PLOS ONE

PIWI-interacting RNAs are differentially expressed during cardiac differentiation of human pluripotent stem cells

Alejandro La Greca, María Agustina Scarafía⬤, María Clara Hernández Cañás⬤, Nelba Pérez, Sheila Castañeda, Carolina Colli, Alan Miqueas Möbbs, Natalia Lucía Santín Velazque, Gabriel Neiman, Ximena Garate, Cyntia Aban, Ariel Waisman, Lucía Natalia Moro, Gustavo Sevlever, Carlos Luzzani, Santiago Gabriel Miriuka⬤*

LIAN, Fleni Institute-CONICET, Buenos Aires, Argentina

* smiriuka@fleni.org.ar

Abstract

PIWI-interacting RNAs (piRNAs) are a class of non-coding RNAs initially thought to be restricted exclusively to germline cells. In recent years, accumulating evidence has demonstrated that piRNAs are actually expressed in pluripotent, neural, cardiac and even cancer cells. However, controversy remains around the existence and function of somatic piRNAs. Using small RNA-seq samples from H9 pluripotent cells differentiated to mesoderm progenitors and cardiomyocytes we identified the expression of 447 piRNA transcripts, of which 241 were detected in pluripotency, 218 in mesoderm and 171 in cardiac cells. The majority of them originated from the sense strand of protein coding and lncRNAs genes in all stages of differentiation, though no evidences of amplification loop (ping-pong) were found. Genes hosting piRNA transcripts in cardiac samples were related to critical biological processes in the heart, like contraction and cardiac muscle development. Our results indicate that these piRNAs might have a role in fine-tuning the expression of genes involved in differentiation of pluripotent cells to cardiomyocytes.

Introduction

Differentiation of pluripotent stem cells (PSC) to cardiomyocytes (CM) was first reported shortly after the characterization of embryonic stem cells (ESC) [1]. Initially, differentiation was non-specific and spontaneously achieved, but in the last 10 years upgraded protocols have been developed significantly improving efficiency and reproducibility in cardiac differentiation [2–4]. These protocols are based in sequentially adding factors (morphogens) and/or inhibitors that modulate Wnt/β-catenin signaling pathways in pluripotent cells. PSC-based models undergo epithelial-to-mesenchymal transition to an early mesoderm progenitor cell (MPC) [2, 5] followed by further committing to cardiac mesoderm and later cardiac progenitor cells (CPC), which may eventually adopt more especialized features. Though this is arguably similar to *in vivo* embryo development, they recapitulate hallmark features of

differentiation thus becoming well suited tools for disease modelling, drug screening and potential cell-based therapies.

Like many other developmental processes, changes associated with differentiation to CM are tightly regulated. Only recently, and mostly due to the advent of next generation sequencing technologies, the scientific community is unveiling the complex regulatory networks governing the shifts in gene expression profiles. Non-coding RNAs (ncRNAs) are critical players in these networks, regulating almost all cellular processes including proliferation, differentiation and death [6, 7]. Although microRNAs (miRNAs) are the most extensively studied in a wide variety of organisms [8–11], other ncRNAs have been identified such as long non-coding RNAs (lncRNAs), small interfering RNAs (siRNAs), circular RNAs (circRNAs) and PIWI-interacting RNAs (piRNAs). Much has been published about these ncRNAs, though piRNAs is one of the least understood. Thought to be initially confined almost exclusively to germinal cell lines [12], piRNAs gained much attention primarily because of an increasing amount of evidence demonstrating that these ncRNAs are not only expressed in somatic cells but they actively participate in gene regulation as well [13–16].

Expression of piRNAs was first described as negatively regulating transposition of repetitive elements thus protecting genome integrity and favouring self-renewal [12, 17]. Reports in numerous organisms showed that they exert their regulatory function through binding a specific clade of the Argonaute (AGO) family -namely PIWI proteins-, resulting in an association which resembles the well-known AGO/miRNA complex [8, 18–20]. Unlike miRNAs, piRNAs are primarily biosynthesized as single-stranded long precursors which are then clived into the 24–34 nucleotide-long mature forms in a Dicer-independent manner [12, 21, 22]. They show a bias for uridine (U) rediudes in 5' ends together with a 2'O-methyl modification at their 3' ends. Germ line piRNAs were also found to be synthesized through a secondary pathway named the Ping-Pong amplification loop, which increases levels of primary piRNAs using target mRNAs as intermediary molecules for processing new piRNA precursors [20, 23]. Of note, these mechanisms seem to be highly conserved across species [18, 24].

In the last few years, many studies have proposed an active participation of PIWI/piRNAs complexes in diverse and critical pathways such as neural development or body regeneration of lower eukaryotes [25, 26]. Furthermore, recent work demonstrated a positive correlation between altered piRNA expression profiles and clinically relevant pathologies. The involvement of specific piRNAs in regulating mRNAs levels of genes related to Alzheimer's disease was described in 2017 [27], while other groups implicated piRNAs in cardiac function and regeneration through modulation of AKT pathway [28, 29]. However, great controversy still remains around expression, function and biosynthetic pathways of somatic piRNAs. Particularly, the potential role of piRNAs in differentiation of pluripotent stem cells to cardiomyocytes has not been formally addressed. Using small RNAseq data generated in our laboratory [11] we characterized the expression profile of small RNAs consistent with piRNAs in three stages of cell differentiation from pluripotency (day 0) to mesoderm (day 3.5) and then contractile cardiocytes (day 21). Results presented here provide evidences supporting the existence of somatic piRNA transcripts and their stage-specific pattern as a mechanism for potentially fine-tuning gene expression during cell differentiation.

Materials and methods

Small RNAseq data

Data samples used in this work (PSC: H9 human embryonic stem cells, MPC: early mesoderm progenitor and CPC:cardiomyocytes) were generated in our laboratory following previously described protocols and are available under accession code GSE108021. Briefly, H9 cells

(H9-hTnnTZ-pGZ-D2 obtained from WiCell) were routinely maintained in co-culture with irradiated primary mouse embryonic fibroblasts. Mesoderm induction (MPC population) was performed by initially seeding cells with mTeSR (Stemcell Technologies, Vancouver, Canada) on plates coated with Geltrex (Thermo Fisher Scientific, MA, USA) and then switching to [2] StemPro®-34 SFM (Thermo Fisher Scientific) supplemented with Activin A only at the first day, BMP4, VEGF and bFGF (Thermo Fisher Scientific) for 3.5 days. At this point, mesoderm progenitors were isolated by FACS with anti-human CD326 (A15782, Molecular Probes, Thermo Fisher Scientific) and anti-human CD56 (564488, BD Bioscience, Franklin Lakes, NJ, USA). CPC population was obtained by formation of embryoid bodies with H9 cells using BMP4, bFGF and Activin A in StemPro®-34 followed by addition of VEGF and Wnt inhibitor, IWR-1. Libraries for small RNA sequencing were prepared with 200 ng of RNA using NEB-Next Small RNA Library Prep Set with modified adaptors and primers compatible for Illumina (New England Biolabs). Single end sequencing was carried out at the TCGB Resources (UCLA Path and Lab Med) using an Illumina HiSeq 2500. Culture conditions and sequencing of small RNAs for these samples are more extensively described in [11].

External data

Human testis small RNA-seq samples from two men of 54 and 37 years old (GSE88414 and GSE88124, respectively) and H1-derived neural progenitor cells (GSM1296459 and GSM1296460) were downloaded from ENCODE (encodeproject.org). RNA-seq data (counts per transcript) from H1 and H9 cardiac differentiation protocols can be found under GEO accesion GSE85331.

Data processing and analyses

Adapters were removed from raw sequencing reads with cutadapt (v1.9.1) keeping reads with a minimum of 20 and up to 50 nt in length. Quality checked (FastQC) processed reads were mapped to human reference genome (GRCh38/hg38) using STAR aligner (v2.5.3a [30]) under mostly default parameters. Mapped reads in output SAM/BAM files were filtered by read length ($23 < RL < 35$) with samtools and custom awk scripting. Resulting reads were intersected (bedtools v2.27.1 [31]) to ncRNAs in a strand specific manner (DASHR [32]) to remove potential misleading alignments. Raw counts on piRNAs were determined with htseq-count matching mapped reads to piRNA coordinates downloaded from piRBase [33]. Counts were then fed into DESeq2 for differential expression analysis ($p < 0.05$ and fdr < 0.1). Soft clustering methods were implemented with R package Mfuzz (v2.42.0 [34]) using parameter m = 1.15. In parallel to our customized pipeline, tools for ping-pong signature detection like ssviz R package and PingPongPro [35] were run following recommendations from authors. Graphics and statistical analyses were performed in R software and deepTools [36]. Further details on custom code is available at https://github.com/sgmiriuka/piRNA_custom_scripting.

Monolayer-directed differentiation protocol

H9 human embryionic stem cells were maintained on Geltrex™ (Thermo Fisher Scientific)-coated dishes using E8 flex defined medium (Thermo Fisher Scientific), replacing it each day. Cells were detached with TrypLE™ Select 1X (Thermo Fisher Scientific) every 4–5 days depending on density. Cardiac differentiation experiments were performed as previously described in [4] with minor adaptations. Briefly, 2.5-3x10^5 cells were seeded on Geltrex-coated 24-well plates with mTeSR™ medium (Stemcell Technologies) and cultured for 72h, replacing medium every 24h. Then, cells were treated with CHIR99021 (SIGMA, St Louis, MO, USA) for 24h at a final concentration of 12μM in RPMI medium (Thermo Fisher Scientific) supplemented with

B27 without insulin (Thermo Fisher Scientific). Untreated pluripotent cells were collected as day zero (D0) of the protocol. On day 3, differentiating cells were treated with IWP2 (Tocris Bioscience, Bristol, UK) at a final concentration of 5μM for 48h. Between days 3 and 4, cells that had not been treated with IWP2 were collected as MPC population (D3.5). From day 12 to 18, media was replaced with low-glucose RPMI (Thermo Fisher Scientific) supplemented with B27 (Thermo Fisher Scientific) and lactate solution (SIGMA) to enrich for cardiac progenitors and later switched back to RPMI supplemented with B27 until day 21 (CPC). When indicated, the protocol was extended to day 30.

Reverse transcription of putative piRNA

To obtain cDNA from piRNA transcripts we adapted a previously described methodology employed for miRNA detection and amplification [37]. Briefly, stem-loop retrotranscription (RT) primers were generated using 6-8nt from the 3' end of every piRNA of interest. Each RT reaction was performed with a maximum of 10 different stem-loop primers including one for RNU6B and hsa-miR-302b as controls. An additional modification was implemented at this step which involved an RT protocol consisting of 40 cycles of 16˚C for 2min, 42˚C for 1min and 50˚C for 1s (adapted from Megaplex stem loop RT system [38]). Detection by qPCR involved forward primers matching the sequence of target piRNAs and a reverse universal primer complementary to the stem-loop region in the second strand synthesized during PCR reaction. Expression data were normalized to RNU6B and analized by ANOVA with randomized block design followed by Tukey's multiple comparison test.

Real time PCR

Total RNA was prepared with TRI-Reagent (SIGMA) following manufacturer's instructions and then reverse transcribed using MMLV reverse transcriptase (Promega, Madison, WI, USA) and random primers for detection of polyadenilated transcripts. Quantitative real time PCR (qPCR) was performed in a StepOne Real Time PCR system (Applied Biosystems, Foster City, CA, USA). Expression was normalized to the geometrical mean of *HPRT1* and *RPL7* housekeeping genes and log2 transformed. Statistical significance for qPCR results was analyzed by ANOVA with randomized block design followed by Tukey's multiple comparison test. Primer sequences are available on request.

Results

Detection and characterization of piRNA

Detection of piRNAs was conducted on small RNAseq samples from three independent experiments consisting of pluripotent stem cells (PSC, day 0), early mesoderm progenitor cells (MPC, day 3.5) and cardiac progenitor cells (CPC, day 21). After aligning reads to human reference genome (hg38), we found that more than half of mapped reads were 20 to 23 nucleotides (nt) long, where the abundant miRNAs were included (Fig 1A, [11]). Considering that the average length of piRNAs in mammals ranges between 24 and 34 nt [12, 22, 39], mapped reads were filtered by length to accommodate to this restriction. Nearly 50–70% of mapped reads were removed from the samples after this initial processing step (S1A and S1B Fig). Then, employing a similar approach as previously published work [15], we filtered out any read that mapped on ncRNAs (DASHR [32]) besides piRNAs (S1C Fig) given that previous publications emphasized on the fact that many identified piRNAs were actually fragments of other types of ncRNAs [40]. Approximately 5–20% of initial mapped reads remained after this
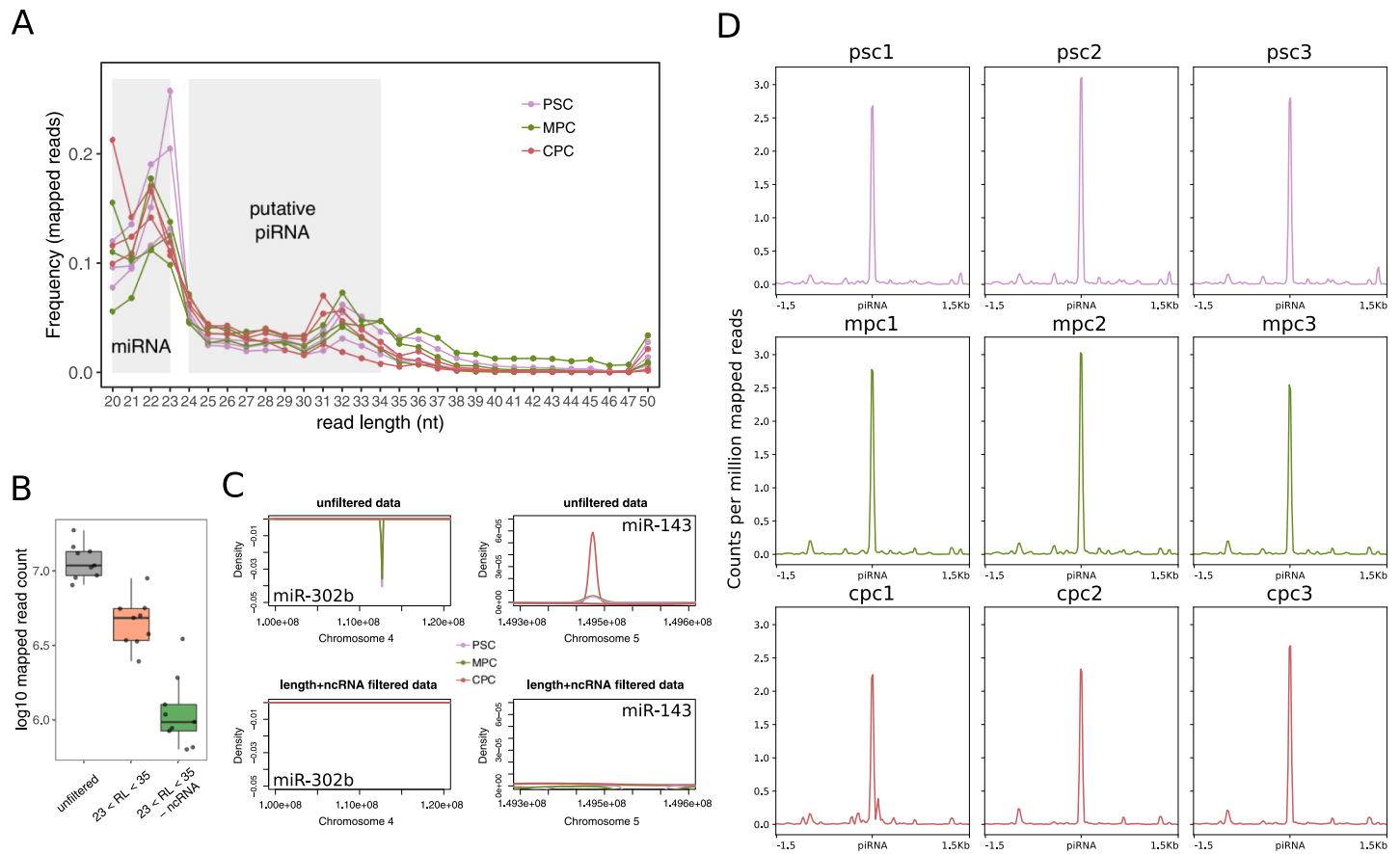
Fig 1. Detection of piRNA transcripts in small RNAseq samples. A) Length frequency of unfiltered mapped reads in three samples from pluripotent cells (PSC), three from mesoderm progenitor cells (MPC) and three from cardiac cells (CPC). Grey areas denote the size range for both microRNAs (miRNA) and putative piwi-interacting RNAs (piRNA). The color key used is indicated in top right corner of the plot. B) Number of mapped reads for all nine samples before processing (unfiltered, grey box), after filtering by read length (23 < RL < 35, coral box) and removing non coding RNA other than piRNAs (-ncRNA, green box). C) Distribution of mapped reads as a function of density on a fragment of chromosome 4 (chr4:100,000,000–120,000,000; −) which includes pluripotency miR-302b and a fragment of chromosome 5 (chr5:149,300,000–149,600,000, +) including cardiac miR-143 for unprocessed alignments (top panels, unfiltered data) and fully processed samples (bottom panels, length+ncRNA filtered). Alignment files from each experimental replicate were merged into one. Color key for the density curves is shown in the graph. D) Analysis of coverage on all putative piRNA loci available in piRbase for fully processed normalized (counts per million) samples.

https://doi.org/10.1371/journal.pone.0232715.g001

step (S1B Fig). Importantly, all nine aligned samples behaved similarly to both filtering steps (Fig 1B), reflecting consistency among experimental replicates.

To verify the elimination of potential misleading contaminants in fully processed alignments, we analyzed the distribution of mapped reads over two well-characterized miRNAs, pluripotency-associated miR-302b and cardiac-expressed miR-143 [11]. As expected, expression of miR-302b was evident in unfiltered data of PSC and MPC populations while miR-143 showed appreciable coverage in unfiltered data of CPC (Fig 1C, top panels). No signal was detected for either of the two genes in processed alignments (Fig 1C, bottom panels). However, these samples showed a strong and sharp coverage signal on known piRNA loci (Fig 1D) confirming that the pipeline employed successfully enriched for reads mapping to these known piRNAs. Henceforth, all analyses were performed on fully processed alignments unless explicitly specified otherwise.

Sequence analysis of reads mapping to known piRNA loci showed that all samples but MPC bore a slight bias for 5' uridine residues (Fig 2A). We corroborated our proceedings by
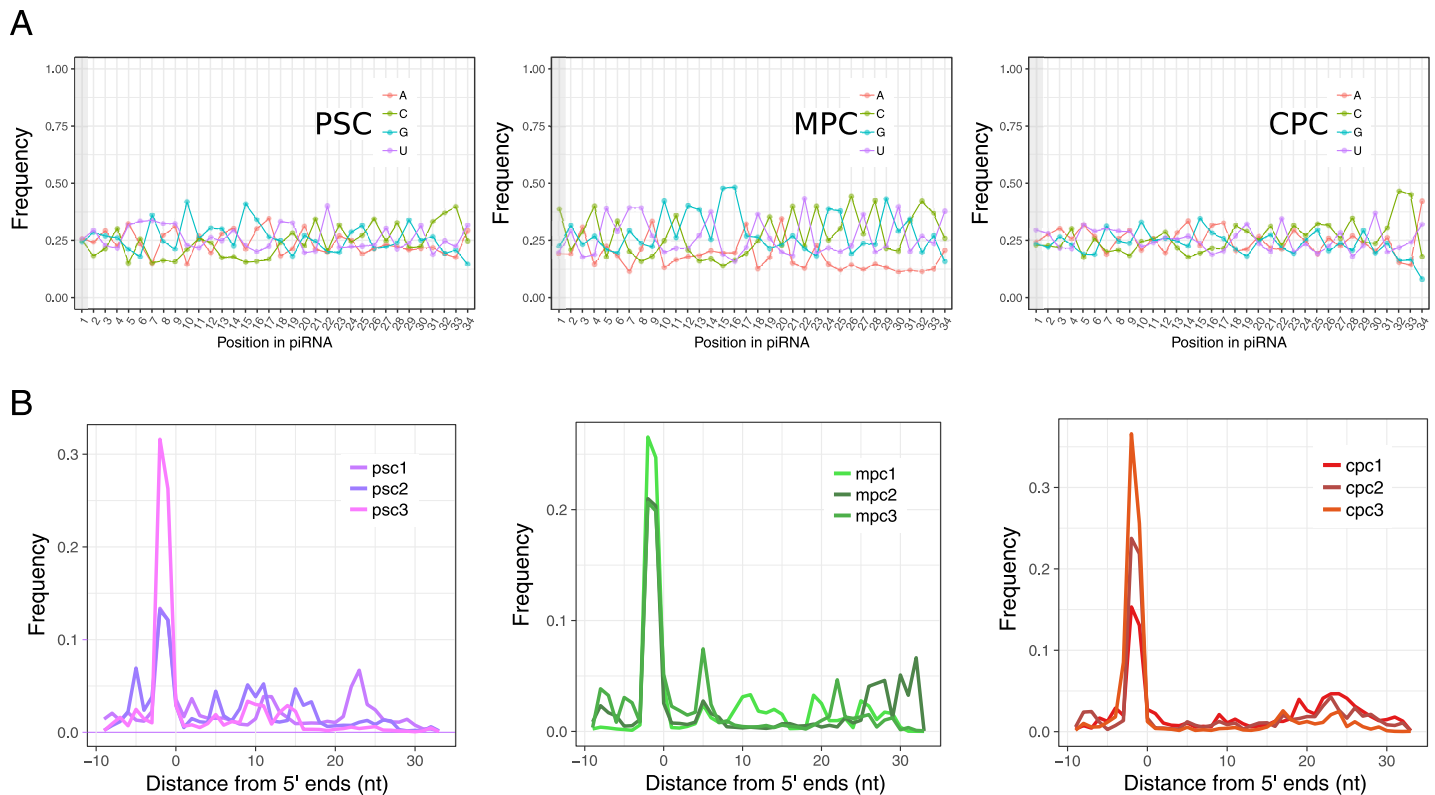
Fig 2. Characterization of reads mapped to putative piRNA loci. A) Frequency of bases per position in reads mapped to piRNA loci after fully processing alignments from PSC, MPC and CPC samples (merged replicates). Position 1 is marked by a vertical grey bar. B) Frequency profile showing overlap between reads mapped to known piRNA loci in sense orientation and complementary reads using an independent approach (ssviz package).

employing the pipeline described above on two small RNA-seq samples from human testis downloaded from the ENCODE project. Indeed, there was a marked preference for uridine at 5' ends in testis samples (S2A and S2B Fig), which suggest that biosynthesis mechanisms of putative piRNAs in our model are only partially conserved compared to germline cells. [26]. In addition, no secondary piRNA production was detected in any of the replicates of our samples given that we did not found evidences of the characteristic 10 nt overlap (ping-pong signature) between 5' ends of sense and antisense mapped reads (Fig 2B).

Expression of piRNAs during cardiac differentiation

To study the expression profile of known piRNAs in differentiating PSC we kept those with an average count among replicates higher than or equal to 3. Normalization by library depth showed equivalent distribution of relative quantifications between samples (S3A Fig), enabling confident identification of 447 piRNAs considering the three cell differentiation stages investigated (S1 Table). Despite some differences between replicates, each stage of cell differentiation was categorically defined by a specif piRNA expression profile (Fig 3A) which was also reflected in Principal Component Analysis results (S3C Fig). These identifying profiles preferentially aggregated PSC and MPC together indicating a greater resemblance between samples of these two cell populations than with CPC.

Of the 447 identified piRNAs, 241 were expressed in PSC, 218 in MPC and 171 in CPC (S1 Table). Differential expression analysis revealed only 30 genes with significant shifts in RNA
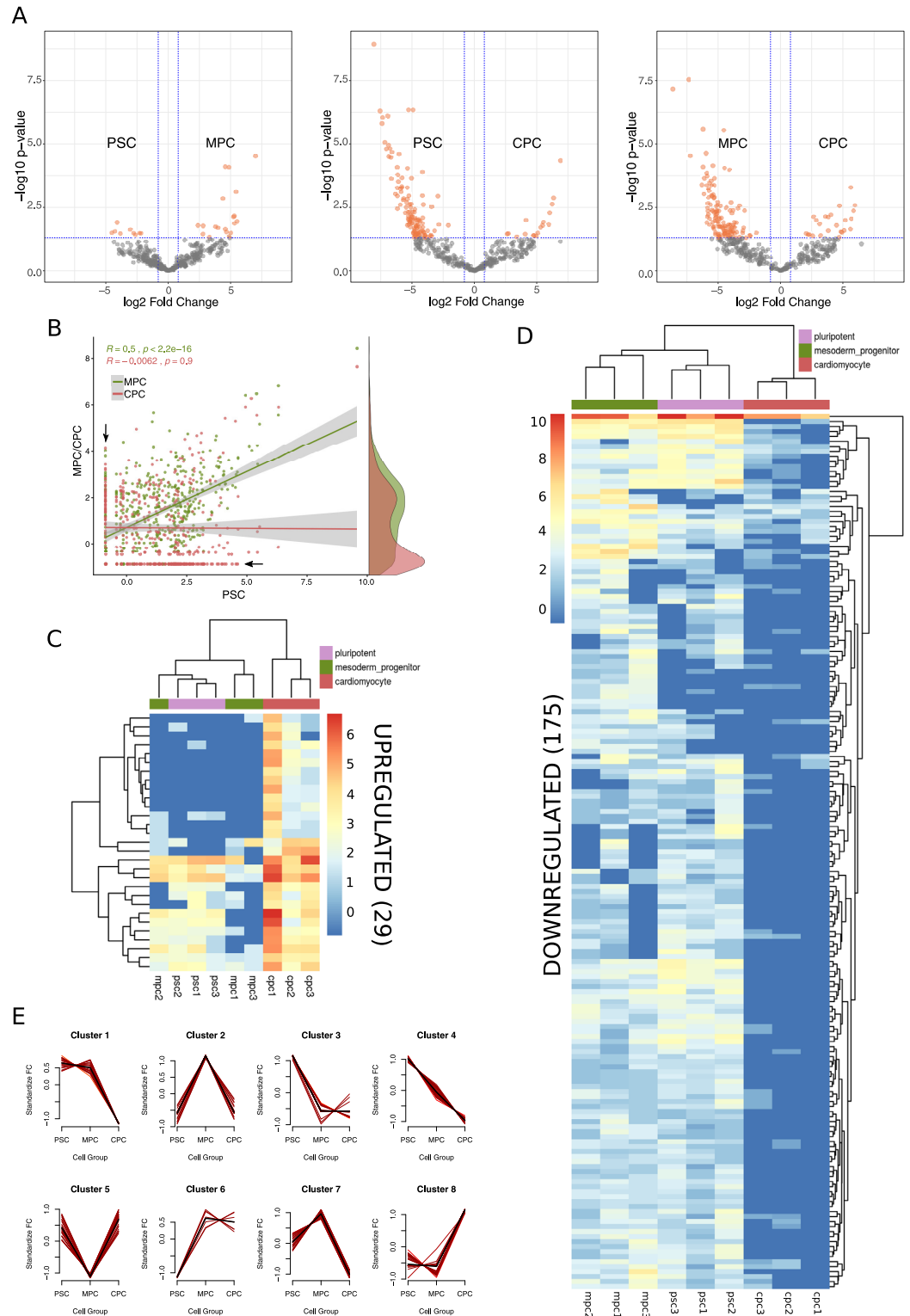
Fig 3. piRNA expression profile in differentiating pluripotent stem cells. A) Differential expression analysis performed on raw counts with DESeq2 package. Significantly different expression values (-1 > = log2FC > = 1; -log10p-value > 1.30)) are represented as orange dots in the volcano plots for the three possible comparisons: MPC vs. PSC (left), CPC vs. PSC (center) and CPC vs. MPC (right). B) Scatter plot of normalized expression data showing Pearson correlation analysis on MPC (green dots) and CPC (red dots) versus PSC. Marginal density plots to the right denote areas of highly abundant data

and black arrows mark the positions of the most DE piRNA transcripts. C) Heatmap shows normalized counts of upregulated piRNA transcripts in CPC considering the three cell populations (PSC, MPC, CPC). Dendrograms resulted from running hard unsupervised clustering algorithms on piRNA (left) and samples (top). D) Heatmap as in c showing downregulated piRNAs in CPC. E) Implementation of soft clustering algorithms (R package MFuzz) produces eight distinct patterns (clusters 1 to 8) of piRNA expression.

levels (-1 > log2FC > 1; -log10p-value > 1.30) for the comparison between PSC and MPC, while 137 were differentially expressed (DE) between PSC and CPC and 153 between MPC and CPC (Fig 3A). Of the total 447 piRNAs, 204 were found to be DE with respect to CPC, 86 of which were shared by MPC and PSC (S3D Fig and S1 Table). These results were consistent with correlation analysis that showed a higher Pearson coefficient for the MPC-PSC pair (R = 0.5, p<2.2e-16) than for CPC-PSC (R = -0.0062, p = 0.9) (Fig 3B), suggesting that PSC bear a greater resemblance to MPC than to CPC not only in the identity of expressed piRNAs, but in their abundance as well. Upregulated piRNAs accounted for 14% of total DE piRNAs in CPC (Fig 3C and S1 Table), far fewer than the downregulated piRNAs (Fig 3D and S1 Table). We validated several piRNA transcripts (piR-4403262, piR-4424378, piR-4193743, piR-2715002, piR-1551388, piR-4091280, piR-2519215, piR-1332287, piR-2413094 and piR-1399886) by qPCR in two independent sets of samples from H9 pluripotent cells differentiated to CM, corroborating our detection pipeline and subsequent DE analysis (S3F Fig).

Differentially expressed piRNAs ranked among the top expressing piRNAs. This is probably due to the fact that highly expressed genes are inherently less sensitive to inter-replicate noise, hence more likely to return a lower p-value for contrasts. Thus, in order to investigate potential patterns underlying expression data which might have been masked from differential expression analysis, we implemented a soft clustering algorithm to data. This approach returned 8 different patterns of piRNA expression (Fig 3E and S2 Table), or Expression Clusters (EC), that reflected two dynamically relevant tendencies: downregulation of piRNAs towards cardiac differentiation (cluster 1 to 4 and 7) and upregulation of piRNAs towards cardiac differentiation (cluster 5, 6 and 8). The former, as was previously observed, encompassed the majority of DE piRNAs. Regardless of the condition (up or downregulated) of piRNAs in CPC, it was clear that a fraction of piRNAs sustained early change (PSC to MPC) while others shifted later in the differentiation process (MPC to CPC). Interestingly, expression profile of human *PIWI* genes changed between day 0 and 21 of differentiation (Fig 4A). Although no conclusive results were obtained for *HIWI* (no expression detected) and *HIWI3* (no significant differences between conditions), *HILI* and *HIWI2* were distinctively and oppositely regulated towards day 21 suggesting that there might be a connection between these *PIWI* genes and cardiac piRNAs. While *HILI* was downregulated all trough differentiation, *HIWI2* showed an increment in mRNA levels. Specific markers of pluripotency and cardiomyocytes were measured at these timepoints to corroborate cell identity (Fig 4B). In addition, analysis of H9 and H1 published RNA-seq data validated downregulation of *HILI* and upregulation of *HIWI2* with cardiac differentiation (S4 Fig, [41]).

Genome distribution of expressed piRNAs

Identified piRNAs were distributed rather uniformly throughout the nuclear genome (Fig 5A), except in chromosome Y for which no data was available given the XX karyotype of H9 embryonic stem cells used in this work. Moreover, Expression Clusters did not seem to follow any particular arrangement in these chromosomes as well (Fig 5A, center of circular plot). Inclusion of the mitochondrial chromosome (chrM) in the analysis revealed that 90 of 447 piRNAs originated from the mitochondrial genome (Fig 5B). This was consistent with previous work
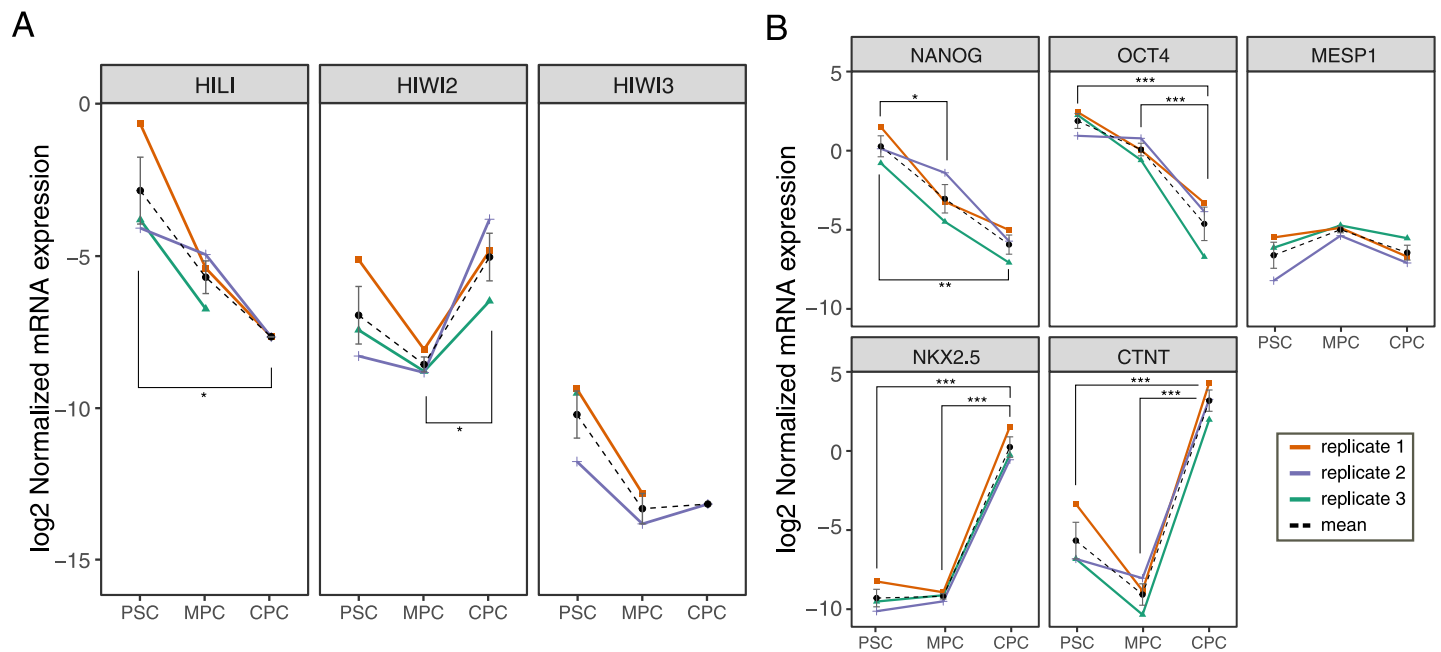
Fig 4. Expression of human PIWI genes in cardiac differentiation. A) Transcript levels of HILI (PIWIL2), HIWI2 (PIWIL4) and HIWI3 were measure by qPCR in H9 pluripotent cells (PSC, D0) and after differentiation to mesoderm (MPC, D3.5) and cardiac progenitors (CPC, D21) using a monolayer-directed protocol previously described (see Materials and methods [4]). B) Stage-specific markers were evaluated by qPCR in samples from A. Pluripotency: *NANOG* and *OCT4*; mesoderm: *MESP1*; cardiomyocytes: *NKX2.5* and *CTNT*. Normalized mRNA levels (log2) of three independent biological replicates are shown for each gene together with the mean of the three replicates ± sem (dashed black line). Statistical significance for the multiple comparisons is indicated in the plot (* $p < 0.05$; *** $p < 0.001$).

in human somatic cancer cell lines reporting the synthesis of piRNAs from mitochondrial genome ([42]). In fact, the chrM was the major contributor of expressed piRNAs in our samples and was mostly dominated by three EC: a) cluster 3, with piRNAs highly expressed in PSC; b) cluster 8, with piRNAs highly expressed in CPC; c) cluster 5, with piRNAs highly expressed in both PSC and CPC.

We corroborated this finding by evaluating the distribution of mapped reads over chrM, and thus eliminated the possibility of errors during counting of reads per transcript (S5A Fig). Despite our pipeline for identification of expressed piRNAs filtered out all reads that mapped to ncRNAs -other than piRNAs- using DASHR database, 90% of mitochondrial piRNAs (81 out of 90) mapped directly to tRNA and rRNA annotations (GENCODE v29) (S5B Fig). DASHR database showed only one annotation in chrM (LSU-rRNA) that corresponded to the large ribosomal subunit RNR2 (S5C Fig). This was not the case in the nuclear genome where no piRNAs were found to map on rRNAs and tRNAs annotated in GENCODE database (S5D Fig). Nonetheless, piRNAs identified in length-filtered data (initial step of filtering, Fig 1B) did not map to nuclear rRNA or tRNA annotations from GENCODE to begin with (S5D Fig), suggesting that this step was sufficient enough to remove reads mapping on them.

Regardless of the chromosome distribution, identified piRNAs localized preferentially on gene annotations (Fig 5C). PSC samples showed that 88.5% of piRNAs were generated from gene features, while the percentage was higher in MPC and CPC samples, with 96.2 and 95.5% respectively.

Protein coding and lncRNA genes hosting piRNAs

Further analysis on genomic distribution of identified piRNAs revealed that nearly 65% of those intersected to gene features originated from coding (53%) and long non-coding (12%)
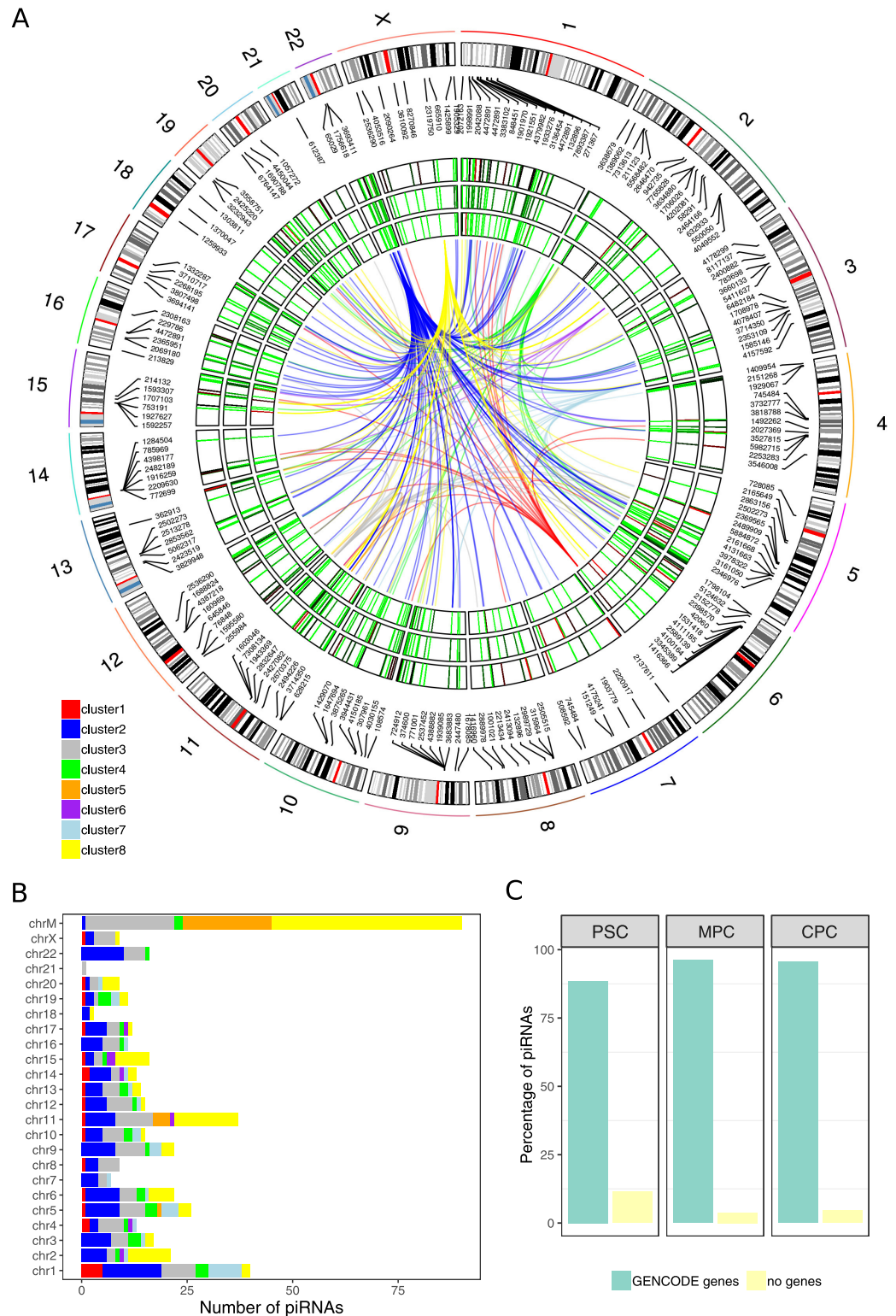
Fig 5. Distribution of piRNA transcripts and expresion clusters in the genome. A) Normalized expression of piRNA transcripts from averaged samples of PSC (outer track), MPC (middle track) and CPC (inner track) displayed lengthwise in a circular representation of human autosomal chromosomes (1 to 22) and chromosome X. Level of expression is depicted as heatmaps inside the tracks and follow a red-green scale for high expression and low expression respectively. The center of the plot shows links between piRNA transcripts that belong to the same Expression Cluster (EC); color key

for the links appears at the bottom left corner of the plot. B) Bars represent number of piRNA expressed in all samples per chromosome. EC membership of piRNAs is indicated with color key as in a (top left corner of barplot). C) Percentage of piRNA transcripts in PSC, MPC and CPC samples that intersected or not with genes from GENCODE database (v29).

annotations (S5D Fig). To test whether these events were random, we shuffled our samples 1500 times (bootstrapping) and analyzed intersection to these features in sense and antisense orientation. Once data was collected, we calculated enrichment on genes as "sample piRNAs" over "shuffled piRNAs" and determined that sense-oriented piRNAs occurred non-randomly on protein coding and long non-coding (lnc) genes (Fig 6A). On the contrary, intersection in antisense had poor fold enrichment values suggesting piRNAs were preferentially located else-where. We observed similar results for piRNAs identified in all three cell differentiation stages studied in this work, as well as in two samples (isogenic replicates) downloaded from ENCODE project corresponding to H1-derived neural progenitor cells (NPC). Both neural samples were handled following the same steps and criteria described before (S6A Fig).

Taking into consideration that reads originated from protein coding and lncRNA features might have been the result of ordinary transcript degradation, we investigated the distribution of reads mapped on such piRNA-hosting genes. Results showed that the percentage of bases covered by sense-oriented reads in these genes was low, with a median value of 0.56% in PSC, 1.50% in MPC, 2.08% in CPC and 3.63% in NPC (S6B Fig). Moreover, coverage was localized to a set of specific piRNAs instead of all piRbase annotations described in any single gene (S6C Fig), proving to be inconsistent with random degradation-produced reads. Coverage by anti-sense-oriented reads was closer to none (S6D Fig) and significantly lower than sense-oriented coverage in all cell population except in MPC (Fig 6B), possibly due to a higher level of disper-sion in values of these samples.

The wide majority of piRNA-hosting protein coding and lncRNA genes harboured a single piRNA transcript with a tendency to augment the number of piRNAs per gene throughout the differentiation process (Fig 6C, pie charts). Like MPC and CPC, NPC exhibited a wider spectrum of piRNAs per gene than undifferentiated pluripotent cells (PSC). A more detailed exploration into CPC results revealed that *MALAT1* (lncRNA gene) and *TTN* (protein coding gene) con-tained the highest number of piRNAs -12 and 6 respectively-, followed by *PLN*, *RPPH1*, *ACTC1* and *AL355075.4* with 4 (Fig 6C, bottom panel). Analysis of *MALAT1* and *TTN* expression by qPCR showed a significant increment from day 3 to 21 and a clear stabilization of mRNA levels between day 21 and 30 of the protocol for both genes (Fig 6D). Expression profiles of piRNAs originated from these genes (S3F Fig; *MALAT1*: piR-4403262 and piR-4424378; TTN: piR-4193743, piR-2715002 and piR-1551388) encompassed significant increments between day 21 and 30, which suggests that they could be involved in a mechanism to modulate the expression of their host.

Using RNA-seq data from H9 cells differentiated to CM [41] we extended our analysis to *ACTC1*, *PLN* and *RPPH1* genes (Fig 6E), which harboured mostly piRNAs from EC 8 (*RPPH1* also contained 1 from EC 2 and one EC 6). While *RPPH1* RNA levels displayed similar expression dynamics to *MALAT1* and *TTN*, *PLN* and *ACTC1* RNA levels increased from day 4 to day 30 practically impervious to piRNA production, though lack of data between day 4 and 30 hindered our analysis for these genes. With respect to *AL355075.4* gene, we found no count data available in RNA-seq samples. However, this annotation overlaps *RPPH1* in sense orientation and it par-tially overlaps protein coding gene *PARP2* in antisense orientation, meaning that piRNAs origi-nated from it could be potentially involved in regulating both genes. In fact, *PARP2* expression dynamic showed a steady descent in transcript levels from day 0 to 30, which is also consistent with the fact that the piRNAs originated from *AL355075.4* were also detected in MPC (S6E Fig).
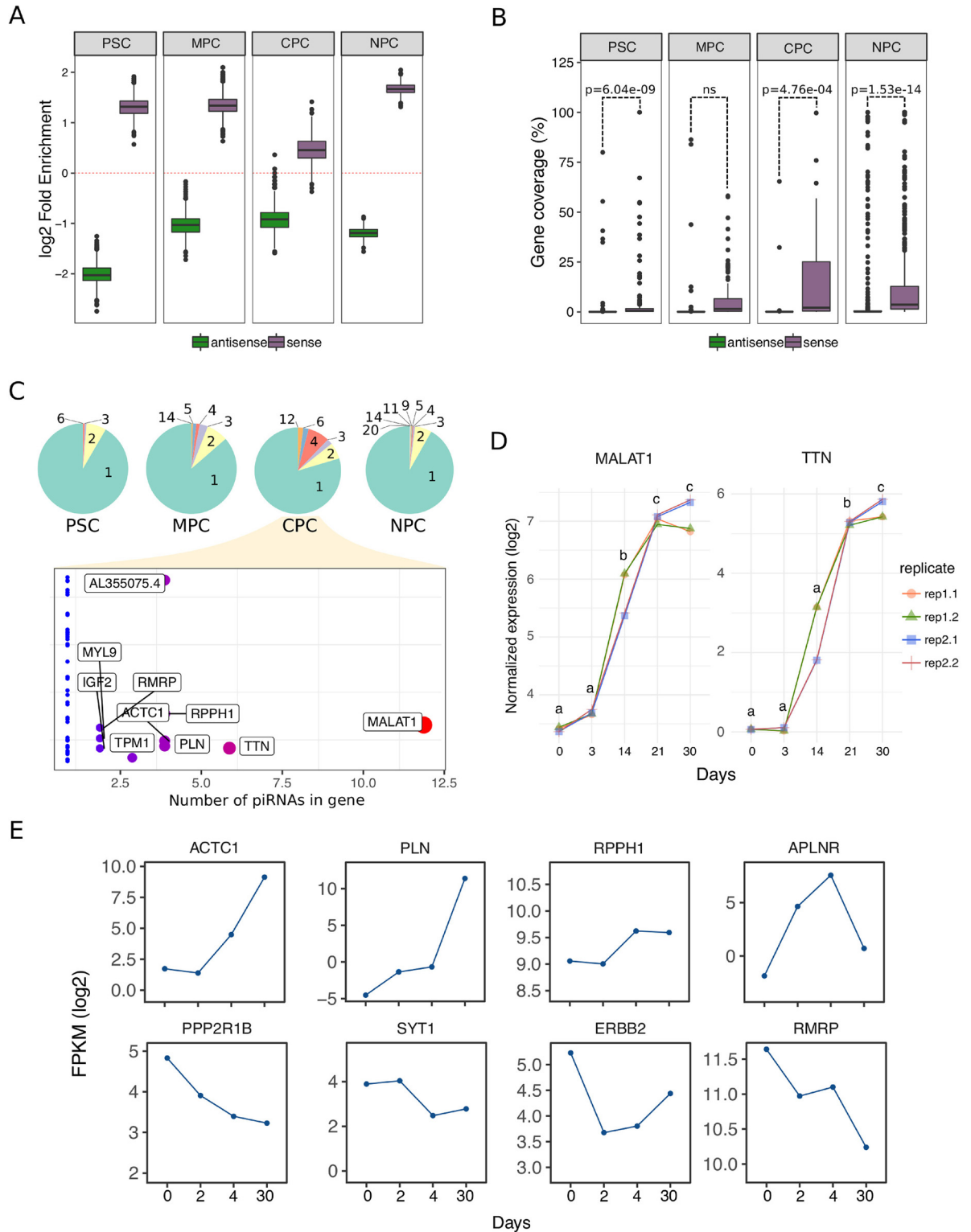
Fig 6. Proteing coding and lncRNA-originated piRNAs. A) Fold enrichment was calculated as number of piRNA transcripts intersected to protein coding and lncRNA genes in sense (purple) and antisense (green) orientation over 1500 different random distributions (bootstrapping). Red dashed line marks the point of no enrichment (log2 1 = 0). NPC: human neural progenitor cells. B) Distribution of percent coverage values on genes in sense (purple) and antisense orientation (green). P-value (p) for statistical analysis is shown in the plots (ANOVA). ns: not significant. C) Pie charts show the proportion of genes containing putative piRNAs. The number of sense piRNA

transcripts per gene in each category is indicated in the charts. Plot at the bottom provides detail on pie chart for CPC labeling genes with two or more putative piRNAs. Mitochondrial genes are not included in the analysis. D) Log2 normalized expression of *MALAT1* and *TTN* genes measured by qPCR during the differentiation protocol extended until day 30. Results of four different RT reactions (rep1.1, rep1.2, rep2.1 and rep2.2) from two independent experiments (rep1 and rep2) are shown and statistically significant differences are displayed as letters in the graphs (p<0.05). E) Log2 normalized counts (FPKM) for eight piRNA-hosting genes during differentiation of H9 pluripotent stem cells (day 0, day 2, day 4 and day 30) to CM. Data was extracted from previously published total RNA-seq experiments [41].

Similar results were found when we studied two genes with high piRNA content (>3) in MPC (Fig 6E) -*APLNR* and *RMRP*- in which almost all piRNAs belong to EC 2.

As we mention before, most piRNA-hosting genes in PSC contained only one piRNA. Interestingly, when we studied the expression profiles of three of these genes -PPP2R1B, SYT1 and ERBB2-, we found that they did not correlate to their piRNA (S3F Fig). Taken together, these evidences suggest that piRNAs may be implicated in fine-tuning mRNA levels of their host gene during differentiation to cardiomyocytes.

Gene ontology analysis on piRNA-hosting genes in differentiated cells

The expression profile of piRNAs proved to be sufficient to clearly discriminate CPC samples not only from PSC and MPC populations, but from neural progenitors (NPC) as well. The comparison between CPC and NPC samples revealed that 52 piRNAs were expressed in both types of differentiated cells, but more importantly the majority were not (Fig 7A). Unshared piRNAs constitute a unique repertoire for each cell population which could possibly reflect upon diverse functional processes. To evaluate this notion, we extracted all protein coding genes which were intersected by at least one piRNA and determined their involvement in any biological process (BP). In search for overrepresented terms (BPs with more genes involved than expected), we found that CPC and NPC showed markedly different terms. The BPs associated to CPC were intimately related to heart development and muscle differentiation and contraction (Fig 7B), while overrepresented BPs in NPC showed a clear inclination towards neurogenesis regulation and neural proliferation and development (Fig 7C). The group of genes intersected by piRNAs shared by both CPC and NPC (52 piRNAs in venn diagram) did not participate in any of the statistically significant overrepresented BPs, meaning that enriched categories for each population are mostly based in their unique collection of piRNAs.

Discussion

Since the first mechanistic evidences of piRNA biogenesis in ovarian follicle cells of *D. melanogaster* were published [14], much information has emerged on the somatic expression and function of this type of regulatory small RNAs. Several publications demonstrated that piRNAs (or piRNA-like RNAs) originate from discrete genomic regions of somatic cells in a wide diversity of species and tissues [15, 27, 28, 43, 44]. In agreement with this line of evidence, we report the expression of 447 small RNAs consistent with piRNAs among three stages of differentiation of pluripotent stem cells to cardiomyocytes using a database-driven approach.

The pipeline leading to the identification and quantification of piRNAs involved two filtering steps that were implemented to avoid innacurate interpretation of results. Firstly, aligned reads shorter than 24nt and longer than 34nt were discarded from further analyses. However, our results showed a higher-than-basal frequency of 36-37nt-long reads, which prompts the issue if these reads should have been kept for further investigation as potential longer piRNAs or perhaps remnants of piRNA precursors. Length restriction answers to one of the hallmark attributes of mature piRNA transcripts, albeit the range seems to vary across species. In fact, in *C. elegans* 21nt-long piRNAs are produced from transcript precursors of 25-27nt in length [45] whose processing machinery is partially unknown. The reason for the range diversity has been
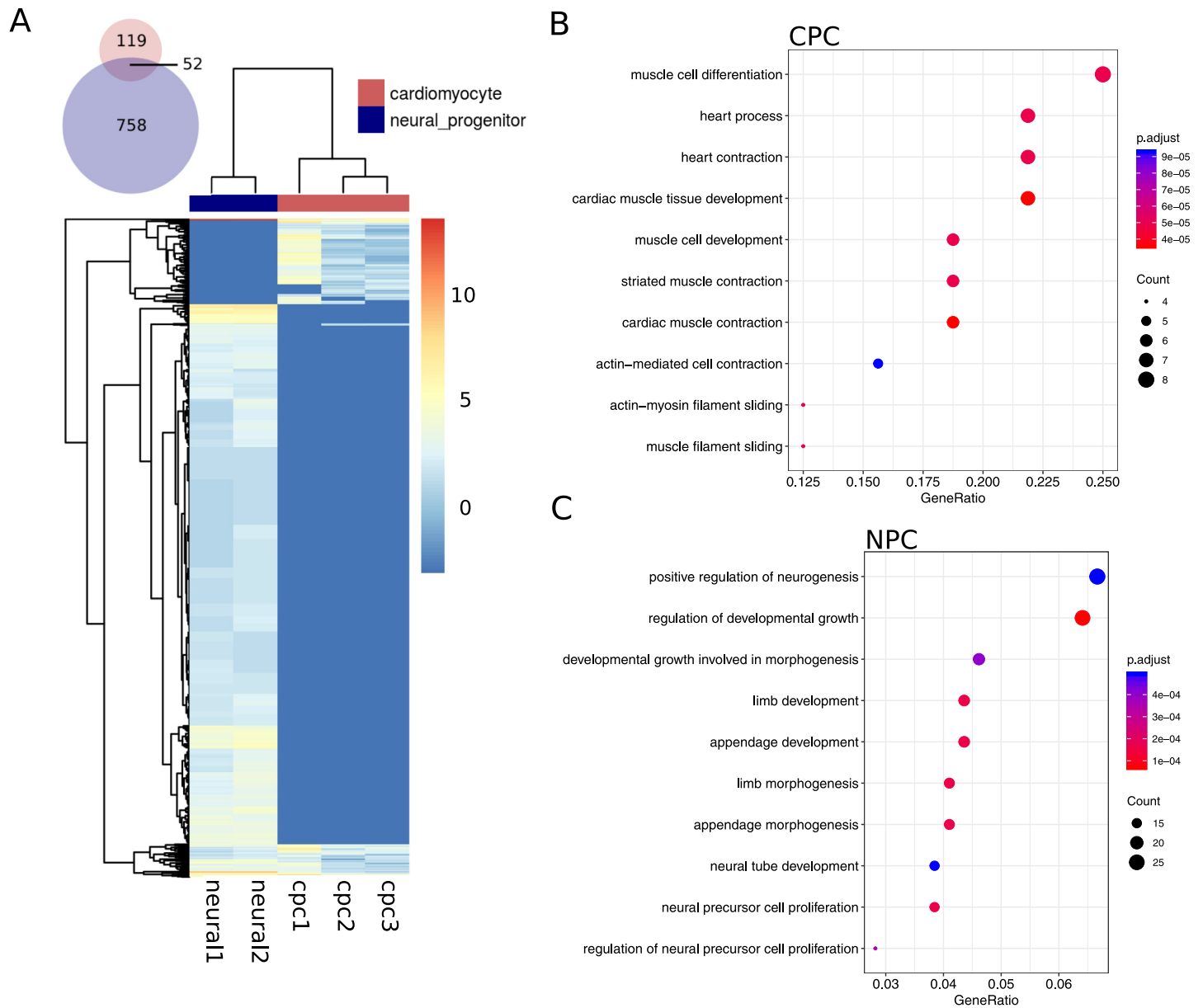
Fig 7. Gene ontology analysis of piRNA transcripts in differentiated cells. A) Expression profiles of piRNA transcripts (log2 normalized counts) in CPC (cpc1, cpc2 and cpc3) and NPC (neural1 and neural2). Color key for sample clustering is displayed at the top right corner of the heatmap. B) Overrepresented biological processes (BPs) on proteing coding genes harbouring piRNA transcripts from CPC samples. C) Overrepresented BPs on proteing coding genes harbouring piRNA transcripts from NPC samples.

convincingly connected to the activity of the proteins involved in their biosynthetic pathway (biogenesis). It is possible that they also contribute to explaining the differences between germ-line and somatic piRNAs considering that diverse sets of enzymes have been reported to be engaged in piRNA synthesis in these cell lineages [45, 46]. Moreover, the differential expression profile of *PIWI* genes observed in our data -and in H9 and H1 external RNA-seq data- points to cell-type specific functions for these proteins and consequently for their piRNA partners.

In a second step, length-filtered reads mapping to small ncRNAs other than piRNAs were removed from samples. The fact that remaining reads in PSC and CPC samples showed a

moderate 5' U bias, whereas MPC samples did not could probably be related to the transitional nature of this cell population. However, none of the samples exhibited the characteristic 10nt-overlap signature of secondary piRNAs, which is a generally accepted feature in germline cells of most animals where piRNAs are synthesized through both primary and secondary (ping-pong loop) pathways [15, 26]. Despite synthesis in somatic cells has been proposed to produce only primary piRNAs, many of the mechanisms underlying piRNA biogenesis -specially in non-gonadal tisues- are not yet fully understood. Conceivably, the filtering steps eliminated potential piRNAs from the three stages, though it has been argued that a considerable amount of annotated piRNAs are actually ncRNA fragments derived from rRNAs, tRNAs and even miRNAs [40]. On the matter, an insightful disscussion by Tosar and collaborators [40], advocating for gonadal piRNAs, suggested that somatic piRNAs mapping to ncRNA fragments are not unquestionably wrong, still further biochemical evidence is needed to include them as such.

An important aspect of this work lies on the identification of a piRNA expression profile associated to each of the cell populations under study. These expression profiles parallel the embryological connection between the stages, where PSC is more closely related to MPC than to CPC [2, 47]. Upon this premise, the piRNAs identified as early-changing could potentially be involved in maintaining pluripotency or in the commitment of pluripotent cells to mesoderm progenitors, which might eventually differentiate to multiple lineages. Late-changing piRNAs, on the contrary, would influence further commitment of mesoderm cells to cardiac progenitors. The fact that six times more piRNAs were downregulated rather than upregulated during differentiation to CPC suggests that piRNA pathways become less relevant in differentiated cells. It is possible that mechanisms evolutionarily linked to the regulation of transposable elements are not as critically conserved in differentiated cells as they do in cells with high proliferation rates or reproductive functions, such as pluripotent cells and germ line cells. For example, it has been proposed that cancer cells might promote piRNA biosynthetic pathways as a mechanism to reduce genome instability caused by increased mitotic and transcriptional activities [48].

The genomic localization of piRNAs included in anyone particular EC was not the same. In fact, piRNA expression appeared to be uniformly scattered across the genome except for the mitochondrial chromosome. The majority of piRNAs identified in the mitochondrial genome mapped to rRNAs or tRNAs and though we have not definitively proved they are truly piRNAs, previous work established a link between tRNA-/rRNA-derived piRNAs, *HIWI2* expression and regulation of metabolic processes in somatic cells [49]. Analogously, the increased levels of piRNAs from mitochondrial tRNAs/rRNAs and the significant upregulation of *HIWI2* (day 21 v. day 3.5, and external H9 RNA-seq data) in CPC could seemingly be connected to the large-scale modifications in CM metabolism [50].

Even though identified piRNAs were dispersed throughout the genome, it was clear that the vast majority of them originated from gene loci. However, it is not yet clear the reason why these piRNAs are generated from the sense strand of their hosting genes. One possibility relies on the capacity of PIWI/piRNA complexes to direct recruitment of DNA and histone methyltransferases, modifying accesibility of transcriptional machinery to chromatin [25, 51]. Available data of DNA or histone methylation status in the three stages of cardiac differentiation is scarse or dissimilar, so preliminary correlation analysis were not conclusive at this point. Nevertheless, further experiments on promoter methylation and H3K9me3 mark deposition ought to be performed to pursuit this possibility. Also, considering that antisense transcripts have been described to positively regulate stability of sense RNA [52], it is possible that sense-originated piRNAs regulate antisense transcript levels in a miRNA-like mechanism. For instance, *TALAM1* -an antisense transcript at the *MALAT1* gene locus- promotes stability and maturation of *MALAT1* RNA by facilitating enzymatic cleavage of its 3' end [52], thus a potential piRNA-mediated downregulation of *TALAM1* would redound to diminished *MALAT1* levels.

In connection with this, there have been reports stating that piRNA biosynthesis from long non-coding genes during spermiogenesis in murine models functions by itself as a degradation mechanism [53, 54], suggesting potential roles of piRNAs other than the canonical slicing-by-complementarity action. Furthermore, studies on mRNA-derived piRNAs from *D. melanogaster* point to possible cis-regulatory activity involving an interaction between the CCR4-NOT deadenylation complex and Aub/Ago3 proteins [54–56].

In sum, the evidences presented here contribute to understanding the dynamic expression of piRNAs during differentiation of pluripotent stem cells to cardiomyocytes and further explore their potential function as post-transcriptional modulators in somatic cells. Together with miRNAs, piRNAs seem to participate in the fine-tuning of transcript levels, adding yet another layer to the complex and intricated networks governing gene expression.

Supporting information

S1 Fig. Read length control in filtered samples and mapping to other non coding RNAs. Related to Fig 1. A) Read length expressed as a function of density for pluripotent (PSC1, PSC2 and PSC3), mesoderm progenitor (MPC1, MPC2 and MPC3) and cardiomyocytes (CPC1, CPC2 and CPC3) samples before filtering (unfiltered) and after filtering (23 < RL < 35 and -ncRNA). Color key is indicated in the plot. B) Number of mapped reads. Reads were counted before processing (1) and after being filtered by length (2) and other ncRNAs (3). Remaining reads after processing are expressed as percentage(%) of unfiltered reads (2/1 and 3/1). C) Analysis of coverage on non coding RNAs loci from DASHR database for fully processed normalized (counts per million) samples.
(TIF)

S2 Fig. Processing of testis samples. Related to Fig 2. A) Number of mapped reads after employing the pipeline described in Fig 1 in human testis samples downloaded from ENCODE (merged replicates). B) Frequency of bases per position in processed mapped reads.
(TIF)

S3 Fig. Analysis of expression data and DE results. Related to Fig 3. A) Boxplot showing reads for the nine samples normalized by library depth and expressed as log2 counts per million (CPM). B) Heatmap in log2 CPM of piRNA transcripts from a. Hard unsupervised clustering was performed on rows (piRNA ID) and columns (sample ID), and is shown as dendrograms. Color keys for heatmap and phenotype are indicated to left and in the top right corner of the graph, respectively. C) Principal Component Analysis performed on DESeq2 normalized counts. The color key is indicated to the right of the plot. D) Overlap of differentially expressed piRNA transcripts in CPC versus MPC (153; green circle) and PSC (137; purple circle). E) Normalized expression of piRNA transcripts upregulated, dowregulated and non-regulated with respect to CPC. F) Ten piRNA transcripts were evaluated by qPCR using a specific retrotranscription protocol designed for small RNAs in day 0, 3, 14, 21 and 30 of cardiac differentiation. piR-4403262 and piR-4424378 originate from *MALAT1*; piR-4193743, piR-2715002 and piR-1551388 are produced from *TTN*; piR-4091280 from *PPP2R1B*; piR-2519215 from *SYT1*; piR-1332287 from *ERBB2*; piR-2413094 and piR-1399886 from intergenic regions. Expression of mir-302b -marker of pluripotency- was analyzed to assess protocol success. Results of four different RT reactions (rep1.A, rep1.B, rep2.A and rep2.B) from two independent experiments (rep1 and rep2) are shown after normalization by small RNA RNU6B. Statistically significant differences are displayed as letters in the graphs (p<0.05).
(TIF)

S4 Fig. Expression profile of human Piwi genes in H1 and H9 embryonic stem cell lines.
Related to Fig 4. Normalized RNA-seq counts (FPKM) from H1 and H9 cell lines were down-
loaded from GEO.
(TIF)

S5 Fig. Expression of piRNA RNAs in mitochondrial chromosome. Related to Fig 5. A) Dis-
tribution of PSC, MPC and CPC mapped reads (merged replicates) as a function of density over
a fraction of mitochondrial chromosome (chrM:1–4000). Profiles above zero correspond to plus
strand and below zero to the minus strand. Color key is located at the bottom left corner of the
plot. B) Coverage profiles in counts per million mapped reads (CPM) on the entire mitochondrial
rRNA extension (MT-rRNA) and center of tRNA (MT-tRNA). Direction of rRNA genes are indi-
cated by 5' and 3'. C) Image captured from IGV software over a portion of the human mitochon-
drial chromosome (chrM:2,184–2,780). The tracks from top to bottom are: piRbase annotated
piRNAs (piRbase), piRNAs identified in our samples (piRNAs), coverage profiles of PSC, MPC
and CPC, DASHR database ncRNA annotations (DASHR) and GENCODE v29 gene annota-
tions (GENCODE v29). D) Fraction of piRNA transcripts mapped to genomic features annotated
in GENCODE v29 database in length-filtered samples (length) compared to length+ncRNA-fil-
tered (length+ncRNA) samples. Color key for features is indicated to the right of the bars.
(TIF)

S6 Fig. Coverage on protein coding and lncRNA genes. Related to Figs 6 and 7. A) NPC sam-
ple processing. Number of mapped reads after employing the pipeline described in N in neural
progenitor samples downloaded from ENCODE project. B) Density estimation of percent (%)
coverage on protein coding genes intersected by piRNA transcripts in sense orientation. Verti-
cal lines indicate medians of each curve. C) Image captured from IGV software portraying
mapped reads of PSC, MPC, CPC and NPC samples on identified piRNA RNAs. Tracks for
piRNA annotation database (piRbase) and gene features (GENCODE v29) are shown. D) Den-
sity estimation as in a, in antisense orientation. E) Left panel shows IGV capture depicting piR-
NAs in *PARP2* vecinity. Expression dynamic of *PARP2* gene in RNA-seq samples from H9
cells differentiated to CM is shown to the right.
(TIF)

S1 Table. Normalized counts for the 447 identified piRNAs-like transcripts. Transcripts are
arranged in descending order of expression (row mean). Differentially regulated piRNA tran-
scripts in CPC are indicated in light blue shading (downregulated) and light red shading
(upregulated).
(PDF)

S2 Table. piRNA transcripts included in Expression Clusters (EC).
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Alejandro La Greca, María Agustina Scarafía, Gabriel Neiman, Ximena
Garate, Cyntia Aban, Ariel Waisman, Lucía Natalia Moro, Carlos Luzzani, Santiago Gabriel
Miriuka.

References

1. Kehat I, Kenyagin-Karsenti D, Snir M, Segev H, Amit M, Gepstein A, et al. Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. J Clin Invest. 2001; 108(3):407–14. https://doi.org/10.1172/JCI12131 PMID: 11489934

2. Evseenko D, Zhu Y, Schenke-Layland K, Kuo J, Latour B, Ge S, et al. Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells. Proc Natl Acad Sci U S A. 2010; 107(31):13742–7. https://doi.org/10.1073/pnas.1002077107 PMID: 20643952

3. Burridge PW, Keller G, Gold JD, Wu JC. Production of de novo cardiomyocytes: human pluripotent stem cell differentiation and direct reprogramming. Cell Stem Cell. 2012; 10(1):16–28. https://doi.org/10.1016/j.stem.2011.12.013 PMID: 22226352

4. Lian X, Zhang J, Azarin SM, Zhu K, Hazeltine LB, Bao X, et al. Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/beta-catenin signaling under fully defined conditions. Nat Protoc. 2013; 8(1):162–75. https://doi.org/10.1038/nprot.2012.150 PMID: 23257984

5. Loh KM, Chen A, Koh PW, Deng TZ, Sinha R, Tsai JM, et al. Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. Cell. 2016; 166(2):451–67. https://doi.org/10.1016/j.cell.2016.06.011 PMID: 27419872

6. Beermann J, Piccoli MT, Viereck J, Thum T. Non-coding RNAs in Development and Disease: Background, Mechanisms, and Therapeutic Approaches. Physiol Rev. 2016; 96(4):1297–325. https://doi.org/10.1152/physrev.00041.2015 PMID: 27535639

7. Devaux Y, Zangrando J, Schroen B, Creemers EE, Pedrazzini T, Chang CP, et al. Long noncoding RNAs in cardiac development and ageing. Nat Rev Cardiol. 2015; 12(7):415–25. https://doi.org/10.1038/nrcardio.2015.55 PMID: 25855606

8. Ha M, Kim VN. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol. 2014; 15(8):509–24. https://doi.org/10.1038/nrm3838 PMID: 25027649

9. Li Q, Gregory RI. MicroRNA regulation of stem cell fate. Cell Stem Cell. 2008; 2(3):195–6. https://doi.org/10.1016/j.stem.2008.02.008 PMID: 18371442

10. Espinoza-Lewis RA, Wang DZ. MicroRNAs in heart development. Curr Top Dev Biol. 2012; 100:279–317. https://doi.org/10.1016/B978-0-12-387786-4.00009-9 PMID: 22449848

11. Garate X, La Greca A, Neiman G, Blüguermann C, Santín Velazque NL, Moro LN, et al. Identification of the miRNAome of early mesoderm progenitor cells and cardiomyocytes derived from human pluripotent stem cells. Sci Rep. 2018; 8(1):8072. https://doi.org/10.1038/s41598-018-26156-3 PMID: 29795287

12. 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.org/10.1038/nature04917 PMID: 16751776

13. Ro S, Park C, Song R, Nguyen D, Jin J, Sanders KM, et al. Cloning and expression profiling of testis-expressed piRNA-like RNAs. RNA. 2007; 13(10):1693–702. https://doi.org/10.1261/rna.640307 PMID: 17698640

14. Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, et al. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell. 2009; 137(3):522–35. https://doi.org/10.1016/j.cell.2009.03.040 PMID: 19395010

15. Yan Z, Hu HY, Jiang X, Maierhofer V, Neb E, He L, et al. Widespread expression of piRNA-like molecules in somatic tissues. Nucleic Acids Res. 2011; 39(15):6596–607. https://doi.org/10.1093/nar/gkr298 PMID: 21546553

16. Ishizu H, Siomi H, Siomi MC. Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. Genes Dev. 2012; 26(21):2361–73. https://doi.org/10.1101/gad.203786.112 PMID: 23124062

17. Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. Developmentally regulated piRNA clusters implicate MILI in transposon control. Science. 2007; 316(5825):744–7. https://doi.org/10.1126/science.1142612 PMID: 17446352

18. Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, et al. Characterization of the piRNA complex from rat testes. Science. 2006; 313(5785):363–7. https://doi.org/10.1126/science.1130164 PMID: 16778019

19. Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, et al. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. Genes Dev. 2006; 20(16):2214–22. https://doi.org/10.1101/gad.1454806 PMID: 16882972

20. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell. 2007; 128(6):1089–103. https://doi.org/10.1016/j.cell.2007.01.043 PMID: 17346786

21. Han YN, Li Y, Xia SQ, Zhang YY, Zheng JH, Li W. PIWI Proteins and PIWI-Interacting RNA: Emerging Roles in Cancer. Cell Physiol Biochem. 2017; 44(1):1–20. https://doi.org/10.1159/000484541 PMID: 29130960

22. Phay M, Kim HH, Yoo S. Analysis of piRNA-Like Small Non-coding RNAs Present in Axons of Adult Sensory Neurons. Mol Neurobiol. 2018; 55(1):483–494. https://doi.org/10.1007/s12035-016-0340-2 PMID: 27966078

23. Sato K, Siomi MC. Piwi-interacting RNAs: biological functions and biogenesis. Essays Biochem. 2013; 54:39–52. https://doi.org/10.1042/bse0540039 PMID: 23829526

24. Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol. 2011; 12(4):246–58. https://doi.org/10.1038/nrm3089 PMID: 21427766

25. Rajasethupathy P, Antonov I, Sheridan R, Frey S, Sander C, Tuschl T, et al. A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. Cell. 2012; 149(3):693–707. https://doi.org/10.1016/j.cell.2012.02.057 PMID: 22541438

26. Ross RJ, Weiner MM, Lin H. PIWI proteins and PIWI-interacting RNAs in the soma. Nature. 2014; 505 (7483):353–359. https://doi.org/10.1038/nature12987 PMID: 24429634

27. Roy J, Sarkar A, Parida S, Ghosh Z, Mallick B. Small RNA sequencing revealed dysregulated piRNAs in Alzheimer's disease and their probable role in pathogenesis. Mol Biosyst. 2017; 13(3):565–576. https://doi.org/10.1039/c6mb00699j PMID: 28127595

28. Vella S, Gallo A, Lo Nigro A, Galvagno D, Raffa GM, Pilato M, et al. PIWI-interacting RNA (piRNA) signatures in human cardiac progenitor cells. Int J Biochem Cell Biol. 2016; 76:1–11. https://doi.org/10.1016/j.biocel.2016.04.012 PMID: 27131603

29. Rajan KS, Velmurugan G, Pandi G, Ramasamy S. miRNA and piRNA mediated Akt pathway in heart: antisense expands to survive. Int J Biochem Cell Biol. 2014; 55:153–6. https://doi.org/10.1016/j.biocel.2014.09.001 PMID: 25220478

30. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635 PMID: 23104886

31. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26(6):841–2. https://doi.org/10.1093/bioinformatics/btq033 PMID: 20110278

32. Leung YY, Kuksa PP, Amlie-Wolf A, Valladares O, Ungar LH, Kannan S, et al. DASHR: database of small human noncoding RNAs. Nucleic Acids Res. 2016; 44(D1):D216–22. https://doi.org/10.1093/nar/gkv1188 PMID: 26553799

33. Wang J, Zhang P, Lu Y, Li Y, Zheng Y, Kan Y, et al. piRBase: a comprehensive database of piRNA sequences. Nucleic Acids Res. 2019; 47(D1):D175–D180. https://doi.org/10.1093/nar/gky1043 PMID: 30371818

34. Futschik ME, Carlisle B. Noise-robust soft clustering of gene expression time-course data. J Bioinform Comput Biol. 2005; 3(4):965–88. https://doi.org/10.1142/s0219720005001375 PMID: 16078370

35. Uhrig S, Klein H. PingPongPro: a tool for the detection of piRNA-mediated transposon-silencing in small RNA-Seq data. Bioinformatics. 2019; 35(2):335–336. https://doi.org/10.1093/bioinformatics/bty578 PMID: 29985981

36. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016; 44(W1):W160–5. https://doi.org/10.1093/nar/gkw257 PMID: 27079975

37. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, et al. Real-time quantification of micro-RNAs by stem-loop RT-PCR. Nucleic Acids Res. 2005; 33(20):e179. https://doi.org/10.1093/nar/gni178 PMID: 16314309

38. Mestdagh P, Feys T, Bernard N, Guenther S, Chen C, Speleman F, et al. High-throughput stem-loop RT-qPCR miRNA expression profiling using minute amounts of input RNA. Nucleic Acids Res. 2008; 36 (21):e143. https://doi.org/10.1093/nar/gkn725 PMID: 18940866

39. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. PIWI-interacting RNAs: small RNAs with big functions. Nat Rev Genet. 2019; 20(2):89–108. https://doi.org/10.1038/s41576-018-0073-3 PMID: 30446728

40. Tosar JP, Rovira C, Cayota A. Non-coding RNA fragments account for the majority of annotated piR-NAs expressed in somatic non-gonadal tissues. Commun Biol. 2018; 1:2. https://doi.org/10.1038/s42003-017-0001-7 PMID: 30271890

41. Liu Q, Jiang C, Xu J, Zhao MT, Van Bortle K, Cheng X, et al. Genome-Wide Temporal Profiling of Transcriptome and Open Chromatin of Early Cardiomyocyte Differentiation Derived From hiPSCs and hESCs. Circ Res. 2017; 121(4):376–391. https://doi.org/10.1161/CIRCRESAHA.116.310456 PMID: 28663367

42. Kwon C, Tak H, Rho M, Chang HR, Kim YH, Kim KT, et al. Detection of PIWI and piRNAs in the mito-chondria of mammalian cancer cells. Biochem Biophys Res Commun. 2014; 446(1):218–23. https://doi.org/10.1016/j.bbrc.2014.02.112 PMID: 24602614

43. Cichocki F, Lenvik T, Sharma N, Yun G, Anderson SK, Miller JS. Cutting edge: KIR antisense transcripts are processed into a 28-base PIWI-like RNA in human NK cells. J Immunol. 2010; 185(4):2009–12. https://doi.org/10.4049/jimmunol.1000855 PMID: 20631304

44. Ng KW, Anderson C, Marshall EA, Minatel BC, Enfield KSS, Saprunoff HL, et al. Piwi-interacting RNAs in cancer: emerging functions and clinical utility. Mol Cancer. 2016; 15:5. https://doi.org/10.1186/s12943-016-0491-9 PMID: 26768585

45. Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, et al. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell. 2006; 127(6):1193–207. https://doi.org/10.1016/j.cell.2006.10.040 PMID: 17174894

46. Zamore PD. Somatic piRNA biogenesis. EMBO J. 2010; 29(19):3219–21. https://doi.org/10.1038/emboj.2010.232 PMID: 20924396

47. Sha Y, Haensel D, Gutierrez G, Du H, Dai X, Nie Q. Intermediate cell states in epithelial-to-mesenchymal transition. Phys Biol. 2019; 16(2):021001. https://doi.org/10.1088/1478-3975/aaf928 PMID: 30560804

48. Liu Y, Zhang J, Li A, Liu Z, He Z, Yuan X, et al. Prediction of cancer-associated piRNA–mRNA and piRNA–lncRNA interactions by integrated analysis of expression and sequence data. Tsinghua Science and Technology. 2018; 23(2):115–125. https://doi.org/10.26599/TST.2018.9010056

49. Keam SP, Young PE, McCorkindale AL, Dang THY, Clancy JL, Humphreys DT, et al. The human Piwi protein Hiwi2 associates with tRNA-derived piRNAs in somatic cells. Nucleic Acids Res. 2014; 42 (14):8984–95. https://doi.org/10.1093/nar/gku620 PMID: 25038252

50. Tohyama S, Hattori F, Sano M, Hishiki T, Nagahata Y, Matsuura T, et al. Distinct metabolic flow enables large-scale purification of mouse and human pluripotent stem cell-derived cardiomyocytes. Cell Stem Cell. 2013; 12(1):127–37. https://doi.org/10.1016/j.stem.2012.09.013 PMID: 23168164

51. Pezic D, Manakov SA, Sachidanandam R, Aravin AA. piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. Genes Dev. 2014; 28(13):1410–28. https://doi.org/10.1101/gad.240895.114 PMID: 24939875

52. Zong X, Nakagawa S, Freier SM, Fei J, Ha T, Prasanth SG, et al. Natural antisense RNA promotes 3' end processing and maturation of MALAT1 lncRNA. Nucleic Acids Res. 2016; 44(6):2898–908. https://doi.org/10.1093/nar/gkw047 PMID: 26826711

53. Vourekas A, Zheng Q, Alexiou P, Maragkakis M, Kirino Y, Gregory BD, et al. Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis. Nat Struct Mol Biol. 2012; 19(8):773–81. https://doi.org/10.1038/nsmb.2347 PMID: 22842725

54. Weick EM, Miska EA. piRNAs: from biogenesis to function. Development. 2014; 141(18):3458–71. https://doi.org/10.1242/dev.094037 PMID: 25183868

55. Saito K, Inagaki S, Mituyama T, Kawamura Y, Ono Y, Sakota E, et al. A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. Nature. 2009; 461(7268):1296–9. https://doi.org/10.1038/nature08501 PMID: 19812547

56. Rouget C, Papin C, Boureux A, Meunier AC, Franco B, Robine N, et al. Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. Nature. 2010; 467(7319):1128–32. https://doi.org/10.1038/nature09465 PMID: 20953170