

An atlas of gastric PIWI-interacting RNA transcriptomes and their utility for identifying signatures of gastric cancer recurrence

Victor D. Martinez¹ · Katey S. S. Enfield¹ · David A. Rowbotham¹ · Wan L. Lam¹

Received: 9 December 2014 / Accepted: 24 February 2015 / Published online: 17 March 2015
© The International Gastric Cancer Association and The Japanese Gastric Cancer Association 2015

Abstract The poor survival and recurrence rate in gastric adenocarcinoma highlights the need for cancer gene discovery. Towards this end, we globally assessed the expression of an emerging class of small non-coding RNAs, called PIWI-interacting RNAs (piRNAs). We analysed the transcriptomes of 358 non-malignant stomach tissue and gastric adenocarcinoma samples, and found that nearly half of the expressed piRNAs were overexpressed in tumours. Our gastric piRNA atlas showed that most piRNAs were embedded in protein-coding sequences rather than known piRNA clusters. Furthermore, we identified a three-piRNA signature associated with recurrence-free survival. In this proof-of-principle study, we demonstrate the potential clinical utility of piRNAs in gastric cancer.

Keywords PIWI-interacting RNAs · Epigenetics · PIWI · Gastric adenocarcinoma · Patient outcome

Introduction

Gastric adenocarcinoma (GA) has a poor 5-year survival, with high rates of relapse, posing an urgent need for biomarker discovery [1]. Small non-coding RNAs, such as

microRNAs, have proven clinical utility, owing to stability in biofluids and formalin-fixed paraffin-embedded material [2]. Recent studies have demonstrated the deregulation of two members of an emerging class of small non-coding RNA, PIWI-interacting RNAs (piRNAs), in a small cohort of GA [3–5].

GA is one of the cancer types selected for profiling by The Cancer Genome Atlas (TCGA), providing a valuable resource for discovery of new cancer genes [6]. Although piRNAs were not one of the dimensions analysed by TCGA, we were able to generate expression profiles for 38 non-malignant stomach tissue samples and 320 GA samples from raw sequencing data using a custom analysis pipeline. We performed an unbiased, global analysis of the 20,821 piRNAs in the human genome to deduce the relationship of deregulated piRNAs with clinicopathological features, and to evaluate a possible role for piRNAs as prognostic biomarkers.

Materials and methods

Samples

A total of 320 GA and 38 non-malignant small RNA sequencing libraries (for processing, see Fig. S1) were obtained from the Cancer Genomics Hub data repository (dbgap project ID 6208). SNP 6.0 copy number profiles were downloaded from (<https://tcga-data.nci.nih.gov/tcga/dataAccessMatrix.htm>). An additional cohort of 25 GA small RNA sequencing libraries was downloaded from Gene Expression Omnibus (series GSE36968) [7]. Rank-normalized expression of the recurrence-free survival (RFS) signature piRNAs (described later) was extracted from nine additional cancer types with available small RNA sequencing data and RFS follow-up.

V. D. Martinez and K. S. S. Enfield contributed equally.

Electronic supplementary material The online version of this article (doi:10.1007/s10120-015-0487-y) contains supplementary material, which is available to authorized users.

✉ Victor D. Martinez
vmartinez@bccrc.ca

¹ Department of Integrative Oncology, BC Cancer Research Centre, 675 West 10th Avenue, Vancouver, BC V5Z 1L3, Canada

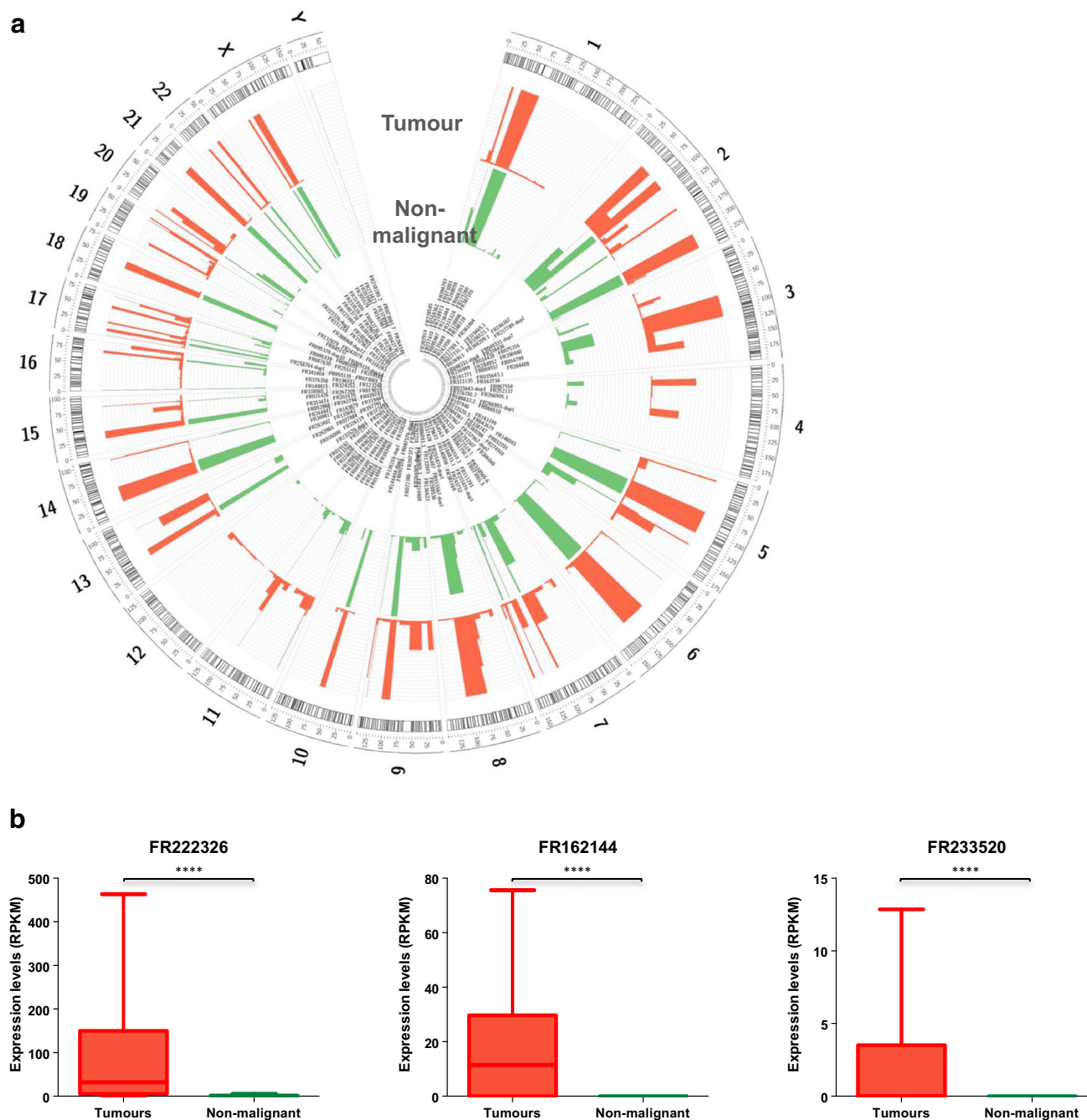


Fig. 1 Expression of PIWI-interacting RNAs (piRNAs) in non-malignant stomach tissue and gastric adenocarcinoma (GA) samples. **a** Circular representation of average expression (reads per kilobase of exon per million mapped reads; *RPKM*) of the 312 piRNAs mapping to 378 piRNA loci in GA (red) and non-malignant stomach tissue (green) [21]. The innermost and outermost rings represent each chromosome, with the names of expressed piRNAs branching outwards from their genomic position. **b** Box and whiskers plots of

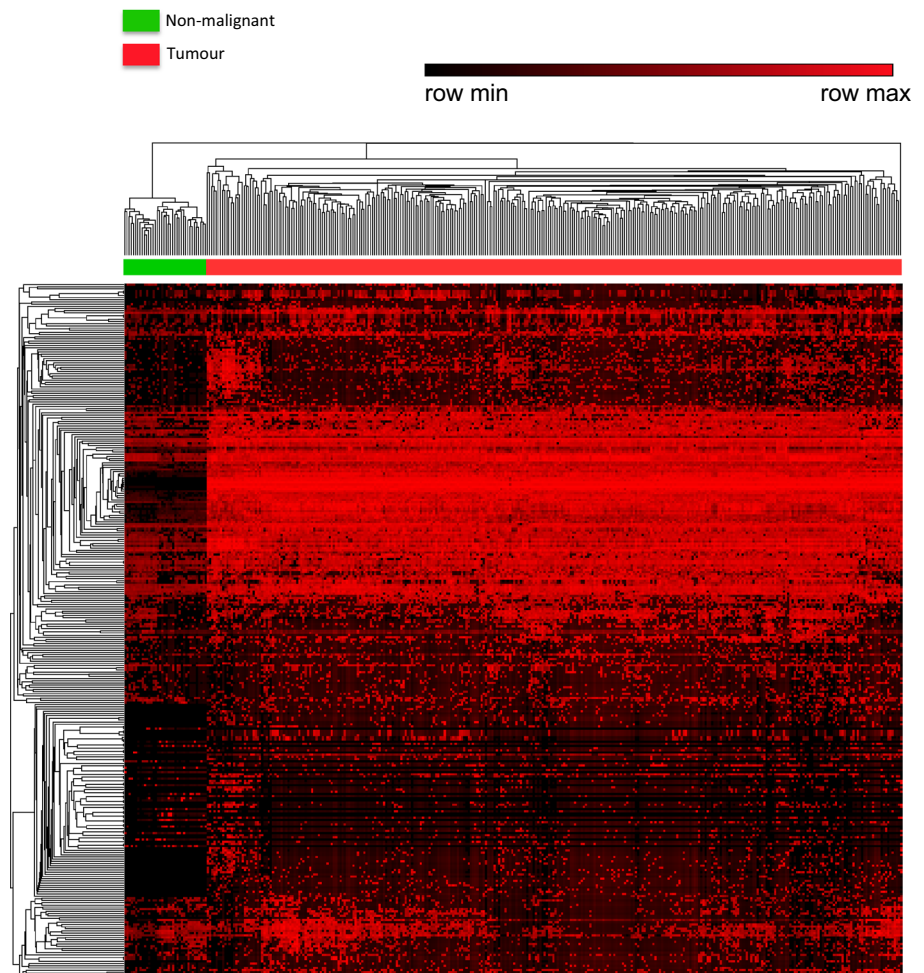
expression values (*RPKM*) of the three examples of piRNAs showing significant average fold change between GA (red bar) and non-malignant stomach tissue (green). The box extends from the 25th to the 75th percentile, with the median shown, whereas the whiskers range from minimum to maximum. Expression between both groups was assessed by the Mann–Whitney *U* test (four asterisks $p < 0.0001$) (colour figure online)

Clustering analysis

Rank-normalized piRNA reads per kilobase of exon per million mapped reads were clustered, using hierarchical

and consensus approaches, in GENE-E (<http://www.broadinstitute.org/cancer/software/GENE-E/index.html>) and GenePattern [8, 9]. Hierarchical clustering was performed using Euclidean distance with average linkage. Consensus

Fig. 2 Unsupervised hierarchical clustering of 312 PIWI-interacting RNAs (piRNAs) expressed in non-malignant stomach tissue and gastric adenocarcinoma (GA) samples. Unsupervised hierarchical clustering using Euclidean distance and average linkage on rank-normalized piRNA expression values is shown. Non-malignant stomach tissue samples are coloured *green*, and GA samples are coloured *red*. Relative expression per piRNA is coloured on a *blue* (minimum) to *red* (maximum) scale (colour figure online)



clustering analysis was performed using the following parameters: $k_{\max} = 5$; clustering algorithm = hierarchical; distance = Euclidean; resampling iterations = 20.

Differential expression analysis

Differentially expressed piRNAs (non-malignant tissues vs. tumours) were identified using the “Comparative Marker Selection” module implemented in GenePattern [9, 10]. Differential expression was assessed through a signal-to-noise ratio test. The nominal p value was estimated using a permutation test (100,000 permutations), and was corrected using the procedure of Benjamini and Hochberg [8]. The expression fold change was calculated by dividing the mean expression value of tumours by the mean expression value in non-malignant tissues.

Survival analysis

Clinical information was obtained from the TCGA data portal. Overall survival (OS) data with at least 1-day

follow-up were available for 282 patients. Sixteen patients died of causes other than GA, and were removed from analysis. Log-rank survival analysis was performed on piRNAs expressed in at least two thirds of samples ($n = 59$ piRNAs) in MATLAB (The MathWorks, Natick, MA, USA); high- and low-expression tertiles were compared. RFS data were available for 240 GA patients (information for the nine additional cancer types assessed for RFS can be found in Table S2). Log-rank survival analysis was performed as for OS. Cox proportional hazard models were evaluated in R (‘survival’ package; R version 3.1.0). Expression values of the piRNAs in the model with the best performance (lowest p value) were transformed into a risk score by multiplying the expression values of each piRNA by their respective Cox proportional hazard coefficient, and then summing their values [11]. Risk scores were ranked, and high- and low-risk tertiles were compared by Kaplan–Meier analysis. In all cases, a p value below 0.05 was considered significant.

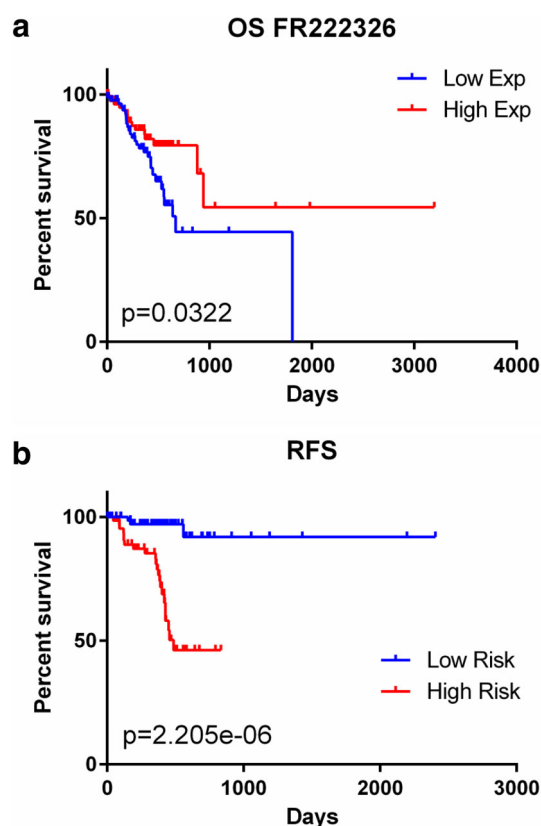


Fig. 3 PIWI-interacting RNAs (piRNAs) significantly associated with gastric adenocarcinoma (GA) patient outcome. Significance was assessed by the log-rank method. **a** Kaplan–Meier plot of GA patient overall survival (OS) stratified by high (red) and low (blue) FR222326 expression. **b** Kaplan–Meier plot of GA patient recurrence-free survival (RFS) stratified by high risk (red) and low risk (blue) scores obtained from the three-piRNA signature. The median time to recurrence for low risk was undefined, and for high risk was 16 months [hazard ratio 10.22 (3.14–15.39)]. Exp expression (colour figure online)

Results

We detected expression of 312 piRNAs encoded at 378 loci: 213 piRNAs in non-malignant stomach tissues and 299 in GA tissues. On the basis of these data, we generated a map of piRNA loci expressed in stomach tissue, and superimposed the malignant piRNA expression pattern onto this gastric transcriptome map (Fig. 1a). Differential expression was observed for piRNAs expressed at low to moderate levels (Fig. 1b, Table S1). Additionally, an unsupervised hierarchical clustering analysis unambiguously separated non-malignant stomach tissue samples from GA samples (Fig. 2), and this was further corroborated by a consensus clustering analysis (Fig. S2). Moreover, we observed that, in contrast to previous

reports, the majority of expressed piRNAs did not originate from known human piRNA clusters [12]. Instead, 70.9 % of the expressed piRNAs were derived from protein-coding sequences.

Remarkably, half ($n = 156$) of the expressed piRNAs were significantly differentially expressed in GA as compared with non-malignant stomach tissue. In fact, 45 displayed GA-specific expression, and 18 were exclusively expressed in non-malignant stomach tissue. Most of the remaining 93 deregulated piRNAs were overexpressed, with only seven undergoing underexpression. We further investigated these differentially expressed piRNAs regarding their association with OS or RFS of GA patients.

Only one piRNA, FR222326, was associated with OS (log-rank $p = 0.0322$) (Fig. 3a), whereas five piRNAs were significantly associated with RFS (Table S2). We evaluated whether or not these RFS-associated piRNAs could be combined into a multi-piRNA signature to better predict RFS in this GA cohort. Using a Cox proportional hazard model, we identified a three-piRNA signature consisting of FR290353, FR064000, and either FR387750 or FR157678, which are sequence variants of the same transcribed locus ($p = 4.913 \times 10^{-5}$) (Table S2). Use of either variant had no effect on statistical output. The Kaplan–Meier plot of piRNA expression risk scores shows the high-risk group to be significantly associated with shorter time to recurrence (log-rank $p = 2.21 \times 10^{-6}$) (Fig. 3b). Notably, these piRNAs were not significantly associated with any other clinicopathological features (data not shown). The piRNA RFS signature was tested in nine additional tumour types. Although it performed well in colon cancer (log-rank $p = 0.0061$), the p values did not approach the significance observed in GA (Table S2).

We further investigated whether DNA copy number levels were associated with expression changes of the five piRNAs associated with RFS in the same TCGA cohort. Copy number alterations at FR381169 ($p = 0.0001$), FR290353 ($p = 0.0294$), and FR064000 ($p = 0.0004$) loci were significantly associated with expression alterations, suggesting genetically selected mechanisms of deregulated piRNA expression in these cases (Fig. 4a). Next, we validated expression levels of these five piRNAs in an independent cohort of GA. FR157678, FR290353, and FR387750 were expressed at similar levels, whereas FR064000 and FR381169 were more lowly expressed (Fig. 4b). If FR064000 is removed from the RFS signature, the remaining piRNAs (FR290353 and FR387750/FR157678) retain the ability to significantly predict RFS (Cox proportional hazard $p = 1.34 \times 10^{-4}$; log-rank $p = 7.84 \times 10^{-5}$).

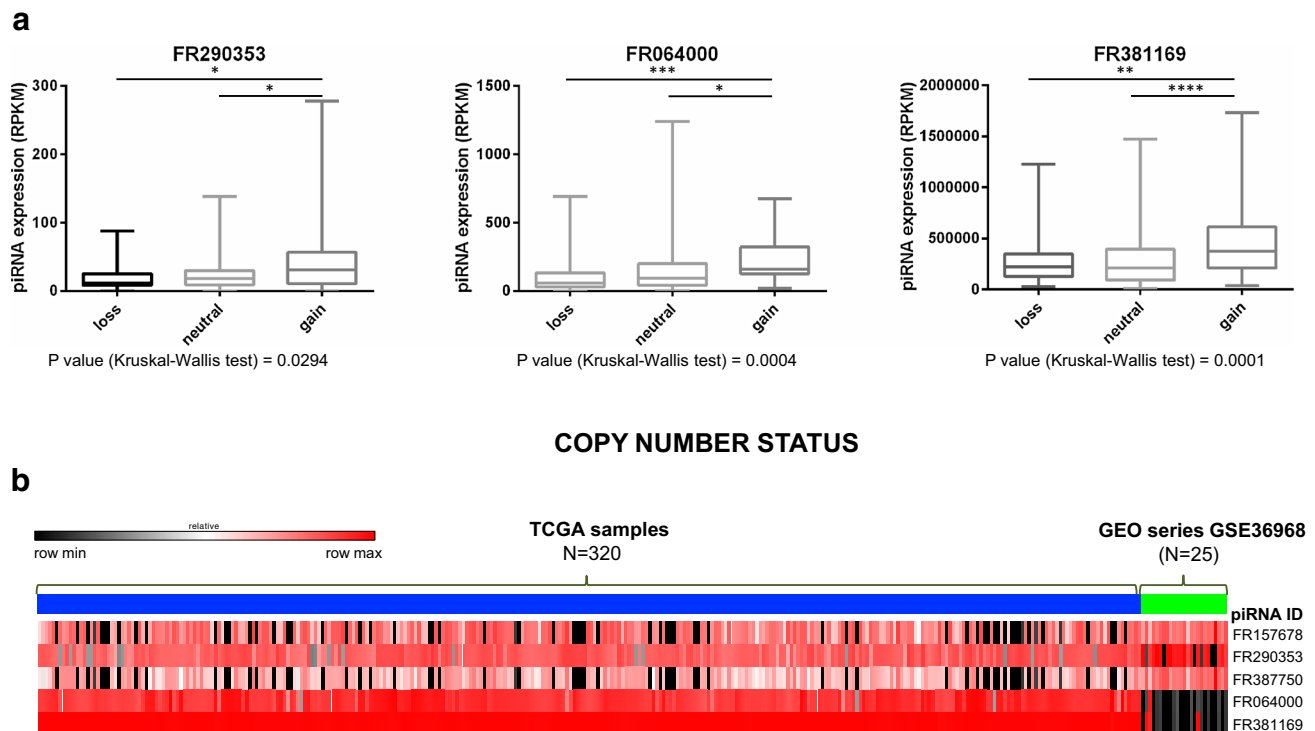


Fig. 4 Validation of expression and DNA-level mechanisms of alteration of recurrence-free survival (RFS)-associated PIWI-interacting RNAs (piRNAs). **a** Box and whiskers plots of \log_2 copy number values for the RFS-associated piRNAs whose expression was significantly associated with copy number status (loss, \log_2 value below -0.2 ; neutral, \log_2 value between -0.2 and 0.2 ; gain, \log_2 value greater than 0.2). The box extends from the 25th to the 75th percentile, with the median shown, whereas the whiskers range from minimum to maximum (Student's t test, one asterisk $p < 0.05$, two

asterisks $p < 0.01$, three asterisks $p < 0.001$, four asterisks $p < 0.0001$; Kruskal-Wallis test, p value is shown). **b** Expression levels for the five RFS-associated piRNAs were compared in two independent cohorts: The Cancer Genome Atlas (TCGA) (blue), and Gene Expression Omnibus (GEO) series GSE36968 (green). Expression levels for both datasets were elucidated through the same analysis pipeline described in the legend for Fig. S1, and were rank normalized prior to comparison. RPKM reads per kilobase of exon per million mapped reads (colour figure online)

Discussion

Recent studies have expanded the function of piRNAs from germline cells to somatic tissues and cancer, including GA [3, 4]. Although efforts have been made to study deregulation of piRNAs in GA, an unbiased analysis of global piRNA expression in gastric tissue was warranted. In this study, we took advantage of the massive sequencing data generated by TCGA by applying a custom analysis pipeline to deduce the piRNA expression patterns in one of the largest cohorts of GA to date.

We detected expression of 312 piRNAs, and remarkably, found that half of these were significantly deregulated in GA. Most of these piRNAs were overexpressed in GA compared with non-malignant stomach tissue, suggesting their importance in GA. Since the function of most piRNAs has not yet been characterized in humans, it is difficult to speculate how their deregulation is mechanistically influencing GA. However, we observed that 70.9 % of these piRNAs were located within protein-coding sequences. Localization of piRNAs within protein-coding sequences

has been associated with *cis*- and *trans*-regulatory effects on protein-coding transcripts in diverse species [13–15].

We have demonstrated piRNAs, like other non-coding RNAs [16–19], are associated with GA patient outcome. FR222326 was significantly associated with OS, and perhaps more impressively, a three-piRNA signature (FR290353, FR064000, FR387750/FR157678) effectively stratified GA patients into low and high risk of recurrence groups. When tested in other cancer types, the RFS signature performed well in colon cancer, suggesting conserved importance to digestive tract malignancies. We did not detect mutations in the RFS-associated piRNA genes; however, we show that DNA copy number is likely one of the genetic mechanisms of deregulation for FR381169, and RFS-signature piRNAs FR290353 and FR064000. (The Illumina HumanMethylation450 BeadChip platform is uninformative for these genes, as they were not covered by any probes.) Although expression of FR064000 did not provide validation in the independent cohort, expression of the remaining RFS-associated piRNAs was able to significantly predict RFS in the TCGA cohort. Although the

clinical utility of piRNAs has not yet been defined, it is highly feasible owing to their small size. Other small RNAs, such as microRNAs, are stable in biofluids, circulating tumour cells, and formalin-fixed paraffin-embedded materials [2]. Considering there are 10–25 times more piRNA species (20,000–50,000) than microRNAs (approximately 2,000) [20], their deregulation is likely at least as relevant. Therefore, piRNAs hold great promise as potential biomarkers.

In summary, we have identified transcribed piRNA loci in non-malignant and malignant stomach tissues, and have characterized malignancy-associated expression patterns of GA. In doing so, we have generated a piRNA transcription atlas of the gastric cancer genome. Furthermore, we use this study as a proof of principle to demonstrate the potential clinical utility of piRNAs in GA patient stratification. We have made the data derived from our analysis publicly available to encourage further investigations of piRNAs in GA.

Acknowledgments K.S.S.E. is supported by a Charles Best Canada Graduate Scholarship from the Canadian Institutes of Health Research. Funding support for this work was through research grants to W.L.L. from the Canadian Institutes of Health Research, the Canadian Cancer Society, the Terry Fox Foundation (Canada) and the National Institutes of Health (USA).

Conflict of interest The authors declare that they have no conflict of interest.

References

- McLean MH, El-Omar EM. Genetics of gastric cancer. *Nat Rev Gastroenterol Hepatol*. 2014;11(11):664–74. doi:10.1038/nrgastro.2014.143.
- Zhu C, Ren C, Han J, Ding Y, Du J, Dai N, et al. A five-microRNA panel in plasma was identified as potential biomarker for early detection of gastric cancer. *Br J Cancer*. 2014;110(9):2291–9. doi:10.1038/bjc.2014.119.
- Cheng J, Deng H, Xiao B, Zhou H, Zhou F, Shen Z, et al. piR-823, a novel non-coding small RNA, demonstrates in vitro and in vivo tumor suppressive activity in human gastric cancer cells. *Cancer Lett*. 2012;315(1):12–7. doi:10.1016/j.canlet.2011.10.004.
- Cheng J, Guo JM, Xiao BX, Miao Y, Jiang Z, Zhou H. piRNA, the new non-coding RNA, is aberrantly expressed in human cancer cells. *Clin Chim Acta*. 2011;412(17–18):1621–5. doi:10.1016/j.cca.2011.05.015.
- Cui L, Lou Y, Zhang X, Zhou H, Deng H, Song H, et al. Detection of circulating tumor cells in peripheral blood from patients with gastric cancer using piRNAs as markers. *Clin Biochem*. 2011;44(13):1050–7. doi:10.1016/j.clinbiochem.2011.06.004.
- The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 2014; 513(7517):202–9. doi:10.1038/nature13480.
- Kim YH, Liang H, Liu X, Lee JS, Cho JY, Cheong JH, et al. AMPK α modulation in cancer progression: multilayer integrative analysis of the whole transcriptome in Asian gastric cancer. *Cancer Res*. 2012;72(10):2512–21. doi:10.1158/0008-5472.CAN-11-3870.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B (Methodol)*. 1995;57(1):289–300. doi:10.2307/2346101.
- Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. *Nat Genet*. 2006;38(5):500–1. doi:10.1038/ng0506-500.
- Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*. 1998;95(25):14863–8.
- Chen HY, Yu SL, Chen CH, Chang GC, Chen CY, Yuan A, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *New Engl J Med*. 2007;356(1):11–20. doi:10.1056/NEJMoa060096.
- Ha H, Song J, Wang S, Kapusta A, Feschotte C, Chen KC, et al. A comprehensive analysis of piRNAs from adult human testis and their relationship with genes and mobile elements. *BMC Genomics*. 2014;15:545. doi:10.1186/1471-2164-15-545.
- Gou LT, Dai P, Yang JH, Xue Y, Hu YP, Zhou Y, et al. Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell Res*. 2014;24(6):680–700. doi:10.1038/cr.2014.41.
- Robine N, Lau NC, Balla S, Jin Z, Okamura K, Kuramochi-Miyagawa S, et al. A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Curr Biol* 2009;19(24):2066–76. doi:10.1016/j.cub.2009.11.064.
- Saito K, Inagaki S, Mituyama T, Kawamura Y, Ono Y, Sakota E, et al. A regulatory circuit for piwi by the large Maf gene traffic jam in *Drosophila*. *Nature*. 2009;461(7268):1296–9. doi:10.1038/nature08501.
- Okugawa Y, Toiyama Y, Hur K, Toden S, Saigusa S, Tanaka K, et al. Metastasis-associated long non-coding RNA drives gastric cancer development and promotes peritoneal metastasis. *Carcinogenesis*. 2014;. doi:10.1093/carcin/bgu200.
- Wu HH, Lin WC, Tsai KW. Advances in molecular biomarkers for gastric cancer: miRNAs as emerging novel cancer markers. *Expert Rev Mol Med*. 2014;16:e1. doi:10.1017/erm.2013.16.
- Zhang Z, Li Z, Li Y, Zang A. MicroRNA and signaling pathways in gastric cancer. *Cancer Gene Ther*. 2014;21(8):305–16. doi:10.1038/cgt.2014.37.
- He Y, Meng XM, Huang C, Wu BM, Zhang L, Lv XW, et al. Long noncoding RNAs: novel insights into hepatocellular carcinoma. *Cancer Lett*. 2014;344(1):20–7. doi:10.1016/j.canlet.2013.10.021.
- Aravin AA, Hannon GJ, Brennecke J. The piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science*. 2007;318(5851):761–4. doi:10.1126/science.1146484.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19(9):1639–45. doi:10.1101/gr.092759.109.