Supplemental Methods

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

Pauliina Paloviita, Christel Hydén-Granskog, Dawit A. Yohannes, Priit Paluoja, Juha Kere, Juha S. Tapanainen, Kaarel Krjutškov, Timo Tuuri, Urmo Võsa and Sanna Vuoristo*

*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

Immunocytochemistry and confocal imaging of human embryos

Four-cell (*n*=3) and 8-cell (*n*=2) embryos were washed briefly in PBS (Corning), fixed in 3.8% PFA at room temperature for 15 min, and washed three times in washing buffer (0.1% (v/v) Tween20 in PBS). Embryos were permeabilised in 0.5% (v/v) Triton X-100 in PBS at room temperature for 15 minutes. After three washes, unspecific primary antibody binding was blocked by incubating the samples in ProteinBlock (Thermo Fisher Scientific) at room temperature for 10 minutes. The samples were incubated in primary monoclonal antibodies (1:100 rabbit anti-DGCR8, clone 3F5 (MA5-24860; Thermo Fisher Scientific) and 4 µg/ml mouse anti-DICER, clone CL0378 (MA5-31353; Thermo Fisher Scientific) diluted in washing buffer at 4˚C overnight. The samples were washed three times and incubated with fluorescent secondary antibodies (donkey anti-rabbit 488 (A21206), and donkey anti-mouse 594 (A21203), both from Thermo Fisher Scientific) diluted 1:500 in washing buffer, for 2 hours at room temperature. The samples were washed twice in washing buffer, and nuclei were labelled with DAPI (1:1000 in washing buffer) at room temperature for 10 minutes. The samples were washed and transferred in washing buffer and imaged on Ibidi 8-well μ slide using a Leica TCS SP8 confocal laser scanning microscope (Leica Systems, Mannheim, Germany). Images were captured with HC PL APO CS2 40X/1.10NA water objective and processed using Fiji (http://fiji.sc). The representative single z-planes images were smoothened using a Gaussian filter (radius=1 pixel kernel).

Sequencing library preparation

Small RNA (sRNA) sequencing (sRNA-seq) library preparation was performed using NEBNext Small RNA Library Prep Set for Illumina kit (New England Biolabs, NEB) with following modifications. The samples were briefly rinsed in $Ca_2^{+/}Mg_2^{+}$ -free PBS and placed in 3 µl of lysis buffer containing 5 mM Tris-HCl, pH 7.5 (Sigma-Aldrich); 0.1% Tween-20 (Sigma-Aldrich); 50 mM KCl (Sigma-Aldrich) and 2.5 units of RiboLock RNase inhibitor (Thermo Fisher Scientific). The samples were stored at -80°C. After thawing the samples, 1 µl of 1:3 diluted 3' SR Adapter (NEB), 1 µl of 0.5 M KCl, 1 µl of 20 µM 5S and 5.8S rRNA masking oligo mixture, and 1 µl of nuclease-free water was added to each sample. The samples were incubated in a preheated thermal cycler at 90°C for 1 min, followed by a 2-minute incubation at 60°C to mask the ribosomal RNA (rRNA). The samples were transferred on ice. 3 µl of 3' ligation enzyme mix (NEB) and 10

µl of 3’ adapter ligation reaction buffer (NEB) was added on each sample. The samples were briefly vortexed and centrifuged before incubation in thermal cycler at 25°C for 60 min. After incubation, 4.5 µl of nuclease-free water and 1 µl of 1:3 diluted SR RT Primer (NEB) were added to each sample. The libraries were incubated at 75°C for 5 min, followed by incubations at 37°C for 15 min and at 25°C for 15 min. During the final incubation step, 5’ SR adapter (NEB) was re-suspended in nuclease-free water. 1:3 diluted adapter was denaturated by incubation in thermal cycler, at 70°C, for 2 min and then immediately placed on ice. Denaturated adapter was used within 30 mins. One microliter of denatured 5’ SR adapter (NEB), 1 µl of 10× ligation reaction buffer (NEB), and 2.5 µl of 5’ ligase enzyme mix (NEB) were added in each sample. The samples were briefly vortexed, centrifuged, and incubated in thermal cycler, at 25°C for 60 mins. After incubation, 8 µl of First Strand Synthesis buffer (NEB), 1 µl of Murine RNase inhibitor (NEB), and 1 µl of ProtoScript II Reverse Transcriptase (NEB) were added to the samples. The libraries were incubated at 50°C for 60 min. The enzyme was inactivated at 75°C for 10 min. The cDNA pool (40 µl) was purified using NucleoSpin Gel and PCR Clean up Columns for gel extraction and PCR clean up (Macherey-Nagel) according to PCR product purification protocol and eluted finally in 30 µl elution buffer. 30 µl 2× Phusion Hot MasterMix (Thermo Fisher Scientific) and 1 µl of each 100 µM primer, universal primer and barcoded primer where NNNNNN is 6 bp specific index sequences in multiplex assay, were added to purified cDNA mixture. All oligonucleotides were desalted. Following PCR was carried out under following conditions: initial denaturation and activation 1 min at 98°C, cycle denaturation 10 s at 98°C, annealing 20 s at 62°C, extension 5 s at 70°C and final extension 5 min at 72°C. In total 19 cycles were performed. The PCR product (80 µl) was purified using NucleoSpin Gel and PCR Clean up Columns for gel extraction and PCR clean up (Macherey-Nagel) according to PCR product purification protocol and eluted finally in 25 µl elution buffer. Small PCR fragments were removed by adding 28 µl AMPureXP beads (Beckman), mixed well and incubated 10 min at room temperature. Beads were captured by magnet and supernatant was completely removed. The pellet was resuspended in 30 µl nuclease-free water and placed back on magnet. Clear supernatant, ready sRNA library, was transferred to a clean tube. The library was quantified by KAPA SYBR FAST qPCR kit (Roche) according to instructions. The sRNA library was sequenced using Illumina MiSeq (Illumina) instrument with read-1 primer for 50 bp and index primer for 6 bp. The oligo and primer sequences are listed in Supplemental Table S7.

Supplemental Table S7. Summary of primer and oligo sequences.

Name	Sequence (5'-3')	Vendor
5S RNA masking oligo	ATCGGCAAAGCCTACAGCACCCGGTATTCCCAGG-biotin	Metabion International AG
5.8S RNA masking oligo	ATCGGCAAAGCCTACAGCACCCGGTATTCCCAGG-biotin	Metabion
universal primer	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGACGA	Metabion
barcoded primer	CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	Metabion
read-1 primer	GATCTACACGTTCAGAGTTCTACAGTCCGACGATC	Sigma-Aldrich
index primer	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC	Sigma-Aldrich

Pre-processing of sRNA-seq data

The pre-processing of the raw sRNA-seq reads and the expression profiling of sRNAs was performed using sRNAbench (Aparicio-Puerta et al. 2019; Rueda et al. 2015). Sequence reads that contained at least the first 10 nucleotides (nt) of the adapter sequence (AGATCGGAAGAGCACACGTCT) were adapter trimmed and kept. Adapter trimmed reads of sizes 17 to 30 nt were used for the expression profiling. We used the genome mode approach in sRNAbench for expression profiling (Rueda et al. 2015). Briefly, the pre-processed reads were aligned using Bowtie (Langmead et al. 2009) to the human genome (GRCh38, primary assembly from Ensembl) allowing one mismatch to determine genomic coordinates for the aligned reads, and further mapped in a hierarchical manner to the following databases: human miRBase (v22) (Kozomara et al. 2018), Ensembl cDNA (hg38), Ensembl non-coding RNA (hg38), RNAcentral version 14 (hg38), and oocyte short piwi-interacting RNAs (piRNAs; os-piRNA) (Yang et al. 2019). Reads were assigned to target reference RNAs only if they align to the reference RNAs and their genomic coordinates lie completely within the genomic

coordinates of their target reference RNA. The remaining unassigned reads of length 17 to 25 nt were aligned to the human genome (GRCh38) for detection of novel microRNAs (miRNAs). To evaluate the distribution of the detected miRNAs, conventional piRNAs, and os-piRNAs across genomic elements, their genomic coordinates were gathered into bed format. For miRNAs and conventional piRNAs these were obtained directly from the miRBase and RNAcentral databases respectively, while for os-piRNAs, genomic coordinates were determined from the alignment genomic coordinates of sequence reads that aligned to the genome and were assigned to os-piRNAs by sRNAbench. The coordinates of these sRNAs were then intersected with GENCODE hg38 annotations (in bedfiles) of coding and noncoding gene elements obtained from UCSC's table browser (Comprehensive gene annotation v23, after filtering for coding and noncoding by Transcript class). Intersection was performed with BEDTools 2.29.0 (Quinlan and Hall 2010) and sRNAs were assigned to one of 6 genomic elements in a successive manner from coding CDS, 5′-UTR, 3′-UTR, intron, and noncoding exon and intron resembling analyses by Yang et al. 2019. In each step only the remaining sRNAs were used in the next step, and with priority of sense over antisense. The remaining sRNAs not assigned to the six gene elements were considered intergenic. To assess the distribution of the intergenic sRNAs in repetitive elements and their flanking regions, similar analyses was performed by intersecting the intergenic sRNAs with repetitive elements downloaded from UCSC, and their flanking regions of 0-1kb and 1-2kb calculated with BEDTools. Intergenic sRNAs overlapping with LINE, LTR, SINE, DNA, and other repeats were determined after which the remaining sRNAs were assigned to flanking regions in LINE-FL1k, LTR-FL1k, SINE-FL1k, DNA-FL1k, other repeat-FL1k and successively in LINE-FL2k, LTR-FL2k, SINE-FL2k, DNA-FL2k, and other repeat-FL2k.

Differential expression analyses

Several samples included in the study were genetically related due to being donated by the same individual or couple. To account for sample relatedness, we utilized R package MACAU 2.0 v1.10 (Sun et al. 2017) in DE analysis of miRNAs and miRNA modification isoforms (isomiRs). Oocytes (n=12) were compared to embryos (n=10) using the Poisson mixed model with filtering set to "TRUE" in MACAU 2.0. We required each miRNA/isomiR to have > 1 reads in at least 5 samples after adding a pseudocount of 1 to be included in the analysis. Phenotypes were coded as 0 and 1 (0, oocytes; 1, embryos) and sequencing batch information was

provided as a covariate (batch 1; n=6, batch 2; n=16). The model incorporated relatedness matrix was construct, by assigning the value 0.5 to samples originating from the same donor/donor couple and 0 to samples from different individuals. Expression changes were regarded as significant when FDR < 0.05, while no cut-off was applied for the effect size (β) or the heritability (h^2) term. miRNA modification ratio differences between oocytes and embryos were determined using a linear model with the modification ratio as the response variable and batch and phenotype as the explanatory variables. Phenotypes and batch were coded as before. A linear mixed random effects model with a random effect term to account for sample relatedness could not be used due to lack of convergence. P-values for the phenotype variable were extracted from each model and adjusted for multiple testing using the Benjamini–Hochberg method (Benjamini and Hochberg 1995). Modifications with FDR < 0.05 were regarded as differentially expressed.

Novel miRNAs

We predicted novel miRNAs with miRDeep2 v0.1.2 (Friedländer et al. 2012) and sRNAbench (Aparicio-Puerta et al. 2019; Rueda et al. 2015). Detection of novel miRNAs was performed using unassigned, pre-processed reads that were 17 to 25 nt long. For miRDeep2, reads from different samples were combined for greatest prediction sensitivity. Mature human miRNA and hairpin sequences from miRBase v22 were provided to detect reads mapping to known miRNAs. Mature miRNAs of all other available species were used to detect novel miRNAs with annotated miRNA homologs in other species. We required a >90% overlap of pre-miRNA coordinate predictions by both software and expression of putative miRNA and/or star sequence in >1 sample with >0 reads assigned by sRNAbench. The novel (pre-)miRNA candidates were further investigated for their confidence based on miRBase high confidence criteria (Kozomara and Griffiths-Jones 2014) and additional quality criteria stated in Fromm et al. (2015).

Human oocyte and pre-implantation embryo mRNA expression data

Previously published human oocyte and embryo RNA-seq data (Yan et al. 2013) was used to analyse gene expression changes between the different developmental stages using pairwise, two-sided Wilcoxon rank sum tests. Expression differences between miRNA 3' adenylating and uridylating genes were assessed using

146 pairwise, paired, two-sided Wilcoxon signed rank tests. P-values were adjusted for multiple testing using the
147 Benjamini–Hochberg method (Benjamini and Hochberg 1995) and expression changes with *FDR*<0.05 were
148 regarded as significant. The RPKM normalised gene expression values from the Yan et al. 2013 (Yan et al.
149 2013) data set were downloaded from the article web-site.

150

151

Supplemental Methods 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.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B* **57**: 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. 2012. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res* **40**: 37–52. https://pubmed.ncbi.nlm.nih.gov/21911355.

Fromm B, Billipp T, Peck LE, Johansen M, Tarver JE, King BL, Newcomb JM, Sempere LF, Flatmark K, Hovig E, et al. 2015. A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome. *Annu Rev Genet* **49**: 213–242. https://pubmed.ncbi.nlm.nih.gov/26473382.

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.

Kozomara A, Griffiths-Jones S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**: D68–D73. https://pubmed.ncbi.nlm.nih.gov/24275495.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25. https://doi.org/10.1186/gb-2009-10-3-r25.

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.

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

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 reveals a class of short PIWI-interacting RNAs in human oocytes. *Nat Commun* **10**: 3389. https://doi.org/10.1038/s41467-019-11312-8.