

# SCIENTIFIC REPORTS



OPEN

## Unique somatic and malignant expression patterns implicate PIWI-interacting RNAs in cancer-type specific biology

Received: 22 January 2015

Accepted: 13 April 2015

Published: 27 May 2015

Victor D. Martinez<sup>1</sup>, Emily A. Vucic<sup>1,\*</sup>, Kelsie L. Thu<sup>1,\*</sup>, Roland Hubaux<sup>1</sup>, Katey S.S. Enfield<sup>1</sup>, Larissa A. Pikor<sup>1</sup>, Daiana D. Becker-Santos<sup>1</sup>, Carolyn J. Brown<sup>1,2</sup>, Stephen Lam<sup>1</sup> & Wan L. Lam<sup>1</sup>

Human PIWI-interacting RNAs (piRNAs) are known to be expressed in germline cells, functionally silencing LINEs and SINEs. Their expression patterns in somatic tissues are largely uncharted. We analyzed 6,260 human piRNA transcriptomes derived from non-malignant and tumour tissues from 11 organs. We discovered that only 273 of the 20,831 known piRNAs are expressed in somatic non-malignant tissues. However, expression patterns of these piRNAs were able to distinguish tissue-of-origin. A total of 522 piRNAs are expressed in corresponding tumour tissues, largely distinguishing tumour from non-malignant tissues in a cancer-type specific manner. Most expressed piRNAs mapped to known transcripts, contrary to “piRNA clusters” reported in germline cells. We showed that piRNA expression can delineate clinical features, such as histological subgroups, disease stages, and survival. piRNAs common to many cancer types might represent a core gene-set that facilitates cancer growth, while piRNAs unique to individual cancer types likely contribute to cancer-specific biology.

Small RNA-guided gene regulation represents a widely conserved mechanism across almost all living organisms<sup>1</sup>. Small RNA-mediated gene silencing typically involves a sequence recognition particle (i.e. small RNA) and a member of the Argonaute protein family, which is composed of Argonaute (Ago) and P-element-induced wimpy testis (PIWI) subfamilies<sup>2</sup>. Based largely on differences in biogenesis and structure, small RNAs can be broadly divided into three groups: small interfering RNAs (siRNAs), microRNAs (miRNAs) and PIWI-interacting RNA (piRNAs). These interact with either the AGO (siRNAs, miRNAs) or PIWI subfamily (piRNAs) to form gene regulatory RNA/protein complexes<sup>2,3</sup>.

The discovery of piRNAs affirmed a fundamental functional role for the small non-protein coding genome<sup>4–9</sup>. piRNAs are small (24–32 nucleotides), single-stranded non-coding RNAs, that have highly conserved functions across species, such as transposon silencing and stem cell maintenance in germline tissues<sup>6,10–12</sup>. Today, conservative estimates for the total number of piRNAs in the eukaryotic genome parallel those of protein-coding genes (~20,000), and largely exceeds the estimated 2,000 miRNA loci<sup>13,14</sup>. Most piRNAs are derived from long, single-stranded RNA precursors transcribed from distinct transposons referred to as “piRNA clusters”<sup>11,15,16</sup>; however, a small fraction of piRNAs are also encoded in intergenic noncoding transcripts as well as protein-coding genes, primarily in the 3′ untranslated regions (3′UTRs)<sup>6,7,10,17–19</sup>.

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

<sup>2</sup>Department of Medical Genetics, University of British Columbia, Vancouver, B. C. V6T 1Z3 Canada. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to V.D.M. (email: vmartinez@bccrc.ca)

Tissue of origin	TCGA code (for tumour types)	Number of normal samples	Number of tumour samples	Number of cases with survival information
Bladder	BLCA	19	261	189
Breast	BRCA	103	1043	909
Colon	COAD	1	421	323
Head and neck	HNSC	43	455	391
Kidney	KIRC	71	529	484
Lung	LUAD	46	497	422
	LUSC	45	467	359
Ovarian	OV	0	480	474
Prostate	PRAD	50	263	204
Stomach	STAD	38	320	265
Thyroid	THCA	59	499	485
Uterine Corpus	UCEC	33	517	471
<b>TOTAL</b>	<b>12</b>	<b>508</b>	<b>5,752</b>	<b>4,976</b>

**Table 1.** TCGA cases and tissue types studied.

In humans, mature piRNAs form complexes with one of the four PIWI proteins (PIWIL1/HIWI, PIWIL2/HILI, PIWIL3, and PIWIL4/HIWI2). Specificity of the piRNA/PIWI riboprotein complex is mediated by the piRNA sequence which targets the complex to sites of complementary DNA or messenger RNA (mRNA); the effects of which are likely mediated by recruited cofactors, in a tissue- and context-specific manner in different species and for different PIWI proteins (reviewed in<sup>15,20</sup>). The piRNA/PIWI complex-mediated transcriptional silencing of specific genomic loci complementary to the piRNA sequence occurs through recruitment of epigenetic machinery and establishment of repressive epigenetic marks<sup>8,9,21–28</sup>. However, piRNAs and PIWI proteins are also found in the cytoplasm, and evidence of piRNA involvement in the regulation of mRNA (i.e., post-transcriptional silencing) is emerging<sup>20</sup>. For example, piRNAs can target the piRNA/PIWI complex to mRNA sequences containing a 3' retrotransposon sequence<sup>20,29,30</sup>. PiRNAs can also bind mRNAs in *trans*, when derived from pseudogenes of target mRNAs, or in *cis*, when encoded within endogenous genes<sup>20,31</sup>. The mechanisms governing piRNA-mediated mRNA degradation in humans are not well understood, although they may involve both small RNA-dependent endonucleases (slicer) and slicer-independent mechanisms for degradation of target RNAs by PIWI proteins and piRNAs, and recruitment of proteins that mediate mRNA deadenylation<sup>20,32</sup>.

The functional role of piRNAs in somatic tissues — from stem cell maintenance, to memory-related synaptic plasticity, to whole body regeneration in diverse species — is a rapidly emerging field of research<sup>8,33–36</sup>. In human cancers, some components of the PIWI biogenesis machinery have been characterized<sup>24,37</sup>; but tumour specific expression of piRNAs are limited to either single gene studies<sup>38,39</sup>, or expression profiles of a few representative tumours<sup>40–42</sup> or cell lines<sup>42–45</sup> for a few types of cancers.

Given the abundant presence of piRNAs in our genomes, the functional conservation of piRNAs across species, their functions in development, and the important physiological roles of even the small number of piRNAs studied to date, further exploration of these molecules in a broad spectrum of normal and diseased somatic tissues is warranted in order to gain further insights into normal and disease-related biology<sup>46</sup>. Towards this aim, we systematically assessed piRNA transcriptomes in 6,260 human tissue samples derived from non-malignant and tumour tissues from multiple organs. We investigated: 1) which piRNAs are expressed in somatic tissues? 2) are they the same ones expressed in germline cells? 3) does piRNA expression follow a tissue-specific pattern? 4) are piRNA expression patterns altered in cancer, and in a tumour-type specific manner? and 5) are piRNA expression patterns related to clinical features important to specific cancer types?

## Results

**piRNAs are expressed across non-malignant human somatic tissues in a shared and tissue-specific manner.** To date, piRNA expression has been primarily described in germline tissues, thus the first goal of this study was to systematically assess piRNA transcriptome patterns across human somatic tissues (Table 1). Briefly, we evaluated the expression of 20,831 unique piRNAs (encoded by 39,545 genomic loci, see Supplementary Table 1) in normal tissues (n = 508) derived from 10 different anatomical sites: bladder, breast, colon, head and neck (different sites), kidney, lung, prostate, stomach, thyroid, and uterine corpus available from The Cancer Genome Atlas (TCGA) consortium. We detected 273 piRNAs expressed in somatic, non-malignant tissues (defined as Reads per kilobase per million

mapped reads (RPKM)  $\geq 1$  in at least 10% of samples). While the number of piRNAs we detect as somatically expressed is small (1.3%) relative to the total pool of piRNAs assessed, our results indicate that piRNAs are expressed in all somatic, non-malignant tissues analyzed. RPKM expression values for all piRNAs are listed in Supplementary Table 2.

We observed intriguing piRNA expression patterns across non-malignant somatic tissues (Fig. 1a). Overall, piRNA expression patterns appeared similar across all samples for both highly and lowly expressed piRNAs. We stress that only one non-malignant colon tissue was available for comparison and therefore colorectal cancer piRNA expression results are likely less robust than for other cancer types. However, when tissue-origin information was overlaid onto the sample dendrogram (denoted by colored bars on top of the heatmap in Fig. 1a), tissue-specific piRNA expression patterns for multiple organs were evident, particularly for thyroid and prostate tissues. Other tissues clustered together moderately (e.g., lung, head and neck or bladder) or were highly heterogeneous with samples clustering with many different tissue types (e.g., stomach and bladder). piRNAs highly expressed across all tissues could suggest a conserved general function for these piRNAs in somatic tissues, while subtle differences in piRNAs expressed at low or moderate levels may be capable of distinguishing tissues based on organ type, possibly indicating a tissue-specific role for these piRNAs, which would be consistent with the literature on somatic piRNAs<sup>8</sup>.

For thyroid—the most homogeneously clustered tissue type—as few as three piRNAs (FR069557, FR066510 and FR184567) were capable of distinguishing a high proportion of thyroid tissues from all other tissue types (Fig. 1b). Expression of FR069557 (the top ranked differentially expressed piRNA between thyroid and all other tissues analyzed) was significantly higher in thyroid compared to all other individual tissues ( $p$  value  $< 0.0001$ , one-way ANOVA, Fig. 1c). Tissue specific expression of piRNAs could implicate a specific functional role for such piRNAs in various somatic tissues.

**piRNA expression distinguishes non-malignant and tumour tissues.** To identify piRNAs that differentiate non-malignant and tumour tissues, we next analyzed piRNA expression in tumour samples ( $n = 5,752$ ) from 12 different tumour types: bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), head and neck squamous cell carcinoma (HNSC), kidney renal clear cell (KIRC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), ovarian serous cystadenocarcinoma (OV), prostate adenocarcinoma (PRAD), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uterine corpus (UCEC). Consistent with a putative role for piRNAs in promoting cell proliferation, we detected a considerably higher number of piRNAs expressed in tumours ( $n = 522$  piRNAs, encoded at 635 piRNA loci) relative to non-malignant tissues ( $n = 273$  piRNAs).

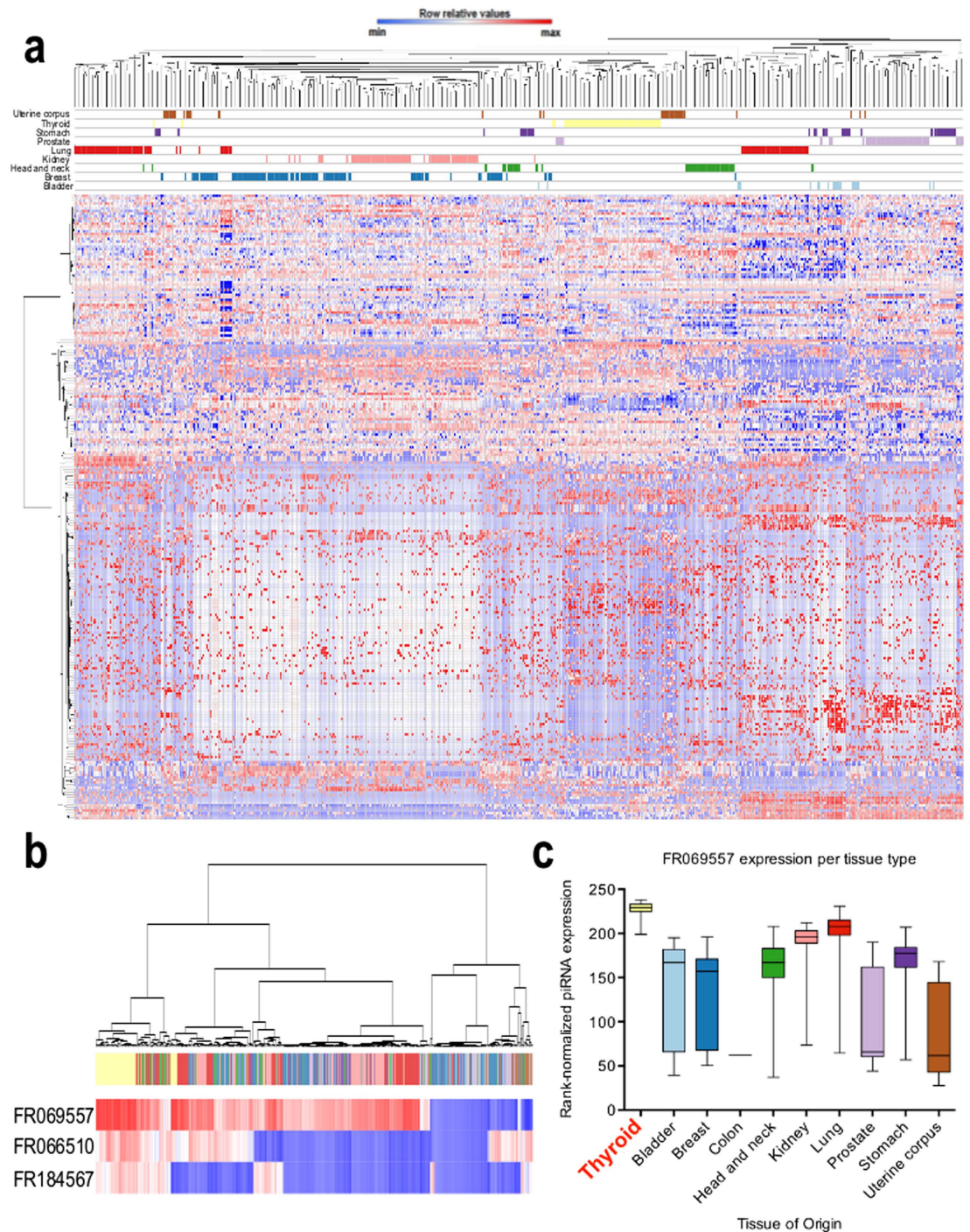
Globally, we noticed that piRNA expression levels were considerably lower in non-malignant tissue compared with tumour tissue, with the exceptions of BRCA and KIRC tumours, which showed levels comparable to those observed in non-malignant tissue (Fig. 2). Of the 522 expressed piRNAs, 324 were significantly differentially expressed between non-malignant and tumour tissues (permutation test, FDR-BH  $p$ -value  $< 0.01$ , Supplementary Table 3), most of which were overexpressed in tumours compared to non-malignant tissue, with 135 piRNAs exclusively expressed in tumours (i.e. not expressed in non-malignant tissues) (Supplementary Table 3).

We also observed intriguing patterns of size distribution amongst non-malignant and tumour tissues (Supplementary Figure 1). For non-malignant tissues, ~40% of expressed piRNAs were almost equally distributed between 29 and 30 nt length. PiRNA size distribution patterns in tumours were strikingly different displaying an enrichment of piRNAs 32 nt long (36–39% of all expressed piRNAs in each tissue), followed by 28 nt (18.5%–26.1%) and 24 nt piRNAs (13.05%–17.2%) — a pattern that was remarkably conserved across all tumour types. Remarkably, over 90% of the piRNA species for the most enriched size (32 nt) were shared amongst all tumour types assessed. While speculative, this strong enrichment could indicate selection of a common tumour-promoting functional role of these piRNAs, such as altered metabolic function or involvement in other cancer hallmarks.

Unsupervised hierarchical clustering of piRNA expression profiles revealed a remarkably distinct segregation of tumour and non-malignant samples within individual tissue types with the exception of PRAD samples (Fig. 3 and Supplementary Figures 2–11). These data suggest that distinct piRNA expression programs are associated with tumorigenesis and support the hypothesis that piRNAs may be functionally important to cancer biology<sup>9,46</sup>.

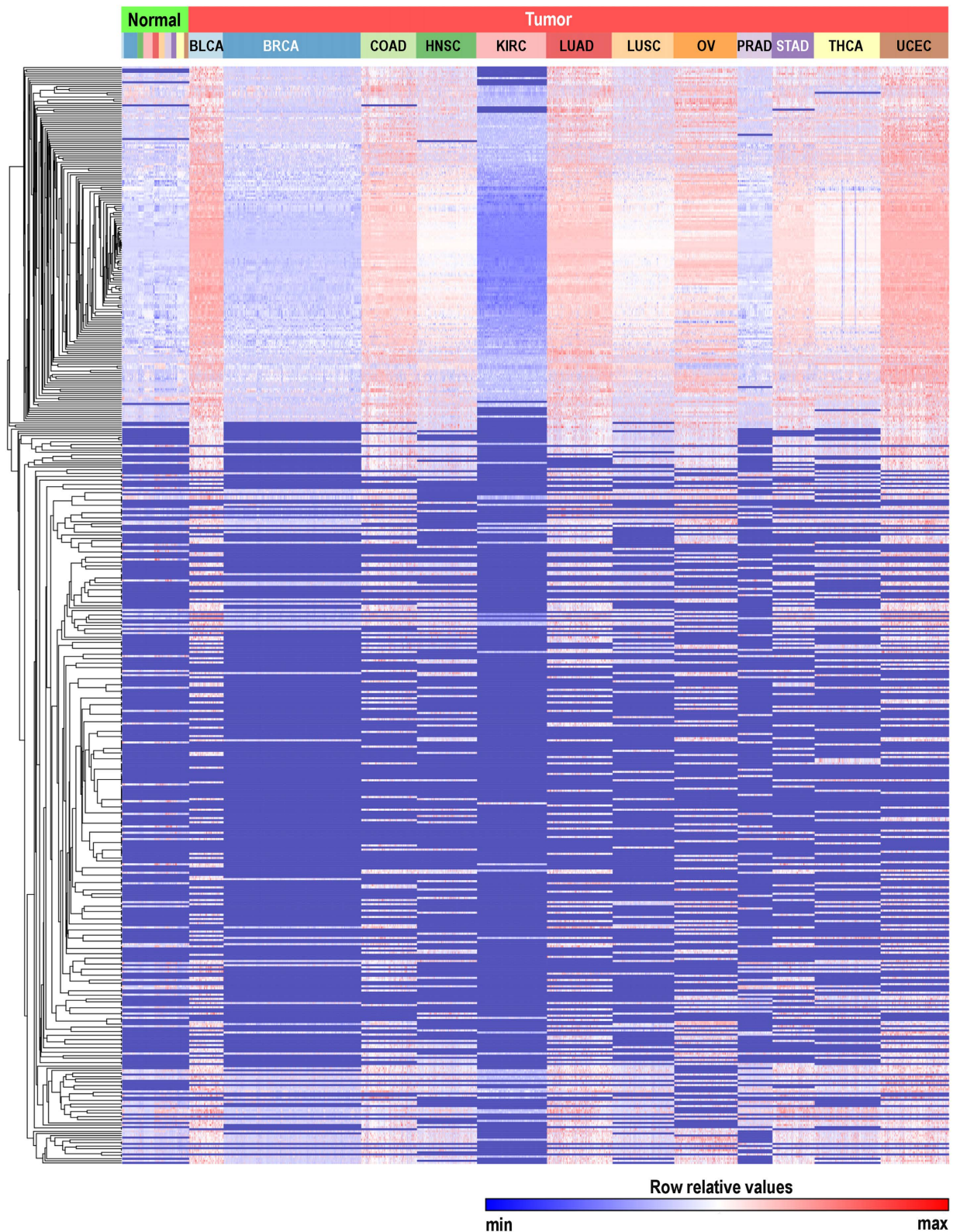
**piRNAs exhibit pan-cancer and tumour-type specific expression patterns.** We next focused our attention on piRNA expression patterns in malignant tissues. As evident in Fig. 4a, most piRNAs were commonly upregulated (relative to non-malignant tissues) across the majority of tumour types. One particularly interesting example we identified was the high level (with the exception of BRCA and KIRC) upregulation of piRNAs mapping to the mitochondrial genome (mt-piRNAs). While expression of some mt-piRNAs such as FR015567 were highly variable across tumour-types, with particularly low expression levels in COAD, THCA and UCEC, other mt-piRNAs such as FR043670, were highly expressed across almost all tumour types (Fig. 4b). The piRNAs observed here to be upregulated in a pan-cancer manner might represent a core set of genes that contribute to oncogenic properties of mt-piRNAs. Additionally, we assessed tumour piRNA expression data derived from independent cancer cohorts, including bladder



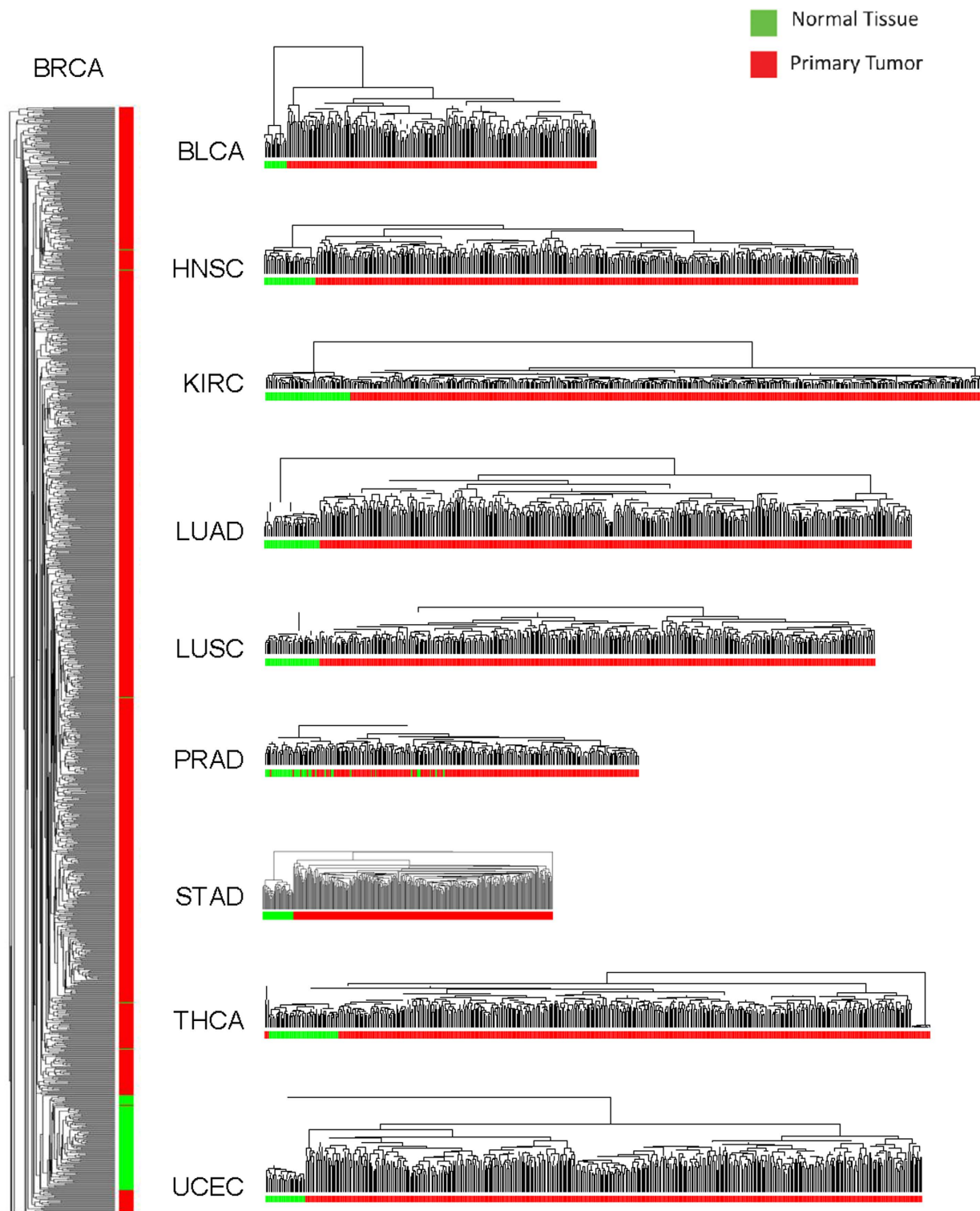


**Figure 1. Genome-wide distribution of piRNA expression in human non-malignant tissues.** Colour coding for somatic tissues noted in Fig. 1A applies to all panels (a–c). **a** Unsupervised hierarchical clustering of rank-normalized piRNA expression levels from 508 non-malignant tissue samples derived from 10 different organs: bladder, breast, colon, head and neck, kidney, lung, prostate, stomach, thyroid and uterine corpus. Tissue types are colour coded and separated by rows under the dendrogram. A total of 273 piRNAs were expressed in non-malignant tissues. Expression levels are coded from low (blue) to high (red) on a row-standardized scale, meaning that expression values (colours) are comparable both across rows. The single colon non-malignant sample analyzed is not shown in this heatmap, to avoid misinterpretation due to a low sample size **b** Top three differentially expressed piRNAs that can discriminate between thyroid and other tissues based on a comparative marker selection analysis<sup>73</sup>. Unsupervised hierarchical clustering was also performed to display the ability to distinguish thyroid (yellow) from other tissues. **c** Box-and-whiskers plots of normalized expression levels of FR069557 across different non-malignant tissue types.



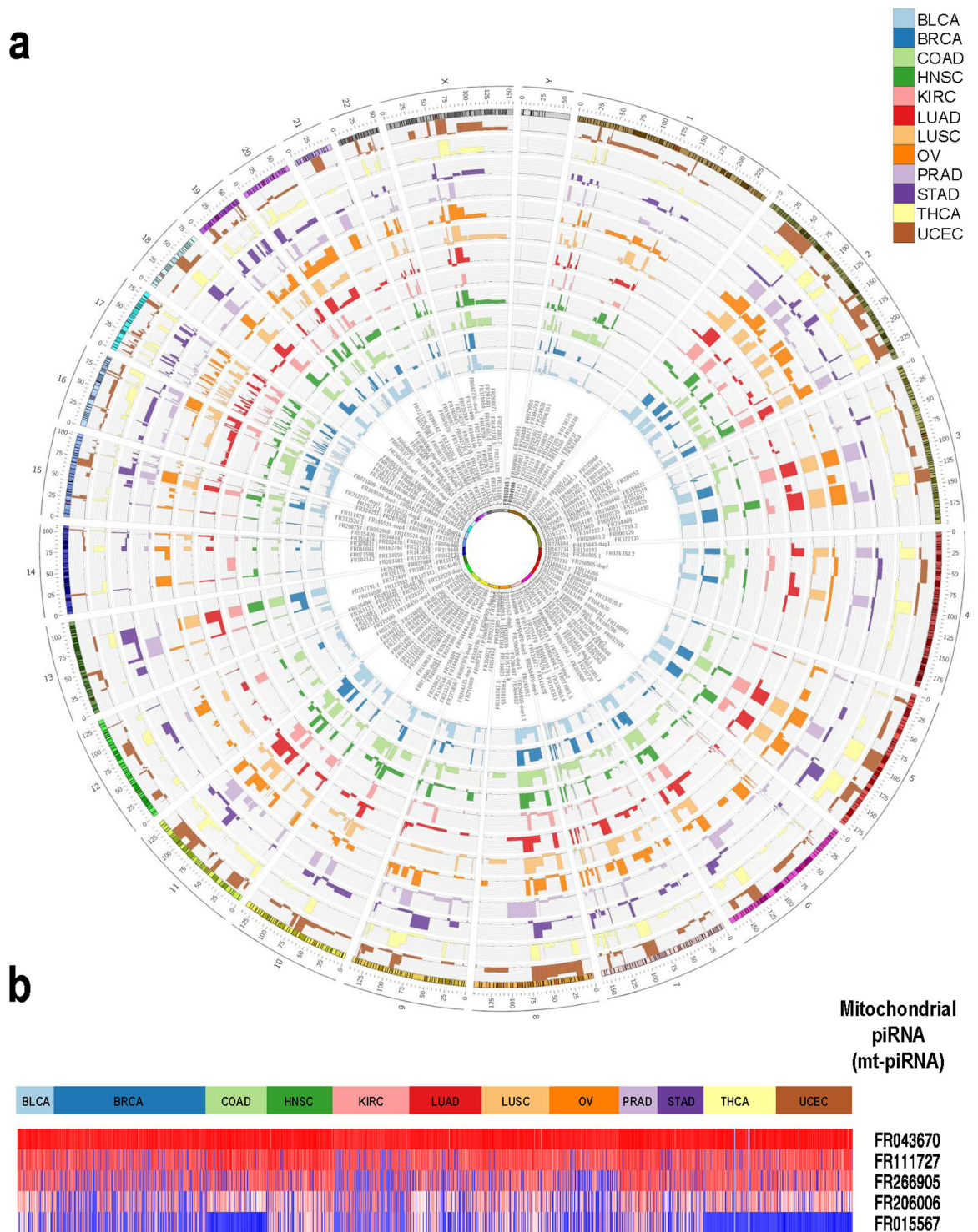


**Figure 2. piRNA expression in non-malignant and tumour tissues.** Expression patterns across all non-malignant (top row green,  $n = 508$ ) and tumour tissues (top row red,  $n = 5,752$ ). The second row indicates source of tissue. Each row in the heatmap denotes the rank-normalized RPKM levels of each of the piRNAs expressed across non-malignant and tumour tissues, in a row-standardized scale. PiRNA expression is clustered according to expression levels. Dark blue represents low expression levels, red denotes higher expression of a given piRNA.

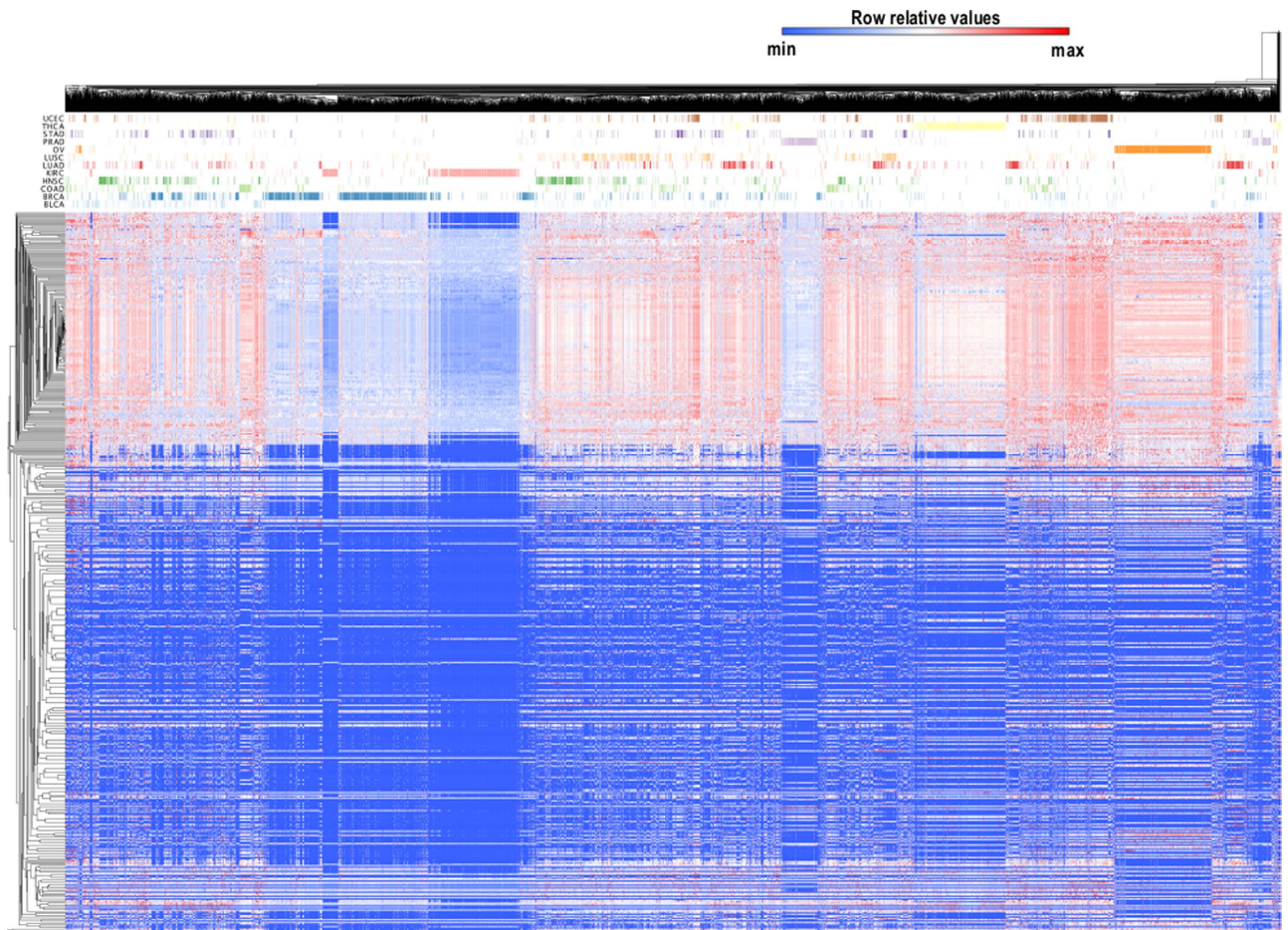


**Figure 3. piRNA expression patterns differentiate between normal and tumour tissues.** Unsupervised hierarchical clustering (average linkage, Euclidean distance) analysis on rank normalized expression values in non-malignant (NM, green) and tumour (T, red) samples derived from the following tissues: bladder (BLCA, NM = 19, T = 260), breast (BRCA, NM = 103, T = 1,043), head and neck (HNSC, NM = 43, T = 455), kidney renal clear cell (KIRC, NM = 71, T = 529), lung adenocarcinoma (LUAD, NM = 46, T = 497), lung squamous cell (LUSC, NM = 45, T = 469), prostate (PRAD, NM = 50, T = 263), stomach (STAD, NM = 38, T = 320), thyroid (THCA, NM = 59, T = 499), and uterine corpus (UCEC, NM = 33, T = 518). COAD and OV were excluded from this analysis because they had one and zero samples derived from normal tissue, respectively.





**Figure 4. Genome-wide expression of piRNA** a) Circular representation of genome-wide expression of piRNAs in 5,752 tumour samples derived from 12 different tissues. An ideogram of human chromosomes is shown in the outer ring, with chromosomes coloured in the same manner than the innermost ring, where the name of the expressed piRNAs are shown considering their chromosome location. Each concentric track between the two ideograms represents a histogram of average RPKM piRNA expression for each tumour type minus the expression of the same piRNA observed in non-malignant tissue (if expressed). Thus, the figure represents tumour-specific expression. The darker line on each track represents an average RPKM value = 0. Bars above this line represent that the expression of a given piRNA is higher in tumours, while bars below correspond to piRNA expressed at higher levels in non-malignant tissue compared to tumours. b) Detailed view of mt-piRNA expression across tumour types. Rank-normalized expression of the five most highly expressed piRNAs is shown.



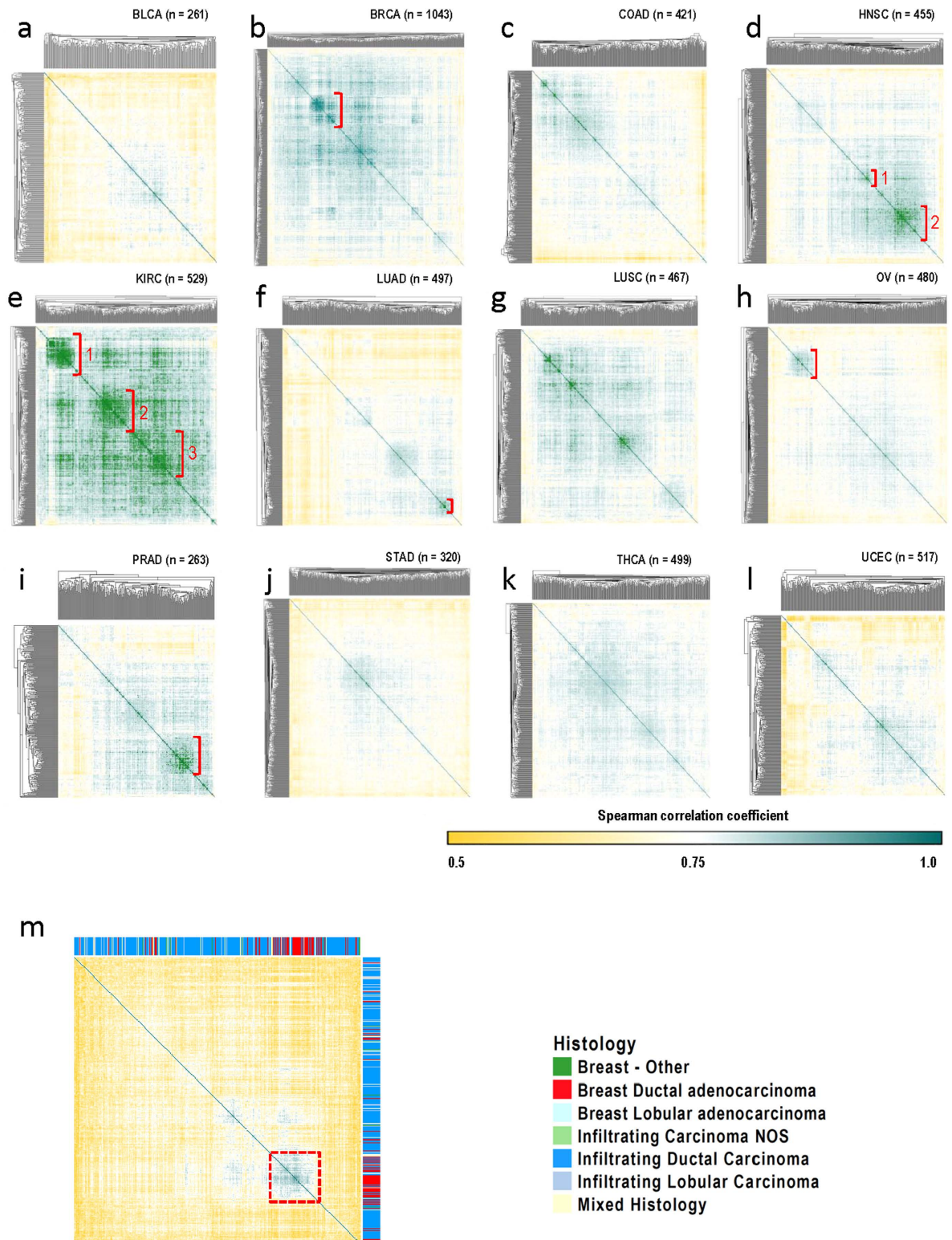
**Figure 5. Unsupervised clustering of piRNA expression in tumour tissues.** An unsupervised hierarchical clustering (average linkage, Euclidean distance) of piRNA expression profiles was performed on tumour samples derived from all analyzed tissues. Tumour types are colour-coded and separated by rows under the dendrogram. Rank-normalized expression values are expressed on a row-standardized scale and range from low (dark blue) to high (red) expression.

(GSE31616,  $n = 10$ ), breast (GSE29173 and GSE28884,  $n = 167$ ), colon (GSE63119,  $n = 50$ ), lung (GEO Accession numbers pending; adenocarcinoma subtype,  $n = 30$ , squamous cell carcinoma subtype,  $n = 30$ ) and stomach (GSE36968,  $n = 25$ ) tumours. This analysis revealed a high degree of concordance with normalized expression levels of the top 50 highest expressed piRNA for each of these tumour types (Supplementary Figure 12).

We next conducted unsupervised hierarchical clustering of all tumour piRNA expression profiles using pan-cancer standardized RPKM values (Fig. 5). We observed that while tumours generally clustered by tissue-of-origin, not all tumours of the same type conform to this pattern, suggesting heterogeneity exists at the level of piRNA expression within some tumour types. For example, PRAD and OV tumours, and to a lesser extent KIRC and THCA, formed discrete clusters (i.e. a low number of samples from other tumour types are observed in clusters dominated by these tumours). For KIRC tumours, distinct clustering appeared to be largely attributable to moderately expressed piRNAs, which were very highly expressed in other tumour types. Conversely, we noted that other tumour types clustered heterogeneously with tumours from other origins. For example, individual BLCA and STAD tumours were scattered among tumours of other origins, with no clusters containing a majority of these tumour samples.

In order to further investigate how similar or divergent piRNA expression patterns were within tumours from the same tissue-of-origin, we compared similarity of samples based on their piRNA expression pattern. To that end, we built similarity matrices for each tumour type, defined by Spearman correlation coefficients (Fig. 6a–l). This analysis revealed tumour sub-group piRNA expression patterns, especially in KIRC, HNSC, PRAD, LUAD, OV, and BRCA.





**Figure 6. Correlation of piRNA expression patterns across samples within the same tumour type. a–l)** Spearman correlation-based matrices for each tumour type. **a)** BLCA, **b)** BRCA, **c)** COAD, **d)** HNSC, **e)** KIRC, **f)** LUAD, **g)** LUSC, **h)** OV, **i)** PRAD, **j)** STAD, **k)** THCA, and **l)** UCEC. Darker green boxes indicate a higher correlation among samples, while yellow indicates a poorer correlation. Red brackets indicate clusters of highly related samples that were selected for correlative analyses with clinical variables. **m)** Spearman correlation-based matrix for 415 breast tumour samples with available histology information. A cluster of highly correlated samples (dashed red square) is significantly enriched ( $p$  value  $< 0.0001$ , Fisher Exact Test) for breast ductal adenocarcinoma histology.

**Clinically-relevant tumour subtypes exhibit different piRNA expression patterns.** We were next interested in assessing whether tumours corresponding to distinct clusters in our similarity matrices for individual tumour types were enriched for clinical features relevant to the corresponding cancer type (Supplementary Table 4). For KIRC, we assessed whether metastasis, nodal status, stage and presence of *Von Hippel-Lindau* (*VHL*) mutation—a common mutation in kidney cancer—were significantly different across the three distinct tumour clusters (Fig. 6e) using multivariate analysis of variance. Intriguingly, KIRC tumour stage was significantly different across the three major sub-clusters based on both univariate ( $p = 0.0346$ ) and multivariate ( $p = 0.03281$ ) analyses. Cluster 1 (in Fig. 6e had the highest proportion and lowest proportion of Stage I and Stage IV tumours, respectively). We next assessed tumour location, alcohol consumption, smoking-status and pack-years, stage and grade and for HNSC tumours, and found nodal status was significantly different (multivariate  $p$  value = 0.002, univariate  $p$  value = 0.0098) between distinct clusters indicated in Fig. 6d (higher fraction of node negative tumours in Cluster 2). For PRAD, we found that Gleason score—a grading system for prostate carcinoma, where higher scores are associated with aggressive growth patterns—was often low, with the majority (72%) of prostate tumours in the PRAD cohort having a Gleason score of 7 or less. The cluster identified in the PRAD similarity matrix (Fig. 6i) has a significantly higher proportion of cases with Gleason scores  $\geq 8$  (multivariate  $p$  value = 0.00752, univariate  $p$  value = 0.0344). We did not find any significant clinical features associated with tumour sub-group clusters evident in other similarity matrices such as those for LUAD and OV (Fig. 6f,h).

Since breast cancer represented the largest tumour set in our analysis, and is a cancer type with well-defined clinical subtypes, we were able to assess whether the highly similar group of tumours indicated by the red dashed box in Fig. 6m was associated with a specific histological subtype within the 415 BRCA tumours that had available histological data. Strikingly, the cluster of highly correlated tumours was significantly enriched for breast ductal adenocarcinoma ( $p$  value < 0.0001, Fisher Exact Test) (Fig. 6m). The piRNA expression signature distinguishing this subtype is listed in Supplementary Table 4. Taken together, it is evident that piRNA expression is associated with clinically relevant features of individual tumour types.

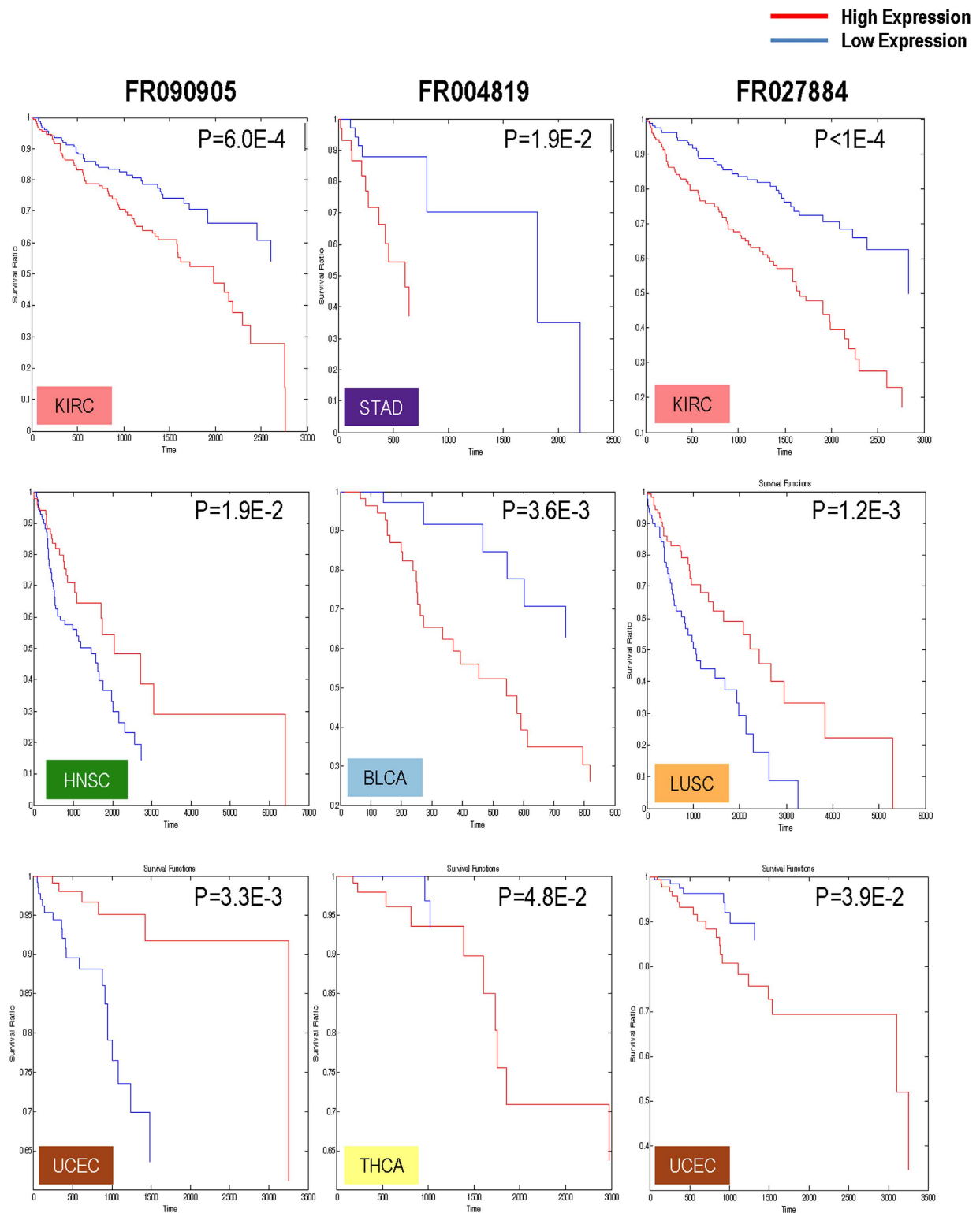
**Prognostic value of piRNA expression.** We next assessed the association between piRNA expression and cancer patient survival on a per cancer-type basis using a log-rank test (Fig. 7). We identified a subset of piRNAs whose expression was associated with survival in at least one tumour type (Supplementary Table 5). These piRNAs had either the same associations in each tumour type (e.g. high levels of FR004819 associated with poorer patient survival in STAD, BLCA, THCA), or had opposite associations in different tumour types. For example, high expression of FR090905 is associated with poor patient survival in KIRC, and improved survival in UCEC. Notably, FR090905 was significantly associated with survival in three different tumour types (KIRC, HNSC, and UCEC), and was nearly significant in LUSC ( $p$  value = 0.065, after false discovery rate correction for multiple testing), highlighting the potential importance of this piRNA in multiple tumour types.

Additionally, we assessed whether the prognostic significance of piRNAs we identified in breast cancer (the largest cancer cohort analyzed) could be replicated in an additional, although smaller cohort of 97 breast tumour samples with clinical follow-up information (GEO accession: GSE29173). Indeed, two piRNAs associated with survival in the TCGA breast cancer cohort (FR378984 and FR025321) validated in this external cohort (Supplementary Table 5).

**Genomic localization of somatically expressed piRNAs.** In contrast to the piRNAs previously described as expressed in germline cells, which predominantly map to piRNA clusters, we found that only 10.8% and 14.3% of the piRNAs expressed in normal somatic and tumour tissues mapped to known human piRNA clusters, respectively. However, our findings were in line with a recent somatic piRNA expression study that described only 14% of piRNAs expressed coordinately with liver regeneration stages mapped to known piRNA genomic clusters<sup>34</sup>. We also noted that 70.4% of expressed piRNAs mapped to known protein-coding and non-coding transcripts, with 209 piRNAs mapping to introns of known genes. Additionally, 11.8% of expressed piRNAs mapped within 100 base pairs up- or downstream of repetitive elements of which 53% were also situated within coding and non-coding transcripts. Of the repetitive element-associated piRNAs, 69.3% corresponded to short interspersed nuclear elements (SINES) and long interspersed nuclear elements (LINES) (Supplementary Table 4). Interestingly, the sequences of 17 piRNAs expressed in tumours overlapped with gene mutations described in the Catalogue of Somatic Mutations in Cancer (COSMIC) (Supplementary Table 6).

**A public piRNA expression atlas to support future piRNA studies.** To facilitate visualization, cross-tissue comparisons, and integration with other genomic features, we have made this information available through a custom UCSC genome browser track displaying normalized mean RPKM piRNA expression levels for these 6,260 samples summarized for each malignant or non-malignant tissue type (accessible at <http://goo.gl/cnMwSv>, Supplementary Figure 13). Users can also complement this information with expression matrices of RPKM values for each expressed piRNA (i.e.,  $\text{RPKM} \geq 1$ ,  $n = 522$ ) for any individual sample (Supplementary Table 2).





**Figure 7. piRNA-based survival analysis.** We assessed the association of piRNA expression with cancer patient survival using a logrank test. We considered the 522 piRNAs exclusively expressed in tumours in 4,976 samples where survival information was available (12 tumour types). Survival associations were only calculated for piRNAs that were detectably expressed in at least two-thirds of the samples for each particular cancer type. Patient outcome was compared in patients with piRNA expression levels ranking in the top and bottom tertiles of expression.

## Discussion

Investigation into the role of piRNAs in the establishment of genetic, epigenetic, and transcriptional patterns—and their involvement in key processes in germline and more recently, somatic cells—is a rapidly growing field of research<sup>8,34,39,40,47,48</sup>. While expression patterns of other small non-coding RNAs (e.g. miRNAs), have been extensively described in a tissue-specific context for multiple eukaryotic organisms and almost all cancer types<sup>49</sup> in humans, malignant and non-malignant piRNA expression patterns are largely uncharted. Thus, we assessed piRNA transcriptomes in over 6,000 normal and tumour tissues from 11 organ sites. Our findings revealed that, 1) piRNAs are expressed in all malignant and non-malignant human somatic tissues analyzed, 2) piRNA expression patterns are specific to malignancies and their clinical groupings, 3) some piRNAs display pan-cancer expression patterns, such as the overexpression of mitochondrial piRNAs, and 4) subgroup-specific piRNA expression patterns from tumours of the same organ exist and correlate with key clinical features relevant to each tumour type. Moreover, we found a subset of cancer-specific piRNAs associated with patient survival either across multiple malignancies or specifically in one tumour type.

Relative to the large number of piRNAs encoded in the genome, only a small proportion of the piRNAs (1.3%) were expressed in non-malignant human somatic tissues. While the number of total miRNA species expressed at any level in somatic tissues is much higher compared to piRNAs, the expression patterns of piRNAs in somatic tissue resembles those of miRNAs, where a small number of unique species represent most of the total expression<sup>49</sup>. It is also possible that, as with other non-coding molecules (e.g., long non-coding RNAs), low expression levels may be sufficient to exert profound biological effects<sup>50</sup>. While very lowly expressed piRNAs (i.e.,  $\leq 1$  RPKM) were filtered from our analysis, the biological importance of these molecules may be elucidated by further interrogation using deep sequencing methods and functional assessment in future studies.

The number of piRNAs expressed in somatic tissue is considerably lower than observed in germline tissues. This is expected, since piRNAs and PIWI proteins play critical roles in germline development and gametogenesis, including germline determination, stem cell maintenance, meiosis, spermatogenesis, and transposon silencing<sup>51</sup>. In the original piRNA discovery paper, Girard *et al.* sequenced candidate piRNA populations from human testes, obtaining 52,099 sequences<sup>7</sup>. Another more recent study that analyzed human testes of 3 individuals obtained 14,608,234 reads, of which 20,121 mapped to piRNAs. The top 10 abundant piRNAs represent 61% of total piRNAs, reinforcing that a large fraction of expressed piRNAs are expressed at low levels<sup>52</sup>.

Most piRNAs are derived from long, single-stranded RNA precursors transcribed from specific genomic loci known as “piRNA clusters”<sup>11,15–18</sup>. In contrast to this pattern of expression, we observed that only 10.8% of the detected piRNAs expressed in non-malignant somatic tissues mapped to known human piRNA clusters. Most expressed piRNAs (70.4%) mapped to known human transcripts, while 11.8% were located within 100 base pairs of repetitive sequences (half of these located within coding and non-coding transcripts). This pattern has been previously observed<sup>7</sup>, and raises the possibility that transcripts hosting piRNAs, particularly those containing SINE or LINE repeats, might either regulate or be regulated by piRNA expression in human tissues.

Overall, clustering of non-malignant tissues based on piRNA expression revealed a high concordance in piRNA expression levels, with most piRNAs being either highly or lowly expressed across all tissues. However, some tissues, such as prostate and thyroid, displayed surprisingly tissue-type specific piRNA expression patterns (Fig. 1). Other tissues, such as stomach and bladder, did not form discrete clusters. These findings highlight a tissue-specific expression pattern, potentially reflecting tissue-specific functions, similar to that found for other small non-coding RNAs<sup>8</sup>. These patterns also highlight the importance of defining baseline tissue-specific expression levels for the interpretation of piRNA expression in tumours. It is noted that, since all non-malignant tissues were derived from patients with cancer, it is possible that tissue-specific patterns observed in this study may deviate to some degree from that of true normal.

Although tumours expressed twice as many piRNAs as non-malignant tissues, the overall number expressed similarly represented a small fraction of the assessed piRNAs (3.05% out of 20,831). This is consistent with what has been previously described for small cohorts of tumours and some cancer cell lines<sup>41,43–45</sup>. Also, similar to that observed in non-malignant tissues, a low percentage (14.3%) of piRNAs expressed in tumour tissues mapped to known human piRNA clusters. Since the majority of the expressed piRNA loci mapped outside of known piRNA genomic clusters in both malignant and non-malignant tissues, piRNA clusters may be predominantly silenced in somatic tissues.

Remarkably, with the exception of prostate adenocarcinoma, tumour-specific piRNA expression patterns appeared distinct from non-malignant tissues (Figs. 2 and 3). The imperfect tumour-normal clustering patterns observed for prostate tumours might be explained, at least in part, by the vast proportion of tumours with low Gleason scores. Interestingly, one of the stark differences found between tumours and non-malignant tissues was the high expression of piRNAs mapping to mitochondrial DNA in tumour tissues (Fig. 4), which is in line with the fact that mt-piRNA precursors exhibit high expression levels in cancer cell types and tumours<sup>53</sup>. However, we cannot rule out the possibility that increased expression of mt-piRNAs may be related to increased mitochondrial copy number in tumours<sup>54</sup>. Altered cellular metabolism in proliferating cells is a well documented phenomenon in cancer, thus it would be interesting to explore the potential metabolic role of these piRNAs in tumour cells<sup>55–58</sup>.



Although some commonalities in piRNA expression were evident across all cancer types, piRNA expression levels were capable of clustering samples by tumour-type for some tissues (Fig. 5). Tissue-specific piRNA clustering was especially evident for PRAD, OV, KIRC, and THCA, whereas other tumour types were more heterogeneous, such as BLCA and STAD. A deeper investigation of similarities in piRNA expression within individual tissues, intriguingly revealed intra-tumour piRNA sub-groups. Sub-groups within KIRC, HNSC, PRAD, and BRCA (Fig. 6) were associated with key clinical features relevant to these cancer types, including tumour stage, nodal status, Gleason score, and histological subtype, respectively. Moreover, we discovered that a subset of piRNAs were significantly associated with survival in at least one tumour type (Fig. 7; Supplemental Table 4). This indicates that piRNAs could represent potential prognostic markers similar to miRNAs, and should be evaluated on a tissue-specific basis in future studies.

Collectively, our analysis of the piRNA transcriptome provides: 1) evidence supporting a potentially tissue-specific role for piRNAs in both tumour and non-malignant somatic tissues that may be clinically relevant, 2) a curated list of piRNAs for prioritization of future piRNA functional interrogation, and 3) a human atlas of 6,260 piRNA transcriptome profiles summarized according to tumour and non-malignant tissue types. As interest in this field grows and sequence detection capacities and technologies improve, new piRNA species and reference resources are sure to rapidly expand. Thus our study represents a first-generation piRNA somatic tumour and non-malignant tissue resource based on expression of piRNAs known to date, and may serve as a useful foundation for comparative studies in future works. Further mechanistic exploration of piRNAs and associated proteins will surely yield important insight into the specificity and establishment of epigenetic patterns, transcriptional regulation, and functional involvement in normal development and human disease. We acknowledge that a major caveat to our study is that it is correlative; however, we hope that our findings may be applied to the prioritization of interesting piRNAs to commence important functional work.

## Methods

**TCGA data acquisition and small-RNA sequencing library processing.** A total of 6,260 samples (508 non-malignant and 5,752 tumour) processed by The Cancer Genome Atlas (TCGA) Research Network (<http://cancergenome.nih.gov/>) were selected for this study. Samples spanned 10 non-malignant tissues and 12 tumour types (lung tumours are divided into LUAD and LUSC, and OV cancer lacked normal samples). The criteria for tissue inclusion were: i) availability for public use as April 2014, ii) carcinoma histology, and iii) at least 100 tumour samples. Small RNA sequencing libraries were generated at Canada's Michael Smith Genome Sciences Centre and sequenced using the Illumina Genome Analyzer and HiSeq2000 platforms. Reads were subjected to quality control in order to exclude non-biological reads (trimming of 3' non-biological adapter sequences) before alignment. A full description of a per-tissue processing can be found in the Synapse archive ([www.synapse.org](http://www.synapse.org)) within the following accessions: syn1461149 (BLCA), syn1461151 (BRCA), syn1461155 (COAD), syn1461156 (HNSC), syn1461159 (KIRC), syn1461166 (LUAD), syn1461168 (LUSC), syn1461171 (OV), syn1461173 (PRAD), syn1461177 (STAD), syn1461178 (THCA), syn1461180 (UCEC). Each library was sequenced to an average depth of 7.42 (range: 0.63–42.10) million reads for tumors and 10.62 (range: 1.42–49.05) for non-malignant tissue. This sequencing strategy has been shown to be generally sufficient to identify moderate-to-low-abundance small RNAs including those exhibiting modest expression differences between samples<sup>59</sup>. We have also calculated the sequencing depth on a per-tissue basis in the revised manuscript (Table 2). Filtered reads were aligned to human genome (GRCh37/hg19) using the Burrows-Wheeler Aligner (BWA) algorithm<sup>60</sup>. Sequences aligned to the human genome (.bam files) were downloaded from the Cancer Genomics Hub ([www.cghub.ucsc.edu/](http://www.cghub.ucsc.edu/)), after receiving authorization from the data access committee (dbgap Project ID: 6208).

**Pre-processing, quantification and selection of piRNAs.** In order to select for reads corresponding to piRNAs, we extracted raw, unaligned read sequences and quality scores (FASTQ format) from the BAM file downloaded from CGHub using the “SamToFastq” tool in the Picard analysis package (<http://picard.sourceforge.net>). Resulting FASTQ files were trimmed based on quality (Phred quality score  $\geq 20$ ) and size (read length  $\geq 21$  bp), in order to enrich for high quality reads corresponding to piRNAs. Samples were realigned to human genome (hg19) using the the Spliced Transcripts Alignment to a Reference (STAR) software, with One mismatch was allowed during re-mapping of sequence reads to the human genome<sup>61</sup> and a custom piRNA reference transcriptome using genomic coordinates for piRNA sequences obtained from the functional RNA database (v3.4)<sup>62,63</sup>. This reference transcriptome considers widely accepted piRNA sequence features, such as sequence bias for Uracil in the first position and presence of PIWI binding domains<sup>7,62</sup>, although known biases for 2-o-methylation in the 3' end were not considered. miRNA annotation references, miRBase mature microRNAs version 20, and miRBase precursor microRNAs version 20, had minimal overlap with regions covered by the piRNA transcriptome reference (i.e. only 0.5% and 0.31% overlap with bases included in the piRNA transcriptome reference, respectively). The locations of piRNA clusters were obtained from piRNABank<sup>64</sup>.

PiRNA expression quantification was performed using PartekFlow™ (Partek Inc., MO, USA). Reads were assigned to piRNA genomic loci based on the Expectation/Maximization (E/M) algorithm<sup>65</sup>. Partek Genome Suite (PGS) was used to further process and filter quantified files. Statistical features for raw

TISSUE	TUMOURS		NON-MALIGNANT	
	Average number (million reads)	Range (million reads)	Average number (million reads)	Range (million reads)
Bladder (BLCA)	9.53	2.31–37.16	22.04	3.3–49.05
Breast (BRCA)	5.62	0.96–33.52	5.45	1.42–25.56
Colon (COAD)	6.95	0.63–42.1	6.07*	6.07–6.07
Head and neck (HNSC)	8.16	1.47–30.59	10.56	3.86–19.19
Kidney (KIRC)	4.6	1.24–25.67	4.76	1.84–11.59
Lung (LUAD)	8.85	0.94–26.36	10.19	2.4–19.19
Lung (LUSC)	6.29	1.31–17.88	13.17	5.35–23.33
Ovarian (OV)	7.94	1.37–25.32	N/A	N/A
Prostate (PRAD)	10.13	2.53–36.41	11.54	4.72–31.07
Stomach (STAD)	8.32	1.28–34.98	12.81	4.13–39.2
Thyroid (THCA)	9.39	2.09–34.4	12.31	4.02–21.17
Uterine corpus (UCEC)	7.77	1.07–36.88	23.2	6.93–40.41

**Table 2.** Sequencing depth per tissue (in millions of reads). \*Only 1 library was available for this tissue

reads are shown in Supplementary Table 7. Reads per kilobase of exon model per million mapped reads (RPKM) was used to scale read count<sup>66</sup>. Following a similar criterion to that used by GENCODE database for non-coding RNAseq-based analysis<sup>50</sup>, we chose a conservative threshold of RPKM  $\geq 1$  for considering a piRNA as “expressed”. Additionally, we applied a frequency filter to define piRNAs as “expressed”, selecting only piRNAs expressed in at least 10% of tumour samples from the same tissue of origin.

**Visualization.** Expression matrices were created using the integrative genome viewer software (IGV)<sup>67,68</sup> and GENE-E software (Available at <http://www.broadinstitute.org/cancer/software/GENE-E/>). Circular representation of genome-wide expression in non-malignant stomach tissue and GA tumours was performed using the CircosPlot<sup>69</sup>.

**Statistical analysis.** In order to perform comparisons between expressions levels, we applied the rank-normalization algorithm implemented in GenePattern, where each RPKM expression value was replaced by its rank within each individual sample. New rank-based values range from 1 (lowest RPKM expression value) to N (highest expression RPKM value, where N is the number of piRNA expressed in a given sample). If two or more piRNAs have the same expression value, their values are replaced by the average of their ranked positions. Unsupervised hierarchical clustering was performed using GENE-E software. Differential expression analyses were performed using the “Comparative Marker Selection” module implemented in Gene Pattern<sup>70,71</sup>. Briefly, this procedure determines the piRNAs that are most closely correlated with either a phenotype, or the significance of that correlation. Differential expression between the non-malignant and tumour samples was assessed through a signal-to-noise (SNR) test, and the significance (nominal P-value) was estimated using a permutation test (10,000 permutations). Nominal P-values were corrected using the Benjamini and Hochberg procedure False Discovery Rate (FDR-BH)<sup>72</sup>. The fold change parameter was calculated by dividing the mean expression value of tumours by the mean expression value in non-malignant tissues. Spearman correlation values of 0.75 or higher were used to call clusters in Fig. 6 heatmaps.

**Survival analysis.** Survival information available for 3,433 cases (Table 1) was pulled from The Cancer Genome Atlas – Data Portal. Information concerning “days to death”, “days to last follow-up” and “vital status” was used for calculations. We assessed whether expression of piRNAs was associated with cancer patient survival using a log-rank test in MATLAB (MATLAB R2010a, The MathWorks Inc., Natick, MA, 2000), considering the 522 piRNAs expressed in tumours. Outcome was compared in patients with piRNA expression levels ranking in the top and bottom tertiles. Log-rank survival analysis was performed on piRNAs that were detectably expressed in at least two-thirds of the samples for each particular cancer type. p-values  $< 0.05$  were considered significant. Kaplan-Meier curves were used to compare survival distribution across patients with high or low expression of a given piRNA.

## References

- van Wolfswinkel, J. C. & Ketting, R. F. The role of small non-coding RNAs in genome stability and chromatin organization. *J. Cell Sci.* **123**, 1825–1839, doi:10.1242/jcs.061713 (2010).



2. Gunawardane, L. S. *et al.* A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **315**, 1587–1590, doi:10.1126/science.1140494 (2007).
3. Wilson, R. C. & Doudna, J. A. Molecular mechanisms of RNA interference. *Annu. Rev. Biophys.* **42**, 217–239, doi:10.1146/annurev-biophys-083012-130404 (2013).
4. Grivna, S. T., Beyret, E., Wang, Z. & Lin, H. A novel class of small RNAs in mouse spermatogenic cells. *Genes. Dev.* **20**, 1709–1714, doi:10.1101/gad.1434406 (2006).
5. Watanabe, T. *et al.* Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes. Dev.* **20**, 1732–1743, doi:10.1101/gad.1425706 (2006).
6. Aravin, A. *et al.* A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203–207, doi:10.1038/nature04916 (2006).
7. Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199–202, doi:10.1038/nature04917 (2006).
8. Ross, R. J., Weiner, M. M. & Lin, H. PIWI proteins and PIWI-interacting RNAs in the soma. *Nature* **505**, 353–359, doi:10.1038/nature12987 (2014).
9. Esteller, M. Non-coding RNAs in human disease. *Nat. Rev. Genet.* **12**, 861–874, doi:10.1038/nrg3074 (2011).
10. Lau, N. C. *et al.* Characterization of the piRNA complex from rat testes. *Science* **313**, 363–367, doi:10.1126/science.1130164 (2006).
11. Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**, 1089–1103, doi:10.1016/j.cell.2007.01.043 (2007).
12. Hartig, J. V., Tomari, Y. & Forstemann, K. piRNAs—the ancient hunters of genome invaders. *Genes. Dev.* **21**, 1707–1713, doi:10.1101/gad.1567007 (2007).
13. Peng, J. C. & Lin, H. Beyond transposons: the epigenetic and somatic functions of the Piwi-piRNA mechanism. *Curr. Opin. Cell Biol.* **25**, 190–194, doi:10.1016/j.ccb.2013.01.010 (2013).
14. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, D68–73, doi:10.1093/nar/gkt1181 (2014).
15. Luteijn, M. J. & Ketting, R. F. PIWI-interacting RNAs: from generation to transgenerational epigenetics. *Nat. Rev. Genet.* **14**, 523–534, doi:10.1038/nrg3495 (2013).
16. Li, X. Z., Roy, C. K., Moore, M. J. & Zamore, P. D. Defining piRNA primary transcripts. *Cell Cycle* **12**, 1657–1658, doi:10.4161/cc.24989 (2013).
17. Robine, N. *et al.* A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Curr. Biol.* **19**, 2066–2076, doi:10.1016/j.cub.2009.11.064 (2009).
18. Saito, K. *et al.* A regulatory circuit for piwi by the large Maf gene traffic jam in *Drosophila*. *Nature* **461**, 1296–1299, doi:10.1038/nature08501 (2009).
19. Grivna, S. T., Pyhtila, B. & Lin, H. MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. *Proc. Natl. Acad. Sci. USA* **103**, 13415–13420, doi:10.1073/pnas.0605506103 (2006).
20. Watanabe, T. & Lin, H. Posttranscriptional Regulation of Gene Expression by Piwi Proteins and piRNAs. *Mol. Cell* **56**, 18–27, doi:10.1016/j.molcel.2014.09.012 (2014).
21. Aravin, A. A. *et al.* A piRNA pathway primed by individual transposons is linked to *de novo* DNA methylation in mice. *Mol. Cell* **31**, 785–799, doi:10.1016/j.molcel.2008.09.003 (2008).
22. Kuramochi-Miyagawa, S. *et al.* DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes. Dev.* **22**, 908–917, doi:10.1101/gad.1640708 (2008).
23. Lee, J. H. *et al.* Stem-cell protein Piwil2 is widely expressed in tumors and inhibits apoptosis through activation of Stat3/Bcl-XL pathway. *Hum. Mol. Genet.* **15**, 201–211, doi:10.1093/hmg/ddi430 (2006).
24. Qiao, D., Zeeman, A. M., Deng, W., Looijenga, L. H. & Lin, H. Molecular characterization of hiwi, a human member of the piwi gene family whose overexpression is correlated to seminomas. *Oncogene* **21**, 3988–3999, doi:10.1038/sj.onc.1205505 (2002).
25. Sharma, A. K. *et al.* Human CD34(+) stem cells express the hiwi gene, a human homologue of the *Drosophila* gene piwi. *Blood* **97**, 426–434 (2001).
26. Ishizu, H., Siomi, H. & Siomi, M. C. Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. *Genes. Dev.* **26**, 2361–2373, doi:10.1101/gad.203786.112 (2012).
27. Aravin, A. A., Hannon, G. J. & Brennecke, J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318**, 761–764, doi:10.1126/science.1146484 (2007).
28. Malone, C. D. & Hannon, G. J. Small RNAs as guardians of the genome. *Cell* **136**, 656–668, doi:10.1016/j.cell.2009.01.045 (2009).
29. Lim, A. K. *et al.* The nuage mediates retrotransposon silencing in mouse primordial ovarian follicles. *Development* **140**, 3819–3825, doi:10.1242/dev.099184 (2013).
30. Rouget, C. *et al.* Maternal mRNA deadenylation and decay by the piRNA pathway in the early *Drosophila* embryo. *Nature* **467**, 1128–1132, doi:10.1038/nature09465 (2010).
31. Kotelnikov, R. N. *et al.* Peculiarities of piRNA-mediated post-transcriptional silencing of Stellate repeats in testes of *Drosophila melanogaster*. *Nucleic Acids Res.* **37**, 3254–3263, doi:10.1093/nar/gkp167 (2009).
32. Darricarrere, N., Liu, N., Watanabe, T. & Lin, H. Function of Piwi, a nuclear Piwi/Argonaute protein, is independent of its slicer activity. *Proc. Natl. Acad. Sci. USA* **110**, 1297–1302, doi:10.1073/pnas.1213283110 (2013).
33. Ortogero, N. *et al.* A Novel Class of Somatic Small RNAs Similar to Germ Cell Pachytene PIWI-interacting Small RNAs. *J. Biol. Chem.*, doi:10.1074/jbc.M114.613232 (2014).
34. Rizzo, F. *et al.* Timed regulation of P-element-induced wimpy testis-interacting RNA expression during rat liver regeneration. *Hepatology* **60**, 798–806, doi:10.1002/hep.27267 (2014).
35. Yan, Z. *et al.* Widespread expression of piRNA-like molecules in somatic tissues. *Nucleic Acids Res.* **39**, 6596–6607, doi:10.1093/nar/gkr298 (2011).
36. Lee, E. J. *et al.* Identification of piRNAs in the central nervous system. *RNA* **17**, 1090–1099, doi:10.1261/rna.2565011 (2011).
37. Lim, S. L. *et al.* Overexpression of piRNA pathway genes in epithelial ovarian cancer. *PLoS One* **9**, e99687, doi:10.1371/journal.pone.0099687 (2014).
38. Cui, L. *et al.* Detection of circulating tumor cells in peripheral blood from patients with gastric cancer using piRNAs as markers. *Clin. Biochem.* **44**, 1050–1057, doi:10.1016/j.clinbiochem.2011.06.004 (2011).
39. Cheng, J. *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.* **315**, 12–17, doi:10.1016/j.canlet.2011.10.004 (2012).
40. Cheng, J. *et al.* piRNA, the new non-coding RNA, is aberrantly expressed in human cancer cells. *Clin. Chim. Acta.* **412**, 1621–1625, doi:10.1016/j.cca.2011.05.015 (2011).
41. Chu, H. *et al.* Identification of novel piRNAs in bladder cancer. *Cancer Lett.*, doi:10.1016/j.canlet.2014.10.004 (2014).
42. Huang, G. *et al.* Altered expression of piRNAs and their relation with clinicopathologic features of breast cancer. *Clin. Transl. Oncol.* **15**, 563–568, doi:10.1007/s12094-012-0966-0 (2013).

43. Law, P. T. *et al.* Deep sequencing of small RNA transcriptome reveals novel non-coding RNAs in hepatocellular carcinoma. *J. Hepatol.* **58**, 1165–1173, doi:10.1016/j.jhep.2013.01.032 (2013).
44. Hashim, A. *et al.* RNA sequencing identifies specific PIWI-interacting small non-coding RNA expression patterns in breast cancer. *Oncotarget* **5**, 9901–9910 (2014).
45. Lu, Y. *et al.* Identification of piRNAs in HeLa cells by massive parallel sequencing. *BMB Rep* **43**, 635–641, doi:10.5483/BMBRep.2010.43.9.635 (2010).
46. Mei, Y., Clark, D. & Mao, L. Novel dimensions of piRNAs in cancer. *Cancer Lett.* **336**, 46–52, doi:10.1016/j.canlet.2013.04.008 (2013).
47. Yan, H. *et al.* piRNA-823 contributes to tumorigenesis by regulating *de novo* DNA methylation and angiogenesis in multiple myeloma. *Leukemia*, doi:10.1038/leu.2014.135 (2014).
48. Rajasethupathy, P. *et al.* A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. *Cell* **149**, 693–707, doi:10.1016/j.cell.2012.02.057 (2012).
49. Hamilton, M. P. *et al.* Identification of a pan-cancer oncogenic microRNA superfamily anchored by a central core seed motif. *Nat. Commun.* **4**, 2730, doi:10.1038/ncomms3730 (2013).
50. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789, doi:10.1101/gr.132159.111 (2012).
51. Thomson, T. & Lin, H. The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. *Annu. Rev. Cell Dev. Biol.* **25**, 355–376, doi:10.1146/annurev.cellbio.24.110707.175327 (2009).
52. Yang, Q. *et al.* MicroRNA and piRNA profiles in normal human testis detected by next generation sequencing. *PLoS One* **8**, e66809, doi:10.1371/journal.pone.0066809 (2013).
53. Kwon, C. *et al.* Detection of PIWI and piRNAs in the mitochondria of mammalian cancer cells. *Biochem. Biophys. Res. Commun.* **446**, 218–223, doi:10.1016/j.bbrc.2014.02.112 (2014).
54. Yu, M. Generation, function and diagnostic value of mitochondrial DNA copy number alterations in human cancers. *Life Sci.* **89**, 65–71, doi:10.1016/j.lfs.2011.05.010 (2011).
55. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674, doi:10.1016/j.cell.2011.02.013 (2011).
56. White, E. Exploiting the bad eating habits of Ras-driven cancers. *Genes Dev.* **27**, 2065–2071, doi:10.1101/gad.228122.113 (2013).
57. Shaw, R. J. & Cantley, L. C. Decoding key nodes in the metabolism of cancer cells: sugar & spice and all things nice. *F1000 Biol. Rep.* **4**, 2, doi:10.3410/B4-2 (2012).
58. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–1033, doi:10.1126/science.1160809 (2009).
59. Morin, R. D. *et al.* Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.* **18**, 610–621, doi:10.1101/gr.7179508 (2008).
60. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760, doi:10.1093/bioinformatics/btp324 (2009).
61. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21, doi:10.1093/bioinformatics/bts635 (2013).
62. Mituyama, T. *et al.* The Functional RNA Database 3.0: databases to support mining and annotation of functional RNAs. *Nucleic Acids Res.* **37**, D89–92, doi:10.1093/nar/gkn805 (2009).
63. Kin, T. *et al.* fRNAdb: a platform for mining/annotating functional RNA candidates from non-coding RNA sequences. *Nucleic Acids Res.* **35**, D145–148, doi:10.1093/nar/gkl837 (2007).
64. Sai Lakshmi, S. & Agrawal, S. piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res.* **36**, D173–177, doi:10.1093/nar/gkm696 (2008).
65. Xing, Y. *et al.* An expectation-maximization algorithm for probabilistic reconstructions of full-length isoforms from splice graphs. *Nucleic Acids Res.* **34**, 3150–3160, doi:10.1093/nar/gkl396 (2006).
66. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628, doi:10.1038/nmeth.1226 (2008).
67. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* **14**, 178–192, doi:10.1093/bib/bbs017 (2013).
68. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26, doi:10.1038/nbt.1754 (2011).
69. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645, doi:10.1101/gr.092759.109 (2009).
70. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868 (1998).
71. Reich, M. *et al.* GenePattern 2.0. *Nat. Genet.* **38**, 500–501, doi:10.1038/ng0506-500 (2006).
72. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300, doi:10.2307/2346101 (1995).
73. Gould, J., Getz, G., Monti, S., Reich, M. & Mesirov, J. P. Comparative gene marker selection suite. *Bioinformatics* **22**, 1924–1925, doi:10.1093/bioinformatics/btl196 (2006).

## Acknowledgments

The authors thank Dr. Kim Lonergan for critical discussions. Funding support for this work is through Grants from the Canadian Institutes for Health Research (CIHR) MOP-97839 and MOP-123273, and the Terry Fox Foundation. K.L.T., E.A.V., K.S.S.E. and L.A.P. are supported by scholarships from Vanier Canada, and CIHR.

## Author Contributions

V.D.M., R.H., E.A.V. and K.L.T. conceived, designed and performed data analysis. K.S.S.E., L.A.P. and D.D.B. collected data and performed analysis. V.D.M., R.H., K.L.T. and D.D.B. performed the statistical analyses. V.D.M., E.A.V., K.L.T. and K.S.S.E. wrote the manuscript. C.J.B., S.L. and W.L.L. are the Principal Investigators. All authors contributed to data interpretation and manuscript preparation.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>



**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Martinez, V. D. *et al.* Unique somatic and malignant expression patterns implicate PIWI-interacting RNAs in cancer-type specific biology. *Sci. Rep.* **5**, 10423; doi: 10.1038/srep10423 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>