Supplemental Files

Small RNA expression and miRNA modification dynamics in human oocytes and early embryos

Pauliina Paloviita[1], Christel Hydén-Granskog[2], Dawit A. Yohannes[1], Priit Paluoja[3], Juha Kere[4,5,6], Juha S. Tapanainen[1,2,7], Kaarel Krjutškov[3,5,8], Timo Tuuri[1,2], Urmo Võsa[9] and Sanna Vuoristo[1*]

[1] University of Helsinki, Department of Obstetrics and Gynecology, Helsinki, Finland

[2] Helsinki University Hospital, Department of Obstetrics and Gynecology, Helsinki, Finland

[3] Competence Centre for Health Technologies, Tartu, Estonia

[4] Karolinska Institutet, Department of Biosciences and Nutrition, Huddinge, Sweden

[5] Stem Cells and Metabolism Research Program, University of Helsinki, Finland

[6] Folkhälsan Research Center, Helsinki, Finland

[7] Department of Obstetrics and Gynecology, University of Oulu and Oulu University Hospital, Medical Research Centre, PEDEGO Research Unit, P.O. BOX 23, FI-90029, Oulu, Finland

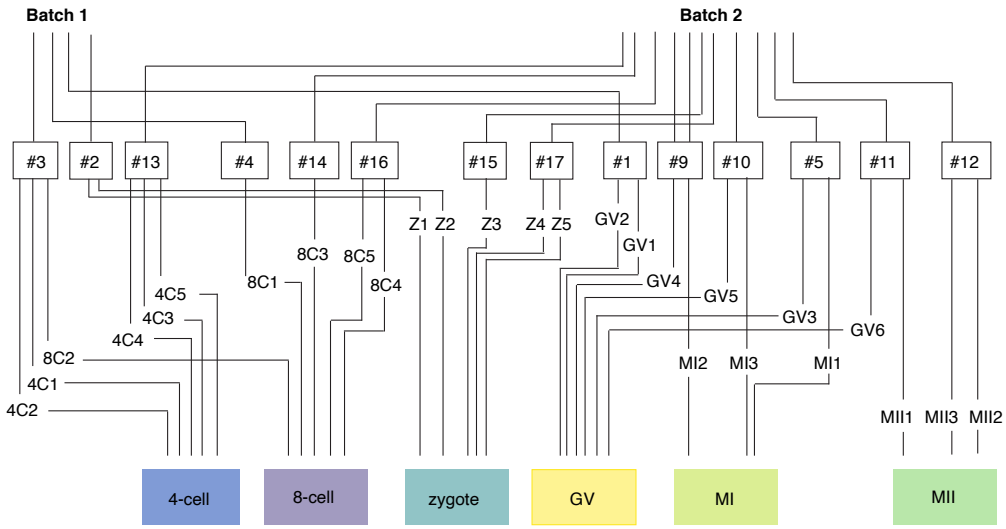
[8] Institute of Clinical Medicine, Department of Obstetrics and Gynecology, University of Tartu, Tartu, Estonia

[9] Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia

*Corresponding author Sanna Vuoristo, Functional Embryo Genomics Lab, Biomedicum Helsinki 1, Haartmaninkatu 8, A430B, FI-00014, University of Helsinki, Helsinki, Finland. Email: sanna.vuoristo@helsinki.fi Phone: +358 9 471 717 30

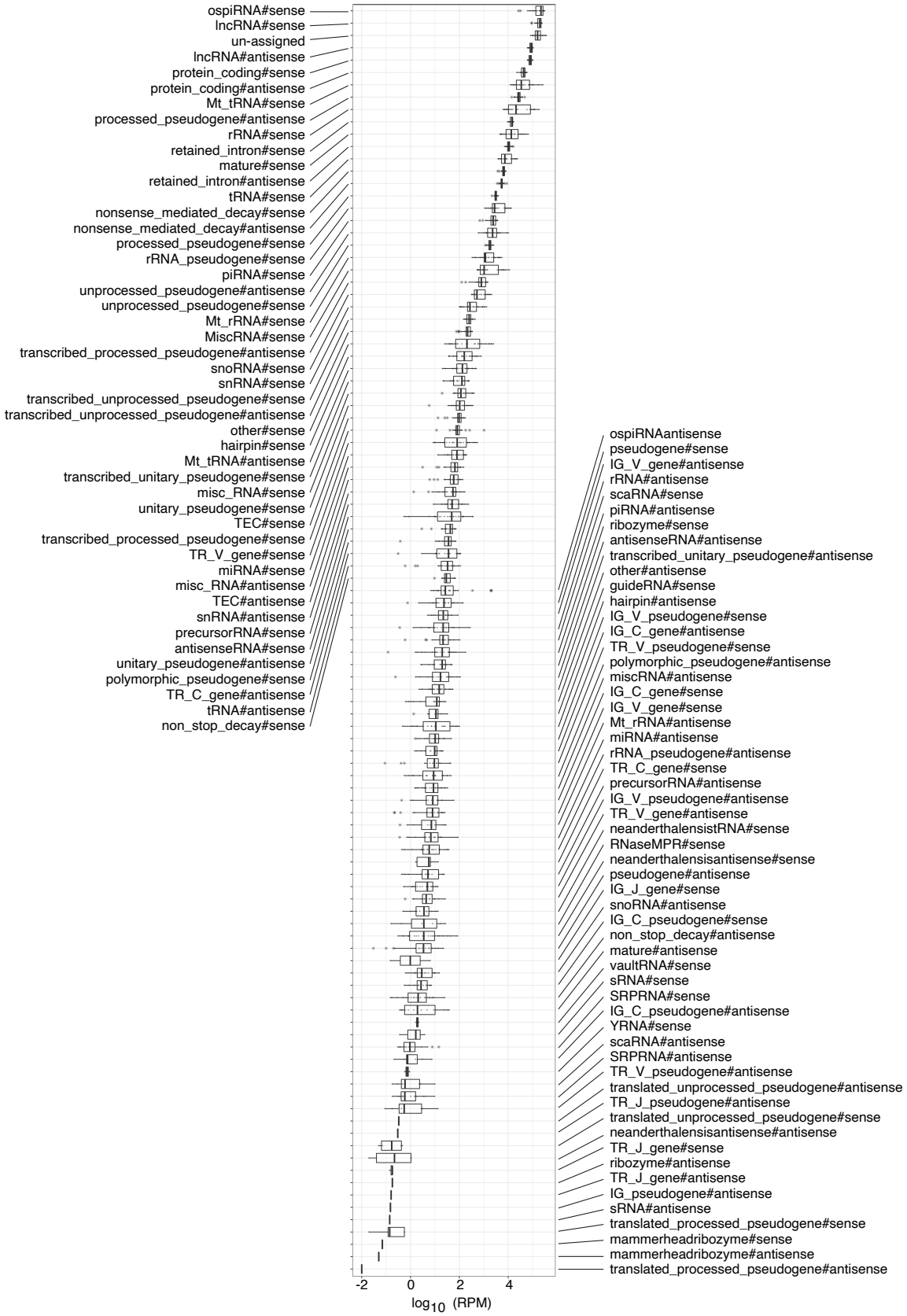
Contents

Supplemental_Fig_S1



Supplemental Fig S1. Small RNA (sRNA) sequencing library structure. sRNA sequencing library batch (Batch 1 or Batch 2), sample relatedness (donor/ donor couples are indicated with # and number), and developmental stage grouping information. Germinal vesicle oocyte, GV; Metaphase I oocyte, MI; Metaphase II oocyte, MII; 4-cell embryo, 4-cell; 8-cell embryo, 8-cell. (*B*)

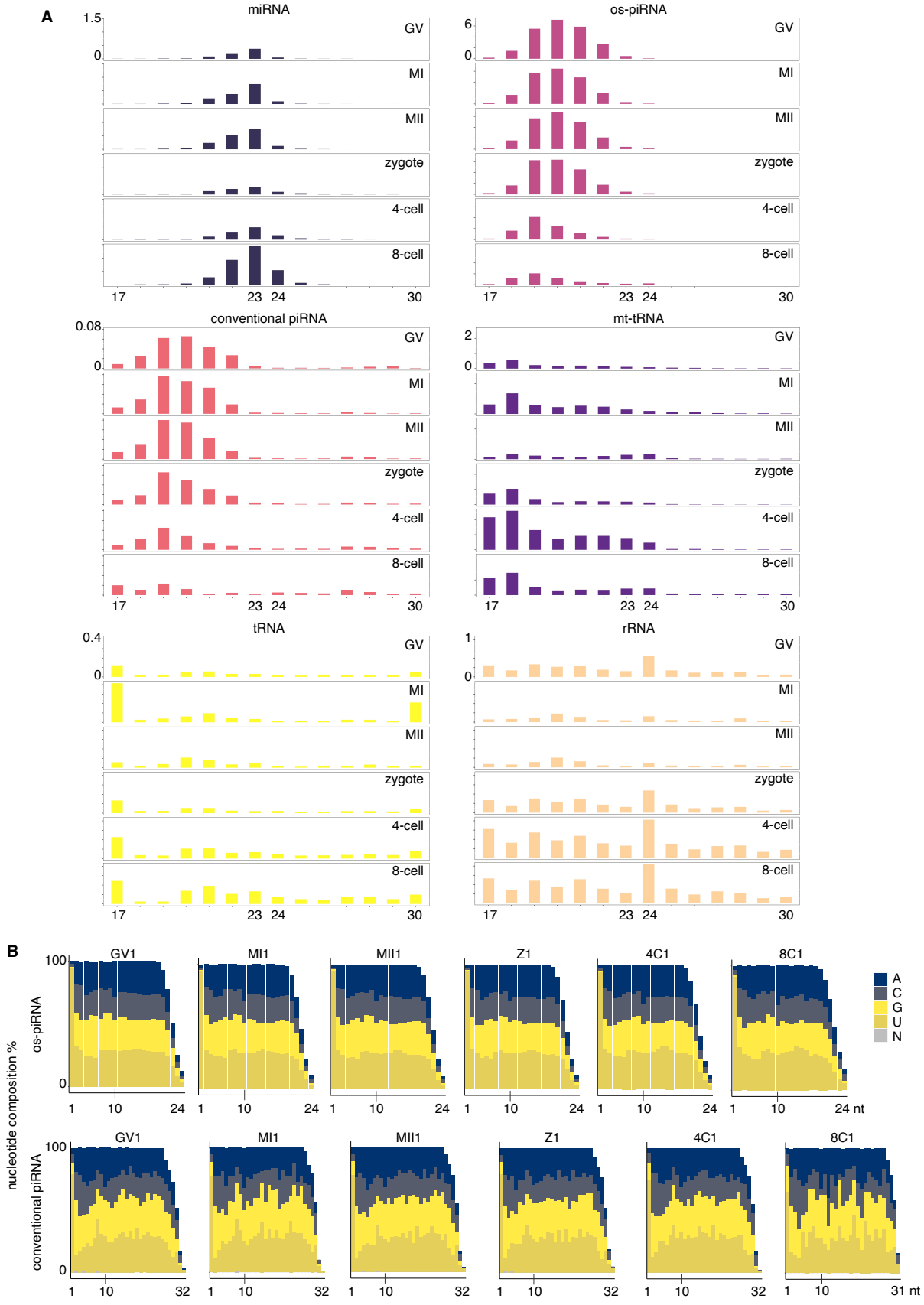
Supplemental_Fig_S2



43

Supplemental Fig S2. Mapping of sRNA-seq libraries. Reads (RPM) mapped to different RNA classes across all samples. The horizontal line in the boxplot indicates the expression median. sRNAbench (Aparicio-Puerta et al. 2019; Rueda et al. 2015) was used with human miRBase (v22) (Kozomara et al. 2018), Ensembl cDNA (hg38), Ensembl non-coding RNA (hg38), and RNAcentral version 14 (hg38) databases, and os-piRNA sequence information (Yang et al. 2019) to map reads against previously identified RNA molecules. Un-assigned reads included mainly reads mapping to mitochondrial transcripts. Naming of the RNA classes follows the database formats with sense and antisense orientations indicated by sRNA bench. Classes “mature” and “hairpin” refer to miRBase miRNAs and pre-miRNAs, respectively.

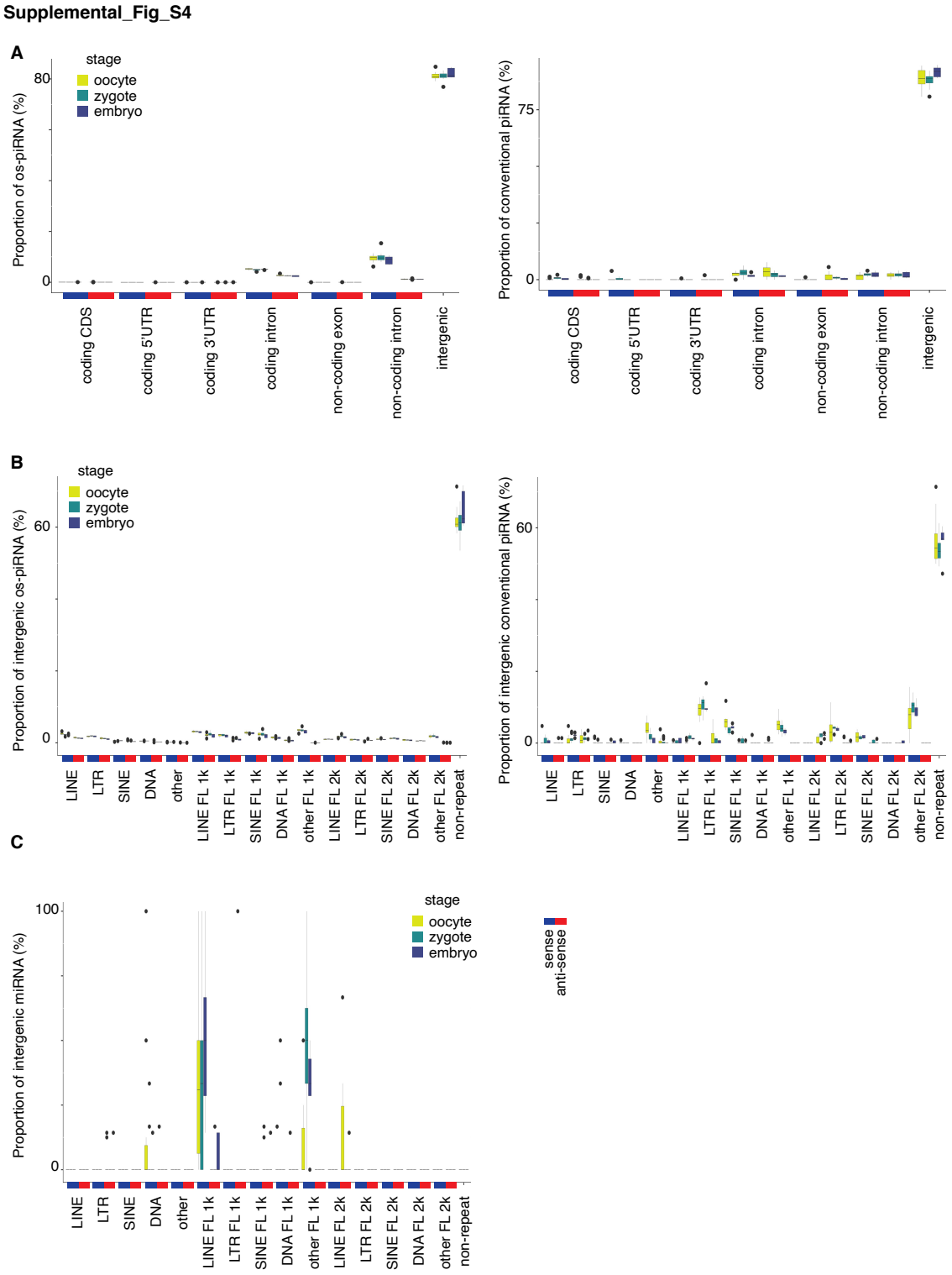
Supplemental_Fig_S3



52

Supplemental Fig S3. sRNA class expression and base position preference.

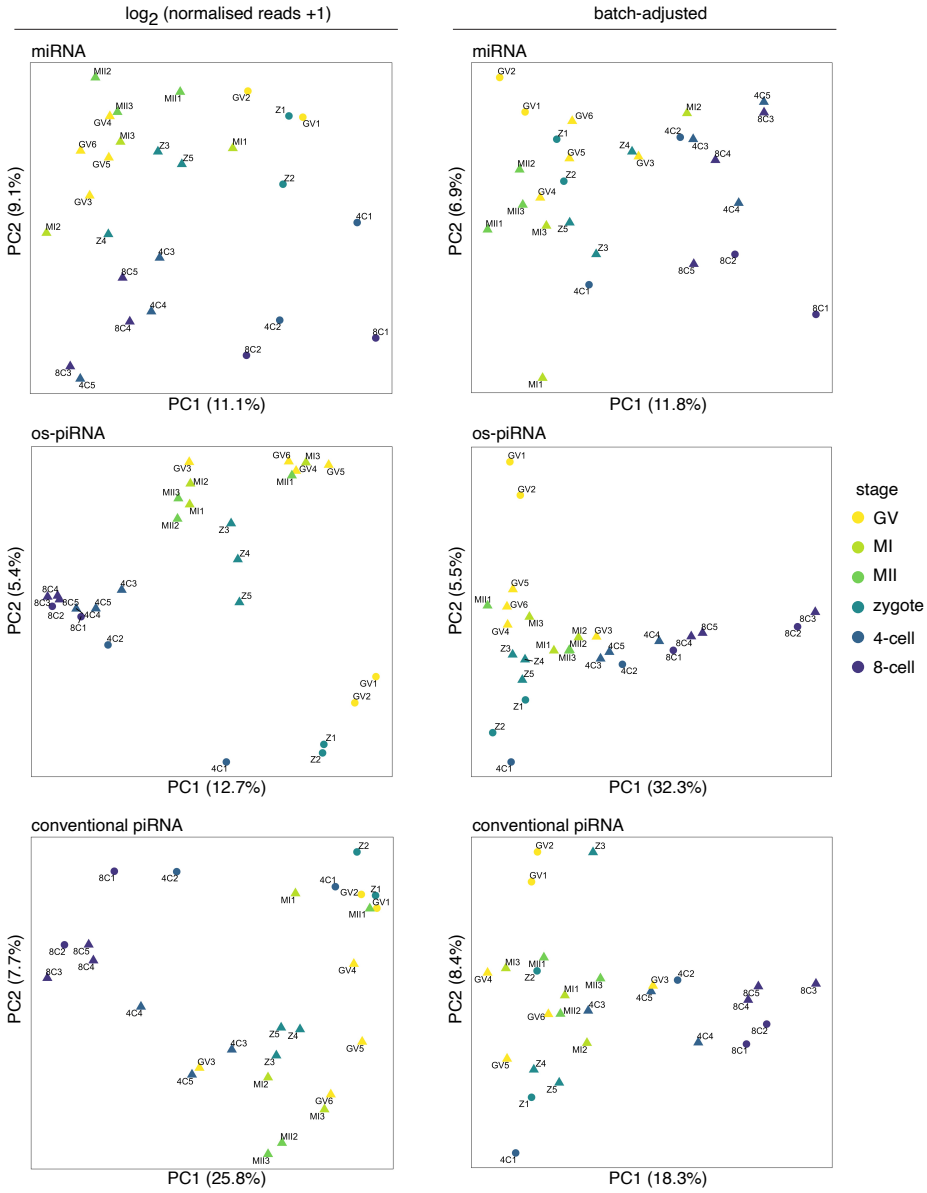
(*A*) sRNA class expression at different nucleotide (nt) lengths (17 -30 nt) in the oocyte maturation stages, zygotes, and cleavage stage embryos. Stage-wise mean proportion of reads mapped to each sRNA class are shown individually. microRNA, miRNA; mitochondrial tRNA, mt-tRNA; oocyte short piRNA, os-piRNA; Piwi-interacting RNA, conventional piRNA; ribosomal RNA, rRNA; transfer RNA, tRNA. (*B*) The relative nucleotide compositions of the conventional piRNAs, and os-piRNAs detected in oocytes, zygotes, and embryos. We used sRNAbench (Aparicio-Puerta et al. 2019; Rueda et al. 2015) with RNAcentral version 14 (hg38) database, and os-piRNA sequence information (Yang et al. 2019) to map and annotate conventional piRNAs and os-piRNAs. Base composition preference of the detected conventional piRNA and os-piRNA molecules are shown according to their sequence information in the databases. A representative plot for a single sample from each developmental stage is shown. The detected os-piRNAs and conventional piRNAs had a 1U and 10A preference, as expected, and a moderate 4A and 5A preference. Germinal vesicle oocyte, GV (*n*=6); Metaphase I oocyte, MI (*n*=3); Metaphase II oocyte, MII (*n*=3); zygote (*n*=5); 4-cell embryo, 4-cell (*n*=5); 8-cell embryo, 8-cell (*n*=5). The number behind the stage symbol indicates the stage-wise replicate number.



Supplemental Fig S4. miRNA, conventional piRNA, and os-piRNA genomic locations. (*A*) os-piRNA (left panel) and conventional piRNA (right panel) locations in genome elements. Both classes were located primarily in intergenic regions. (*B*) Intergenic os-piRNA (left panel) and conventional piRNA (right panel) locations in repeat elements. Both classes were located mainly within non-repeat elements and in lesser

numbers LINE or LTR associated regions. (*C*) The intergenic miRNAs were mainly associated with LINE and other repeat class flanking elements. (*A*-*C*) DNA transposon, DNA; Long interspersed nuclear element, LINE; Long terminal repeat, LTR; Other repeat class, Other; Short interspersed nuclear element, SINE. The repeat element flanking regions of 0-1kb and 1-2kb are indicated with FL1k and FL2k, respectively. The horizontal line in the boxplot indicates the median value. A blue bar indicates the sense and a red bar the antisense orientations of the intersected genomic loci. Intersection was performed with BEDTools 2.29.0 (Quinlan and Hall 2010) with GENCODE hg38 annotations obtained from UCSC's table browser (Comprehensive gene annotation v23).

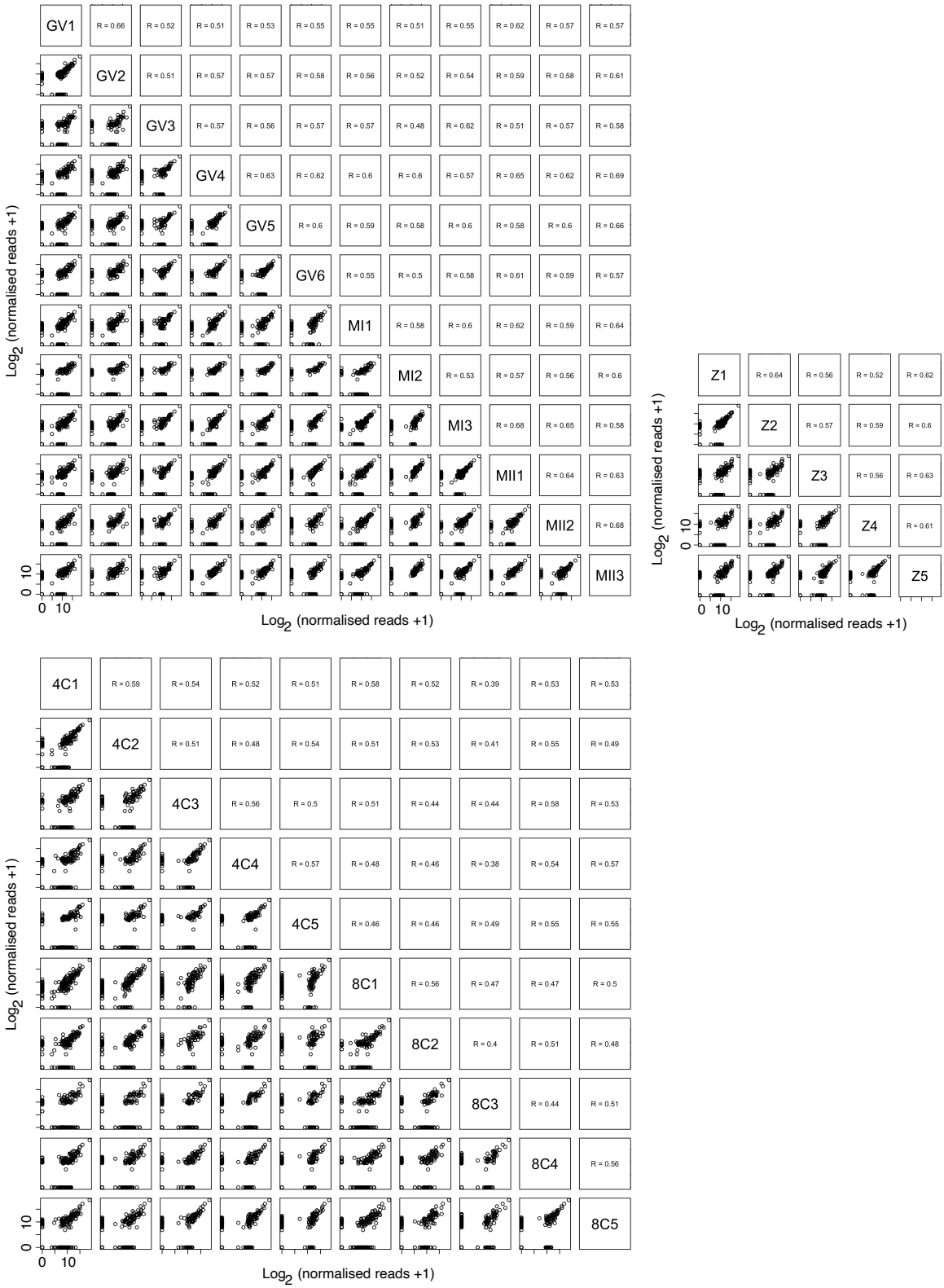
Supplemental_Fig_S5



Supplemental Fig S5. Principal component analysis (PCA) of miRNAs, os-piRNAs, and conventional piRNAs. PCA was performed using R FactoMiner package (Lê et al. 2008) with log2 transformed sRNA class library size normalised reads (left) as well as voom-normalised and limma batch-adjusted reads (Law et al. 2014; Ritchie et al. 2015) (right). miRNAs, os-piRNAs, and conventional piRNAs with >0 observations in at least 3 samples were included in the analyses. Prior to batch-adjustment the samples clustered according to sequencing batch and mainly by developmental stage within the two batches. Samples were batch-

adjusted using broad sample grouping (oocyte, zygote, and embryo) due to an unbalanced library design in the case of the different oocyte maturation stages, and thus some of the developmental stage-wise clustering may be lost upon correction. The batch-adjusted datasets clustered according to developmental stage, with clearest stage-wise separation between oocytes and embryos, for miRNAs and conventional piRNAs, and oocytes, zygotes, and embryos for os-piRNAs. There does not seem to be significant separation in the clustering of 3-pronuclear and the 2-pronuclear zygotes, while there is some general dispersion of samples belonging to the same developmental stage.

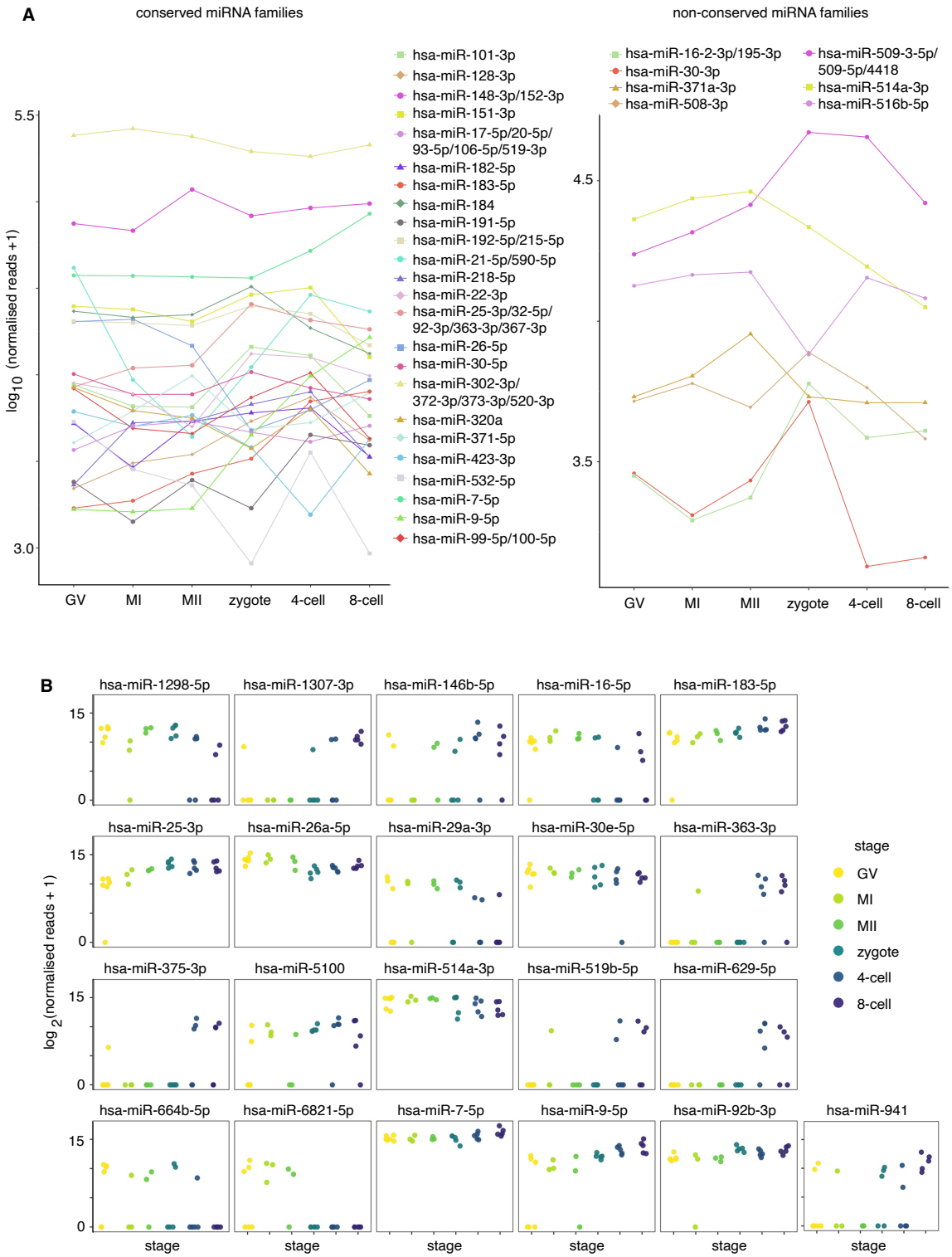
Supplemental_Fig_S6



109

Supplemental Fig S6. Correlation analysis of miRNA expression. Pairwise correlations of oocytes (top left), zygotes (top right), and embryos (bottom) were assessed using normalized miRNA reads. R represent the Pearson correlation coefficient. Only miRNAs detected in both samples to be compared were used for calculation of R. Samples of the same stage had a R in the range of 0.38-0.69. The scatter plots show normalized miRNA expression of all detected miRNAs (>0 observations in at least 3 samples) for all sample pairs.

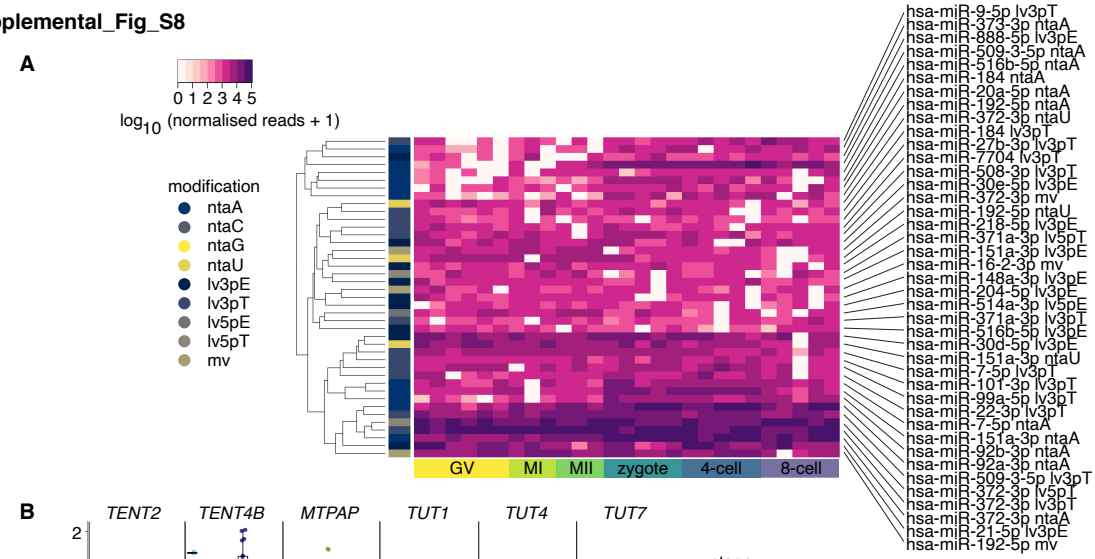
Supplemental_Fig_S7

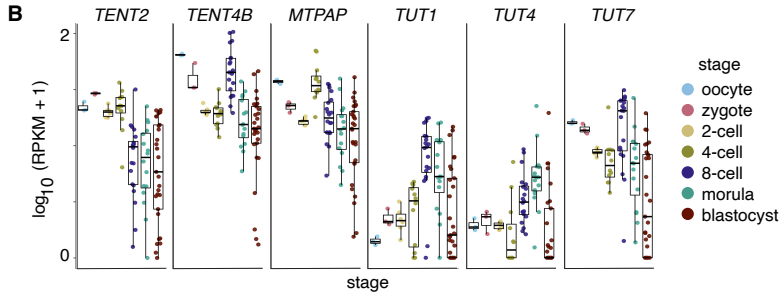


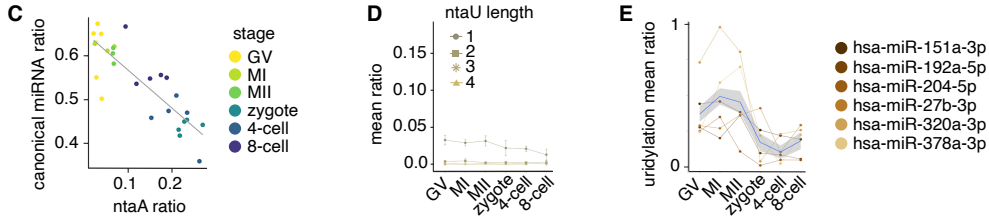
Supplemental Fig S7. Highly and differentially expressed (DE) miRNAs. (*A*) Expression of conserved and non-conserved miRNA families of the highly expressed miRNAs in oocytes, zygotes, and embryos.

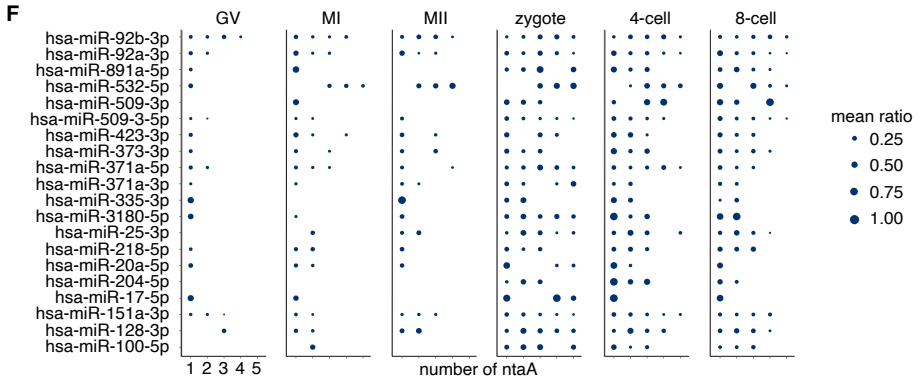
Normalised reads of the highly expressed miRNAs belonging to each miRNA family are aggregated and presented as the stage-wise mean. (*B*) DE miRNA expression levels in different developmental stages. Germinal vesicle oocyte, GV; Metaphase I oocyte, MI; Metaphase II oocyte, MII; 4-cell embryo, 4-cell; 8-cell embryo, 8-cell.

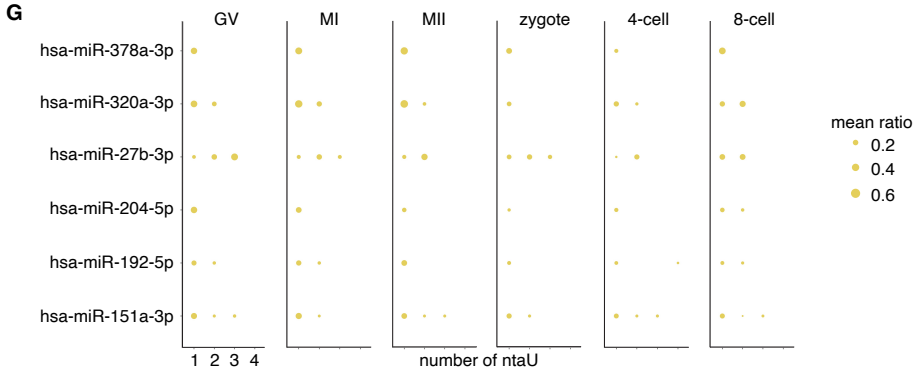
Supplemental_Fig_S8







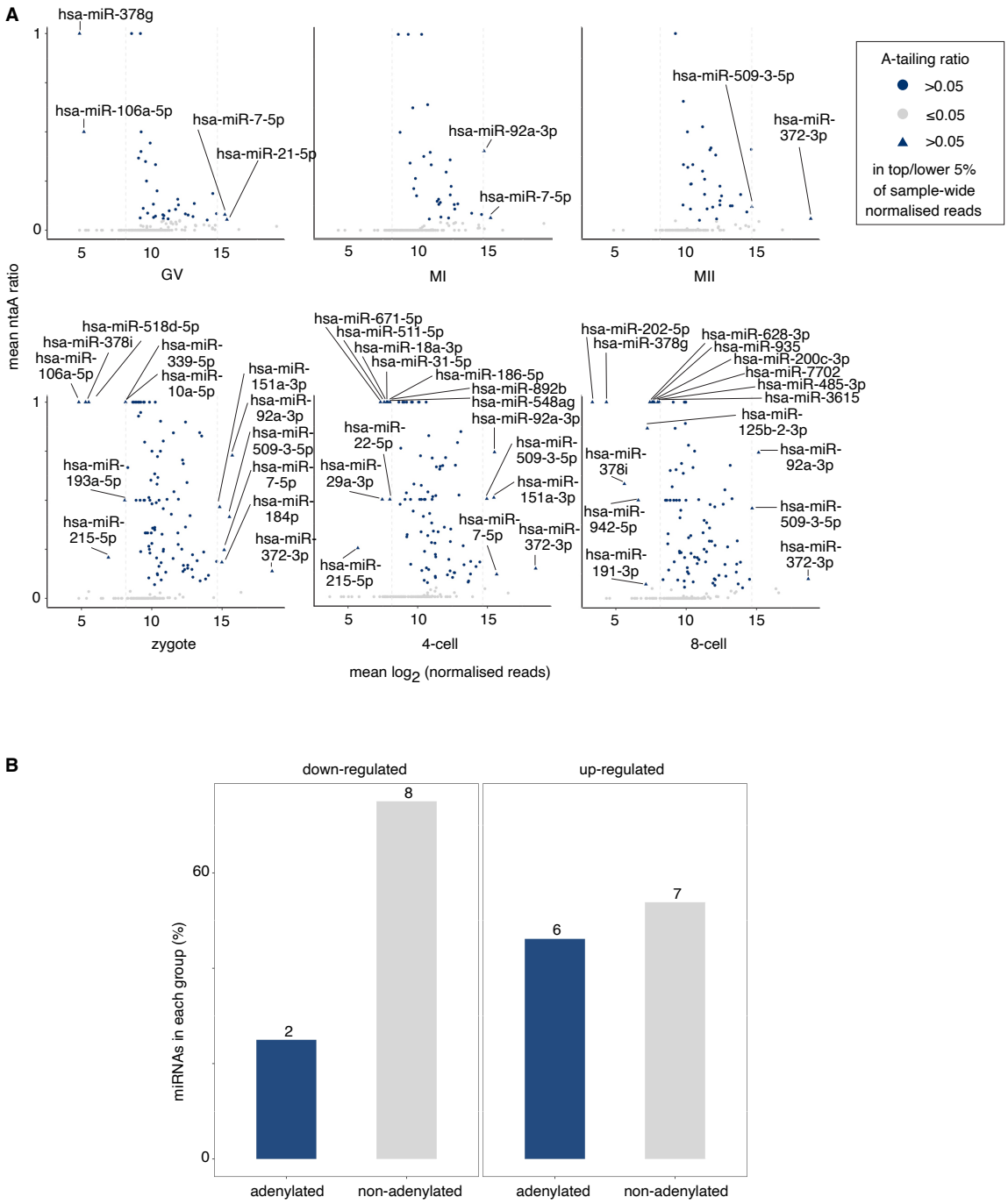




123

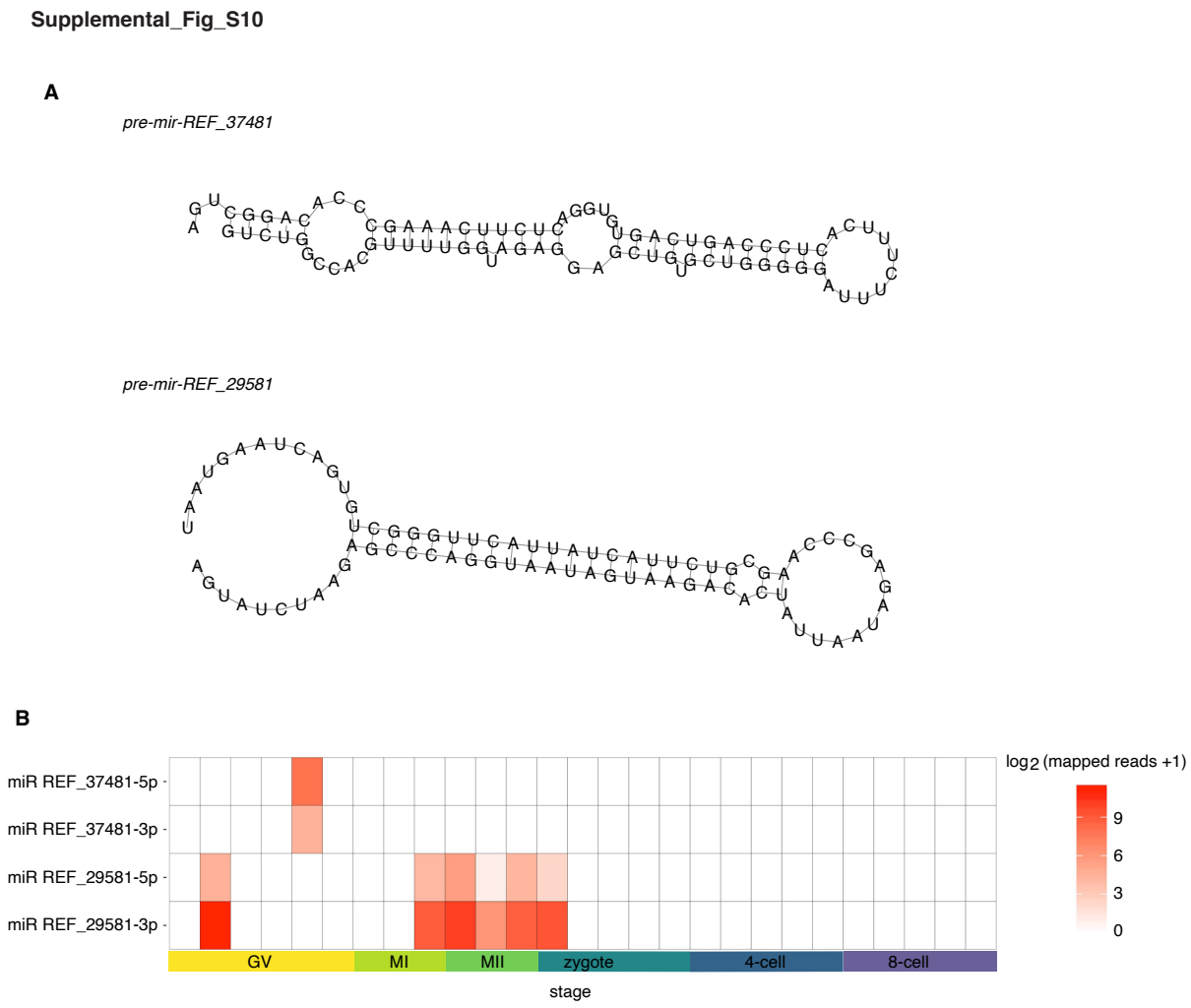
Supplemental Fig S8. isomiR expression in early human development. (*A*) Heatmap of miRNA modification isoforms (isomiRs) with highest expression levels (top 20% according to median, detected in > 80% of samples) across samples. Euclidean distance and average linkage were used to organise isomiRs according to expression. Germinal vesicle oocyte, GV; Metaphase I oocyte, MI; Metaphase II oocyte, MII; 4-cell embryo, 4-cell; 8-cell embryo, 8-cell. (*B*) Small RNA adenylation (*TENT2,* former *PAPD4; TENT4B*, former *PAPD5*; *MTPAP*, former *PAPD1*) and uridylation (*TUT1, TUT4, TUT7*) factor mRNA levels in human oocytes and pre-implantation embryos (Yan et al. 2013). Adenylation factors were expressed at higher levels than uridylation factors throughout development (*FDR* <0.01, paired Wilcoxon signed rank test, two-sided). The horizontal line in the boxplot indicates the expression median. *(C)* Sample-wise canonical sequence and 3' adenylated (A-tailed) miRNA read ratios in different developmental stages. We found both the canonical sequence and A-tailing modification classes to be differentially expressed (DE; *FDR*< 0.05) between oocytes (*n*=10) and embryos (*n*=12), using a linear model to account for batch. (*D*) miRNA 3' uridylation (U-tailing) modification length in different developmental stages. Modification lengths of up to 5 nucleotides are shown. (*E*) Modification dynamics of miRNAs with highest 3' uridylation (U-tailing) ratios. A total of 6 miRNA species were analysed (mean normalised reads > 1000 and mean modification ratio > 0.33 in at least one of the stages). The data points represent the developmental stage-wise mean U-tailing ratio. The locally estimated scatterplot smoothing (loess) curve deviated from the overall U-tailing modification trend (Fig. 4C) with the oocyte stages having higher modification ratios in the 6 analysed miRNA species. (*F*-*G*) miRNA 3' adenylation (*F*) and uridylation (*G*) modification length for miRNAs with highest modification levels. Size of the circle indicates stage-wise mean modification ratio. (*C*-*G*) Mean stage-wise miRNA modification ratio and (*D*) the 95% confidence interval are shown.

Supplemental _Fig_S9



Supplemental Fig S9. miRNA 3' adenylation and miRNA abundance. (*A*) Scatter plot of mean stage-wise microRNA (miRNA) 3' adenylation (A-tailing) ratio and miRNA abundance in human oocyte, zygote, and embryo stages. All 260 mature miRNA species included in the analyses are shown. A-tailed miRNAs (stage-wise mean of A-tailed reads > 0.05; blue) with the stage-wise mean abundance in the top or lower 5% (separated by dotted grey lines) are named. There was no clear correlation between miRNA A-tailing ratio

and miRNA abundance. Several highly A-tailed miRNAs were found in abundant levels across multiple developmental stages indicating that A-tailing may selectively stabilize specific miRNAs. Germinal vesicle oocyte, GV; Metaphase I oocyte, MI; Metaphase II oocyte, MII; 4-cell embryo, 4-cell; 8-cell embryo, 8-cell. *(B)* The differentially expressed miRNAs that were determined using MACAU 2.0 v1.10 (Sun et al. 2017) were defined as adenylated if the mean sample-wide A-tailing percentage was over 5%. Approximately half of the miRNAs that were up-regulated in embryos were also A-tailed while for the embryo down-regulated miRNAs A-tailing was observed for a quarter of the miRNAs. The number of miRNAs belonging to each category are indicated above the columns.



Supplemental Fig S10. Novel human pre-miRNAs. *(A)* Putative novel pre-microRNA hairpin sequences predicted by sRNAbench. RNAfold web server (Lorenz et al. 2011) with parameters corresponding to RNAfold -d2 –noLP in RNAfold 2.4.13 was used for putative pre-miRNA hairpin structure prediction. *(B)* Expression of novel miRNA 5p and 3p sequences corresponding to the identified putative pre-miRNAs in human oocytes, zygotes, and embryos. Log2 transformation of 5p and 3p sequence mapped reads for each putative novel miRNA are shown. Sample-wise miRNA library sizes are given in Supplemental Table S1A. Germinal vesicle oocyte, GV; Metaphase I oocyte, MI; Metaphase II oocyte, MII; 4-cell embryo, 4-cell; 8-cell embryo, 8-cell.

Supplemental Table S1A. Summary of small RNA sequencing data

Sample ID	Batch	Group	Group broad	Donor	Raw reads	High-quality reads (17-30 nt)	Library mapped reads	Percentage of library mapped reads (%)	miRNA mapped reads	os-piRNA mapped reads	conventional piRNA mapped reads
GV1	Set1	GV	oocyte	ind1	14390516	6521775	5597577	85.8	34358	1710042	14556
GV2	Set1	GV	oocyte	ind1	17446476	5439398	4567119	84.0	27687	1200777	11690
GV3	Set2	GV	oocyte	ind5	1299576	638516	541653	84.8	11619	170955	1799
GV4	Set2	GV	oocyte	ind9	4020883	2462699	2215189	89.9	22234	723607	8181
GV5	Set2	GV	oocyte	ind10	5248672	2755762	2393791	86.9	23430	715814	9343
GV6	Set2	GV	oocyte	ind11	3336892	2237355	2052009	91.7	14754	695480	7712
MI1	Set2	MI	oocyte	ind5	4302443	1616543	1347384	83.3	50587	311213	4504
MI2	Set2	MI	oocyte	ind9	1569799	795870	713464	89.6	6245	226257	2474
MI3	Set2	MI	oocyte	ind10	3854054	2123318	1903643	89.7	28508	561476	7742
MII1	Set2	MII	oocyte	ind11	4994479	3306109	2905818	87.9	34868	915120	9963
MII2	Set2	MII	oocyte	ind12	1600520	933011	813317	87.2	24678	238752	3441
MII3	Set2	MII	oocyte	ind12	1895320	1011003	881240	87.2	24906	245553	3176
Z1	Set1	zygote	zygote	ind2	11980763	4735847	3902966	82.4	25455	1002231	10722
Z2	Set1	zygote	zygote	ind2	19855711	7400529	6160517	83.2	35589	1518912	14660
Z3	Set2	zygote	zygote	ind15	2877658	1759571	1537988	87.4	21856	503525	5024
Z4	Set2	zygote	zygote	ind17	2792556	1672546	1456960	87.1	18507	446720	4787
Z5	Set2	zygote	zygote	ind17	3067169	1790849	1550200	86.6	26235	485618	4353
4C1	Set1	4_cell	embryo	ind3	19184546	6197366	4818929	77.8	77471	922877	12178
4C2	Set1	4_cell	embryo	ind3	6822202	1736592	1300611	74.9	30557	172738	2222
4C3	Set2	4_cell	embryo	ind13	2358708	923508	756504	81.9	12449	168000	2691
4C4	Set2	4_cell	embryo	ind13	1787880	659105	530300	80.5	12733	70490	1242
4C5	Set2	4_cell	embryo	ind13	5293336	1785489	1298012	72.7	8194	141663	2097
8C1	Set1	8_cell	embryo	ind4	15100759	2738667	1876748	68.5	99550	87496	2492
8C2	Set1	8_cell	embryo	ind3	4572029	738621	469565	63.6	26993	19649	507
8C3	Set2	8_cell	embryo	ind14	742206	183249	119349	65.1	12117	11441	274
8C4	Set2	8_cell	embryo	ind16	940055	347491	278466	80.1	14967	53410	794
8C5	Set2	8_cell	embryo	ind16	906686	352863	263577	74.7	17537	43594	754

Small RNA sequencing library information with Sample ID, sequencing batch, sample grouping, and donor information are shown for each sample. Sequencing library size metrics include number of raw reads,

number of high-quality reads of length 17-30 nucleotides (nt) after adapter removal, percentage of library mapped reads as well as number of reads mapped to miRNAs, os-piRNAs, and conventional piRNAs.

Supplemental Table S1B. Summary of detected miRNA, conventional piRNA, and os-piRNA species

Small RNA class	**Number of detected species**
miRNA	260
os-piRNA	22851
conventional piRNA	153

Number of detected (>0 reads in at least 3 samples) miRNA os-piRNAs, and conventional piRNA species in human oocytes, zygotes, and embryos are shown.

Supplemental Table S6. Significantly differentially expressed miRNA modifications in oocytes (*n*=12) and embryos (*n*=10).

miRNA modification	term	estimate	std.error	statistic	p-value	FDR
ntaA	Biol_group_broadembryo	0.1404	0.0158	8.8960	3.34E-08	3.34E-07
ntaG	Biol_group_broadembryo	0.0062	0.0011	5.4083	3.22E-05	1.07E-04
ntaU	Biol_group_broadembryo	-0.0189	0.0027	-6.8930	1.42E-06	7.10E-06
lv3pT	Biol_group_broadembryo	-0.0254	0.0072	-3.5531	2.12E-03	5.31E-03
exact	Biol_group_broadembryo	-0.0858	0.0274	-3.1274	5.55E-03	1.11E-02
lv3pE	Biol_group_broadembryo	-0.0234	0.0092	-2.5397	2.00E-02	3.33E-02
lv5pE	Biol_group_broadembryo	-0.0025	0.0010	-2.4347	2.49E-02	3.56E-02

Significantly differentially expressed miRNA modifications (*FDR* < 0.05) between human oocytes (*n*=12) and embryos (*n*=10) were determined using a linear model adjusting for batch. Biological group parameter estimates, standard error (std. error), test statistic (statistic), p-value, and FDR are shown. Table is sorted by FDR. Exact (canonical) mature sequence, exact. 3' trimmed, lv3pT; 3' elongated, lv3pE; 5' elongated, lv5pE; Non-templated nucleotide addition (NTA) of A, ntaA; NTA of G, ntaG; NTA of U, ntaU.

Supplemental References

Aparicio-Puerta E, Lebrón R, Rueda A, Gómez-Martín C, Giannoukakos S, Jaspez D, Medina JM, Zubkovic A, Jurak I, Fromm B, et al. 2019. sRNAbench and sRNAtoolbox 2019: intuitive fast small RNA profiling and differential expression. *Nucleic Acids Res* **47**: W530–W535. https://pubmed.ncbi.nlm.nih.gov/31114926.

Kozomara A, Birgaoanu M, Griffiths-Jones S. 2018. miRBase: from microRNA sequences to function. *Nucleic Acids Res* **47**: D155–D162. https://doi.org/10.1093/nar/gky1141.

Law CW, Chen Y, Shi W, Smyth GK. 2014. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**: R29. https://doi.org/10.1186/gb-2014-15-2-r29.

Lê S, Josse J, Husson F. 2008. FactoMineR : An R Package for Multivariate Analysis. *J Stat Softw* **25**: 1–18. https://www.jstatsoft.org/v025/i01.

Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**: 26. https://doi.org/10.1186/1748-7188-6-26.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842. https://pubmed.ncbi.nlm.nih.gov/20110278.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**: e47–e47. https://pubmed.ncbi.nlm.nih.gov/25605792.

Rueda A, Barturen G, Lebrón R, Gómez-Martín C, Alganza Á, Oliver JL, Hackenberg M. 2015. sRNAtoolbox: an integrated collection of small RNA research tools. *Nucleic Acids Res* **43**: W467–W473. https://pubmed.ncbi.nlm.nih.gov/26019179.

Sun S, Hood M, Scott L, Peng Q, Mukherjee S, Tung J, Zhou X. 2017. Differential expression analysis for RNAseq using Poisson mixed models. *Nucleic Acids Res* **45**: e106. http://www.ncbi.nlm.nih.gov/pubmed/28369632.

Yan L, Yang M, Guo H, Yang L, Wu J, Li R, Liu P, Lian Y, Zheng X, Yan J, et al. 2013. Single-cell RNA-

215 Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol* **20**:
216 1131–1139. http://dx.doi.org/10.1038/nsmb.2660.

217 Yang Q, Li R, Lyu Q, Hou L, Liu Z, Sun Q, Liu M, Shi H, Xu B, Yin M, et al. 2019. Single-cell CAS-seq
218 reveals a class of short PIWI-interacting RNAs in human oocytes. *Nat Commun* **10**: 3389.
219 https://doi.org/10.1038/s41467-019-11312-8.

220