA/PIWI complex recruits the silencing machinery required to induce new DNA methylation events (red lollipop) nearby the targeted region, repressing transcript expression. At the RNA level, miRNAs, together with a RNA-induced silencing complex (RISC), bind to a 3' untranslated region (UTR) target sequence through base complementarity, which results in translational repression or mRNA degradation.

to predict LUAD patient outcome (Figure 1B) [8]. We applied this signature to our discovery dataset of LUAD and assessed the ability to predict patient overall survival (OS). Patient risk scores, indicating risk of death, were derived from a Cox proportional hazard model composed of these miRNAs. LUAD patients were divided into low, intermediate, and high risk groups and subjected to log-

rank survival analysis. While this miRNA signature is able to classify low risk patients in the discovery dataset, the intermediate and high risk groups are not well segregated (Figure 4A). In the external dataset, the miRNA signature is better able to stratify LUAD patient risk groups and achieves marginal significance (low risk vs. high risk $p = 0.051$) (Supplementary Figure S2).

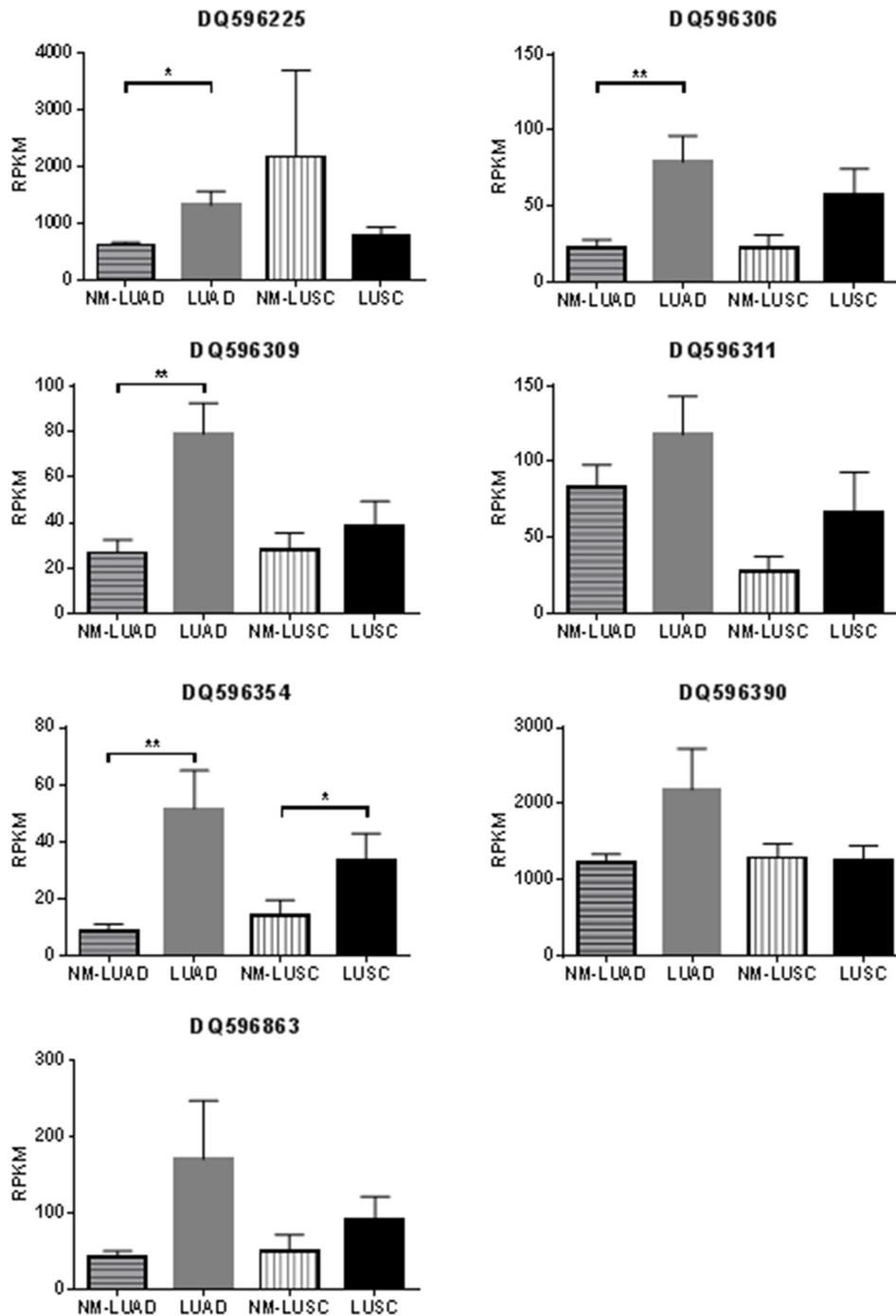


Figure 3: Histograms of piRNAs expressed in the discovery dataset (BCCA). Histograms display mean RPKM expression plus SEM in 84 paired non-malignant lung (NM-LUAD) and lung adenocarcinoma samples (LUAD), and 34 paired non-malignant lung (NM-LUSC) and lung squamous cell carcinoma (LUSC) samples. Significant p -values resulting from paired sign-rank analyses of gene expression are indicated as follows: * $p < 0.05$ ** $p < 0.01$.

Next, we investigated if piRNAs expressed from the *DLK1-DIO3* locus could predict LUAD patient outcome. Just as the consideration of multiple miRNAs produced a predictive signature, we hypothesize the consideration of multiple piRNAs expressed from this locus could result in a similar signature. Interestingly, while piRNA expression alone was unable to significantly predict OS in univariate or multivariate analysis, the incorporation of piRNA expression into the miRNA signature improved the stratification of patients into risk groups. The final survival model was selected by adding different combinations of the seven expressed piRNAs to the miRNA Cox proportional hazard model, and the model with the lowest *p*-value was used to calculate patient risk scores. The final survival model included the three-miRNA signature and four piRNAs encoded at this locus (DQ596306, DQ596309, DQ596390, and DQ596863), and will be referred to as the miRNA+piRNA signature.

Approximately one-third of patients from each risk group are reclassified by the miRNA+piRNA signature (Figure 4C). Low risk LUAD patients had significantly improved outcome compared to both high ($p = 0.002$) and intermediate ($p = 0.015$) risk groups (Figure 4B). In the external cohort, high-risk LUAD patients had significantly worse outcome compared to both low ($p = 0.037$) and intermediate ($p = 0.011$) risk groups (Supplementary Figure S2). In the external dataset the Kaplan-Meier curves of the low and medium risk groups were overlapping; suggesting the miRNA+piRNA signature is better able to categorize intermediate risk patients into either high

or low risk groups. When the new low and intermediate risk groups are combined in this dataset, the OS prediction improves ($p = 0.004$) (Figure 7, Supplementary Figure S2). A family-wise error rate (FWER) adjustment was applied to the *p*-values using the stringent Bonferroni method in order to test the robustness of the signature. Even after adjustment, the majority of the miRNA+piRNA signature *p*-values passed the new significance threshold (Figure 7).

The miRNA+piRNA signature is able to predict overall survival of lung squamous cell carcinoma patients

The previously-described miRNA signature has not been assessed in the other major subtype of NSCLC, LUSC. In both our discovery and external datasets, LUSC patient risk groups as defined by the miRNA signature did not have significantly different OS outcomes (Figure 5A). Similarly, the LUSC patient risk groups stratified by piRNA expression did not have significantly different OS. However, as was shown in LUAD, the miRNA+piRNA signature was also able to classify LUSC patients into risk groups with distinct OS outcomes in both the discovery (Figure 5B) and external datasets (Figure 7, Supplementary Figure S3). *P*-values remained significant after Bonferroni adjustment in the external dataset. All but one of the intermediate risk LUSC patients were reclassified into either high or low risk groups by the miRNA+piRNA signature in the discovery dataset (Figure 5C). Furthermore, the intermediate and high risk

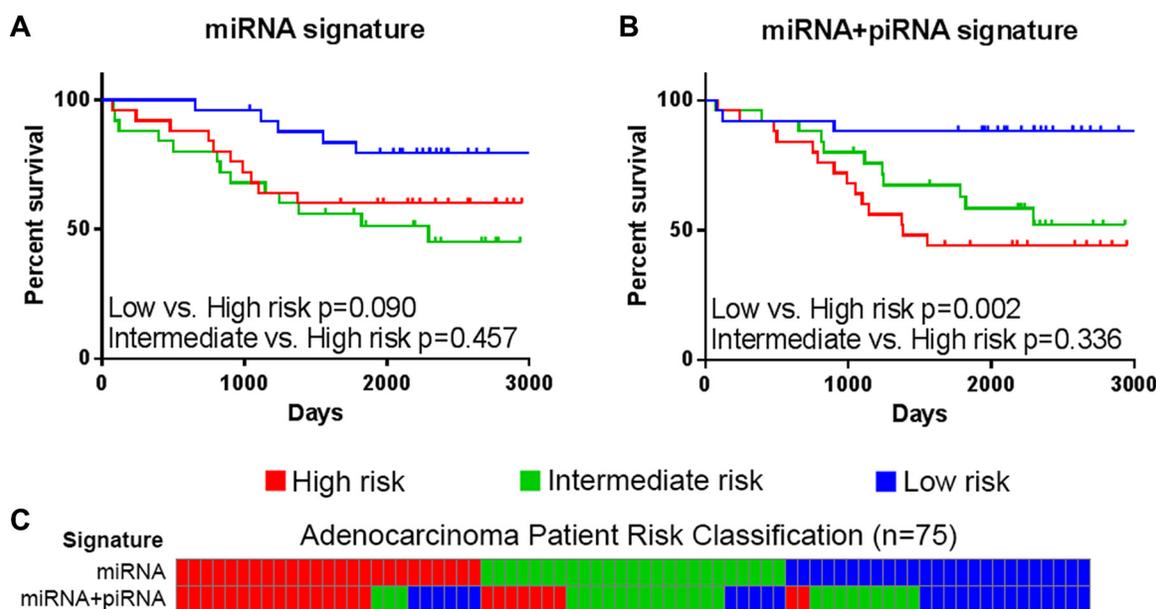


Figure 4: Overall survival of risk groups as defined in the lung adenocarcinoma discovery dataset (BCCA) ($n = 75$). Kaplan-Meier curves of high (red), intermediate (green), and low (blue) risk groups as defined by (A) the miRNA signature and (B) the miRNA+piRNA signature are shown. Log-rank *p*-values of select survival comparisons are shown. (C) Patients are ordered by their miRNA signature-based risk classification (top) in order to illustrate the re-classification that occurs when the miRNA+piRNA signature (bottom) is applied to the dataset.

highlighting the potential of piRNA-based prognostic markers across a variety of tumour types.

In summary, our results provide further evidence of the involvement of imprinting-regulated small ncRNAs

into NSCLC biology. The tissue-specific expression patterns of piRNAs and other small ncRNAs warrants further studies in order to establish their role across a wider spectrum of tissue types.

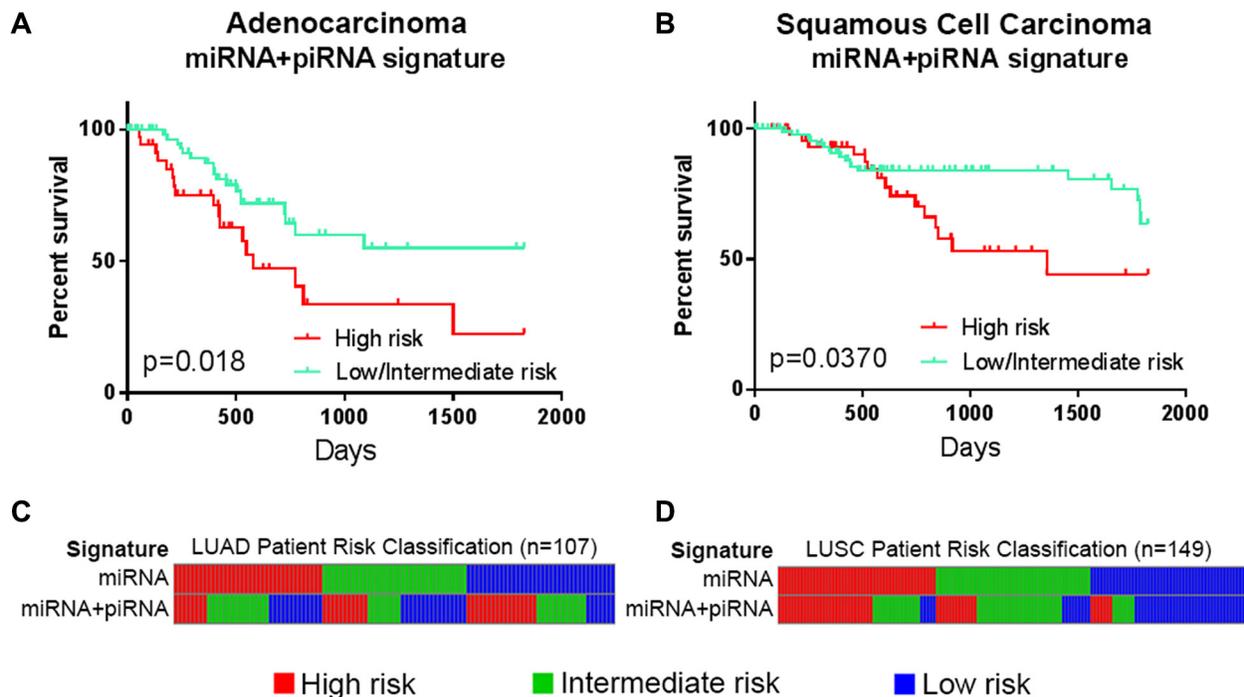


Figure 6: Performance of small ncRNA-based signatures predicting recurrence-free survival in non-small cell lung cancer. Risk groups as defined in the external dataset of (A) lung adenocarcinoma (LUAD, $n = 107$) and (B) lung squamous cell carcinoma (LUSC, $n = 149$). Kaplan-Meier curves of high risk groups (red) compared to the combined low and intermediate risk groups (turquoise) as defined by the miRNA+piRNA signature are shown. (C) LUAD patients and (D) LUSC patients are ordered by their miRNA signature-based risk classification (top) in order to illustrate the re-classification that occurs when the miRNA+piRNA signature (bottom) is applied to the dataset. Intermediate and low risk patients are represented by green and blue bars, respectively.

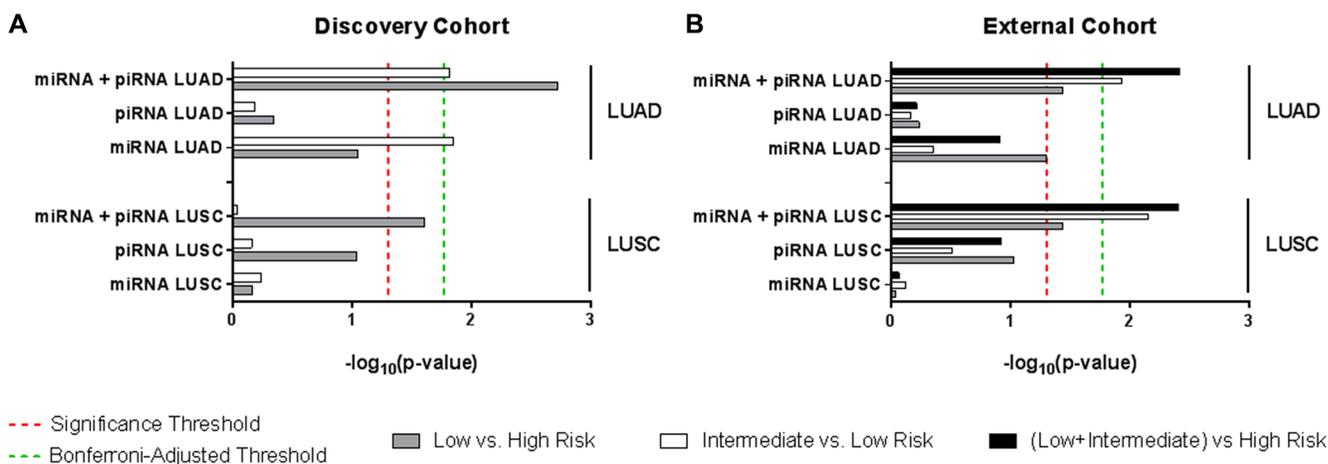


Figure 7: Log-rank p -value summary for overall survival predictions. Bar lengths represent the $-\log_{10}(p\text{-value})$ of each signature for LUAD (top) and LUSC (bottom) patients from (A) the discovery cohort and (B) the external cohort. Comparison across different risk groups are as follows: low vs. high risk (grey bars), intermediate vs. low risk (white bars), low vs. intermediate risk (black bars). Significance thresholds are established at $p\text{-value} = 0.05$ (red dashed line indicates $-\log_{10} 0.05$), and at $p\text{-value} = 0.017$ (Bonferroni-adjusted $p\text{-value}$) (green dashed line indicates $-\log_{10} 0.017$).

MATERIALS AND METHODS

Clinical cohorts and sample purity estimates

A description of discovery and external datasets and their clinical features can be found in Table 1.

Discovery cohort

Paired tumour and non-malignant samples were obtained from the British Columbia Cancer Agency (BCCA) under Research Ethics Board approval. Tumour sections were microdissected to achieve > 80% cell purity, as directed by a pathologist [21, 22].

External cohort

In The Cancer Genome Atlas (TCGA) dataset, purity estimates for tumour samples were publically available (<http://cancergenome.nih.gov/>). A purity cut-off of > 70% was applied according to previously published studies [23]. This was to make expression profiles more comparable between datasets, and to reduce contaminating sequences derived from alternative cell types since piRNA and miRNA expression is highly tissue specific.

RNA extraction and small RNA sequencing

For our discovery cohort (BCCA), total RNA was extracted using Trizol reagent (ThermoFisher, Waltham, MA, USA), according to the manufacturer's instructions, and eluted in RNase-free water. RNA concentration and quality was determined using a NanoDrop™ 2100 spectrophotometer, and samples were stored at -80°C. Sequencing analysis protocol was performed in the same manner for both the discovery and external (TCGA) cohorts [24]. Small RNA sequencing libraries were generated at Canada's Michael Smith Genome Sciences Centre and sequenced using Illumina HiSeq2000 instruments. For miRNA expression levels in our discovery cohort, reads were aligned using the Burrows-Wheeler Aligner (Version 0.5.7) and quantified against a miRNA annotation reference (miRBase Mature microRNAs Version 20). miRNA expression levels for the external cohort (TCGA) were accessed and retrieved in January 2015 using the TCGA data portal.

piRNA expression was determined as previously described [13]. Briefly, reads were first subject to quality control to exclude non-biological artifacts. Then, unaligned reads (in FASTQ format) were trimmed by size (retained reads \geq 23 bp) and quality score (Phred quality scores \geq 20) in order to enrich for high-quality reads mapping to piRNAs. Using the PartekFlow™ platform (Partek Inc., MO, USA), high-quality reads were mapped to the human genome (GRCh37/hg19) using the Spliced Transcripts Alignment to a Reference (STAR) aligner [25]. Reads were quantified by an Expectation/Maximization (E/M)

algorithm [26] using a piRNA-specific annotation file generated from the piRNABank database (<http://pirnabank.ibab.ac.in/>) [27]. Partek Genome Suite (PGS) was used to further process and filter quantified files. Reads per kilobase of exon model per million mapped reads (RPKM) was used to scale and normalize read count [28].

Small non-coding RNA differential expression analysis

Small ncRNAs were considered expressed if they had a scaled/normalized expression value \geq 1 in at least 10% of both the discovery and external datasets. In the discovery cohort, small ncRNA expression from the paired tumour and non-malignant lung samples were compared by the sign-rank test, and between histological subtypes by the Mann Whitney *U*-test. In the external cohort, all two-group comparisons were performed using the Mann Whitney *U*-test. Significance threshold was established at *p*-value \leq 0.05.

Survival analysis

Univariate analysis: Cases were grouped based on piRNA expression tertiles, and survival analysis was conducted by log-rank test. For piRNAs with expression of 0 RPKM in > 1/3 of samples, cases were dichotomized into those with RPKM = 0 and those with RPKM > 0.

Cox proportional hazard model: Samples that had complete miRNA expression (RPKM), piRNA expression (RPKM), and survival data (overall survival or recurrence-free survival) were considered for Cox proportional hazard models. In addition to miRNAs previously associated with lung cancer patient outcome (*miR-370*, *miR-376a*, *miR-411*), Cox proportional hazard models including combinations of the seven expressed piRNAs, and combinations of the miRNAs and piRNAs were analyzed. The model with the lowest *p*-value was chosen for further analysis. Patient risk scores were generated per model by multiplying the expression value of a given gene by its hazard coefficient, and then summing the transformed gene expression values per sample [21]. Risk scores were ranked and divided into tertiles of high, intermediate, and low risk. Risk group Kaplan-Meier survival curves were then compared using the log-rank method. Significance threshold was established at *p*-value \leq 0.05. Raw *p*-values were then adjusted using the Bonferroni method, resulting in an adjusted *p*-value cut-off of \leq 0.017.

ACKNOWLEDGMENTS AND FUNDING

This work was supported by a grant from the Canadian Institutes for Health Research (CIHR grant number FDN - 143345). KSSE, SHYK and GLS are supported by CIHR Frederick Banting and Charles Best Canada Graduate Scholarship.

CONFLICTS OF INTEREST

None.

REFERENCES

1. Brannan CI, Bartolomei MS. Mechanisms of genomic imprinting. *Current opinion in genetics & development*. 1999; 9:164–170. doi:10.1016/S0959-437X(99)80025-2.
2. Jelinic P, Shaw P. Loss of imprinting and cancer. *The Journal of pathology*. 2007; 211:261–268. doi:10.1002/path.2116.
3. Hirasawa R, Feil R. Genomic imprinting and human disease. *Essays in biochemistry*. 2010; 48:187–200. doi:10.1042/bse0480187.
4. Kagami M, Sekita Y, Nishimura G, Irie M, Kato F, Okada M, Yamamori S, Kishimoto H, Nakayama M, Tanaka Y, Matsuoka K, Takahashi T, Noguchi M, et al. Deletions and epimutations affecting the human 14q32.2 imprinted region in individuals with paternal and maternal upd(14)-like phenotypes. *Nature genetics*. 2008; 40:237–242. doi:10.1038/ng.2007.56.
5. Lempiainen H, Couttet P, Bolognani F, Muller A, Dubost V, Luisier R, Del Rio Espinola A, Vitry V, Unterberger EB, Thomson JP, Treindl F, Metzger U, Wrzodek C, et al. Identification of Dlk1-Dio3 imprinted gene cluster noncoding RNAs as novel candidate biomarkers for liver tumor promotion. *Toxicological sciences*. 2013; 131:375–386. doi:10.1093/toxsci/kfs303.
6. Benetatos L, Hatzimichael E, Londin E, Vartholomatos G, Loher P, Rigoutsos I, Briasoulis E. The microRNAs within the DLK1-DIO3 genomic region: involvement in disease pathogenesis. *Cellular and molecular life sciences*. 2013; 70:795–814. doi:10.1007/s00018-012-1080-8.
7. Valdmanis PN, Roy-Chaudhuri B, Kim HK, Sayles LC, Zheng Y, Chuang CH, Caswell DR, Chu K, Zhang Y, Winslow MM, Sweet-Cordero EA, Kay MA. Upregulation of the microRNA cluster at the Dlk1-Dio3 locus in lung adenocarcinoma. *Oncogene*. 2015; 34:94–103. doi:10.1038/onc.2013.523.
8. Nadal E, Zhong J, Lin J, Reddy RM, Ramnath N, Orringer MB, Chang AC, Beer DG, Chen G. A MicroRNA cluster at 14q32 drives aggressive lung adenocarcinoma. *Clinical cancer research*. 2014; 20:3107–3117. doi:10.1158/1078-0432.CCR-13-3348.
9. Watanabe T, Tomizawa S, Mitsuya K, Totoki Y, Yamamoto Y, Kuramochi-Miyagawa S, Iida N, Hoki Y, Murphy PJ, Toyoda A, Gotoh K, Hiura H, Arima T, et al. Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrfl locus. *Science*. 2011; 332:848–852. doi:10.1126/science.1203919.
10. Ng KW, Anderson C, Marshall EA, Minatel BC, Enfield KS, Saprunoff HL, Lam WL, Martinez VD. Piwi-interacting RNAs in cancer: emerging functions and clinical utility. *Molecular cancer*. 2016; 15:5. doi:10.1186/s12943-016-0491-9.
11. Busch J, Ralla B, Jung M, Wotschovsky Z, Trujillo-Arribas E, Schwabe P, Kilic E, Fendler A, Jung K. Piwi-interacting RNAs as novel prognostic markers in clear cell renal cell carcinomas. *Journal of experimental & clinical cancer research*. 2015; 34:61. doi:10.1186/s13046-015-0180-3.
12. Martinez VD, Vucic EA, Thu KL, Hubaux R, Enfield KS, Pikor LA, Becker-Santos DD, Brown CJ, Lam S, Lam WL. Unique somatic and malignant expression patterns implicate PIWI-interacting RNAs in cancer-type specific biology. *Scientific reports*. 2015; 5:10423. doi:10.1038/srep10423.
13. Martinez VD, Enfield KS, Rowbotham DA, Lam WL. An atlas of gastric PIWI-interacting RNA transcriptomes and their utility for identifying signatures of gastric cancer recurrence. *Gastric cancer*. 2016; 19:660–665. doi:10.1007/s10120-015-0487-y.
14. Huang G, Hu H, Xue X, Shen S, Gao E, Guo G, Shen X, Zhang X. Altered expression of piRNAs and their relation with clinicopathologic features of breast cancer. *Clinical & translational oncology*. 2013; 15:563–568. doi:10.1007/s12094-012-0966-0.
15. Cui L, Lou Y, Zhang X, Zhou H, Deng H, Song H, Yu X, Xiao B, Wang W, Guo J. Detection of circulating tumor cells in peripheral blood from patients with gastric cancer using piRNAs as markers. *Clinical biochemistry*. 2011; 44:1050–1057. doi:10.1016/j.clinbiochem.2011.06.004.
16. Cheng J, Guo JM, Xiao BX, Miao Y, Jiang Z, Zhou H, Li QN. piRNA, the new non-coding RNA, is aberrantly expressed in human cancer cells. *Clinica chimica acta*. 2011; 412:1621–1625. doi:10.1016/j.cca.2011.05.015.
17. Fu A, Jacobs DI, Zhu Y. Epigenome-wide analysis of piRNAs in gene-specific DNA methylation. *RNA biology*. 2014; 11:1301–1312. doi:10.1080/15476286.2014.996091.
18. Fu A, Jacobs DI, Hoffman AE, Zheng T, Zhu Y. PIWI-interacting RNA 021285 is involved in breast tumorigenesis possibly by remodeling the cancer epigenome. *Carcinogenesis*. 2015; 36:1094–1102. doi:10.1093/carcin/bgv105.
19. Molina-Pinelo S, Salinas A, Moreno-Mata N, Ferrer I, Suarez R, Andres-Leon E, Rodriguez-Paredes M, Gutekunst J, Jantus-Lewintre E, Camps C, Carnero A, Paz-Ares L. Impact of DLK1-DIO3 imprinted cluster hypomethylation in smoker patients with lung cancer. *Oncotarget*. 2016 Jul 15. doi: 10.18632/oncotarget.10611. [Epub ahead of print].
20. Firmino N, Martinez VD, Rowbotham DA, Enfield KS, Bennewith KL, Lam WL. HPV status is associated with altered PIWI-interacting RNA expression pattern in head and neck cancer. *Oral oncology*. 2016; 55:43–48. doi:10.1016/j.oraloncology.2016.01.012.
21. Becker-Santos DD, Thu KL, English JC, Pikor LA, Martinez VD, Zhang M, Vucic EA, Luk MT, Carraro A, Korbelik J, Piga D, Lhomme NM, Tsay MJ, et al. Developmental transcription factor NFIB is a putative target of oncofetal miRNAs and is associated with tumour aggressiveness in lung adenocarcinoma. *The Journal of pathology*. 2016. doi:10.1002/path.4765.

22. Vucic EA, Thu KL, Pikor LA, Enfield KS, Yee J, English JC, MacAulay CE, Lam S, Jurisica I, Lam WL. Smoking status impacts microRNA mediated prognosis and lung adenocarcinoma biology. *BMC cancer*. 2014; 14:778. doi:10.1186/1471-2407-14-778.
23. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nature communications*. 2015; 6:8971. doi:10.1038/ncomms9971.
24. Chu A, Robertson G, Brooks D, Mungall AJ, Birol I, Coope R, Ma Y, Jones S, Marra MA. Large-scale profiling of microRNAs for The Cancer Genome Atlas. *Nucleic acids research*. 2016; 44:e3. doi:10.1093/nar/gkv808.
25. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013; 29: 15–21. doi:10.1093/bioinformatics/bts635.
26. Xing Y, Yu T, Wu YN, Roy M, Kim J, Lee C. An expectation-maximization algorithm for probabilistic reconstructions of full-length isoforms from splice graphs. *Nucleic acids research*. 2006; 34:3150–3160. doi:10.1093/nar/gkl396.
27. Sai Lakshmi S, Agrawal S. piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic acids research*. 2008; 36:D173–177. doi:10.1093/nar/gkm696.
28. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*. 2008; 5:621–628. doi:10.1038/nmeth.1226.