Received: 6 June 2022 | Revised: 29 September 2022 | Accepted: 1 October 2022

DOI: 10.1111/andr.13312

ANDROLOGY WILEY

ORIGINAL ARTICLE

Small non-coding RNA sequencing reveals global dysregulation of piwi-interacting RNA (piRNA) expression in gonadal malignant germ cell tumours

Sean Laidlaw[1] | Luz Alonso-Crisostomo[2] | Shivani Bailey[2] | Harpreet K. Saini[3] | Attila Molnár[4] | James C. Nicholson[5,6] | Anton J. Enright[2] | Cinzia G. Scarpini[2] | Raheleh Rahbari[1,#] | Nicholas Coleman[2,7,#] | Matthew J. Murray[2,5] | on behalf the Children's Cancer and Leukaemia Group[1]

[1]Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK

[2]Department of Pathology, University of Cambridge, Cambridge, UK

[3]EMBL-European Bioinformatics Institute, Cambridge, UK

[4]Department of Plant Sciences, University of Cambridge, Cambridge, UK

[5]Department of Paediatric Haematology and Oncology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

[6]Department of Paediatrics, University of Cambridge, Level 8, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

[7]Department of Histopathology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

Correspondence
Matthew J. Murray, Department of Pathology, University of Cambridge, Cambridge, CB2 1QP, UK.
Email: mjm16@cam.ac.uk

[#]Raheleh Rahbari and Nicholas Coleman are senior authors.

Funding information
NIHR Cambridge Biomedical Research Centre, Grant/Award Number: BRC-1215-20014; St Baldrick's Foundation, Grant/Award Number: 358099

Abstract

Background: Analyses of small non-coding RNA (ncRNA) expression in malignant germ cell tumours (GCTs) have focused on microRNAs (miRNAs). As GCTs all arise from primordial germ cells, and piwi-interacting RNAs (piRNAs) have important roles in maintaining germline integrity via transposon silencing, we hypothesised that malignant GCTs are characterised by fundamental piRNA dysregulation.

Aims: We undertook global small ncRNA sequencing in malignant GCTs, in order to describe small ncRNA expression changes for both miRNAs and piRNAs.

Materials and methods: We performed small ncRNA next generation sequencing on a representative panel of 47 samples, comprising malignant GCT (n = 31) and control (n = 16) tissues/cell lines. Following quality control and normalisation, filtered count reads were used for differential miRNA and piRNA expression analyses via *DESeq2*. Predicted mRNA targets for piRNAs were identified and utilised for pathway enrichment analyses.

Results: Overall, miRNAs and piRNAs comprised 21.9% and 43.0% of small ncRNA species, respectively. There were 749 differentially expressed miRNAs in malignant GCTs, of which 536 (72%) were over-expressed and 213 (28%) under-expressed. The top-ranking over-expressed miRNAs were exclusively from the miR-371~373 and miR-302/367 clusters. The most significantly under-expressed miRNAs were miR-100-5p, miR-214-3p, miR-125b-5p and *let-7* family members, including miR-202-3p. There were 1,121 differentially expressed piRNAs in malignant GCTs, of which 167 (15%) were over-expressed and 954 (85%) under-expressed. Of note, of the top-20 differentially expressed piRNAs, 16 were over-expressed, of which piR-hsa-2506793 was both top-ranking and most abundant. Mobile element (ME; i.e., transposon)-associated

piRNAs comprised 166 (15%) of the 1,121 differentially expressed piRNAs, of which 165 (>99%) were down-regulated. The remaining 955 (85%) non-ME-associated piRNAs may have wider cellular roles. To explore this, predicted mRNA targets of differentially expressed piRNAs identified putative involvement in cancer-associated pathways.

Conclusion: This study confirms previous miRNA observations, giving credence to our novel demonstration of global piRNA dysregulation in gonadal malignant GCTs, through both ME and non-ME-associated pathways, which likely contributes to GCT pathogenesis.

KEYWORDS
germ cell tumour, microRNA, ncRNA, piRNA, sequencing

1 | INTRODUCTION

Germ cell tumours (GCTs) are clinically and pathologically complex neoplasms that occur from the neonatal period through to late adulthood and at gonadal (testicular/ovarian) and extragonadal sites.[1] Despite good overall outcomes, GCTs remain an area of unmet clinical need as they are the most common malignancy in young adult men, their global incidence is steadily increasing,[2] and the testicular form represents the highest adult cause for average years of life lost per person dying of cancer.[3] Malignant GCTs are typically classified into seminoma/germinoma and nonseminomatous tumours (e.g., yolk sac tumour [YST] and embryonal carcinoma [EC]).[4] Testicular GCTs (TGCTs) in adults are believed to originate from the precursor lesion germ cell neoplasia in situ (GCNIS).[5] For ovarian GCTs, although the peak in incidence is in early adulthood, similar to TGCTs, they have a much lower rate, namely three per million person years risk, compared with 80 for their testicular counterparts.[6] Despite their heterogeneity, the germ cell theory of tumorigenesis states that all GCTs arise from totipotent primordial germ cells (PGCs).[7] Studies of a class of small, non-coding RNAs (ncRNAs) termed microRNAs (miRNAs; typically, 18–24 nucleotides [nt] in length) have identified dysregulated expression that is conserved across patient age, anatomical site, and histological subtype of malignant GCTs.[4] Specifically, over-expression of microRNAs from the miR-371~373 and miR-302/367 clusters is observed in all malignant GCTs, but not the GCT subtype teratoma,[4] a finding that has led to miRNAs from these clusters being identified as highly sensitive and specific circulating biomarkers for malignant GCT diagnosis.[5,8,9] Furthermore, over-expression of the miR-371~373 and miR-302/367 clusters in malignant GCTs resulted in global down-regulation of mRNAs, which were involved in cancer-associated and signalling pathways, demonstrating the functional significance of this dysregulation.[4] These miRNA findings are consistent with those seen in murine development, where the embryonic stem cell (ESC)-specific miR-290-295 cluster (corresponding to the miR-371~373 cluster homologue in humans[10]) is expressed at high levels in PGCs and spermatogonia.[11] In addition, the high levels of miR-302/367 cluster expression seen in malignant GCTs is also characteristic of ESCs but not other somatically differentiated cells.[12]

Another class of small ncRNAs, termed P-element induced wimpy testis (piwi)-interacting RNAs (piRNAs; typically 25–31 nt), occurs in both the testis[13,14] and ovary.[15] PiRNAs associate with piwi family proteins and are known to play critical roles in a diverse range of cellular functions, including silencing of transposable or mobile elements (MEs) and gene expression regulation. This piRNA function is important, as through functional experiments exploring deletion of the miRNA biogenesis enzyme Dicer, it has been shown that miRNAs are crucial for murine PGC and spermatogonia proliferation, but dispensable for ME silencing in developing germ cells.[11] In addition to ME silencing, piRNAs have roles in epigenetic regulation including DNA methylation[16,17] and imprinting.[14] As piRNAs are predominantly found in the germline, their roles are believed to be critical for the maintenance of genomic stability and PGC function. These piRNA functions are of particular interest as it has been shown that GCNIS cells have an active DNA demethylation pathway, and the resultant hypomethylated genome may contribute to subsequent invasive TGCT development.[18] Furthermore, altered piwi protein expression has been detected in TGCTs.[19,20] Of note, in a murine model system, spermatogenesis has been shown to be regulated by piRNA-directed cleavage of mRNA transcripts[21] However, studies of piRNA expression in human GCTs and/or testis have been small and limited.[20,22–24] Rounge et al. undertook small RNA sequencing and demonstrated global loss of piRNA expression in TGCTs compared with normal testis.[23] Of note, precursor GCNIS samples clustered with testis samples, and no histological subtype difference in piRNA expression was observed.[23] In contrast, Gainetdinov et al. used small RNA sequencing to show that both TGCTs and GCNIS lose piRNA expression compared with healthy testis.[24]

Here, we use next generation sequencing (NGS) to investigate small ncRNA expression in malignant GCTs of both testicular and ovarian origin. We first re-confirmed our[4] and other investigators[25,26] observations regarding miRNA expression in these tumours, validating the method used for small ncRNA profiling. We then

demonstrated global downregulation of piRNA expression in malignant GCTs of both testicular and ovarian origin and highlight mechanisms through which these observations may contribute to GCT pathogenesis.

2 | MATERIALS AND METHODS

Ethical approvals. The study received Multicentre (East-Midlands/Derby REC ref: 08/H0405/22+5; formerly Trent- =REC ref: 02/4/071) and Local Research Ethics Committee (ref:01/128) approvals and was performed under Children's Cancer and Leukaemia Group (CCLG) studies CCLG 2002 BS03 and CCLG 2020 BS02.

Samples for NGS. We performed small ncRNA NGS on two groups, representing a total of 47 samples (Table 1). The first set of 13 samples comprised 10 malignant GCTs, one immature teratoma and two gonadal controls (one testis and one ovary). Of the 10 malignant GCTs, four had YST histology (two testicular; two ovarian), four seminoma/germinoma histology (all ovarian origin) and two embryonal carcinoma (EC; both testicular origin) (Table 1). Note that we use the term 'seminoma/germinoma' to refer to testicular seminoma or ovarian dysgerminoma, given their typically identical histology. We supplemented these data with a second set of 34 samples including cell lines and additional controls. This second set comprised six gonadal controls (three testes and three ovaries, Origene total RNA; 1 μg input starting material) and four cell lines and their derived extracellular vesicles (EVs) from cell line culture media. Of these four, three were representative malignant GCT cell lines (namely TCam2 [seminoma],[27] GCT44 [YST][28] and 2102Ep [EC][29]], cultured as previously described.[30] The fourth was the human testis fibroblast cell line Hs1.Tes (ATCC; CRL-7002), cultured as per ATCC recommendations and used as a control cell line. In brief, cells were cultured at 37°C in 5% CO_2 in cell line-specific medium containing 10% standard fetal calf serum (FCS) and 1% penicillin/streptomycin unless otherwise stated (TCam2: Roswell Park Memorial Institute 1640 medium (RPMI-1640); GCT44: Dulbecco's Modified Eagle's Medium (DMEM) with GlutaMAX-I [ThermoFisher Scientific] with sodium pyruvate; 2102Ep: DMEM with GlutaMAX-I without sodium pyruvate and 10% heat-inactivated USA FCS; Hs1.Tes: DMEM). These four cell lines were run in biological triplicate (n = 12 samples), with each replicate consisting of 2.5 × 10^6 cells. All cell lines were authenticated by short-tandem-repeat profiling,[31] as previously reported.[4,30] EVs were derived from their respective cell lines and extracted from culture media after 48 h incubation by ultracentrifugation, consistent with strict guidelines for EV isolation, purification and quality control (QC) assessment from the International Society for Extracellular Vesicles https://www.isev.org/).[32] These four cell line-derived EVs were run in biological quadruplicate (n = 16 samples), and each replicate was derived from six T175 (175 cm^2) flasks.

Sample and library preparation (set-1). Total RNA was initially isolated from the 13 fresh frozen biopsy or resection samples (samples 1–13) using TRIzol reagent (ThermoFisher Scientific), with phase separation followed by isopropanol precipitation,[4,30,33] from which the small RNA fraction was then isolated from 1 μg of total RNA (total RNA concentration range 635–5000 ng/μl) using the mirVana miRNA isolation kit (Ambion; AM1560), as per the manufacturer's instructions. In brief, samples were first lysed using the denaturing lysis solution, stabilising the RNA and inactivating RNases. Lysates were then extracted using acid-phenol:chloroform, with two subsequent ethanol filtration steps (at one-third and then two-thirds ethanol volume to aqueous phase), resulting in purification of small RNAs < 200 nt. Next, 3′ RNA adapter ligation was performed, using T4-RNA-ligation buffer, T4-RNA-ligase and RNaseOUT (all ThermoFisher Scientific), as per the manufacturer's instructions. For this work, a 10 μM concentration of seven unique adapters was used, each containing unique 4 nt sequences within their 8 nt barcode (Table S1). These barcodes used Hamming code, which allowed error detection and correction, as described,[34] and were further modified to ensure desirable properties, that is, no polynucleotide sequences or high G/C content, as described.[35] These seven unique adapters allowed the 13 samples to be run in multiplexed NGS lanes (Table 1). Ligated RNA was then loaded and run on a 15% TBE-Urea gel (Bio-Rad; 161–1117) for 30 min at 150 V. Following ethidium bromide staining of the gel, bands were cut corresponding to products of length 45–62 nt (i.e., 18–35 nt small RNAs of interest plus 27 nt 3′ adapter) using ultraviolet light guidance. Ligated RNA was then eluted from the gel bands using 0.3 M NaCl at 4°C overnight. The eluate was then added into Costar Spin-X Cellulose acetate filters (Corning; 8163) and centrifuged, prior to addition of additional 0.3 M NaCl and further centrifugation. The resultant eluate was transferred to new microtubes (Sigma ClearView; T4816-250EA), 3 μl glycoblue (Ambion) added and the RNA pellet obtained using isopropanol and ethanol extraction with sequential centrifugation. Samples next underwent 5′ universal RNA adapter ligation and RNA elution, as per the 3′ ligation approach but using a generic universal primer (Table S1). Samples were loaded onto a 10% TBE-Urea gel (Bio-Rad Mini-PROTEAN-II; 4565035) and run for 30 min at 150 V. Gel bands were cut using ultraviolet light guidance corresponding to 71–88 nt (previous 45–62 3′ ligated product plus 26 nt 5′adapter) and RNA eluted. Samples then underwent reverse transcription (RT) using a 21 nt 3′ RT primer (Table S1) and the SuperScript-III First-Strand Synthesis System (ThermoFisher Scientific; 18080051), as per the manufacturer's instructions. The resulting copy-DNA (cDNA) libraries underwent PCR using the Phusion High-Fidelity DNA polymerase (ThermoFisher Scientific; F-530XL) and associated reagents, as per the manufacturer's instructions, using Solexa small RNA PCR Primers 1 and 2 (Table S1). PCR conditions were 98°C for 30 s (s); then 25 cycles of 98°C for 10s, 58°C for 30s, 72°C for 20s; then 72°C for 5 min. Samples were then loaded into a 10% TBE gel (Bio-Rad; 161–1182) and bands of interest (93-110 nt) cut out using ultraviolet guidance from which cDNA was eluted using 0.5 M ammonium acetate-based elution buffer. The cDNA libraries were then assessed and quantified using an Agilent 2100 Bioanalyser and a molar equivalent of the 13 samples mixed in their two groups for running as two 'test' lanes (SL251–SL252; Table 1) as single-end 50 nt reads on a Solexa Genome Analyzer sequencer (GAIIx; Illumina), as per the manufacturer's recommendations. Following initial QC and assessment of these two pilot 'test' lanes

TABLE 1 Details of the malignant germ cell tumour tissue samples, cell lines and extracellular vesicle specimens used in this study and non-malignant control samples

Set-1						
Sample number	**Sample type**	**Sample code***	**Primary site**	**Patient age (years)**	**Solexa 8 nt barcode**	**Comment**
1	YST	YST-3	Testicular	2	A	NGS test lane SL251-A; final SL287-A
2	YST	YST-4	Ovarian	13	C	NGS test lane SL251-C; final SL286-C
3	YST	YST-10	Ovarian	0	D	NGS test lane SL251-D; final SL286-D
4	YST	YST-6	Testicular	4	E	NGS test lane SL251-E; final SL286-E
5	Germinoma	Sem-25	Ovarian	13	F	NGS test lane SL251-F; final SL288-F
6	Germinoma	Sem-21	Ovarian**	11	G	NGS test lane SL251-G; final SL287-G
7	Germinoma	Sem-22	Ovarian	14	H	NGS test lane SL251-H; final SL287-H
8	Germinoma	Sem-13	Ovarian	12	A	NGS test lane SL252-A; final SL288-A
9	EC	EC-26	Testicular***	15	C	NGS test lane SL252-C; final SL289-C
10	EC	EC-28	Testicular	13	D	NGS test lane SL252-D; final SL289-D
11	IT	IT-31	Abdominal	0	F	NGS test lane SL252-F; final SL289-F
12	Testis	N-43	NA	27	G	NGS test lane SL252-G; final SL-290-G
13	Ovary	N-42	NA	7	H	NGS test lane SL252-H; final SL291-H
Set-2						
Sample number	**Sample type**	**Sample code***	**Primary site**	**Patient age (years)**	**Biological replicates**	**Comment**
14	Ovary	NA	NA	31	Triplicate	Total RNA; OriGene; catalogue CR560115
15	Ovary	NA	NA	42		Total RNA; OriGene; catalogue CR561006
16	Ovary	NA	NA	45		Total RNA; OriGene; catalogue CR561036
17	Testis	NA	NA	59	Triplicate	Total RNA; OriGene; catalogue CR560016
18	Testis	NA	NA	79		Total RNA; OriGene; catalogue CR560548
19	Testis	NA	NA	79		Total RNA; OriGene; catalogue CR561118
20	Hs1.Tes	NA	Cell line	NA	Triplicate	Testis stromal fibroblasts; replicate 1
21	Hs1.Tes	NA	Cell line	NA		Testis stromal fibroblasts; replicate 2
22	Hs1.Tes	NA	Cell line	NA		Testis stromal fibroblasts; replicate 3

(Continues)

TABLE 1 (Continued)

Set-2						
Sample number	**Sample type**	**Sample code***	**Primary site**	**Patient age (years)**	**Biological replicates**	**Comment**
23	TCam2	CL-39	Cell line	NA	Triplicate	Seminoma cell line; replicate 1
24	TCam2	CL-39	Cell line	NA		Seminoma cell line; replicate 2
25	TCam2	CL-39	Cell line	NA		Seminoma cell line; replicate 3
26	GCT44	CL-36	Cell line	NA	Triplicate	YST cell line; replicate 1
27	GCT44	CL-36	Cell line	NA		YST cell line; replicate 2
28	GCT44	CL-36	Cell line	NA		YST cell line; replicate 3
29	2102Ep	CL-40	Cell line	NA	Triplicate	EC cell line; replicate 1
30	2102Ep	CL-40	Cell line	NA		EC cell line; replicate 2
31	2102Ep	CL-40	Cell line	NA		EC cell line; replicate 3
32	Hs1.Tes	NA	EVs	NA	Quadruplicate	Testis stromal fibroblasts; replicate 1
33	Hs1.Tes	NA	EVs	NA		Testis stromal fibroblasts; replicate 2
34	Hs1.Tes	NA	EVs	NA		Testis stromal fibroblasts; replicate 3
35	Hs1.Tes	NA	EVs	NA		Testis stromal fibroblasts; replicate 4
36	TCam2	NA	EVs	NA	Quadruplicate	Seminoma cell line; replicate 1
37	TCam2	NA	EVs	NA		Seminoma cell line; replicate 2
38	TCam2	NA	EVs	NA		Seminoma cell line; replicate 3
39	TCam2	NA	EVs	NA		Seminoma cell line; replicate 4
40	GCT44	NA	EVs	NA	Quadruplicate	YST cell line; replicate 1
41	GCT44	NA	EVs	NA		YST cell line; replicate 2
42	GCT44	NA	EVs	NA		YST cell line; replicate 3
43	GCT44	NA	EVs	NA		YST cell line; replicate 4
44	2102Ep	NA	EVs	NA	Quadruplicate	EC cell line; replicate 1
45	2102Ep	NA	EVs	NA		EC cell line; replicate 2
46	2102Ep	NA	EVs	NA		EC cell line; replicate 3
47	2102Ep	NA	EVs	NA		EC cell line; replicate 4

Abbreviations: EC, embryonal carcinoma; EVs: extracellular vesicles; IT, immature teratoma; NA: not applicable; YST, yolk sac tumour.

*see reference PMID 20332240.

**Lymph node metastasis.

***Lung metastasis.

(see below), five 'final' NGS lanes were run (SL286-SL291; Table 1). As expected, samples within these final lanes generally contained greater read depth, with 147,216 median reads per sample (range: 87,184-464,681) compared with 64,153 per sample (range: 13,033-178,495) for the test lanes.

Sample and library preparation (set-2). Testis and ovary samples were purchased as total RNA from Origene, with 1 μg of total RNA input used (samples 14–19). Small RNA was extracted from cell line, EV and the gonadal control samples using the miRNeasy Mini and Serum/Plasma kits (Qiagen) as per the manufacturer's recommendations, and as described[36] (Table 1). Library preparation was performed by Cambridge Genomic Services (https://www.cgs.path.cam.ac.uk/; University of Cambridge, UK). In brief, RNA quality and concentration were determined using an Agilent 2100 Bioanalyser and Agilent 4200 TapeStation

(Agilent Technologies). As expected, the median small RNA concentration in set-2 was 1003 ng/μl for cell lines (samples 20–31) and lower at 76.8 ng/μl for EVs (samples 32–47). All 34 set-2 samples passed RNA QC. Small RNA libraries were prepared with a NEXTFLEX Small RNA-Seq Kit v3 (PerkinsElmer; NOVA-5132-05), containing NEXTFLEX four nucleotide barcode add-ons, using the gel-free size selection and bead clean-up protocol, according to the manufacturer's instructions. Samples were then run on a NextSeq 500 machine (Illumina) for 22 cycles, using the 75-cycle NextSeq 550Dx High Output Reagent Kit v2.5 (Illumina; 20028870).

Small RNA NGS QC, alignment, quantification and clustering. NGS analysis of the raw sequenced reads initially involved using cutadapt[37] to remove adapter sequences and QC using the fastqc library,[38] with filtering out of low quality basecalls, and removal of barcodes, using a custom Perl script and Reaper tool.[39] During removal of barcodes, reads were assigned to their relevant samples. Initial QC assessment for miRNA expression showed that identical samples in set-1 from the 'test' and 'final' lanes clustered tightly together (Figure S1), demonstrating technical reproducibility of different NGS runs. Due to their greater depth, read counts from final lanes were subsequently used for all downstream analyses. Next, batch assessment for miRNA expression demonstrated that set-1 and set-2 samples occupied similar principal component analysis (PCA) space (Figure S2) and could therefore, following batch normalisation (see below), be integrated for combined analyses. The DASHR database of human small ncRNAs was used to define small ncRNA families,[40,41] except for miRNA, which was made from grouping 'miR-3p', 'miR-5p', 'miR-5p3pno' and 'miRNAprimary' RNAs, and transfer RNA (tRNA) family, which was made from grouping 'tRNA', 'tRF3' and 'tRF5' RNAs. Filtered reads were then mapped against all known mature miRNA sequences in the miRNA reference database miRBase (https://www.mirbase.org/; v22.1; 2019)[42] using Chimira[43] and against piRNA sequences (see below). Filtering was performed on both miRNA and piRNA counts matrices to remove miRNA and piRNA sequences found in two or less individual samples. *R* and *RStudio* (versions 4.1.2 and 2021.09.0-build-351, respectively) were used to analyse and normalise raw counts generated for each sample utilising *DESeq2*, as described.[44] Small RNA clustering was performed using *DESeq2* convenience functions using dimensionality reduction methods to reduce the data to two dimensions and enable visualisation via generation of PCA plots.

MiRNA alignment. The tool Chimira[43] was used to align 18–24 nt small RNA sequences to human miRNAs in miRBase,[42] with up to two sequence mismatches permitted, as described.[43] Importantly, in Chimira, the two permitted mismatches are within the 'local' alignment of the core miRNA sequence, such that any 5′ or 3′ post-transcriptional modifications would not be excluded from identification.[43]

PiRNA alignment. Consistent with other piRNA studies,[45] we identified piRNAs by filtering reads to retain only those conforming to a strict definition. In the first instance, reads were aligned to the human reference genome build hg38 using the STAR aligner and all reads filtered to remove sequences <24 nt and >34 nt, as described.[45] Misleading alignments (those that intersected with other small ncRNA families) were removed by intersecting the coordinates of the filtered reads to those of known other ncRNAs reported in the DASHR database.[40,41] The retained sequences were therefore of appropriate length and not matching DASHR coordinates for members of other ncRNA families. From these, we identified as piRNAs those that were reported in the piRNA database piRBase,[46,47] which were then quantified using the tool *featureCounts*[48] Finally, as a further filtering step, the resultant list of differentially expressed piRNAs (see below) were also compared directly against all human miRNA hairpin sequences extracted from miRBase, using Chimira.

NGS differential expression analysis. The filtered read counts were input into the *DESeq2 R* library, for which the ashr shrinkage estimator was utilised.[49] For each of the differential expression analyses, cut-offs were absolute $\log_2$ fold change ($\log_2$FC) $\geq$1 or $\leq$-1, and adjusted-p < 0.05 (Benjamini-Hochberg method[50]). Volcano plots were produced from the matrices of differential expression results using the *EnhancedVolcano R* package. For plotting the heatmaps of the differentially expressed genes, the filtered read counts were normalised using a variance stabilising transformation.[51] Where both sample sets were used, potential batch effects were corrected using the limma[52] *R* library. To determine differentially expressed piRNAs that could distinguish GCT subtypes, pairwise differential expression analysis was performed with *DESeq2* between each subtype. The resultant lists of piRNAs then underwent recursive feature elimination using the *FeatureTerminatoR R* package,[53] to obtain the most discriminatory piRNAs of potential clinical relevance.

Predicted mRNA targets of piRNAs. To predict potential mRNA targets of piRNAs, the bioinformatic tool miRanda[54] was used. This algorithm was originally developed for miRNAs, by calculating base-pair similarity with the 3′ untranslated region (UTR) of potential mRNA targets, as well as the thermodynamic stability of the predicted duplex. However, this methodology may also be utilised for piRNAs, as reported extensively elsewhere.[55–58] Consequently, using this identical method, we generated outputs of all possible mRNA targets for each piRNA for predicted duplexes that passed QC thresholds (score $\geq$ 200 and -20 kcal/mol). Our bespoke python script collapsed the list of possible mRNA-piRNA duplexes to a single best-fit mRNA transcript for each piRNA, based on the highest scoring possibility reported by miRanda. RNA counts of different transcripts corresponding to the same gene were collapsed to obtain single count values for each gene. Resulting counts of putative target mRNAs were analysed for pathway enrichment with the *pathfindR R* library using the STRING protein–protein interaction network,[59] so as to establish lists of putatively targeted KEGG pathways[60] that were over-represented in malignant GCT samples compared with controls.

3 | RESULTS

3.1 | Overview

Distribution of NGS count data by read length revealed that the 47 individual samples in set-1 and set-2 (Figure S3/S4, respectively) had good representation of reads corresponding to miRNAs (typically

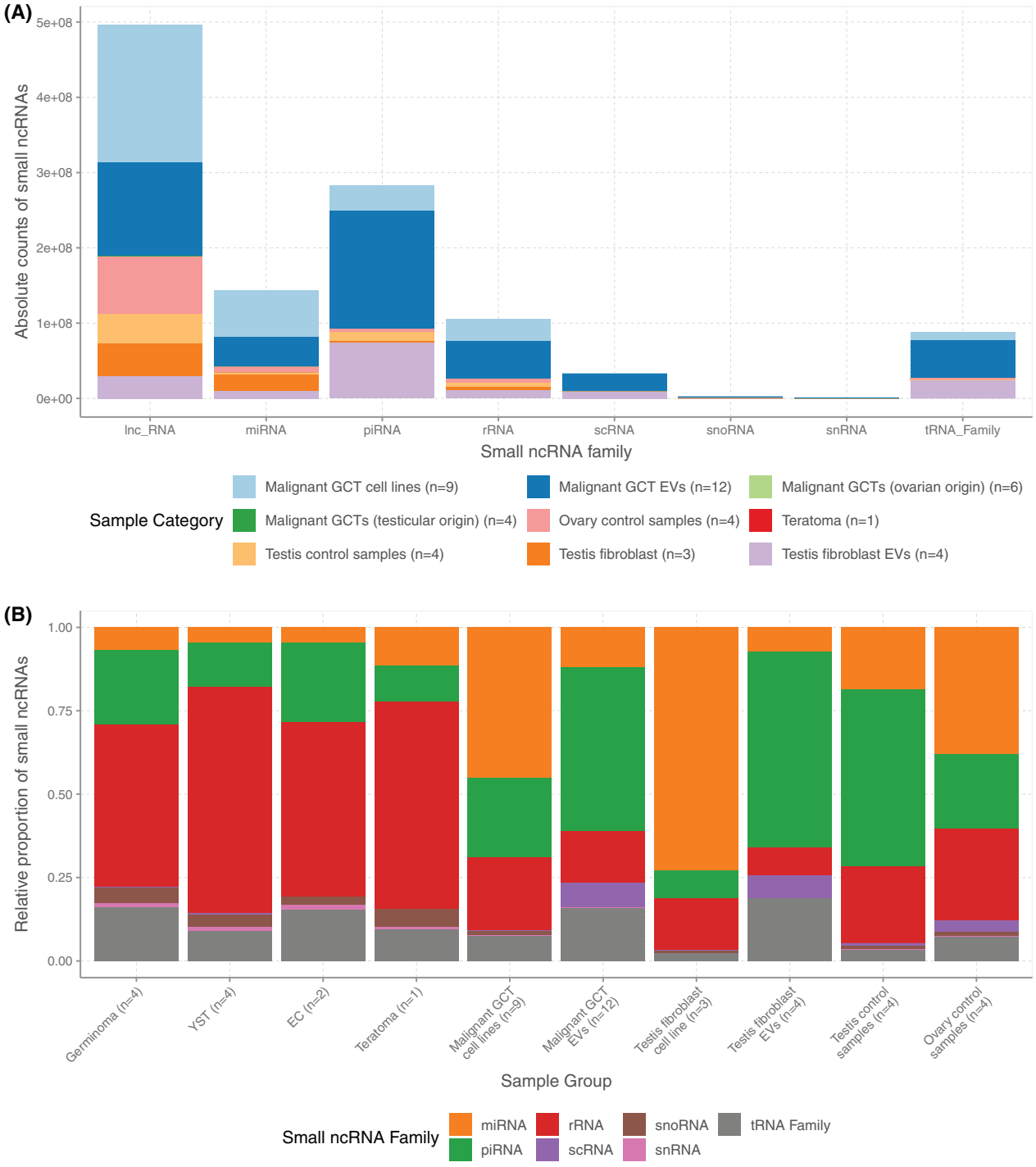

FIGURE 1 Stacked bar charts showing small non-coding RNA (ncRNA) fractions in malignant germ cell tumours and control samples. (A) Absolute read count per sample type, aligned to each family of small ncRNA. (B) Relative proportion of reads per small ncRNA family, per sample type

18–24 nt) and piRNAs (25-31 nt). The largest absolute number of small ncRNA reads were likely fragments of longer RNAs, as they mapped to long ncRNAs (lncRNAs; typically >200 nt) (Figure 1A and Figure S5A). PiRNAs and then miRNAs were the next most abundant class of small ncRNAs (Figure 1A), respectively. As lncRNAs were not the focus of our current study, these were removed from subsequent analyses. Relative proportions of the remaining true small ncRNA fraction were then calculated across the two sets (Figure 1B and Figure S5B). The proportion of miRNAs and piRNAs comprising the true small ncRNA fraction were noted to be relatively depleted and enriched, respectively, in the EV populations derived from both the malignant GCTs and Hs1.Tes (fibroblast) cell lines (Figure 1B).

3.2 | MiRNA expression counts

Initial PCA for global miRNA expression on the NGS data for the tissue samples in set-1 (n = 13) and the extended sample set in set-2 (n = 34) showed clear segregation by sample type. For set-1, these differences were reflected in the first and second principal components (PC1 and PC2) showing 22% and 18% variance, respectively (Figure S6A). For set-2, PC1 distinguished malignant GCT samples from non-malignant samples with 40% variance and PC2, distinguishing different histological subtypes, displayed 21% variance (Figure S6B). When combining both sets, comprising a total of 47 samples, malignant samples (n = 31; comprising malignant GCT tissues, representative cell lines and their extracellular vesicles [EVs]) were completely segregated from non-malignant samples (n = 16; including normal gonadal controls, testis fibroblasts and their EVs) (Figure 2A). This difference between malignant and non-malignant samples comprised the majority (34%) of the variance seen in this set (PC1). Furthermore, clustering also occurred by histological subtype, accounting for 16% of variance seen on PC2 (Figure 2A). Of note, the teratoma sample clustered closely with normal gonadal controls (Figure 2A). PCA excluding the testis fibroblasts also showed very robust segregation between histological subtypes, with PC1 and PC2 comprising 32% and 16% of the variance, respectively, confirming it was appropriate to include the testis fibroblast samples with the non-malignant samples group for subsequent comparisons (Figure S7). This robust segregation by global miRNA profiles on NGS data is entirely consistent with our previously reported findings using different methodologies (miRNA array and PCR) for malignant versus non-malignant GCT samples[4] and by GCT histological subtypes.[61]

Next, we sought to establish the qualitative nature of miRNA expression in these sample sets. Comparison of the malignant GCT tissue and cell line samples with non-malignant samples identified 749 differentially expressed miRNAs, of which 536 (71.6%) and 213 (28.4%) were up-regulated and down-regulated, respectively, in malignant GCTs (Figure S8/Table S2). A representative heatmap showing the top-20 differentially expressed miRNAs robustly segregated all malignant GCT samples from the non-malignant control group (Figure 2B). Of the top-20 ranking differentially expressed miRNAs in this analysis by adjusted p-value, 14 were up-regulated and six down-regulated (Table 2A). Of particular note, 13 of the 14 upregulated miRNAs in this top-20 comparison were from the miR-371~373 and miR-302/367 clusters (Table 2A). Of the six down-regulated miRNAs in this top-20 list, miR-100-5p, miR-214-3p and miR-125b-5p were also noted to be present (Table 2A). Tumour suppressor *let-7* miRNA family members *let-7b-3p*, *let-7b-5p* and miR-202-3p were also in this top-ranking downregulated list (Table 2A). Lists of differentially expressed miRNAs for individual malignant GCT subtypes were also derived versus the non-malignant group, with 729, 547 and 514 miRNAs identified for SEM, YST, and EC, respectively (Table S3). Taken together, our miRNA profiling data for malignant GCTs and individual subtypes are entirely consistent with previously published studies, for example,[4] thereby validating the approach we have taken for piRNA profiling.

3.3 | PiRNA expression counts

We next assessed piRNA expression, another class of short ncRNAs believed to have important biological function and relevance to germ cell tumorigenesis. PCA of global piRNA expression counts demonstrated clustering of malignant GCT samples, regardless of histological subtype, away from gonadal control samples, particularly the testis samples (Figure S9). This accounted for 53% of the variance for PC1, with 14% variance between histological subtypes (PC2) (Figure S9). We next assigned the 10 malignant GCT samples according to their anatomical site of primary disease, with six being of ovarian origin and four testicular (Table 1); comparison of these two groups showed no piRNA expression difference, and therefore the 10 malignant GCTs were considered together for subsequent downstream analyses. Qualitatively, comparison of the 31 malignant GCT tissue and cell line samples with the 16 non-malignant samples identified 1,121 differentially expressed piRNAs. Of these, 167 (14.9%) and 954 (87.8%) were up-regulated and down-regulated in malignant GCTs, respectively (Figure 3A/Table S4). Importantly, further filtering of this list of 1,121 differentially expressed piRNAs using Chimira against miRBase miRNA hairpin sequences demonstrated that there were no perfect (100% identity) miRNA matches. Only 41 of 1,121 differentially expressed piRNAs (3.7%) displayed a single mismatch and therefore might alternatively represent miRNAs with a single post-transcriptional modification (Table S4), although only a proportion of these 41 had the mismatch at the 3′ end where miRNA modifications most commonly occur. Of note, whilst the majority of piRNAs were down-regulated in malignant GCTs, in the top-20 ranking differentially expressed piRNAs (by adjusted p-value), 16 were up-regulated, and only four down-regulated (Table 2B). As expected, given that the number of piRNAs assessed (26,324; Figure 3A) was far greater than the number of corresponding miRNAs (2,501; Figure S8), the abundance of the vast majority of individual differentially expressed piRNAs (Table S4) was lower than for the corresponding differentially expressed miRNA list (Table S2). As there were a greater number of piRNAs than miRNAs interrogated, the top-50 differentially expressed piRNAs (with read counts >10) were selected to generate a representative heatmap, which partially segregated malignant GCT samples from the non-malignant control group (Figure 3B). Accordingly, the heatmap dendrogram divided into two main branches, with malignant GCT samples predominantly on the left (23 of 26 samples) and non-malignant control samples predominantly clustering on the right (13 of 21 samples) (Figure 3B). Of note, altering the number of differentially expressed piRNAs in this analysis did not improve the segregation further, likely reflecting the internal heterogeneity of individual malignant GCT tissues regarding piRNA expression – that is, compared with the miRNA analysis and established common miRNA expression changes, there were more private piRNA expression changes than shared.

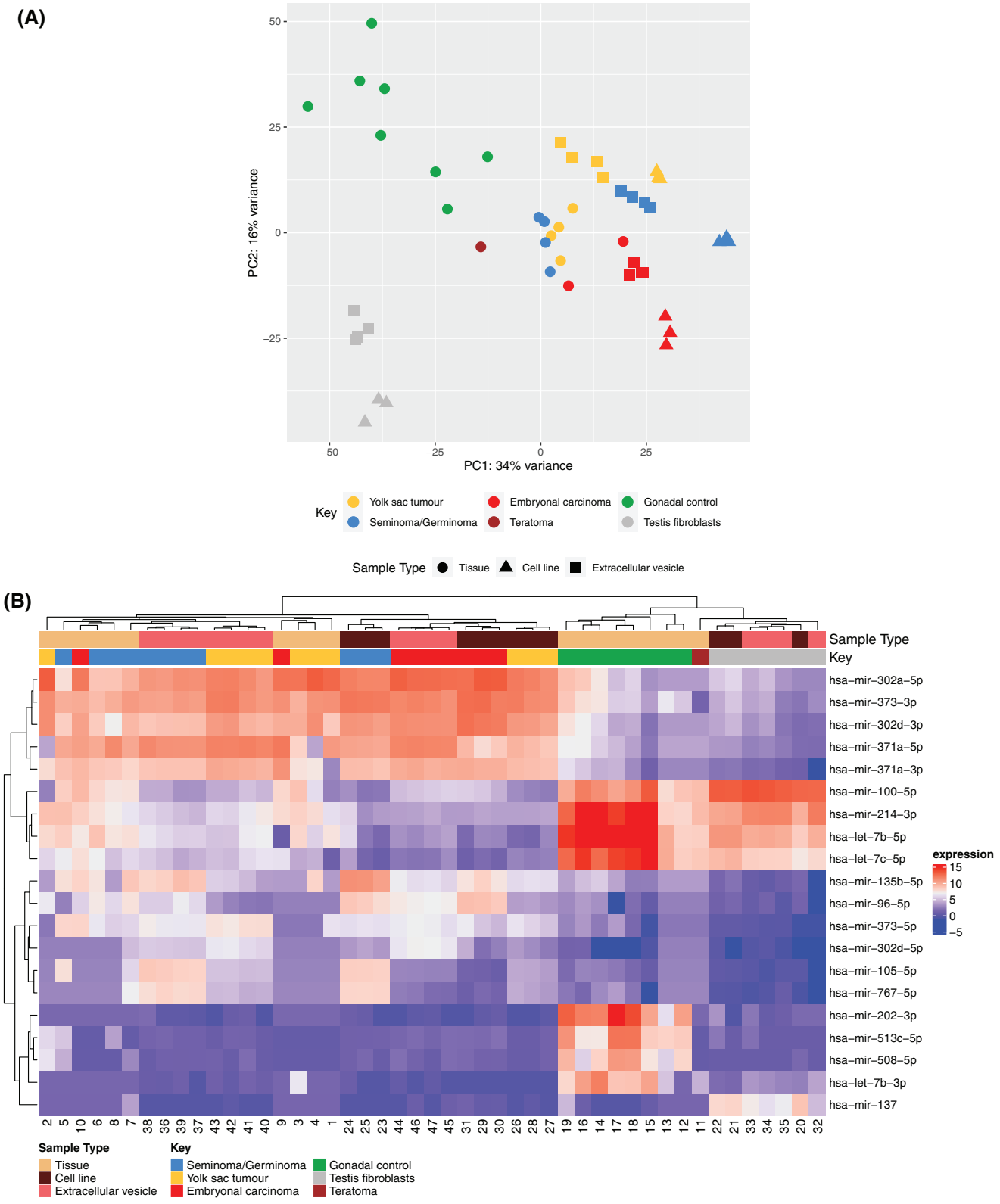

FIGURE 2 MiRNA expression analysis in malignant germ cell tumours across all samples in set-1 and set-2. (A) Principal component analysis of global miRNA expression across all samples in set-1 and set-2. Histological types are represented by colours and sample types by different shapes. (B) Heatmap of the top-20 differentially expressed miRNAs between malignant germ cell tumours (GCTs) and non-malignant control samples. Numbers at the base of each column in the heatmap relate to the sample numbers in Table 1.

TABLE 2 List of the top-20 ranking differentially expressed (A) miRNAs and (B) piRNAs in malignant germ cell tumours compared with non-malignant control samples, by adjusted *p*-value. Chromosomal loci for the miRNAs (from miRBase) and piRNAs (from piRBase), mean read count, and log_2 fold change (FC) are also listed.

A

Top-20 differentially expressed miRNAs

Rank	miRNA	Chromosomal locus for stem loop sequence (miRBase v22.1)	Mean read count	Log_2FC	Adjusted *p*-value
1	hsa-miR-373-5p	chr19: 53788705–53788773	714	6.59	1.23 E-51
2	hsa-miR-371a-3p	chr19: 53787675–53787741	7616	6.65	6.86 E-49
3	hsa-miR-371a-5p	chr19: 53787675–53787741	48,861	7.10	2.99 E-47
4	hsa-miR-202-3p	chr10: 133247511–133247620	2481	-10.20	9.35 E-46
5	hsa-miR-302d-5p	chr4: 112648004–112648071	374	6.72	7.73 E-45
6	hsa-miR-100-5p	chr11: 122152229–122152308	15,131	-6.48	6.43 E-43
7	hsa-miR-372-5p	chr19: 53787890–53787956	393	5.98	6.43 E-43
8	hsa-miR-373-3p	chr19: 53788705–53788773	12,836	6.40	3.93 E-42
9	hsa-miR-302d-3p	chr4: 112648004–112648071	33,450	6.61	9.57 E-42
10	hsa-miR-302b-5p	chr4: 112648485–112648557	353	5.81	1.87 E-41
11	hsa-miR-214-3p	chr1: 172138798–172138907	3179	-7.57	2.56 E-39
12	hsa-miR-302a-5p	chr4: 112648183–112648251	80,543	6.49	3.64 E-39
13	hsa-miR-302c-5p	chr4: 112648363–112648430	4817	6.07	9.18 E-39
14	hsa-miR-125b-5p	chr11: 122099757–122099844	25,696	-6.27	2.53 E-36
15	hsa-miR-302c-3p	chr4: 112648363–112648430	7455	6.24	3.33 E-36
16	hsa-miR-302b-3p	chr4: 112648485–112648557	119,052	6.29	3.77 E-34
17	hsa-miR-302a-3p	chr4: 112648183–112648251	21,928	6.28	3.20 E-33
18	hsa-miR-363-3p	chrX: 134169378–134169452	2021	5.57	3.28 E-32
19	*hsa-let-7b-3p*	chr22: 46113686–46113768	291	-7.40	4.67 E-32
20	*hsa-let-7b-5p*	chr22: 46113686–46113768	15,568	-7.48	6.14 E-32

B

Top-20 differentially expressed piRNAs

Rank	piRNA	Chromosomal locus for piRNA sequence (piRBase v3.0)	Mean read count	Log_2FC	Adjusted *p*-value
1	piR-hsa-2506793	chr4: 112648389–112648433	4228	5.97	6.12 E-34
2	piR-hsa-2212452	chr16: 28164511–28164555	106	5.65	1.66 E-30
3	piR-hsa-3357867	chr9: 112786058–112786101	372	5.39	2.24 E-27
4	piR-hsa-665793	chr12: 10969350–10969393*	105	5.24	6.29 E-26
5	piR-hsa-4394510	chr19: 53702775–53702819	356	5.74	5.61 E-22
6	piR-hsa-1686332	chr5: 75359497–75359541	58	-5.63	7.01 E-21
7	piR-hsa-371272	chr15: 88608110–88608154	28	3.92	8.29 E-19
8	piR-hsa-6444742	chr9: 28092883–28092926	80	4.74	6.26 E-18
9	piR-hsa-4417776	chr22: 46113707–46113750	32	-5.26	6.89 E-18
10	piR-hsa-5589150	chrX: 115905137–115905181	26	4.15	9.89 E-18
11	piR-hsa-2545800	chr12: 6964144–6964188	1396	5.60	1.31 E-17
12	piR-hsa-4396881	chr7: 93483934–93483978	98	4.82	4.48 E-17
13	piR-hsa-3969142	chr14: 56817517–56817561	13	3.53	1.36 E-16
14	piR-hsa-4469586	chr19: 53741358–53741402	124	5.26	1.36 E-16
15	piR-hsa-4412353	chr19: 53674220–53674264	519	4.97	1.39 E-15

(Continues)

TABLE 2 (Continued)

B Top-20 differentially expressed piRNAs					
Rank	piRNA	Chromosomal locus for piRNA sequence (piRBase v3.0)	Mean read count	Log_2FC	Adjusted *p*-value
16	piR-hsa-4441775	chr19: 53725446–53725490 **	343	5.19	1.94 E-15
17	piR-hsa-2458636	chr3: 183856370–183856413	14	3.69	3.00 E-15
18	piR-hsa-4169104	chr8: 4919051–4919094	25	-3.73	3.00 E-15
19	piR-hsa-6193980	chr1: 31067879–31067923	46	-4.06	5.16 E-15
20	piR-hsa-58861	chr17: 45097848–45097892	23	3.93	9.76 E-15

Abbreviations: has, human short ncRNA (either miRNA or piRNA) – i.e., abbreviation in miRBase and piRBase for homosapiens; Log_2FC, log_2 fold-change.
*Total nine piRNA matches, first hit given.
**Total two piRNA matches, first hit given.

3.4 | PiRNA expression for discriminating malignant GCT subtypes

Given the heterogeneity described above in individual malignant GCT tissue samples, set-2 samples (n = 34), which predominantly comprised very uniform cell line samples and associated EVs, were utilised through successive differential expression analyses to identify putative subtype-specific piRNA profiles (Figure 4). Successive iterations of random forest-based feature reduction produced a list of four up-regulated piRNAs (namely piR-hsa-4396881 [mean read count 98; log_2FC +4.82], piR-hsa-3620401 [69; log_2FC +2.21], piR-hsa-3736525 [51; log_2FC +4.33], and piR-hsa-6444742 [80; log_2FC +4.74; ranked 8th in the top-20 list]) from five piRNAs overall, which discriminated malignant GCTs (n = 21) from the non-malignant control group (n = 13) but also between the differing malignant GCT subtypes (Figure S10).

3.5 | PiRNA mapping to the human genome, including to MEs in malignant GCTs

MEs/transposons are DNA sequences that can move around the genome and alter the activity of neighbouring genes. Repression of MEs is therefore essential to maintain the integrity of the germline and may have relevance to germ cell tumorigenesis. Consequently, the list of 1,121 differentially expressed piRNAs identified above was interrogated and 166 (14.8%) had alignment that overlapped with the 469 MEs in piRBase (i.e., piRNA mapping provided evidence that 166/469 [35.4%] of all known MEs are dysregulated in malignant GCTs [Figure 5A]). Apart from a single piRNA, namely piR-hsa-14799, 165 of these were downregulated in malignant GCTs compared with non-malignant controls (Table S5). These 166 ME-associated piRNAs aligned to long interspersed nuclear elements (LINEs; n = 56), short interspersed nuclear elements (SINEs; n = 23) and long terminal repeats (LTRs; n = 87) (Table S5). The remaining 955 (85.2%) differentially expressed piRNAs were non-ME associated (Figure 5A) and likely to have functional roles other than ME silencing/repression. Next, we sought to further confirm our classification of piRNAs into those which are ME associated (also termed 'pre-pachytene', produced by the classical or 'ping-pong' cycling mechanism) and non-ME associated (or 'pachytene').[62] We therefore undertook additional analysis for a '1U:10A' bias indicative of ME associated, pre-pachytene piRNAs.[62] To do this, we compared the frequency of each base (T/U, C, G, A) at positions 1 and 10 within our lists of differentially expressed piRNA sequences, as described.[62] As expected, the 166 ME-associated piRNAs had the highest 1U:10A skew, with 81.9% of these piRNA sequences starting with a U (uridine) at position 1 (1U), and 31.3% having an A (adenine) at position 10 (10A) (Table S6). The non-ME-associated piRNAs had the lowest 1U:10A skew, as anticipated (Table S6). We next confirmed that this global downregulation of ME associated piRNAs in malignant GCTs occurred in each histological subtype (seminoma/germinoma, YST, EC) (Figure 5B) and by anatomical site of primary tumour (testicular and ovarian) (Figure 5C).

3.6 | Potential other functional consequences of piRNA dysregulation in malignant GCTs

Finally, we attempted to identify the potential functional impact of piRNA dysregulation in malignant GCTs by aligning the list of 1,121 differentially expressed piRNAs to the human reference transcriptome using full mRNA sequences, and constructed a pathway enrichment network. The top-10 resultant KEGG pathways with the greatest enrichment in tumour-derived piRNA targets were predominantly related to cellular signalling (JAK-STAT, RAP1, calcium, and RAS) and cancer-associated pathways (prostate cancer, gastric cancer, non-small cell lung cancer) (Figure 6).

4 | DISCUSSION

GCTs are clinically and pathologically complex neoplasms with incidence that varies by anatomical site.[6] Despite overall good outcomes, clinical questions remain. For example, patients with International

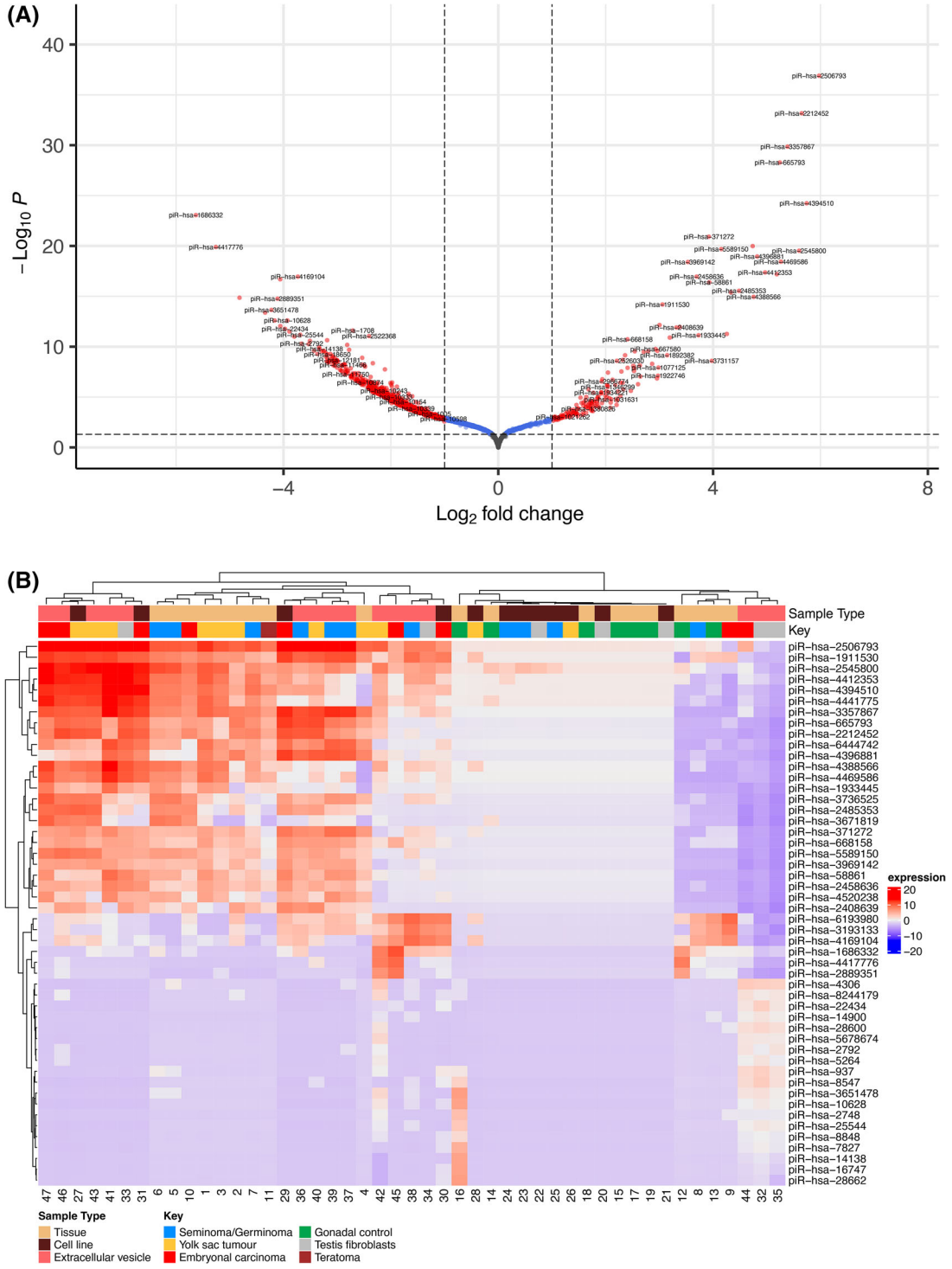

FIGURE 3 PiRNA expression analysis in malignant germ cell tumours across all samples in set-1 and set-2. (A) Volcano plot showing significantly differentially expressed piRNAs in malignant germ cell tumours (GCTs) (red circles) compared with the non-malignant control group. (B) Heatmap of the top-50 differentially expressed piRNAs between malignant GCTs and non-malignant control samples, with read count >10. Numbers at the base of each column in the heatmap relate to the sample numbers in Table 1.

Germ Cell Consensus Classification (IGCCC) poor-risk disease have a 41% 5-year progression-free survival,[63] which even with improvements in management over time have only recently modestly improved to 54%.[64] Consequently, an increased comprehensive molecular understanding of these under-investigated tumours may lead to improved diagnosis and prognostication within such risk-groups, may identify novel therapeutic opportunities, and ultimately impact upon clinical outcomes.

Recently, studies of small ncRNAs have been facilitated by the development of technologies including NGS. Over the last decade,

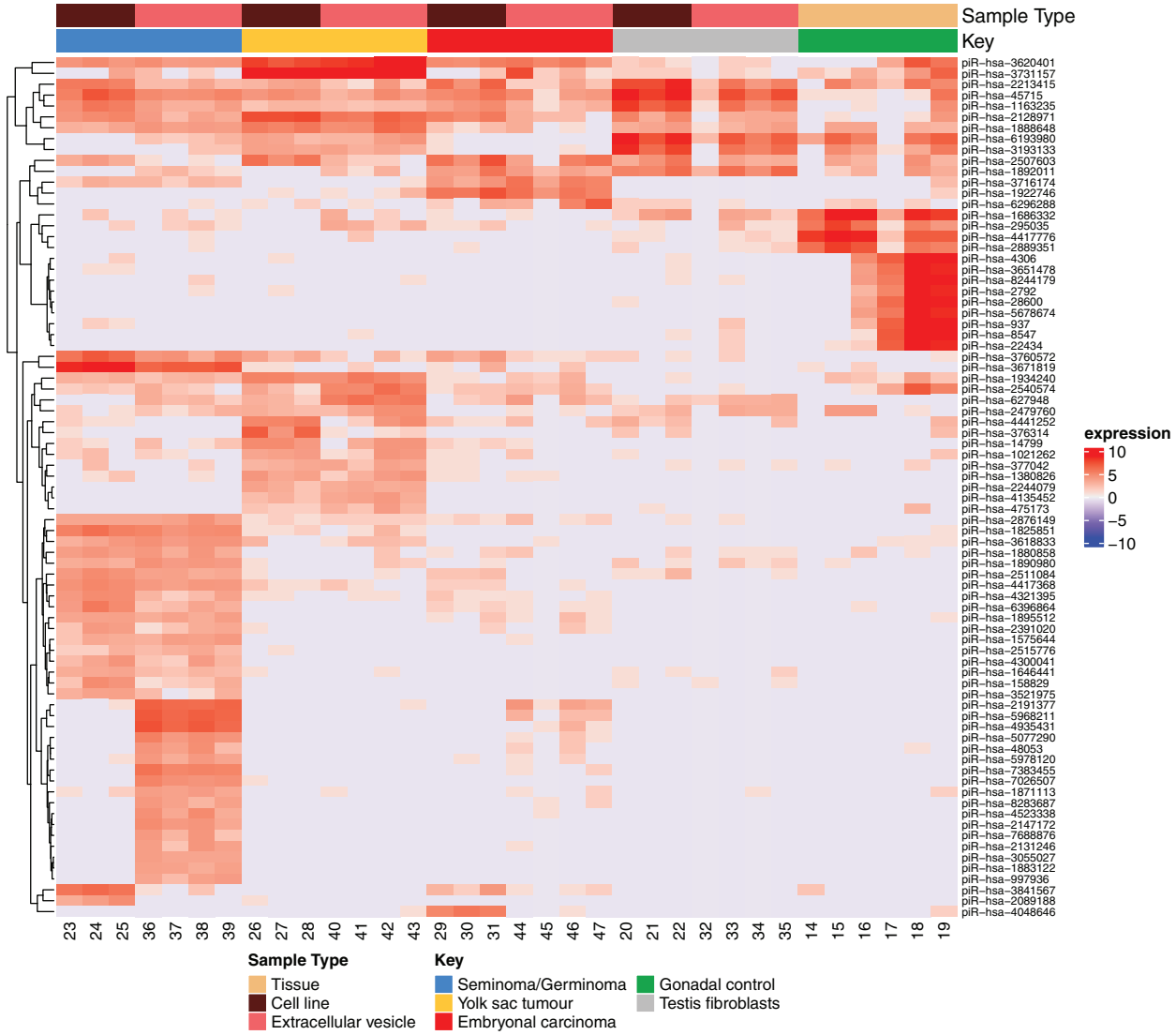

FIGURE 4 Identification of putative malignant germ cell tumour (GCT) subtype-specific piRNAs. Heatmap of subtype-specific piRNA profiles across set-2 samples. Numbers at the base of each column in the heatmap relate to the sample numbers in Table 1.

over-expression of miRNAs from the miR-371~373 and miR-302/367 clusters has been shown in all malignant GCTs, regardless of patient age, anatomical site, and histological subtype.[4] Of note, these miRNAs are not expressed in the GCT subtype teratoma, which formed a non-malignant control group with gonadal control samples.[4] This has led to circulating miRNAs from these clusters becoming promising biomarkers for malignant GCT diagnosis, disease-monitoring, and surveillance.[5,8,9] Here, our small ncRNA NGS data confirms these previously reported miRNA findings, which used different RNA extraction methods and methodologies (miRNA array and PCR),[4] demonstrating that our current approach was both appropriate and valid for downstream piRNA analyses. Furthermore, tumour-suppressor *let-7* miRNA family members (including *let-7b-3p*, *let-7b-5p* and miR-202-3p) were down-regulated in malignant GCTs, consistent with our previous studies.[4,30] We also noted down-regulation of miR-100-5p and miR-125b-5p in malignant GCT cases, consistent with previous work,[4] as well as miR-214-3p. The impact of these dysregulated miRNAs as biomarkers and their potential functional and therapeutic consequences in malignant GCTs has been described elsewhere.[65] Malignant GCT subtype miRNAs were also identified that confirmed our previous findings, for example, the miR-182/183 cluster in SEM, miR-375 in YST, and chromosome 19 miRNA cluster (C19MC) miRNAs in EC.[4]

Importantly, our NGS data comprehensively assessed another class of small ncRNAs, termed piRNAs. PiRNAs are known to be present in both testis[13,14] and ovary[15] and have important roles in silencing of MEs and epigenetic regulation including DNA methylation[16,17] and imprinting.[14] As piRNAs are predominantly found in the germline, these roles are believed to be critical for the maintenance of genomic stability and PGC function. To date, studies of piRNA expression in human GCTs and/or testis have been small, conflicting and limited.[20,22–24] Here, we demonstrate global down-regulation of piRNA expression in malignant GCTs of both testicular and ovarian origin, consistent with previous reports for TGCTs alone.[23] Such observations are likely to contribute to GCT pathogenesis. As piRNAs are

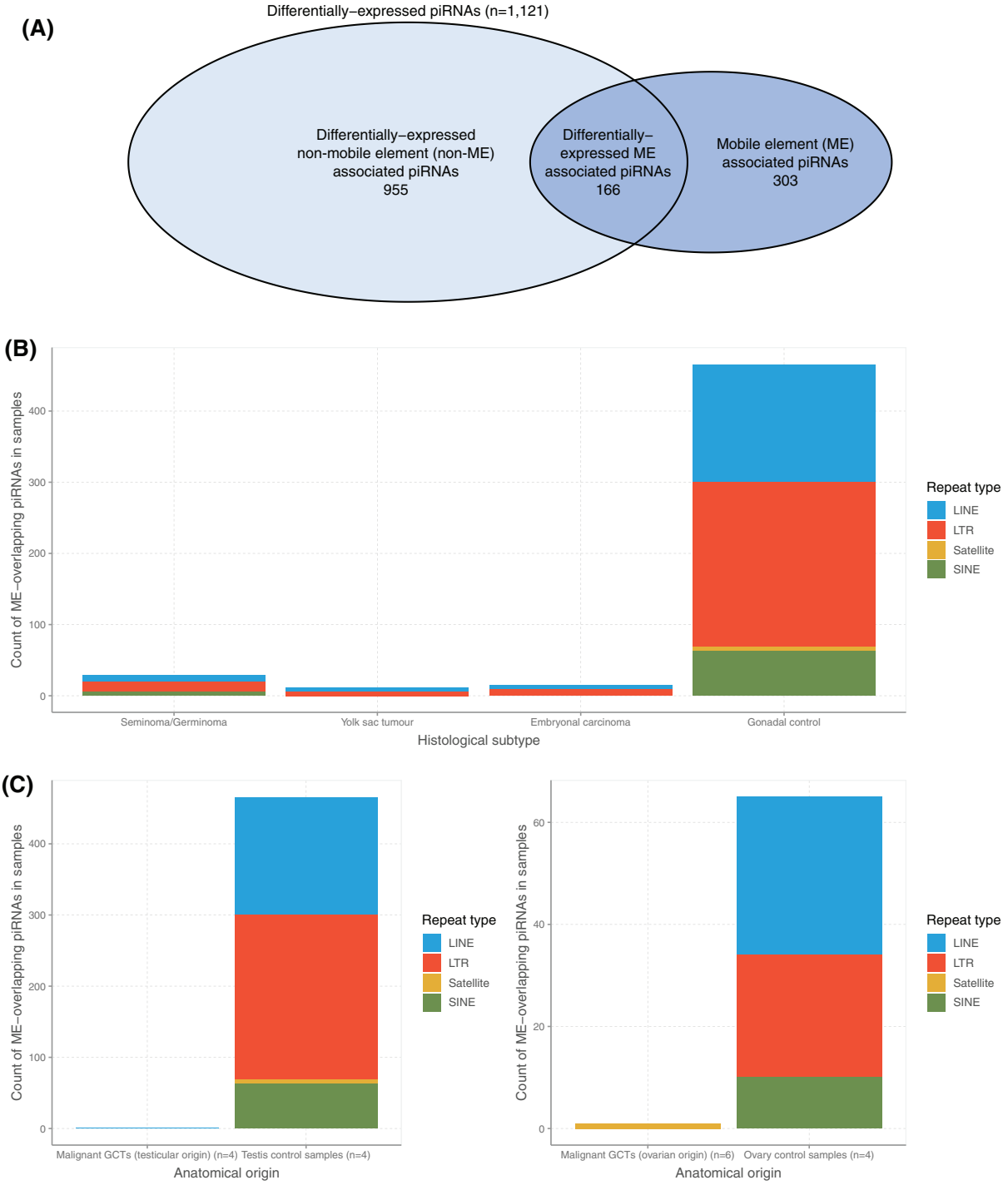

FIGURE 5 Identification of mobile element (ME)-associated and non-ME-associated differentially expressed piRNAs in malignant germ cell tumours (GCTs). (A) Venn diagram showing that 166 of 1,121 differentially expressed piRNAs in malignant GCTs overlap mobile element (ME) co-ordinates ($n = 469$). (B) Count of ME overlapping piRNAs per malignant GCT histological subtype, compared with combined gonadal control samples. (C) Count of ME overlapping piRNAs between malignant GCTs of testicular origin and testis control samples (left) and of ovarian origin and ovary control samples (right).

present in both the testes and ovaries in multiple species, such as drosophila[66] but also vertebrates such as zebrafish,[67] mouse[68] and human,[68] this highly conserved finding is again consistent with this class of short ncRNAs having important functional roles. However, defining such roles has remained under-investigated and enigmatic, including whether a differential role exists between piRNAs of the testis and ovary. For example, piwi knockout in mice, which results in the removal of the piwi/piRNA pathway, is associated with phenotypic defects in males, but not females, in contrast to findings in drosophila and zebrafish.[68] Consequently, direct demonstration of the functional

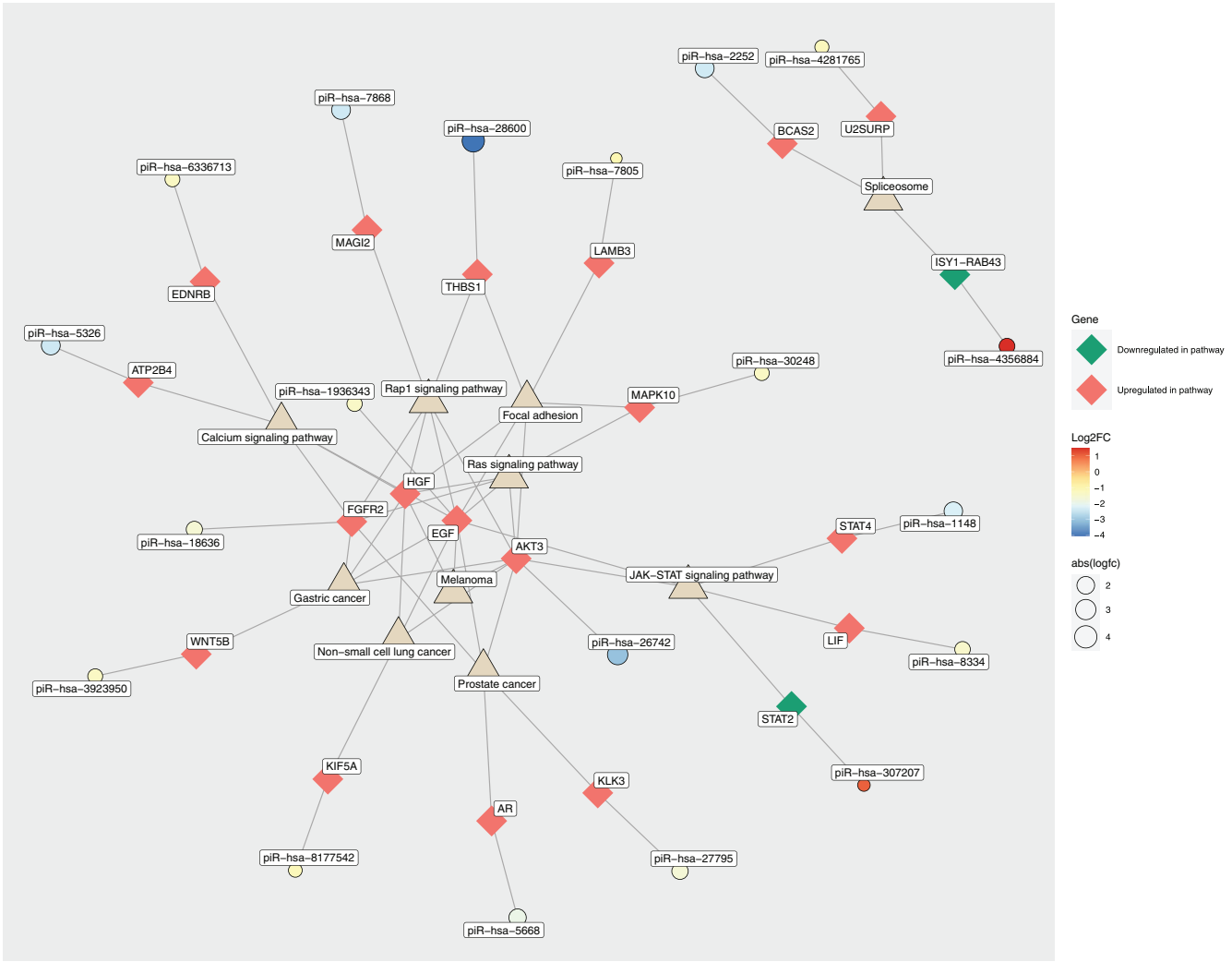

FIGURE 6 Network diagram of the top-10 enriched KEGG pathways from putative mRNA targets by differentially expressed piRNAs in malignant germ cell tumours

significance of the down-regulation of piRNAs in human malignant GCTs of both testicular and ovarian origin on germline integrity, or other roles, remains to be determined.

Further work identified a putative up-regulated four-piRNA classifier to segregate malignant GCTs from non-malignant controls, and between histological subtypes. In total, ~15% of the differentially expressed piRNAs in malignant GCTs were ME associated, of which all except one was down-regulated. This is of particular interest, given the reported piRNA down-regulation and associated global DNA hypomethylation of LINE-1 repetitive sequences described in malignant TGCTs.[20] Such global hypomethylation of LINE-1 elements has also been described in GCTs of childhood when compared with gonadal controls,[69,70] where it was highlighted that the potential mechanism by which this may affect tumour development is worthy of further investigation.[69] These data suggest that the ME-associated piRNA down-regulation seen in malignant GCTs may be associated with increased transposon mutagenesis and GCT pathogenesis.

In addition to transposon silencing through ME-associated piRNAs, multiple studies have demonstrated additional functional roles for piRNAs in post-transcriptional gene silencing, targeting mRNAs,[71–75] in a similar mechanism to miRNA-mediated gene repression.[76,77] For example, this mechanism has been studied in lung carcinoma, where silenced piR-55490 expression was restored, resulting in reduced cell proliferation through the targeting of mTOR 3′UTR mRNA.[78] Here, we identified that differentially expressed piRNAs in malignant GCTs may have a role through mRNA targeting, via cellular signalling and cancer-associated pathways. Of note, previous studies directly quantifying mRNA expression in both seminoma and nonseminomatous GCTs have found upregulation of genes involved in the above pathways,[79] suggesting a potential regulatory role of these differentially expressed non-ME-associated piRNAs in malignant GCTs.

Strengths of this work include the wide range of tissues, cell lines and EVs used, including testicular and ovarian tumour and control samples, which allow previous limited observations to be extended. Confirmation of previous miRNA findings, despite different RNA extraction

and RNA quantification methods, gives confidence to our novel piRNA findings. Limitations of this work are that potential biological roles for these differentially expressed piRNAs are bioinformatically predicted, rather than based on functional experimentation. However, whilst these analyses have identified a substantial number of differentially expressed piRNAs in malignant GCTs, many of which are of relatively low abundance, it is not possible to exclude potential functional roles for them based on our current limited knowledge. The agnostic study described here therefore provides a robust foundation for such in vitro and in vivo work. In summary, our unbiased NGS study of small ncRNAs in a representative panel of malignant GCT tissues, cell lines and controls has identified novel piRNA findings that warrant further exploration. Future work should seek to explore the potential functional roles this dysregulated piRNA expression has in malignant GCT pathogenesis.

AUTHOR CONTRIBUTIONS

Study conceptualisation: RR, NC and MJM. Obtaining experimental data: LAC, SB, AM and MJM. Data curation, interpretation and analysis: SL, HKS, AJE, RR, NC and MJM. Supervision: RR, NC and MJM. Writing - original manuscript draft: SL and MJM. Writing - review and editing, approval of final manuscript: SL, LAC, SB, HKS, AM, JCN, AJE, CGS, RR, NC and MJM.

ACKNOWLEDGEMENTS

The authors thank the CCLG Tissue Bank and CCLG Principal Treatment Centres for the collection and provision of tissue samples (under CCLG project numbers 2002 BS 03 and 2020 BS 02). The CCLG Tissue Bank is funded by Cancer Research UK and CCLG. Some tissues were acquired from the Human Research Tissue Bank at Cambridge University Hospitals NHS Foundation Trust, whom we must additionally thank for their provision of histological services, and which is supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014*). We recognise support from the St Baldrick's Foundation (MJM and NC; grant number 358099). We are grateful for the support from the Max Williamson Fund and from Christiane and Alan Hodson, in memory of their daughter Olivia. We thank Professor David Baulcombe, Department of Plant Sciences, University of Cambridge, in whose laboratory the set-1 samples were prepared for sequencing. The funders were not involved in the study design, data collection or interpretation, nor the decision to submit for publication.

CONFLICT OF INTEREST

The authors have no conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Sean Laidlaw https://orcid.org/0000-0001-7658-1428
Matthew J. Murray https://orcid.org/0000-0002-4480-1147

REFERENCES

1. Oosterhuis JW, Looijenga LHJ. Human germ cell tumours from a developmental perspective. *Nat Rev Cancer*. 2019;19(9):522-537. https://doi.org/10.1038/s41568-019-0178-9
2. Znaor A, Skakkebaek NE, Rajpert-De Meyts E, et al. Global patterns in testicular cancer incidence and mortality in 2020. *Int J Cancer*. 2022;151(5):692-698. https://doi.org/10.1002/ijc.33999
3. Howlader N, Noone AM, Krapcho M, et al. *SEER Cancer Statistics Review, 1975–2017, National Cancer Institute*. National Cancer Institute; 2020.
4. Palmer RD, Murray MJ, Saini HK, et al. Malignant germ cell tumors display common microRNA profiles resulting in global changes in expression of messenger RNA targets. *Cancer Res*. 2010;70(7):2911-2923. https://doi.org/10.1158/0008-5472.CAN-09-3301
5. Almstrup K, Lobo J, Morup N, et al. Application of miRNAs in the diagnosis and monitoring of testicular germ cell tumours. *Nat Rev Urol*. 2020;17(4):201-213. https://doi.org10.1038/s41585-020-0296-x
6. Murray MJ, Fern LA, Stark DP, Eden TO, Nicholson JC. Breaking down barriers: improving outcomes for teenagers and young adults with germ cell tumours. *Oncol Rev*. 2009;3:201-206.
7. Teilum G. Classification of endodermal sinus tumour (mesoblatoma vitellinum) and so-called "embryonal carcinoma" of the ovary. *Acta Pathol Microbiol Scand*. 1965;64(4):407-429.
8. Murray MJ, Huddart RA, Coleman N. The present and future of serum diagnostic tests for testicular germ cell tumours. *Nat Rev Urol*. 2016;13(12):715-725. https://doi.org10.1038/nrurol.2016.170
9. Leao R, Albersen M, Looijenga LHJ, et al. Circulating MicroRNAs, the next-generation serum biomarkers in testicular germ cell tumours: a systematic review. *Eur Urol*. 2021;80(4):456-466. https://doi.org10.1016/j.eururo.2021.06.006
10. Wu S, Aksoy M, Shi J, Houbaviy HB. Evolution of the miR-290-295/miR-371-373 cluster family seed repertoire. *PLoS One*. 2014;9(9):e108519. https://doi.org10.1371/journal.pone.0108519
11. Hayashi K, Chuva de Sousa Lopes SM, Kaneda M, et al. MicroRNA biogenesis is required for mouse primordial germ cell development and spermatogenesis. *PLoS ONE*. 2008;3(3):e1738.
12. Barroso-del Jesus A, Lucena-Aguilar G, Menendez P. The miR-302-367 cluster as a potential stemness regulator in ESCs. *Cell Cycle*. 2009;8(3):394-398. https://doi.org10.4161/cc.8.3.7554
13. Girard A, Sachidanandam R, Hannon GJ, Carmell MA. A germline-specific class of small RNAs binds mammalian piwi proteins. *Nature*. 2006;442(7099):199-202. https://doi.org10.1038/nature04917
14. Sun YH, Lee B, Li XZ. The birth of piRNAs: how mammalian piRNAs are produced, originated, and evolved. *Mamm Genome*. 2021;33(2):293-311. https://doi.org10.1007/s00335-021-09927-8
15. Williams Z, Morozov P, Mihailovic A, et al. Discovery and characterization of piRNAs in the human fetal ovary. *Cell Rep*. 2015;13(4):854-863. https://doi.org10.1016/j.celrep.2015.09.030
16. Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science*. 2007;316(5825):744-747. https://doi.org10.1126/science.1142612
17. Aravin AA, Sachidanandam R, Bourc'his D, et al. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell*. 2008;31(6):785-799. https://doi.org10.1016/j.molcel.2008.09.003
18. Kristensen DG, Nielsen JE, Jorgensen A, Skakkebaek NE, Rajpert-De Meyts E, Almstrup K. Evidence that active demethylation mechanisms maintain the genome of carcinoma in situ cells hypomethylated in the adult testis. *Br J Cancer*. 2014;110(3):668-678. https://doi.org10.1038/bjc.2013.727
19. Qiao D, Zeeman AM, Deng W, Looijenga LH, Lin H. Molecular characterization of hiwi, a human member of the piwi gene family whose overexpression is correlated to seminomas. *Oncogene*. 2002;21(25):3988-3999. https://doi.org10.1038/sj.onc.1205505

20. Ferreira HJ, Heyn H. Epigenetic loss of the PIWI/piRNA machinery in human testicular tumorigenesis. *Epigenetics*. 2014;9(1):113-118.
21. Goh WS, Falciatori I, Tam OH, et al. piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis. *Genes Dev*. 2015;29(10):1032-1044. https://doi.org10.1101/gad.260455.115
22. Ha H, Song J, Wang S, et al. A comprehensive analysis of piRNAs from adult human testis and their relationship with genes and mobile elements. *BMC Genomics*. 2014;15:545. https://doi.org10.1186/1471-2164-15-545
23. Rounge TB, Furu K, Skotheim RI, Haugen TB, Grotmol T, Enerly E. Profiling of the small RNA populations in human testicular germ cell tumors shows global loss of piRNAs. *Mol Cancer*. 2015;14:153. https://doi.org10.1186/s12943-015-0411-4
24. Gainetdinov IV, Skvortsova YV, Kondratieva SA, Klimov A, Tryakin AA, Azhikina TL. Assessment of piRNA biogenesis and function in testicular germ cell tumors and their precursor germ cell neoplasia in situ. *BMC Cancer*. 2018;18(1):20. https://doi.org10.1186/s12885-017-3945-6
25. Gillis AJ, Stoop HJ, Hersmus R, et al. High-throughput microRNAome analysis in human germ cell tumours. *J Pathol*. 2007;213(3):319-328.
26. Port M, Glaesener S, Ruf C, et al. Micro-RNA expression in cisplatin resistant germ cell tumor cell lines. *Mol Cancer*. 2011;10:52. https://doi.org1476-4598-10-52
27. de Jong J, Stoop H, Gillis AJ, et al. Further characterization of the first seminoma cell line TCam-2. *Genes Chromosomes Cancer*. 2008;47(3):185-196.
28. Pera MF, Blasco Lafita MJ, Mills J. Cultured stem-cells from human testicular teratomas: the nature of human embryonal carcinoma, and its comparison with two types of yolk-sac carcinoma. *Int J Cancer*. 1987;40(3):334-343. https://doi.org10.1002/ijc.2910400309
29. Damjanov I, Andrews PW. Ultrastructural differentiation of a clonal human embryonal carcinoma cell line in vitro. *Cancer Res*. 1983;43(5):2190-2198.
30. Murray MJ, Saini HK, Siegler CA, et al. LIN28 expression in malignant germ cell tumors downregulates let-7 and increases oncogene levels. *Cancer Res*. 2013;73(15):4872-4884. https://doi.org/10.1158/0008-5472.CAN-12-2085[pii]
31. Masters JR, Thomson JA, Daly-Burns B, et al. Short tandem repeat profiling provides an international reference standard for human cell lines. *Proc Natl Acad Sci U S A*. 2001;98(14):8012-8017.
32. Thery C, Witwer KW, Aikawa E, et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the international society for extracellular vesicles and update of the MISEV2014 guidelines. *J Extracell Vesicles*. 2018;7(1):1535750. https://doi.org10.1080/20013078.2018.1535750
33. Palmer RD, Barbosa-Morais NL, Gooding EL, et al. Pediatric malignant germ cell tumors show characteristic transcriptome profiles. *Cancer Res*. 2008;68(11):4239-4247.
34. Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat Methods*. 2008;5(3):235-237. https://doi.org10.1038/nmeth.1184
35. Craig DW, Pearson JV, Szelinger S, et al. Identification of genetic variants using bar-coded multiplexed sequencing. *Nat Methods*. 2008;5(10):887-893. https://doi.org10.1038/nmeth.1251
36. Murray MJ, Scarpini CG, Coleman N. A circulating MicroRNA panel for malignant germ cell tumor diagnosis and monitoring. *Methods Mol Biol*. 2021;2195:225-243. https://doi.org10.1007/978-1-0716-0860-9_15
37. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal*. 2011. https://doi.org/10.14806/ej.17.1.200
38. Andrews S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. ScienceOpen; 2010.
39. Davis MP, van Dongen S, Abreu-Goodger C, Bartonicek N, Enright AJ. Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods*. 2013;63(1):41-49. https://doi.org10.1016/j.ymeth.2013.06.027
40. Leung YY, Kuksa PP, Amlie-Wolf A, et al. DASHR: database of small human noncoding RNAs. *Nucleic Acids Res*. 2016;44(D1):D216-D222. https://doi.org10.1093/nar/gkv1188
41. Kuksa PP, Amlie-Wolf A, Katanic Z, Valladares O, Wang LS, Leung YY. DASHR 2.0: integrated database of human small non-coding RNA genes and mature products. *Bioinformatics*. 2019;35(6):1033-1039. https://doi.org10.1093/bioinformatics/bty709
42. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res*. 2019;47(D1):D155-D162. https://doi.org10.1093/nar/gky1141
43. Vitsios DM, Enright AJ. Chimira: analysis of small RNA sequencing data and microRNA modifications. *Bioinformatics*. 2015;31(20):3365-3367. https://doi.org10.1093/bioinformatics/btv380
44. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. https://doi.org10.1186/s13059-014-0550-8
45. La Greca A, Scarafia MA, Hernandez Canas MC, et al. PIWI-interacting RNAs are differentially expressed during cardiac differentiation of human pluripotent stem cells. *PLoS One*. 2020;15(5):e0232715. https://doi.org10.1371/journal.pone.0232715
46. Wang J, Zhang P, Lu Y, et al. piRBase: a comprehensive database of piRNA sequences. *Nucleic Acids Res*. 2019;47(D1):D175-D180. https://doi.org10.1093/nar/gky1043
47. Wang J, Shi Y, Zhou H, et al. piRBase: integrating piRNA annotation in all aspects. *Nucleic Acids Res*. 2022;50(D1):D265-D272. https://doi.org10.1093/nar/gkab1012
48. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930. https://doi.org10.1093/b0ioinformatics/btt656
49. Stephens M. False discovery rates: a new deal. *Biostatistics*. 2017;18(2):275-294. https://doi.org10.1093/biostatistics/kxw041
50. 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;57:289-300.
51. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11(10):R106. https://doi.org10.1186/gb-2010-11-10-r106
52. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. https://doi.org10.1093/nar/gkv007
53. Hutson G. *FeatureTerminatoR: Feature Selection Engine to Remove Features with Minimal Predictive Power [R package]*. CRAN; 2021. https://CRAN.R-project.org/package=FeatureTerminatoR
54. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human MicroRNA targets. *PLoS Biol*. 2004;2(11):e363. https://doi.org10.1371/journal.pbio.0020363
55. Rajan KS, Velmurugan G, Gopal P, et al. Abundant and altered expression of PIWI-Interacting RNAs during cardiac hypertrophy. *Heart Lung Circ*. 2016;25(10):1013-1020. https://doi.org10.1016/j.hlc.2016.02.015
56. Zuo Y, Liang Y, Zhang J, et al. Transcriptome analysis identifies piwi-interacting RNAs as prognostic markers for recurrence of prostate cancer. *Front Genet*. 2019;10:1018. https://doi.org10.3389/fgene.2019.01018
57. Wang A, Liu J, Zhuang X, et al. Identification and comparison of piRNA expression profiles of exosomes derived from human stem cells from the apical papilla and bone marrow mesenchymal stem cells. *Stem Cells Dev*. 2020;29(8):511-520. https://doi.org10.1089/scd.2019.0277

58. Singh G, Mallick B. Predicting sequence and structural features of effective piRNA target binding sites. *J Mol Recognit*. 2022;35(4):e2949. https://doi.org10.1002/jmr.2949
59. Franceschini A, Szklarczyk D, Frankild S, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2013;41:D808-D815. https://doi.org10.1093/nar/gks1094
60. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30. https://doi.orggkd027
61. Murray MJ, Saini HK, van Dongen S, et al. The two most common histological subtypes of malignant germ cell tumour are distinguished by global microRNA profiles, associated with differential transcription factor expression. *Mol Cancer*. 2010;9:290. https://doi.org/10.1186/1476-4598-9-290
62. De Fazio S, Bartonicek N, Di Giacomo M, et al. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature*. 2011;480(7376):259-263. https://doi.org10.1038/nature10547
63. International Germ Cell Consensus Classification: a prognostic factor-based staging system for metastatic germ cell cancers. International Germ Cell Cancer Collaborative Group. *J Clin Oncol*. 1997;15(2):594-603.
64. Gillessen S, Sauve N, Collette L, et al. Predicting outcomes in men with metastatic nonseminomatous germ cell tumors (NSGCT): results from the IGCCCG update consortium. *J Clin Oncol*. 2021;39(14):1563-1574. https://doi.org10.1200/JCO.20.03296
65. Murray MJ, Coleman N. MicroRNA dysregulation in malignant germ cell tumors: more than a biomarker? *J Clin Oncol*. 2019;37(16):1432-1435. https://doi.org10.1200/JCO.19.00578
66. Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*. 2006;313(5785):320-324. https://doi.org10.1126/science.1129333
67. Houwing S, Kamminga LM, Berezikov E, et al. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell*. 2007;129(1):69-82. https://doi.org10.1016/j.cell.2007.03.026
68. Roovers EF, Rosenkranz D, Mahdipour M, et al. Piwi proteins and piRNAs in mammalian oocytes and early embryos. *Cell Rep*. 2015;10(12):2069-2082. https://doi.org10.1016/j.celrep.2015.02.062
69. Jeyapalan JN, Noor DA, Lee SH, et al. Methylator phenotype of malignant germ cell tumours in children identifies strong candidates for chemotherapy resistance. *Br J Cancer*. 2011;105(4):575-585. https://doi.org10.1038/bjc.2011.218
70. Amatruda JF, Ross JA, Christensen B, et al. DNA methylation analysis reveals distinct methylation signatures in pediatric germ cell tumors. *BMC Cancer*. 2013;13:313. https://doi.org10.1186/1471-2407-13-313
71. Rouget C, Papin C, Boureux A, et al. Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. *Nature*. 2010;467(7319):1128-1132. https://doi.org10.1038/nature09465
72. Watanabe T, Lin H. Posttranscriptional regulation of gene expression by piwi proteins and piRNAs. *Mol Cell*. 2014;56(1):18-27. https://doi.org10.1016/j.molcel.2014.09.012
73. Gou LT, Dai P, Yang JH, et al. Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell Res*. 2014;24(6):680-700. https://doi.org10.1038/cr.2014.41
74. Hashim A, Rizzo F, Marchese G, et al. RNA sequencing identifies specific PIWI-interacting small non-coding RNA expression patterns in breast cancer. *Oncotarget*. 2014;5(20):9901-9910. https://doi.org10.18632/oncotarget.2476
75. Tan L, Mai D, Zhang B, et al. PIWI-interacting RNA-36712 restrains breast cancer progression and chemoresistance by interaction with SEPW1 pseudogene SEPW1P RNA. *Mol Cancer*. 2019;18(1):9. https://doi.org10.1186/s12943-019-0940-3
76. Liu Y, Dou M, Song X, et al. The emerging role of the piRNA/piwi complex in cancer. *Mol Cancer*. 2019;18(1):123. https://doi.org10.1186/s12943-019-1052-9
77. Zhang P, Kang JY, Gou LT, et al. MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes. *Cell Res*. 2015;25(2):193-207. https://doi.org10.1038/cr.2015.4
78. Peng L, Song L, Liu C, et al. piR-55490 inhibits the growth of lung carcinoma by suppressing mTOR signaling. *Tumour Biol*. 2016;37(2):2749-2756. https://doi.org10.1007/s13277-015-4056-0
79. Qin G, Mallik S, Mitra R, et al. MicroRNA and transcription factor co-regulatory networks and subtype classification of seminoma and non-seminoma in testicular germ cell tumors. *Sci Rep*. 2020;10(1):852. https://doi.org10.1038/s41598-020-57834-w

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Laidlaw S, Alonso-Crisostomo L, Bailey S, et al. Small non-coding RNA sequencing reveals global dysregulation of piwi-interacting RNA (piRNA) expression in gonadal malignant germ cell tumours. *Andrology*. 2022;1-18. https://doi.org/10.1111/andr.13312