nature biotechnology

Single-cell sequencing of the small-RNA transcriptome

Omid R Faridani[1,6], Ilgar Abdullayev[1,2,6], Michael Hagemann-Jensen[1,3], John P Schell[4,5], Fredrik Lanner[4,5] & Rickard Sandberg[1,2]

Little is known about the heterogeneity of small-RNA expression as small-RNA profiling has so far required large numbers of cells. Here we present a single-cell method for small-RNA sequencing and apply it to naive and primed human embryonic stem cells and cancer cells. Analysis of microRNAs and fragments of tRNAs and small nucleolar RNAs (snoRNAs) reveals the potential of microRNAs as markers for different cell types and states.

The ability to profile mRNAs in single cells has transformed the molecular identification and the characterization of cell types, states and rare cellular phenotypes[1,2]. However, current single-cell methods are restricted to long RNAs[1–3], primarily mRNAs. Several classes of small RNAs exist in mammalian cells, including microRNAs (miRNAs)[4], and RNA fragments derived from tRNAs (tsRNAs, tRNA-derived small RNAs) and snoRNAs (sdRNAs, sno-derived RNAs)[5,6], some of which have been shown to have tissue[7], cancer-type[8,9] or cell-state-specific[7] expression. Profiling of small RNAs has required large amounts of sample, but a more comprehensive understanding of their roles necessitates analysis of their expression patterns in single cells.

Here we describe a method for sequencing the small-RNA transcriptome in single cells (**Supplementary Fig. 1**). We constructed libraries by ligating adapters to all RNA species harboring 5′ phosphate and 3′ hydroxyl groups regardless of their size. Masking oligonucleotides were designed to evade highly abundant 5.8S rRNA. Also, an enzymatic digestion step was applied to reduce the formation of adaptor dimers. Unique molecular identifiers (UMIs)[10] were added to the 5′ adapters to counteract PCR stochasticity[11] and to enable RNA molecule counting. No experimental size-selection step was used (simplifying automation) and small RNAs were identified computationally (see below and Online Methods).

First, we characterized the method by sequencing small RNAs from individual naive and primed human embryonic stem cells (hESCs). In addition, we also sequenced individual HEK293FT cells to compare with bulk small-RNA sequencing. A computational pipeline (**Supplementary Software**) was developed to assign sequenced reads to the different small-RNA classes (using Gencode, mirBase and GtRNAdb), to convert reads to molecules using the adjacency method and to separate small RNAs from their precursor RNA species (**Supplementary Figs. 2** and **3**). Among the small regulatory RNAs (18–40 nucleotides (nts) long), we focused on miRNAs, tsRNAs and sdRNAs (**Fig. 1a** and **Supplementary Fig. 4a**). On average, we captured 3,800 miRNA, 3,500 tsRNA and 600 sdRNA molecules per cell (**Fig. 1b**). Next we investigated the length distributions of the small RNAs and observed that miRNAs, as expected, were ~22 nts long. Notably, different length distributions were observed for tsRNAs (mainly 33 and 19 nts) and sdRNAs (mainly 29 and 38 nts) (**Fig. 1c**), suggesting capture of the processed small RNAs and not intermediate degradation products. mRNA-derived small RNAs also showed distributions of around 18 and 31 nts (**Supplementary Fig. 4b**).

Aligning the small RNAs to their respective precursors demonstrated, as expected, that mature miRNAs covered the two arms of the pre-miRNA hairpins. The tsRNA molecules were mainly derived from the two regions of the mature tRNA, the 5′ and 3′ tRNA halves[12], and from the region positioned after the 3′ end of the mature tRNA corresponding to the tRNA-derived RNA fragment-1 (tRF-1)[13] (**Fig. 1d**). The sdRNAs were mainly derived from the 5′ ends of snoRNAs. This was further evidence that our single-cell libraries had captured small RNAs and not random degradation products. In addition to small regulatory RNAs, our single-cell libraries contained sequence reads likely coming from longer transcripts including snoRNAs, tRNAs and, to a lesser extent, pre-miRNAs (**Fig. 1e**,**f**). These RNAs serve as precursors for sdRNA, tsRNA and miRNA, respectively. Expression of sdRNA and tsRNA correlated moderately with the expression of their corresponding precursor transcripts (**Fig. 1g**). However, miRNA and pre-miRNA correlated poorly, possibly a result of short half-lives and secondary structures of pre-miRNAs.

Comparison of miRNAs in naive and primed hESCs showed that 60% of miRNAs were differentially expressed (**Fig. 1h** and **Supplementary Table 1**), consistent with miRNA profiling of naive and primed cells in bulk (**Supplementary Fig. 5**)[14]. Notably, the miR-302 family was more highly expressed in primed hESCs whereas the miR-371-3 cluster was more highly expressed in naive hESCs (**Fig. 1h**). The miR-302 family was reported to regulate cell cycle and apoptosis in hESCs[15] whereas, the miR-371-3 cluster was suggested to be crucial for the maintenance of hESC pluripotency[16] and negatively correlated with neural differentiation in human pluripotent stem cells[17]. In contrast to miRNAs, few tsRNAs or sdRNAs were significantly differentially expressed in naive and primed ESCs (**Supplementary Fig. 6**).

Next, we assessed the heterogeneity of miRNA expression in primed and naive hESCs (**Fig. 1i**), which revealed that the miR-375 and the cluster of miR-371-3 varied in expression across individual primed hESCs, but not naive ESCs, with high expression in a few primed ESCs (**Supplementary Fig. 7**). Variable expression of the miR-371-3 cluster has been noted in comparisons across human pluripotent stem cell lines[17] but not within a hESC population. Moreover, we examined cell-to-cell variability in the expression of sdRNAs and tsRNAs, and found that specific tsRNAs showed variable expression in naive and primed hESCs (**Supplementary Fig. 8**).

[1]Ludwig Institute for Cancer Research, Stockholm, Sweden. [2]Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, Sweden. [3]Respiratory Medicine Unit, Department of Medicine, Solna & Center for Molecular Medicine, Stockholm, Sweden. [4]Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden. [5]Division of Obstetrics and Gynecology, Karolinska University Hospital, Stockholm, Sweden. [6]These authors contributed equally to this work. Correspondence should be addressed to O.R.F. (omid.faridani@licr.ki.se) or R.S. (rickard.sandberg@ki.se).

Received 29 April; accepted 18 September; published online 31 October 2016; doi:10.1038/nbt.3701

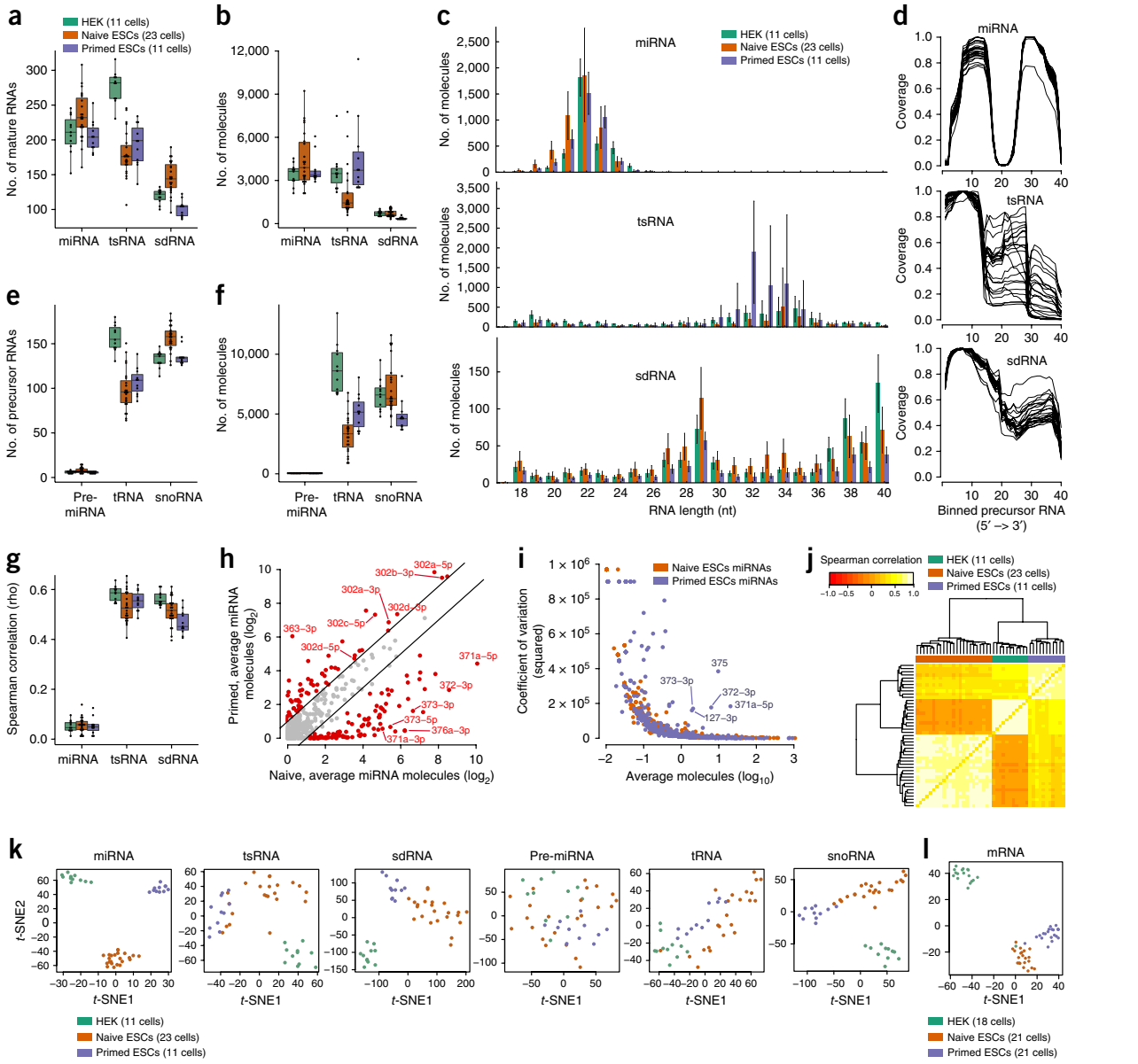

Figure 1 Characterization of small RNAs in HEK293FT, naive and primed hESCs. (**a**) Number of mature miRNAs, tsRNAs and sdRNAs detected per individual cell and cell type. (**b**) Absolute miRNA, tsRNA and sdRNA molecules detected per individual cell and cell type. (**c**) Size distributions of small RNAs profiled in single cells and per cell type. Error bars denote s.d. (**d**) Coverage of small RNAs (detected in naive hESCs) across the length of the precursor RNAs (divided into 40 bins). (**e**) Numbers of pre-miRNAs, tRNAs and snoRNAs detected per individual cell and cell type. (**f**) Absolute numbers of pre-miRNAs, tRNAs and snoRNAs molecules detected per individual cell and cell type. (**g**) Spearman correlations between the expression of small RNAs and precursors within individual cells, per cell type. (**h**) Scatter plot of miRNA molecules detected in naive and primed hESCs (97 and 90 cells, respectively). Genes with statistically significant (adjusted *P*-value < 0.05) and >2-fold differences in molecule counts between cell types were colored red. (**i**) miRNA expression variation within primed and naive hESCs (97 and 90 cells, respectively). (**j**) Hierarchical clustering of miRNA abundance profiles per cell, using pairwise Spearman correlations. (**k**) *t*-SNE maps of cells calculated from abundance profiles of small RNAs and precursor RNAs, including genes expressed (≥ 1 molecule) in at least two cells. (**l**) *t*-SNE map of cells calculated from single-cell mRNA sequencing on the same cell types (including genes expressed at ≥ 1 read per kilobase gene model and million mapped reads (RPKM) in at least two cells).

To determine whether single-cell small-RNA expression profiles could distinguish cell types, we first performed hierarchical clustering of the single-cell miRNA profiles. This revealed high concordance within cells of the same type and robust separation of cell types (**Fig. 1j**). Notably, we observed that primed hESCs clustered closer to HEK293FT than to naive hESCs, mainly owing to a set of miRNAs that were exclusively expressed in naive hESCs (**Supplementary Fig. 9**).

Using *t*-distributed stochastic neighbor embedding (*t*-SNE), the single-cell miRNA profiles robustly separated HEK293FT cells, naive hESCs and primed hESCs (**Fig. 1k**), whereas the sparser profiles of pre-miRNAs did not. These separations were not driven by sequencing depth or the number of miRNAs expressed (data not shown). The abundance profiles of sdRNAs and snoRNAs also indicated cell-type specificity, in line with the variable expression of snoRNAs across cell types[18], however, the cell-type specificity was not as pronounced as that of miRNAs. Profiles of tsRNAs and tRNAs on the other hand could not robustly separate cells of the various types. As a control, mRNA profiles generated using Smart-seq2 (ref. 19) could also separate cell types (**Fig. 1l**). To further examine the clustering capability of single-cell miRNA profiling, we generated

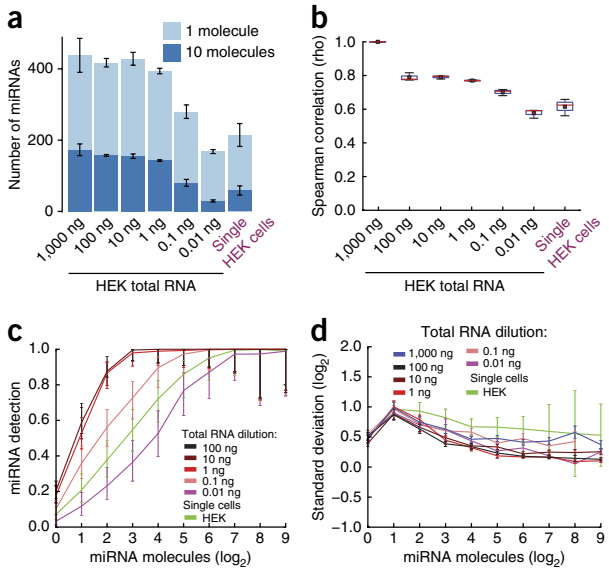

Figure 2 Comparison of miRNA profiling in diluted RNA and single cells. (**a**) Mean number of mature miRNAs detected in serial dilutions of HEK293FT RNA and in single HEK293FT cells. Mature miRNAs detected in more than 10 molecules or 1 molecule were colored dark blue and light blue, respectively. Mean and s.d. computed from three replicates (per RNA dilution) or 20 individual HEK293FT cells. (**b**) Spearman correlations of the miRNA abundance profiles, comparing profiles on 1,000 ng with that of dilutions and single cells. Median and mean denoted by red line and square, respectively. (**c**) Fraction of mature miRNAs reproducibly detected in replicate dilutions or single cells. Binned according to the number of molecules detected in 1,000 ng HEK293FT RNA. Mean and 90% confidence intervals were reported. (**d**) s.d. in miRNA expression estimated within replicates per RNA dilutions and in single cells, binned according to the molecule counts.

and sequenced small-RNA libraries from single glioblastoma cells as well as more naive and primed hESCs and HEK293FT cells. Again, all could be robustly separated (**Supplementary Fig. 10**). We did not detect apparent differences between naive and primed hESCs profiled in different batches (**Supplementary Fig. 10**). Together, miRNA abundance profiles possess considerable potential to assign cell types, and single-cell miRNA profiling may therefore have unrecognized potential to decode cellular heterogeneity within complex tissues.

To determine the sensitivity and quantitative accuracy of miRNA sequencing from single cells, we diluted microgram amounts of HEK293FT total RNA down to nanogram (ng) and picogram (pg) levels and performed small-RNA sequencing. The number of detected mature miRNAs was fairly constant (450 miRNAs expressed ≥1 molecule and 170 miRNAs expressed ≥10 molecules) until 1 ng total RNA; below that technical losses were observed. At dilutions to 10 pg, close to single-cell levels, we detected 40% of the mature miRNAs, which is comparable to the detection of miRNAs in individual HEK293FT cells (**Fig. 2a**). As a control, we compared miRNA profiles obtained from 1,000 ng HEK293FT RNA with those of several dilutions and with single cells. The Spearman correlations stayed relatively high down to 1 ng RNA and declined moderately with lower dilutions (**Fig. 2b** and **Supplementary Fig. 11**). This confirmed the robustness in our single-cell method. We also compared our data with results obtained using a commercial kit on 1 µg HEK293FT total RNA, which showed that the methods performed comparably (**Supplementary Fig. 12**). Finally, we evaluated the reproducibility of the method for miRNA detection in dilutions of HEK293FT RNA and in single HEK293FT cells, by calculating the percentage of miRNAs detected in replicate dilutions or cells (**Fig. 2c**). The reproducibility in gene detection started to decrease at ≤ 0.1 ng HEK293FT RNA. Variation in miRNA levels increased for genes with low expression (**Fig. 2d**). Importantly, the biological variation of miRNA expressions in HEK293FT cells was well above technical variation, even for genes expressed at low levels.

In summary, we have developed a single-cell protocol for counting small-RNA molecules. Notably, the protocol does not include RNA size selection steps, which makes the method suitable for automation. Exploring the small RNAs from various cell types revealed that tRNA halves were abundant in hESCs and HEK293FT cells. Our data on miRNA expression demonstrated the ability of miRNAs to robustly separate cell types even at the single-cell level, which paves the way for using small-RNA sequencing to decode cell heterogeneity in complex samples. A previous microarray-based study found that global miRNA profiling performed more effectively in the classification of poorly differentiated tumors than did mRNA profiling[20]; however, the experiment was performed on bulk RNA extracted from whole tumors. We therefore foresee application of the method in analyzing the heterogeneity of tumor tissues.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. SRA: SRP074776; GEO: GSE81287.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank J. Muhr and D. Topcic (Ludwig Institute for Cancer Research, Stockholm) for providing glioblastoma cells, and W. Kang for constructive comments on the computational pipeline. This work was supported by Knut and Alice Wallenberg Foundation (2015.0096) (F.L.), the Swedish Research Council (F.L. and R.S.), Ragnar Söderberg Foundation (F.L.), the Swedish Foundation for Strategic Research (F.L. and R.S.), the Bert L. and N. Kuggie Vallee Foundation (R.S.) and the European Research Council (648842) (R.S.).

AUTHOR CONTRIBUTIONS

O.R.F. conceived the study, developed the method, interpreted the results and prepared the manuscript. I.A. constructed the bioinformatics pipeline, analyzed the data, prepared the figures and contributed to manuscript text. M.H.-J. helped optimize the protocol and constructed libraries, J.P.S. and F.L. contributed naive and primed embryonic stem cells. R.S. supervised the development of the method and the computational analyses, and prepared the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Sandberg, R. *Nat. Methods* **11**, 22–24 (2014).
2. Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C. & Teichmann, S.A. *Mol. Cell* **58**, 610–620 (2015).
3. Fan, X. *et al. Genome Biol.* **16**, 148 (2015).
4. Ha, M. & Kim, V.N. *Nat. Rev. Mol. Cell Biol.* **15**, 509–524 (2014).
5. Taft, R.J. *et al. RNA* **15**, 1233–1240 (2009).
6. Thompson, D.M., Lu, C., Green, P.J. & Parker, R. *RNA* **14**, 2095–2103 (2008).
7. Sharma, U. *et al. Science* **351**, 391–396 (2016).
8. Hayes, J., Peruzzi, P.P. & Lawler, S. *Trends Mol. Med.* **20**, 460–469 (2014).
9. Landgraf, P. *et al. Cell* **129**, 1401–1414 (2007).
10. Kivioja, T. *et al. Nat. Methods* **9**, 72–74 (2011).
11. Kebschull, J.M. & Zador, A.M. *Nucleic Acids Res.* **43**, 143–152 (2015).
12. Fu, H. *et al. FEBS Lett.* **583**, 437–442 (2009).
13. Lee, Y.S., Shibata, Y., Malhotra, A. & Dutta, A. *Genes Dev.* **23**, 2639–2649 (2009).
14. Sperber, H. *et al. Nat. Cell Biol.* **17**, 1523–1535 (2015).
15. Zhang, Z. *et al. Stem Cell Reports* **4**, 645–657 (2015).
16. Rosa, A., Papaioannou, M.D., Krzyspiak, J.E. & Brivanlou, A.H. *Dev. Biol.* **391**, 81–88 (2014).
17. Kim, H. *et al. Cell Stem Cell* **8**, 695–706 (2011).
18. Cavaillé, J. *et al. Proc. Natl. Acad. Sci. USA* **97**, 14311–14316 (2000).
19. Picelli, S. *et al. Nat. Methods* **10**, 1096–1098 (2013).
20. Lu, J. *et al. Nature* **435**, 834–838 (2005).

ONLINE METHODS

Cell cultures. HEK293FT cells (Invitrogen R700-07) were cultured in DMEM containing glucose and glutamine (Gibco), supplemented with 10% FBS (Stem Cell Technology) and 100 µg/mL PenStrep (Gibco). Glioblastoma cells (KS4, JM3, JM4) were cultured on poly-L-ornithine- and laminin (Sigma-Aldrich)-coated plates in Human NeuroCult NS-A Proliferation Media, 2 µg/mL heparin, 10 ng/mL FGF, 20 ng/mL EGF (all from Stem Cell Technology) and 100 µg /mL PenStrep (Gibco). Cells were passaged using TrypLE (Gibco) and DTI (Gibco) every 3–10 d. Medium was replaced every 3–4 d. U87 cells were cultured in DMEM containing glucose and glutamine (Gibco), supplemented with 10% FBS (Stem Cell Technology) and 100 µg /mL PenStrep (Gibco). Cells were passaged as described above. All cells were mycoplasma tested.

hESC cultures. *Naive hESCs.* hESC line WA09 (female) (WiCell, Madison, WI) was transitioned from conventional primed conditions to 5i/L/FA naive state by closely following the conversion protocol described in Theunissen *et al.*[21]. Naive hESCs were maintained and expanded in a medium containing 50:50 N2B27/neurobasal (GIBCO) with 1 mM glutamine (Invitrogen), 1% NEAA (Invitrogen), 0.1 mM β-mercaptoethanol (Sigma), and freshly supplemented with 10 µM Y-27632 ROCKi (Stemgent), 1 µM PD0325901 MEKi (Sigma), 1 µM IM-12 GSK3i (Sigma), SB590885 BRAFi (Sigma), 1 µM WH-4-023 SRCi (Sigma), 20 ng/ml Activin A (R&D), 0.5% KOSR (Invitrogen), 8 ng/ml bFGF (R&D), 50 µg/ml BSA (Sigma) and 20 ng/ml hLIF (Millipore)[21]. Cells were propagated on irradiated E12.5 MEFs (GIBCO) seeded at 250k cells/well of a 6-well plate (9.5 cm^2/well), and once colonies reached confluence, they were dissociated and expanded by single cell using Accutase (GIBCO). WA09 (H9) cells were mycoplasma tested and validated by immunocytochemistry. Based on our Smart-seq2 data, our cells exhibit transcriptional profiles resembling previous work on these cells. This cell line was obtained from WiCell Research Institute, Inc., from which we hold a MTA (license). The conventional primed hESC line WA09 (WiCell, Madison, WI) was maintained on a layer of mouse embryonic fibroblasts (MEF), in media consisting of DMEM: F12 (Invitrogen), 20% KSR (Invitrogen), 4 ng/ml bFGF (R&D), 2 mM glutamine (GIBCO), 0.1 mM non-essential amino acids (Invitrogen), 0.1 mM β-Mercaptoethanol (Sigma). Once colonies reached confluence, they were mechanically dissociated by passage through a 26-gauge needle.

Cells sorting. Cell suspensions were stained with propidium iodide (PI) to mark dead cells. Single cells were sorted into 96-well plates using a BD FACSAria III instrument.

Oligonucleotides and primers. Sequences and modifications are available in **Supplementary Table 2**.

Small-RNA library preparation and sequencing. Prior to sorting, 3 µl of lysis buffer (0.13% Triton-X-100, 4 units (u) recombinant RNase Inhibitor, Takara) was deposited in each well of the 96-well collection plate. After single cell sorting, the plates were transferred to −80 °C for long storage. After thawing the plate, 1 µl of 5.8S rRNA masking oligo (5 pmol) was added and the whole plate was incubated at 72 °C for 20 min. Then, 2 µl of 3′ adapter ligation reaction was added (20 pmol 3′ adapter oligo, 8.33% PEG 8000, 50 u T4 RNA Ligase 2, truncated KQ, NEB, 0.83× T4 RNA ligase buffer, NEB, 4 u recombinant RNase Inhibitor, Takara) and the reaction was incubated at 30 °C for 6 h followed by 4 °C for 10 h. Next, 3 µl of RT primer and enzymes for removal of free adapters were added (200 pmol RT primer, 2.5 u Lambda exonuclease, NEB, 10 u 5′ deadenylase, NEB) and the reaction was incubated at 30 °C for 15 min followed by 37 °C for 15 min. Next, 2 µl of 5′ adapter ligation reaction was added (45 pmol 5′ adapter oligo, 0.68 mM Tris-buffered ATP, Thermo Fisher, 4 u T4 RNA ligase, Thermo Fisher, 0.23× T4 RNA ligase buffer, NEB) and the mix was incubated at 37 °C for 1 h. Reverse transcription reaction was performed by adding 7 µl of RT reaction (1.28× Taq DNA Polymerase PCR Buffer, Thermo Fisher, 8.33 mM DTT, 0.42 mM/each dNTP, 4 u recombinant RNase Inhibitor, 150 u Superscript II reverse transcriptase, Thermo Fisher). The PCR amplification was carried out by adding 35 µl of the reagents (0.94× Phusion HF buffer, 1 u Phusion Hot Start II DNA Polymerase, Thermo Fisher, 0.12 mM/each dNTP, 1.89 µM RP1 primer) and incubating at 98 °C for 30 s followed by 13 cycles of 98 °C for 10 s, 60 °C for 30 s and 72 °C for 30 s and a final incubation at 72 °C for 5 min. At last, 1 µl of the amplified product was transferred to a fresh tube and to a 25 µl of second PCR reaction consisting of 2 µM indexed primers, 0.2 µM RP1 primer, 1× Phusion HF buffer, 0.5 u Phusion Hot Start II DNA Polymerase, Thermo Fisher, 0.2 mM/each dNTP, followed by a 30 s incubation at 98 °C, 13 cycles of 98 °C for 10 s, 67 °C for 30 s and 72 °C for 30 s and a final incubation at 72 °C for 5 min. The amplified libraries of single cells were then pooled and purified with DNA Clean & Concentrator-5 (Zymo research). All samples were sequenced 1× 43 bp or 1× 51 bp using the Illumina HiSeq 2500 instrument. Library preparations were performed for serial dilutions of HEK RNAs following the above procedure. For NEBNext library preparation of HEK293FT RNAs, we followed the manufacturer's protocol, except the size selection step. Sequencing statistics for all libraries used in this study can be found in **Supplementary Table 3**.

Read alignments and gene-expression estimation. First, sequences in the FastQ file corresponding to UMIs were removed from sequence reads and appended to the read name (for later analyses). Adapter and polyA sequences were removed from reads using cutadapt v1.8.1, with the minimum overlap between adapter and the 3′ of the read set to 1 nt. Reads shorter than 18 nts after adapter trimming were discarded. Trimmed reads were aligned to the human genome (hg38) using STAR[22] v2.4.0 with the following parameters--outSAMstrandField intronMotif --outFilterMismatchNoverLmax 0.04 --outFilterMatchNmin 18 --outFilterScoreMinOverLread 0 --outFilterMatchNminOverLread 0 --alignIntronMax 1. We did not allow any mismatches within the first 25 nt and allowed only 1 mismatch in the remaining part of the sequence reads (25–41 nts). Spliced alignments were disabled and hard/soft-clipping was disabled for the 5′ of the read. Soft-clipping of up to 3 nts was allowed at the 3′ end of the read in order to account for RNA editing and 3′ tRNA modifications such as addition of CCA. Reads mapping with insertions or deletions were removed. We collapsed PCR amplicons and counted RNA molecules using the adjacency network approach (dedup_umi.py at https://github.com/CGATOxford/UMI-tools), where reads with UMIs having a single Hamming distance from another UMI were collapsed. We separated RNA molecule counts into two categories based on the extent of the alignment of our sequenced read with the human genome. Reads aligning over the full sequence (41 nts) were assigned as precursor molecules, whereas reads aligning with 40 or less nts were assigned as a potential small RNA. Potential small RNAs were further scrutinized to find instances where trimmed 3′ ends matched the genomic sequence after the alignment (cases where actual RNA sequence matched adapter sequence, false-positive adapter trimming). Consequently, these molecules were instead assigned to the precursor RNA.

Then we estimated expression for annotation transcripts in the following databases (Gencode V22, Mirbase V21 and GtRNAdb). During the quantification procedure, molecules were hierarchically assigned to annotated biotypes in the following order: Mirbase miRNAs, tRNAs from GtRNAdb, small RNA biotypes from Gencode (such as snoRNAs, snRNAs, rRNAs), and lastly the remaining Gencode transcripts such as protein coding and lincRNAs. We counted molecules from both intronic and exonic regions of the protein coding and lincRNAs in an attempt to capture small RNAs transcribed from these regions. We used a weighing approach to assign molecules of sequences mapping to multiple genomic locations, by dividing the number of molecules by the number of annotated locations. Finally, for miRNAs we collapsed molecules supporting the same miRNA that were expressed from different genomic loci (i.e., where the microRNAs gene was present in multiple genomic locations). Finally, we excluded libraries with fewer than one million reads (before alignment) and if the number of expressed transcripts was more than two s.d. away from the median number of expressed transcripts obtained within the cell population (to filter out potential doublets and low-quality libraries). All analyses in the manuscript were carried out on small-RNA molecule counts (i.e., amplicon corrected values), except where otherwise noted. The complete computational pipeline is included in **Supplementary Software** and has been archived on github: https://github.com/eyay/smallseq.

Dimensionality reduction analysis and sample distance correlation. Molecule count tables were separated according to RNA biotypes. We used transcripts detected in two or more cells and cells that had a Spearman correlation above 0.2 to at least one other cell. The *t*-distributed stochastic

neighborhood embedding (t-SNE)[23] was used to assess the separating ability of each small RNA class and was performed using the parameters perplexity = 50, epoch = 50, initial dimensions of 100 and maximum iteration of 1,000. Hierarchical clustering was performed using $\log_{10}$ transformed miRNA expression values and using complete linkage as distance measure between clusters. A pseudocount of 1 was added before $\log_{10}$ conversion. Spearman correlations were calculated pairwise using the miRNA $\log_{10}$ abundance profiles from the individual cells.

Differential expression analysis with SCDE and miRNA abundance scatter. Differential expression of miRNAs in naive (n = 97 cells) and primed hESCs (n = 90) was determined using SCDE (single cell differential expression analysis)[24] with default parameters except requiring a minimum of 100 genes to consider for the test (parameter min.size.entries = 100 to call scde.error.models function). miRNAs with no expression in naive or primed hESCs were excluded before analysis. The Z scores and corrected Z scores (cZ) to adjust for the multiple testing were converted into two-sided P-values using pnorm function in R. The average miRNA abundance in naive and primed hESCs was plotted in a scatter plot, where significantly differentially expressed miRNAs (SCDE adjusted P-value < 0.05 and fold change > 2) were colored red.

Analysis of sensitivity and technical variation between HEK293FT cells and total RNA serial dilutions. MicroRNA gene detection analysis (**Fig. 2c**) was performed over pairs of HEK293FT total RNA technical replicates or single cells. MicroRNA genes were binned by abundances detected in the 1,000 ng RNA sample. The mean across all possible pairs of samples within a dilution or single cells was reported together with the s.d. calculated using the adjusted Wald method. Variation analysis (**Fig. 2d**) was similarly calculated on pairs of samples, binning miRNAs by the mean of $\log_2$ molecule counts. In both figures, miRNAs were considered detected if they had a minimum molecular count of 0.2 in both samples.

Analysis of heterogeneity. Samples have been depth normalized so that the distribution of total number of molecules between naive and primed hESCs are similar. Squared coefficient of variation was computed as a function of average number of molecular counts in order to assess the heterogeneity.

Data analysis. Statistical analysis and data visualization were performed using in-house scripts written in Python and R programming languages.

21. Theunissen, T.W. *et al. Cell Stem Cell* **15**, 471–487 (2014).
22. Dobin, A. *et al. Bioinformatics* **29**, 15–21 (2013).
23. Van der Maaten, L. & Hinton, G. *Journal of Machine Learning.* **9**, 2579–2605 (2008).
24. Kharchenko, P.V., Silberstein, L. & Scadden, D.T. *Nat. Methods* **11**, 740–742 (2014).

doi:10.1038/nbt.3701