

## Single-cell microRNA/mRNA co-sequencing reveals non-genetic heterogeneity and novel regulatory mechanisms

**Nayi Wang<sup>1,2,3</sup>, Ji Zheng<sup>2,3,4</sup>, Zhuo Chen<sup>1,2,3</sup>, Yang Liu<sup>2,3</sup>, Burak Dura<sup>1</sup>, Minsuk Kwak<sup>1</sup>, Juliana Xavier-Ferrucio<sup>3,5</sup>, Yi-Chien Lu<sup>3,5</sup>, Miaomiao Zhang<sup>2,4,6</sup>, Christine Roden<sup>2,3</sup>, Jijun Cheng<sup>2,3</sup>, Diane Krause<sup>3,5</sup>, Ye Ding<sup>7</sup>, Rong Fan<sup>1,3\*</sup> & Jun Lu<sup>2,3,8,9\*</sup>**

<sup>1</sup>Department of Biomedical Engineering, Yale University, New Haven, CT 06520, USA

<sup>2</sup>Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06510, USA

<sup>3</sup>Yale Stem Cell Center, Yale Cancer Center, New Haven, CT, 06520, USA

<sup>4</sup>Southwest Hospital, Third Military Medical University, Chongqing, 400038, China

<sup>5</sup>Department of Laboratory Medicine, Yale University School of Medicine, New Haven, Connecticut 06510, USA

<sup>6</sup>School of Medicine, Jiangsu University, Zhenjiang, Jiangsu, 212013, China

<sup>7</sup>Wadsworth Center, New York State Department of Health, Albany, New York 12208, USA

<sup>8</sup>Yale Center for RNA Science and Medicine, New Haven, CT, 06520, USA

<sup>9</sup>Yale Cooperative Center of Excellence in Hematology, New Haven, CT, 06520, USA

\* Correspondence should be addressed to:

Rong Fan, 55 Prospect Street, Room 213, New Haven, CT 06520.

Telephone: 203-432-9905; Fax: 203-432-6441;

Email: [rong.fan@yale.edu](mailto:rong.fan@yale.edu)

or

Jun Lu, 10 Amistad Street, Room 237C, New Haven, CT 06520.

Telephone: 203-737-3426; Fax: 203-785-4305;

Email: [jun.lu@yale.edu](mailto:jun.lu@yale.edu)

## ABSTRACT

Co-measurement of multiple omic profiles from the same single cells opens up the opportunity to decode molecular regulation that underlie intercellular heterogeneity in development and disease. Here, we present co-sequencing of microRNAs and mRNAs in the same single cells using a half-cell genomics approach. This method demonstrates good robustness (~95% success rate) and reproducibility ( $R^2=0.93$  for both miRNAs and mRNAs), and yields paired half-cell miRNA and mRNA profiles that could be independently validated. Linking the level of miRNAs to the expression of predicted target mRNAs across 19 single cells that are phenotypically identical, we observe that the predicted targets are significantly anti-correlated with the variation of abundantly expressed miRNAs, suggesting that miRNA expression variability alone may lead to non-genetic cell-to-cell heterogeneity. Genome-scale analysis of paired miRNA-mRNA co-profiles further allows us to derive and validate new regulatory relationships of cellular pathways controlling miRNA expression and variability.

## INTRODUCTION

Small RNAs have emerged as important non-coding regulators in diverse biological settings. The most widely studied small RNAs are miRNA that regulate protein-coding expression post-transcriptionally<sup>1-3</sup>. MiRNA expression profiles in tissues or cell populations are highly informative to reveal cellular states, such as in human cancers<sup>2, 4, 5</sup> and identify cellular mechanisms<sup>6, 7</sup>. The advances of single cell genomics have raised the prospect that single-cell miRNA profiles could add a new dimension to small RNA research by facilitating the molecular understanding of intercellular heterogeneity. Unlike single-cell RNAseq techniques, which have matured considerably in the past years<sup>8-17</sup>, single-cell small RNA sequencing has been demonstrated very recently and applied to differentiating naive versus primed human embryonic stem cells and examining intercellular heterogeneity of miRNA expression<sup>18</sup>, with the latter confirming previous findings of single-cell miRNA measurements by lower throughput methods such as quantitative RT-PCR or fluorescence *in situ* hybridization(FISH)<sup>19-21</sup>. However, given the roles of small RNAs to modulate gene expression post-transcriptionally, for example, via miRNA-mediated degradation of target transcripts, it is important to measure both small RNAs and mRNAs at the genome-scale in order to decipher the mechanism underlying observed intercellular miRNA and mRNA heterogeneity. We thus envisioned that obtaining miRNA and mRNA profiles from the same single cells could empower the study on how miRNAs modulate non-genetic cell-to-cell variability post-transcriptionally and how miRNA variability could be modulated via protein-coding genes.

Over the past years, technology breakthroughs in single-cell omics have enabled genome-wide profiling of genetic mutations<sup>22</sup>, copy number variation<sup>23, 24</sup>, DNA methylation<sup>24, 25</sup>, chromatin-accessibility<sup>26</sup>, and gene expression at the transcriptional level<sup>27</sup>. Combining two or three of these measurements on the same single cells has been demonstrated<sup>24, 28-34</sup>, but almost exclusively on co-analysis of genomic DNA and mRNAs to correlate pre-transcriptional alterations to mRNA gene expression. Genome-wide profiling of both mRNAs and post-transcriptional regulators such as miRNAs remains challenging, due in part to the incompatibility of current mRNA-seq and small RNA sequencing protocols<sup>35</sup>. Strategies such as

separation of miRNAs from mRNAs are not yet reliable for single-cell co-measurement of both types of RNAs at the genome-scale.

We report here the co-sequencing of single-cell small RNA and mRNA transcriptomes using a half-cell genomics approach, which has been built upon our prior success in direct capture and profiling of miRNAs in low-quantity whole cell lysate<sup>2</sup> and the rigorous validation of half-cell sampling to reliably represent the whole cell RNA expression levels. While previous efforts have examined the relationships between miRNA and mRNA expression profiles in bulk samples<sup>36-40</sup>, we demonstrate that decoding single-cell-level small RNA and mRNA profiles is feasible to reveal regulatory mechanisms both upstream and downstream of intercellular miRNA heterogeneity. This work demonstrates the first study of single-cell miRNA and mRNA co-sequencing at the genome-scale and opens new opportunities to investigate how the changes in small RNA expression in single cells contribute to non-genetic cell-to-cell variability and how to perturb such relationships for controlling cellular function in a heterogeneous population.

## RESULTS

### Strategy to co-profiling of single-cell miRNAs and mRNAs

To achieve the goal of profiling small RNAs and mRNAs from the same single cells, we utilized a half-cell genomics approach in which a single cell was lysed and the lysate was split evenly into two half-cell fractions, with each fraction subjected to either miRNA or mRNA transcriptome sequencing (**Figure 1a, b & c**). For profiling miRNA expression in half cells, we modified our previous method<sup>2</sup> to develop a protocol for generating small RNA sequencing library from single cells or half-cell-level materials. In this protocol, single cell lysate was heated to release miRNAs for sequential ligation reactions with 3'- and 5'- molecular adaptors and then the ligated products were reverse transcribed, and amplified with polymerase chain reaction (PCR) for library preparation and next-generation sequencing (**Figure 1b**). For profiling mRNAs, we used SMART-Seq (v.4), a commercialized kit for manual processing of low-input or single-cell polyadenylated mRNAs to generate full-length libraries for RNA sequencing (**Figure 1c**). As detailed below, we successfully developed, combined, and validated these methods

to generate sequencing libraries from half-cell materials and the half-cell miRNA and mRNA profiles faithfully capture single-cell transcriptomes.

### **Evaluating the profiling of half-cell miRNA expression**

Although single-cell omics technologies have been developed to quantify intercellular heterogeneity via comparing measured molecular profiles between individual cells and the bulk data, it remains challenging to determine how much variability is true biological heterogeneity and how much is attributable to technical noise associated with the processing of single cells. The best experiment to address this question is to measure variability between two half-cell materials split evenly from the same single cell. We first evaluated this approach to profiling miRNAs. A single K562 human acute myeloid leukemia cell was transferred to and lysed in a PCR tube. The lysate (total volume ~ 10 $\mu$ L) was eventually split into two halves into two PCR tubes to independently perform small RNA capture, amplification, and preparation of sequencing libraries as described in **Figure 1b**. The sequencing data from each of these two halves were compared against each other for miRNA expression, which gave rise to an  $R^2$  value of 0.930 (**Figure 2a**), indicating low technical noise and high consistency.

Of note, initial tests with even splitting (1:1 in vol) of single-cell lysate prepared with standard cell lysis protocol did not yield even splitting of miRNAs in two half-cell samples. Interestingly, we observed selective enrichment or depletion of a group of miRNAs in one half-cell sample presumably due to the binding to intracellular proteins. This remains to be further investigated. In this work, we modified the lysis protocol in several steps (see **Methods**) including introducing freeze/thaw and heat treatment, which eventually led to a robust procedure to evenly split miRNAs in two half-cell samples (**Supplementary Figure 1**). We further examined the validity of this half-cell approach using different cell types and another measurement technique. Small RNA libraries were generated from two halves of a single murine hematopoietic progenitor BaF3 cell, or single K562 and 293T cells, and measured with a Luminex bead array assay. The results again exhibited a high degree of concordance between these two half cells (**Supplementary Figure 2a**). While two half-cell samples gave rise to high concordance in miRNA profile, single cells obtained from the same cell population exhibited a higher degree of variability by both qRT-PCR analysis and expression

profiling (**Supplementary Figures 3, 4**),

Upon validating the consistency of the half-cell approach to capture single-cell miRNA profiles, we turned to measure intercellular variability by sequencing 20 half-cell K562 miRNA samples. The other halves of the same cells were processed for RNAseq (to be described in next section). The sequencing data from 19 single cells passed the quality check for both miRNAs and mRNAs. The one that failed the quality assessment had very low levels of mappable reads in both miRNA and mRNA data presumably due to the poor quality of this cell itself (*e.g.*, cell damaged in the capture and transfer process). Principal component analysis (PCA) of these 19 samples and 2 half-cell samples derived from the same cell further confirmed high concordance in two half-cell splits and the variability among single cells is higher, suggesting the existence of intercellular heterogeneity of miRNA expression (**Supplementary Figure 5**). Overall, the success rate (95%) to obtain quality sequencing libraries for both miRNA and mRNA using the half-cell genomics approach was excellent.

The intercellular variability of miRNA expression could be directly observed in our data either by a correlation matrix (**Supplementary Figure 6**) or by quantifying the standard deviation of log<sub>2</sub>-transformed miRNA expression versus the mean expression across 19 half cells (**Figure 2b**). As expected, low-expression-level miRNAs showed inherently large standard deviations in log<sub>2</sub> data. The variation of high abundance miRNAs gradually decreased as the expression level increased. MiR-92a-3p was detected as the most abundant miRNA in our K562 cell sample, in agreement with previous literature<sup>40</sup>. Comparing intercellular variability across miRNAs, both miR-146b-5p and let-7i-5p showed higher variability than miRNAs with comparable mean expression levels (for miR-146b-5p, std=0.377; for let-7i-5p, std=0.939). For example, let-7a-5p, a miRNA in the same family as let-7i-5p, had comparable mean expression to let-7i-5p, but with much lower variability (for let-7a-5p, std=0.351). These results showed the direct evidence of intercellular miRNA expression heterogeneity, which differed substantially for individual miRNAs.

To validate the observed intercellular heterogeneity, we performed single-cell qRT-PCR to measure four selected miRNAs, let-7a-5p, let-7i-5p, miR-146b-5p, and miR-92a-3p, in a separate batch of cells (**Figure 2c**). Indeed, the variations from single-cell qRT-PCR paralleled those observed in half-cell small RNA sequencing data: miR-92a-3p and let-7a-5p show relatively lower levels of variation (standard deviation std=0.83 and std=1.51 respectively), miR-146b-5p an intermediate level of variation

(std =2.57), and let-7i-5p a high level of variation (std =5.64). (**Figure 2c**). We further determined that these variations could not be fully attributed to genetic differences but instead associated with non-genetic cellular heterogeneity. We generated 17 K562 single-cell clones and measured the same miRNAs using qRT-PCR for each clone (~40,000 cells per clone). One would expect that if the miRNA expression heterogeneity in single cells was driven by genetic differences between cells, large variation would be manifested similarly between single-cell clonal populations because each clone was derived from a single cell and cells within a clone share the genetic variations present in the initial single cell. In contrast, we observed low levels of variability cross these single-cell clones (for miR-92a-3p, std=0.48; for let-7i-5p, std=0.68; for miR-146b-5p, std=0.60) (**Figure 2d**). Collectively, these data support that the half-cell small RNA profiles could faithfully capture single-cell miRNA variability, which allowed us to use half-cell small RNA sequencing to interrogate non-genetic intercellular miRNA heterogeneity.

### Evaluating the profiling of half-cell mRNA expression

As described above, the 20 single K562 cells processed for half-cell small RNA sequencing were also analysed for mRNA transcriptomes using the remaining half-cell lysates. SMART-seq was reported to be of high coverage, low bias, and good reproducibility compared to other methods when single-cell data were compared to population, spike-in controls or theoretical predictions<sup>41</sup>. However, the true technical variability cannot be accessed unless the same cell can be analysed multiple times for RNAseq, which has not been practical. Herein, we conducted SMART-seq with two halves of a single K562 cell and the gene expression profiles were compared (**Figure 3a**), which yielded a Pearson correlation  $R^2$  value of 0.930, which provided a direct evidence about the level of technical noise. Similar results were observed with human 293T and murine NIH3T3 cells (**Supplementary Figure 2b**). These data supported the validity to use half-cell sequencing data to represent single-cell mRNA transcriptome.

We next examined the variation of mRNA expression across 19 half K562 cells (**Supplementary Figure 7**). The plot (**Figure 3b**) showing mRNA variation across 19 half cells as a function of the mean gene expression level followed the anticipated trend that the log<sub>2</sub> variation became lower as the levels of gene expression increased. Unsupervised consensus clustering was conducted to detect transcriptional states with similar profiles and resulted in three major clusters (**Figure 3c**). The first cluster of cells



was enriched for genes associated with translational processes but depleted for genes associated with processes such as acetylation. The second cluster of cells were opposite, depleted for translation-related genes but enriched for the acetylation process (**Supplementary Table 1**). The third cluster of cells contain both signatures of the first and second clusters and are likely more differentiated cells with the increased expression of erythroid lineage genes.

In order to further validate the observed clusters, we used high-throughput single-cell 3'-end mRNA-seq technology<sup>42-44</sup> to generate the mRNA transcriptome data from hundreds of single K562 cells. Quality check was performed and compared against published single-cell RNA-seq data for the same cell line (**Supplementary Figures 8 and 9**). Unsupervised consensus clustering analysis again yielded three major clusters with the proportions (40.55% : 42.83% : 16.62%) similar as the 19 single-cell data (36.84% : 47.37% : 15.79%) (**Figures 3c and 3d**). The analysis using t-distributed stochastic neighbour embedding (**t-SNE**) was performed to visualize the major clusters and the gene expression distribution (**Supplementary Figure 10**). Pathway analysis further confirmed that these clusters were enriched for the same biological processes, notably the translational activities and the acetylation pathway (**Figures 3c, 3d and Supplementary Tables 1 & 2**). Therefore, the mRNA transcriptional phenotypes detected by the half-cell genomics approach and single-cell 3' mRNA-seq were in good agreement, justifying the validity to use half-cell mRNA profiles to capture single-cell transcriptome and to examine single-cell gene expression heterogeneity.

### **miRNA and miRNA target relationship on the single-cell level**

Having both miRNA and mRNA profiles obtained from the same single cells, we reasoned that if miRNA expression heterogeneity contributed to the control of single-cell level mRNA expression heterogeneity, one would expect that the mRNA targets to be anti-correlated with miRNA expression levels. It has been known that the level of global miRNA expression is low in cancer cells<sup>4</sup>, particularly cancer cell lines such as K562, we focussed first on the most abundantly detected miRNA miR-92a-3p. We ranked all mRNAs based on their Pearson correlation to the miRNA, and asked whether predicted targets of miR-92a-3p were more anti-correlated to the miRNA than non-targets. Using two independently defined miRNA target predictions (TargetScan and MSigDB), we observed that both target sets were significantly enriched toward the



negative correlation scores (**Figure 4a**). In contrast, predicted targets of another miRNA, miR-125a-5p, were not enriched for correlation with miR-92a-3p (**Figure 4a**). These data support that the single-cell-level expression of miRNA targets was overall more negatively correlated to miRNA expression, and suggest that miRNA expression variation contributes to the control of mRNA variability.

While examining the predicted targets of miR-92a-3p, we noticed an interesting relationship between the level of target expression and the likelihood of the target being negatively correlated with the miRNA. Specifically, for targets of higher abundance (mean  $\log_2(\text{RPKM}) > 4$ ), we observed a much stronger and significant enrichment toward negative correlation than targets with lower expression (**Figure 4b**). This relationship could also be observed when analysing correlation between miR-125a-5p and its predicted targets, as well as between miR-26a-5p and its targets (**Figure 4c, 4d**). In both cases, only targets with mean  $\log_2(\text{RPKM})$  over 4 showed significant enrichment toward negative correlation with the corresponding miRNA, whereas targets with lower expression were not. Indeed, if all the targets of miR-125a-5p and miR-26a-5p were analysed without categorization based on their expression levels, we could not detect any significant enrichment. This was obviously due to predicted targets of lower expression diluting out the signal for abundantly expressed targets, the latter of which were of a small fraction among all predicted targets. See Discussion for the possible explanations of this observed correlation. Taken together, the data above support that miRNA heterogeneity may contribute to shaping the mRNA variability in single cells, particularly for abundantly expressed targets.

The three miRNAs mentioned above, including miR-92a-3p, miR-125a-5p and miR-26a-5p, shared the common properties that they were of relatively high abundance in K562 cells and they had a reasonably long list of predicted targets (>500) to allow effective stratification by expression levels. We thus examined miRNAs with lower abundance to determine if targeting relationships could also be observed in paired half-cell miRNA and mRNA profiles. For let-7a-5p, which was weakly expressed in K562 cells (**Figure 2b, 2c**), we did not observe significant enrichment of targets toward negative correlation with the miRNA, even though a trend could be seen for predicted targets with abundant expression (**Figure 4e**). This was not surprising, as it has been reported that functional targeting requires at least several hundred miRNA molecules per cell<sup>45</sup>. Likewise, we did not observe significant targeting relationships for let-7i-5p

and miR-146b-5p, presumably due to low abundance (for let-7i-5p) or a short list of predicted targets (for miR-146b-5p).

Taken together, the data above demonstrate the feasibility of examining functional miRNA targeting directly in paired half-cell miRNA and mRNA profiles, with the targeting relationship on single-cell level being dependent on both the level of miRNA expression and the level of its targets. The data suggest the possibility for abundantly expressed miRNAs to influence global transcriptomic phenotypes in single cells.

### **Single-cell miRNA and mRNA profiles reveal regulatory relationships controlling miRNA expression**

It is conceivable that heterogeneity in cells' molecular states may contribute to miRNA expression heterogeneity. We thus attempted the searching for new regulatory relationships that control miRNA gene expression utilizing the cell-to-cell variability of miRNA and mRNA in single cells. We specifically focused on the two highly variable miRNAs, miR-146b-5p and let-7i-5p, and queried the mRNA genes (e.g., using gene set enrichment analysis (GSEA) based upon the MSigDB database<sup>46</sup>) or molecular connectivity (e.g, using the Connectivity Map (CMap)<sup>47</sup>) in order to reveal cellular pathways and chemical perturbations that are capable of modulating the expression of the corresponding miRNA. (**Figure 5a**).

For miR-146b-5p, we ranked mRNA genes according to their correlation with the miRNA, and performed GSEA querying with MSigDB<sup>46, 48</sup>. The most noticeable association was a negative relationship between multiple gene sets for protein translation and miR-146b-5p (**Figure 5b and Supplementary Table 3**). Interestingly, protein translation was also identified as a molecular process differentially expressed between the major mRNA transcriptomic subtypes (**Figure 3c**). We hypothesized that a common upstream regulator controls cellular translation and miR-146b-5p expression in opposite ways. Given that the AKT pathway positively regulates cellular translation, we thus tested the hypothesis that inhibition of AKT could elevate miR-146b-5p expression. We treated K562 cells with an allosteric AKT inhibitor MK-2206, which significantly increased miR-146b-5p expression (**Figure 5d**). A similar increase was observed in MCF-7 cells (**Supplementary Figure 11a**). In contrast, MK-2206 did not affect miR-92a-3p or let-7i-5p (**Figure 5d, 5h**), and the inhibition of the hedgehog pathway by cyclopamine or the inhibition of HDAC by Trichostatin A did not

significantly affect miR-146b-5p expression in K562 cells (**Figure 5d, 5g**), indicating the specificity at the levels of both miRNA and cellular pathway. Upon AKT inhibition of K562 cells, we noticed that the intercellular variability of miR-146b-5p expression was reduced (**Supplementary Figure 11b**, for control  $\text{std}=2.41$ , for AKT inhibition  $\text{std}=1.50$ ;  $p<0.006$  see Methods). We further noticed that the AKT inhibitor elevated the expression of mature miR-146b-5p, but decreased the expression of the primary transcript of miR-146b (**Figure 5i**), whereas no change was observed for the pri-let7i or mature let-7i-5p (**Figure 5h**), suggesting that the AKT pathway specifically inhibits the post-transcriptional maturation of miR-146b. Similar results were observed when K562 cells were treated with two additional AKT inhibitors (**Supplementary Figure 11c, 11d**). These data indicate that a novel regulatory relationship for miR-146b-5p expression could be uncovered by analysing the correlation between miRNA and mRNA profiles in single cells, and support a possible role for AKT to control intercellular variability of miRNA expression.

For mRNAs correlated with let-7i-5p, we did not find statistically significant positive or negative enrichment for any gene set in MSigDB after multiple-hypotheses correction (**Supplementary Table 3**). Instead, we defined the positive and negative transcriptomic signatures of let-7i-5p by using mRNAs correlated and anti-correlated with this miRNA respectively, and used these signatures to query the CMap, a database of gene expression responses in the presence of chemical perturbation<sup>47</sup>. Two histone-deacetylase (HDAC) inhibitors, including Trichostatin A, were scored as top hits (**Figure 5c and Supplementary Table 4**), whose gene expression responses in CMap database were negatively enriched for the let-7i-5p signature, leading to the hypothesis that inhibiting HDAC could decrease let-7i-5p expression. This was again experimentally verified by treating K562 cells with Trichostatin A, which led to a significant downregulation of let-7i, with no effect on miR-146b-5p (**Figure 5e, 5g**). MK-2206 and cyclopamine that target other pathways had no effect on the expression of let-7i-5p (**Figure 5e, 5h**). Trichostatin A did not alter the level of primary let-7i transcript, suggesting post-transcriptional control on let-7i-5p expression (**Figure 5f**). Taken together, the above results support that informative regulatory relationships for miRNA expression could be revealed through single-cell miRNA and mRNA co-profiling.

## DISCUSSION

Despite the recent advances in single-cell functional genomics and the studies reported on separating DNA and RNA for co-measurement of pre-transcriptional alterations (e.g., DNA methylome) and mRNA gene expression, to our best knowledge, this is the first report on co-analysis of post-transcriptional regulators (miRNAs) and the mRNA transcriptome in single cells. This was realized with a half-cell genomics approach, which overcomes the challenge to separate and capture small RNAs from mRNA without introducing material loss and technical variation<sup>35</sup>. We evaluated the validity of sequencing half-cell RNA materials to measure single-cell miRNA or mRNA expression profiles, and further combined half-cell miRNA and mRNA profiles from same single cells to perform multi-omic analysis. Having the genome-wide miRNA and mRNA profiles from the same single cells permits direct assessment of whether intercellular mRNA transcriptomic heterogeneity is in part shaped by the intercellular variability of the corresponding miRNA regulators. We observed that several abundantly expressed miRNAs, including miR-92a-3p, miR-125a-5p and miR-26a-5p, were more inversely correlated with its predicted targets than non-targets, supporting the above notion. Predicted targets with abundant expression levels were more strongly anti-correlated with the miRNA than predicted targets with lower expression. Why was a stronger anti-correlation observed only for predicted targets of high abundance? One possibility is that the higher technical noise (on log<sub>2</sub> transformed data) for genes with low expression levels resulted in the lower degree of statistical significance even if there were correlation between miRNA and targets. Another possibility is that protein-coding genes with low expression levels are more prone to regulation by miRNAs of both high and low abundance, diluting the correlation with any given single miRNA. In contrast, protein-coding genes with high expression levels are susceptible to regulation only by abundantly expressed miRNAs, leading to the preservation of correlation with the corresponding miRNA. This second possibility is consistent with recent findings that support the efficacy of miRNA-mediated regulation being a function of the ratio between miRNAs and targets<sup>49-52</sup>. Overall, our experiments opened up the opportunity to interrogate single cells for post-transcriptional regulation of protein-coding genes to modulate mRNA cell-to-cell variability.

In addition to assessing miRNA-based regulation of mRNA heterogeneity, the availability of paired half-cell mRNA and miRNA profiles allowed us to derive informative and verifiable hypotheses on regulatory relationships between cellular

pathways and miRNA expression that may contribute to intercellular heterogeneity of miRNA expression. Notably, the expression of miR-146b-5p could be regulated by an AKT inhibitor, which is well recognized to control protein translation. This regulation seems to occur at the level of miRNA maturation, because the primary transcript was reduced yet the mature miRNA was increased. This effect cannot be explained by previous findings of AKT controlling AGO2<sup>53</sup>, which would predict similar responses to AKT inhibition by all miRNAs. The lack of effect of the AKT inhibitor on other miRNAs (miR-92a-3p and let-7i-5p) and the decrease of primary miR-146b transcript by AKT inhibition indicate that this modulation occurs at a separate step from AGO2 regulation. The exact mechanism requires further studies to elucidate. Similarly, we demonstrated that the let-7 family member let-7i-5p could be downregulated by HDAC inhibition. HDAC inhibition is often associated with activation of gene transcription, which did not seem to occur with let-7i, as the primary transcript of let-7i was not altered, again supporting the existence of post-transcriptional regulation of let-7i-5p by HDAC. It is thus interesting that post-transcriptional regulation of both mRNA and miRNA can be revealed on the single-cell level through co-analysis of mRNA and miRNA transcriptomes.

While our study demonstrated the feasibility and utility of obtaining paired half-cell mRNA and miRNA profiles, our current approach has limited throughput in terms of the number of cells per run. Integrating the protocols demonstrated in this work in microfluidic systems could potentially improve on the throughput. Additionally, while this study focused on miRNAs, the half-cell small RNA sequencing data also contain information of other small RNA species, such as tRNA fragments that have been shown to be functionally important in human cancers and other diseases<sup>54-56</sup>. In addition, small RNAs mappable to other genomic regions could be readily detected in our half-cell profiles (**Supplementary Table 5**). We envision that further development and application of half-cell co-profiling of small RNAs and mRNAs could lead to novel mechanistic insights and far-reaching impact on small RNA biology.

## MATERIALS AND METHODS

### Cell culture

The human K562 cell line was from American Type Culture Collection (ATCC), and was maintained in RPMI 1640 media supplemented with 10% heat-inactivated fetal bovine serum (FBS, Life Technologies # 16000044) at the cell density between  $2 \times 10^5$  and  $2 \times 10^6$  per ml at 37°C with 5% CO<sub>2</sub>. The murine BaF3 hematopoietic cell line was from ATCC. BaF3 cells were cultured in RPMI 1640 medium containing 10% FBS, 1% of 100 × Penicillin-Streptomycin-Glutamine (Life Technologies # 10378016) and 3 ng ml<sup>-1</sup> of recombinant murine IL-3 (Peprotech #213-13). MCF-7 cells were cultured in DMEM media (Life Technologies, #11995065) with 10% FBS and 1% of Penicillin-Streptomycin-Glutamine. K562 cells were tested and showed sensitivity to Imatinib. BaF3 cells were tested and showed sensitivity toward IL3 withdrawal. MCF-7 cells were tested and showed sensitivity toward estrogen receptor inhibition. Cells were acquired mycoplasma-free, but have not been specifically tested for mycoplasma in this study.

For deriving K562 single cell clones, single K562 cells were hand-picked under microscope (see Cell Lysis for details) and maintained in single wells of a 96-well plate. Visual inspection was performed to ensure that only a single cell was present per well. Expanded single cell clones were harvested after ~20 days from the initial single-cell plating and further expanded for total RNA preparation.

For chemical treatment experiments, K562 cells were plated at 50,000 cells/well in a 96-well plate. Cells were treated with 10 μM of cyclopamine, 10 μM MK-2206, 1 μM Trichostatin A, 5 μM Afuresertib (AFU) or 10 μM Ipatasertib (IPA). All chemicals were from Sigma Aldrich, except for AFU and IPA which were from Selleckchem (GSK2110183 and GDC-0068). Chemicals were dissolved in DMSO, and DMSO was added to the control group. Cells were harvested after 24 hours for total RNA extraction.

### Oligonucleotides and primers

Sequences and modifications are available in **Supplementary Table 6**.

### Cell Lysis and Pre-treatment



Cells were washed once with cold PBS (Corning, #21-031-CV) and diluted to < 1000 cells/ml. Prior to picking single cells, 10  $\mu$ l of lysis buffer (0.25% Triton-X100, Sigma #T8787, 4 units (U) recombinant RNase Inhibitor, Takara #2313A) was deposited in each PCR tube. All single cell experiments were performed using 8-strip PCR tubes (USA Scientific #1402-4700) and 10  $\mu$ l filter tips (Denville #P1096-FR). To ensure single cell resolution, we performed most experiments by hand-picking of single cells under microscope. The diluted cell mixture was placed on the cover of a petridish, and single cells without close neighbours were manually picked using a 10  $\mu$ l pipette, with each cell contained within less than 1  $\mu$ l of volume. Single cells were then added into the lysis buffer. The single-cell lysates were stored at  $-80^{\circ}\text{C}$ . After thawing, the lysates were incubated at  $72^{\circ}\text{C}$  for 20 minutes or  $75^{\circ}\text{C}$  for 5 minutes to release small RNAs for reaction. The lysate was then split into two  $\sim 5$   $\mu$ l halves, with each half transferred into an empty PCR tube.

### **Small RNA library preparation and sequencing**

Preparation of the 3' adaptor: 5'-phosphorylated 3' adaptor oligonucleotides were ordered from IDT. Pre-adenylation of 5'-phosphorylated 3' adaptor was performed by using the 5' DNA adenylation kit (NEB, #E2610S) and purified using the Nucleotide Removal kit (Qiagen, #28304), both following the manufacturer's protocol. The concentration of the pre-adenylated adaptor was estimated by analysis on an 15% denaturing polyacrylamide gel (made with American Bioanalytical Sequel NE, #AB13021, AB13022) and comparison of band intensity to known quantities of synthesized oligonucleotides. The pre-adenylated 3'-adaptor (referred to below as 3' adaptor) was stored at  $-80^{\circ}\text{C}$  in aliquots. 5' Adaptor was ordered from Dharmacon. All other oligonucleotides were from IDT. Purification of oligonucleotides was performed by the synthesis companies.

Small RNA library preparation: Single-cell or half-cell lysate (5  $\mu$ l) was subjected to 3' adaptor ligation by adding 5.24  $\mu$ l of 3' adapter ligation reaction mixture (1 pmole 3' adapter, 1.66  $\mu$ l 50% PEG 8000, NEB #M0373S, 250 U T4 RNA Ligase 2 truncated KQ, NEB #M0373S, 0.83  $\mu$ l 10 $\times$  T4 RNA ligase buffer, NEB #M0373S, 20 U recombinant RNase Inhibitor, Takara, #2313A) and the reaction was incubated at  $30^{\circ}\text{C}$  for 6 hours followed by  $4^{\circ}\text{C}$  for 10 hours. Next, 5  $\mu$ l adaptor-digestion mixture was added (10 pmole RT primer, 5 U Lambda exonuclease, NEB



#0262S, 5 U 5' deadenylase, NEB #M0331S) and the reaction was incubated at 30 °C for 15 min followed by 37 °C for 15 min. Next, 5 µl 5' adapter ligation reaction mixture was added (10 pmole 5' adapter oligo, 10 U T4 RNA ligase 1, Thermo Fisher #EL0021, 1 µl T4 RNA ligase buffer, Thermo Fisher #EL0021, 2 µl 50% PEG 8000, NEB #M0373S) and incubated at 37 °C for 1 hour. Reverse transcription reaction was performed in three steps using M-MLV reverse transcriptase (RT) (Invitrogen #28025013). First, 5 µl of RT reaction mix was added (10 pmole RT primer) and the reaction was incubated at 65 °C for 5 minutes. Second, the mix of 2.1 µl 5x First Strand buffer (Invitrogen, #28025013), 0.3 µl H<sub>2</sub>O and 0.8 µl of 0.1M DTT (Invitrogen #28025013) was added and incubated at 42 °C for 30 minutes. Third, 3.3 µl of RT reaction was added (0.6µl M-MLV reverse transcriptase, 1 µl 5x First Strand buffer, Invitrogen #28025013, 0.3 µl H<sub>2</sub>O, 0.8 µl of 0.1M DTT, Invitrogen #28025013 and 0.6 µl 10mM dNTP mix, Invitrogen #18427088) and the reaction was incubated at 42 °C for 2 hours. The first PCR amplification was carried out by adding 35 µl of the reagents (3.5 µl 10xThermoPol Reaction Buffer, NEB #M0267S, 0.7 µl 10mM dNTP mix Invitrogen #18427088, 35 pmole RT primer oligo, 35 pmole PCR primer oligo, 0.5 µl Taq DNA Polymerase NEB #M0267S, 23.3 µl H<sub>2</sub>O) and incubating at 98 °C for 30 seconds followed by 13 cycles of 98 °C for 10 seconds, 60 °C for 30 seconds and 72 °C for 30 seconds and a final incubation at 72 °C for 5 minutes. Next, 1 µl of the amplified product was transferred to a fresh tube and to 25 µl of second PCR reaction (consisting of 10 µM indexed primer, 10 µM 5' Illumina PCR primer, 2.5 µl 10x ThermoPol Reaction buffer, 0.5 µl Taq DNA Polymerase NEB #M0267S, 0.5 µl 10mM dNTP mix, Invitrogen #18427088), followed by a 30 second incubation at 98 °C, 13 cycles of 98 °C for 10 seconds, 67 °C for 30 seconds and 72 °C for 30 seconds and a final incubation at 72 °C for 5 minutes. PCR products of ~140 bp were gel-purified after the second PCR on a 8% non-denaturing acrylamide gel. Gels were prepared by using Protein Gel Mix (American Bioanalytical #AB00283). Samples were prepared by adding 1/5 volume of 6X Gel loading dye purple (NEB #B7024S). Electrophoresis was performed at 140V for 80 minutes. After electrophoresis, the gel was stained in 15 ml water containing 1.5 µl GelStar Nucleic Acid Gel Stain (Lonza #50535) for 15 minutes. Bands were cut, eluted in 0.3 M sodium chloride overnight and precipitated.

All barcoded small RNA libraries were sequenced on an Illumina HiSeq 2000

instrument. Alternatively, for Luminex-based detection, the second PCR was performed with a 5' biotinylated primer, and detected on the Luminex platform as previously published<sup>57</sup>.

### **SMART-Seq mRNA profiling library preparation and sequencing**

We followed the manufacturer's protocol of SMART-Seq v4 Ultra Low Input RNA kit for sequencing (Clontech #634889). Then, Nextera DNA Library Preparation Kit (Illumina #FC-131-1096) was used to fragment and prepare the sequencing library. The libraries were sequenced on an Illumina HiSeq 2000 instruments with paired end 100 base reads.

### **Massively parallel single cell 3'-enriched RNA-seq**

Massively parallel single cell 3'-end mRNA sequencing was performed by adopting the DropSeq procedure<sup>37</sup> to microwell arrays, similar to demonstrated before<sup>38,39</sup>. Microwell arrays were fabricated in polydimethylsiloxane (PDMS) using standard soft lithography. Each array consists of 15,000 – 50,000 microwells, and microwells are sized 55 micron in diameter and 45 micron in depth to ensure capture of only one barcoded bead with poly-T-containing probe in each well by size exclusion. Cells were loaded into the device such that only 5-10% of the wells were loaded with single cells to prevent cell doublets and verified by microscopy before proceeding further. Once cell loading was completed, beads were loaded into the array at high density to ensure that each well receives a single bead (>95% bead occupancy). Following bead loading, lysis buffer was introduced into the array, and the array was then immediately sealed by introducing fluorinated oil. The cell lysates and beads were then incubated for an hour to capture the released mRNA molecules onto barcoded beads. Beads were then removed from the devices by centrifuging the devices upside down to push the beads out of microwells and flushing them out into an Eppendorf tube. The beads are then processed with subsequent reverse transcription, exonuclease treatment, amplification and library preparation as in the Dropseq protocol, and sequenced using paired end sequencing. After sequencing, read alignment and generation of gene expression matrices were performed as described in the DropSeq protocols<sup>42</sup>.

### **Small RNA sequence analysis**

Small RNA reads were analysed by a custom perl pipeline that has been described<sup>58</sup>. This pipeline was based on and modified from the miRDeep2 package<sup>56</sup>. Specifically, after removing adaptor sequences, collapsing reads with the same sequence, and removal of reads with less than 15 bases in length, small RNA reads were mapped to miRNA precursors from miRBASE release 21 using miRDeep2, allowing zero mismatches. Mapped reads were quantified by miRDeep2. In addition, reads were also mapped to several categories of known RNA species sequentially, using bowtie2. Reads were first mapped to human miRNA precursors, with unmapped reads used in the next step to map to snoRNAs. Similar sequential mapping was performed for snRNA, scRNA, tRNA, rRNA and piRNAs. Unmapped reads left from the above mapping were also mapped to the genome. However, we found that there are often sequence variants of the ligation adaptors with perfect mapping to the genome, thus creating artefacts. We thus did not utilize the mapping to the genome to quantify the mapping results. Overall, among the 19 half cells, total reads that map to miRNA account for  $30.18 \pm 7.62\%$  of all reads mapping to the above mentioned RNA species. Reads that map to snoRNA, snRNA, scRNA, tRNA, rRNA and piRNA represent  $8.23 \pm 2.41\%$ ,  $1.02 \pm 0.33\%$ ,  $2.17 \pm 1.07\%$ ,  $6.11 \pm 0.88\%$ ,  $52.28 \pm 9.27\%$  and  $\sim 0.01\%$  respectively.

Normalization of miRNA reads were performed by dividing the reads quantified by miRDeep2 with total miRNA reads quantified by bowtie 2. This is because some reads were counted multiple times when miRDeep2 assigned them to miRNA precursors, especially for miRNAs that map to multiple genomic loci. The resultant data reflect the fraction of a given miRNA within all miRNAs, which we refer to as miRNA expression or miRNA levels in figures. The validity of using this fraction-based data was supported by single-cell qRT-PCR validation. Normalized data were further applied with a minimal expression level of  $10e-4$  (i.e. all values lower than this threshold were set to this level), and log2 transformed for further analyses.

### **RNAseq data analysis**

RNAseq data from the half-cell profiles were processed using GenePattern software. Specifically, K562 RNAseq data were aligned using Tophat to hg19 genome assembly, using gtf files based on UCSC gene definition downloaded from the GenePattern server. Aligned reads were quantified using the CuffDiff program, by using UCSC gtf definition for hg19, to obtain RPKM values. Among the 19 half K562

cell profiles, the number of genes detected is  $11209 \pm 592$  per half-cell. Data were applied with a minimal expression level of  $10e^{-3}$ , and  $\log_2$  transformed for further analyses.

To identify similar groups of cells, consensus clustering was performed using matlab codes. Specifically, for each iteration, 80% of samples were randomly sampled, and subjected to hierarchical clustering. Hierarchical clustering was performed with average linkage and Pearson correlation metrics after row-based centring and normalization (subtracting mean of each row and dividing by standard deviation of each row). A total of 50 iterations was performed, and consensus clusters of 2 to 5 were manually examined. Three clusters were selected based on the clustering results for further analyses.

To identify differential gene expression between clusters, permutation was performed to shuffle sample labels, with a total of 5000 iterations, and using ttest score as a metric. Nominal p values were obtained by quantifying the fraction of permutations that reached a more extreme t-test score. False discovery rates were then computed based on the nominal p values using the Benjamini-Hochberg method. For 19 K562 half-cell data, the differential gene expression was defined with  $FDR < 0.1$ . For ~400 K562 single cells, the differential gene expression was defined with  $FDR < 0.05$ . Enrichment of pathways was performed by DAVID analysis with the corresponding groups of differentially expressed genes.

### **Co-analysis of miRNA and mRNA**

To analyse the relationship between miRNA and miRNA targets, mRNA genes were ranked by the Pearson correlation to miRNA expression among the 19 K562 half-cells. Initial analyses were performed using Gene Set Enrichment Analysis, using miRNA target gene set defined in MSigDB, with 1000 or more permutations on gene labels, and with Pearson correlation as weights during the enrichment analyses. In addition, miRNA target genes defined in TargetScan release 7.1 were also examined, using the same enrichment algorithm as in GSEA.

For examining miRNA-target relationships using cumulative distribution functions, the predicted miRNA targets were obtained from TargetScan release 7.1, and filtered to eliminate predicted targets with poor total context scores. Specifically, we only retained predicted targets that had total context scores  $< -0.1$ . Targets were further

categorized according to mean log<sub>2</sub> expression (measured in RPKM). P values were calculated based on the Kolmogorov-Smirnov test, using matlab's kstest2 function.

For any given miRNA, the rank list of mRNA according to the correlation was also used to query MSigDB for potential pathways or gene sets. To query the Connectivity map (cMap), the ranked mRNAs were separated into positively correlated genes and negatively correlated genes, with a cutoff of 0.45 or -0.45 for the correlation score. These two lists of gene symbols were then mapped to the probe names of U133A chip to query CMap.

### **Single-cell Quantitative Real Time RT-PCR**

Single cell qRT-PCR was performed following a published study<sup>19</sup> with modifications. For preparation of cell lysate for qRT-PCR, prior to picking single cells, 4 µl of lysis buffer (1% Triton X-100, 4 U recombinant RNase Inhibitor, Takara #2313A) and 2e-19 mole of synthetic hsa-miR-371-5p, Ambion) was deposited in each PCR tube (TempAssure PCR 8-Stripes, USA Scientific #1402-4700). The spiked-in hsa-miR-371-5p permitted final data normalization with a miRNA not expressed endogenously in K562 cells. K562 cells were first washed once with cold PBS. K562 single cells were manually picked by using a 10 µl pipette in 1µl volume and added to lysis buffer. Cell lysate was stored in -80 °C, thawed and incubated at 75 °C for 5 minutes and subjected to miScript RT reactions (Qiagen, by adding 2 µl 5x miScript HiSpec buffer, 1 µl 10x nucleotides, 1 µl miScript RT enzyme mix). RT product was diluted 20 times (v/v) with water, and 1 µl diluted RT product was used for Pre-PCR reaction in 10 µl total volume with Power SYBR-green PCR master mix (Applied Biosystems) for miRNAs. For a given experiment, all relevant miRNA-specific primers (2 pmole each), as well as that of the spike-in control were added. See **Supplementary Table 6** for specific primers used. qPCR reaction was then performed after pre-PCR and the dilution of pre-PCR product 20 times with water (v/v). The total volume for the reaction was 10 µl (5 µl Power SYBR-green PCR master mix, 2pmole miScript universal primer, 2pmole miRNA specific primer, 2 µl diluted pre-PCR product). The expression of each miRNA was quantified by comparative Ct method ( $2^{-\Delta Ct}$ ), using the detected levels of hsa-miR-371-5p as a control.

### **RNA Extraction and bulk population qRT-PCR**

Total RNA was extracted using the TRIzol Reagent (Life Technologies #15596018) following the manufacture's protocol. For qRT-PCR of bulk population samples, 10 ng of total RNA was used for each biological sample. U6 small RNA was used as a control for qRT-PCR of miRNAs. All RT, qPCR and data analysis details were following the procedure in above sections.

### **Statistical Analysis**

Student's t-test (unequal variance, two-tailed) was used in analysis of chemical treatment experiments. Number of replicates were determined after an initial experiment with three biological replicates to gauge the data variation. For analysis of correlation between miRNAs and its predicated targets, Kolmogorov-Smirnov test was used (two-sided) through Matlab `kstest2` function. Statistics of Gene Set Enrichment Analysis was obtained from the GSEA program output, using the familywise-error rate. For evaluating the variation of miR-146 with AKT inhibitor treatment, p value was calculated using custom matlab codes by performing random permutations. Specifically, for both control and AKT inhibitor groups, 15 samples per group were randomly drawn to calculate standard deviations. A total of 1000 permutations were performed. P value was calculated as the probability of rejecting the hypothesis that the AKT inhibitor group has lower variation than the control group, by dividing the number of random instances, in which the standard deviation of the AKT inhibitor group is larger or equal to that of the control group, by the total number of random instances. P value of 0.006 was observed. Similar significant results were obtained with random sample size from 13 to 18.

### **Data Availability**

Next-generation sequencing data have been deposited to GEO under series GSE114071.



## References

1. Pritchard, C.C., Cheng, H.H. & Tewari, M. MicroRNA profiling: approaches and considerations. *Nat Rev Genet* **13**, 358-369 (2012).
2. Wang, N.Y., Cheng, J.J., Fan, R. & Lu, J. Capture, amplification, and global profiling of microRNAs from low quantities of whole cell lysate. *Analyst* **142**, 3203-3211 (2017).
3. Bushati, N. & Cohen, S.M. MicroRNA functions. *Annu Rev Cell Dev Bi* **23**, 175-205 (2007).
4. Lu, J. et al. MicroRNA expression profiles classify human cancers. *Nature* **435**, 834-838 (2005).
5. Calin, G.A. & Croce, C.M. MicroRNA signatures in human cancers. *Nat Rev Cancer* **6**, 857-866 (2006).
6. Volinia, S. et al. A microRNA expression signature of human solid tumors defines cancer gene targets. *P Natl Acad Sci USA* **103**, 2257-2261 (2006).
7. Croce, C.M. Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet* **10**, 704-714 (2009).
8. Tang, F.C. et al. RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat Protoc* **5**, 516-535 (2010).
9. Picelli, S. Single-cell RNA-sequencing: The future of genome biology is now. *Rna Biol* **14**, 637-650 (2017).
10. Tang, F.C. et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* **6**, 377-U386 (2009).
11. Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat Rev Genet* **14**, 618-630 (2013).
12. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-181 (2014).
13. Navin, N. et al. Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90-U119 (2011).
14. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. *Cell Rep* **2**, 666-673 (2012).
15. Islam, S. et al. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res* **21**, 1160-1167 (2011).
16. Ramskold, D. et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol* **30**, 777-782 (2012).
17. Islam, S. et al. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods* **11**, 163-+ (2014).
18. Faridani, O.R. et al. Single-cell sequencing of the small-RNA transcriptome. *Nat Biotechnol* **34**, 1264-1266 (2016).
19. Tang, F.C. et al. 220-plex microRNA expression profile of a single cell. *Nat Protoc* **1**, 1154-1159 (2006).
20. Condorelli, G. microRNA: Medical Evidence From Molecular Biology to Clinical Practice Foreword. *Adv Exp Med Biol* **888**, V-Vi (2015).
21. Lu, J. & Tsourkas, A. Imaging individual microRNAs in single mammalian cells in situ. *Nucleic Acids Res* **37** (2009).
22. Zong, C.H., Lu, S.J., Chapman, A.R. & Xie, X.S. Genome-Wide Detection of Single-Nucleotide and Copy-Number Variations of a Single Human Cell. *Science* **338**, 1622-1626 (2012).



23. Baslan, T. et al. Genome-wide copy number analysis of single cells. *Nat Protoc* **7**, 1024-1041 (2012).
24. Macaulay, I.C., Ponting, C.P. & Voet, T. Single-Cell Multiomics: Multiple Measurements from Single Cells. *Trends Genet* **33**, 155-168 (2017).
25. Lorthongpanich, C. et al. Single-Cell DNA-Methylation Analysis Reveals Epigenetic Chimerism in Preimplantation Embryos. *Science* **341**, 1110-1112 (2013).
26. Clark, S.J., Lee, H.J., Smallwood, S.A., Kelsey, G. & Reik, W. Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. *Genome Biol* **17** (2016).
27. Levsky, J.M., Shenoy, S.M., Pezo, R.C. & Singer, R.H. Single-cell gene expression profiling. *Science* **297**, 836-840 (2002).
28. Klein, C.A. et al. Combined transcriptome and genome analysis of single micrometastatic cells. *Nat Biotechnol* **20**, 387-392 (2002).
29. Taniguchi, Y. et al. Quantifying E-coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science* **329**, 533-538 (2010).
30. Clark, S.J. et al. scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat Commun* **9** (2018).
31. Angermueller, C. et al. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods* **13**, 229-232 (2016).
32. Macaulay, I.C. et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat Methods* **12**, 519-+ (2015).
33. Hu, Y.J. et al. Simultaneous profiling of transcriptome and DNA methylome from a single cell. *Genome Biol* **17** (2016).
34. Hou, Y. et al. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res* **26**, 304-319 (2016).
35. Roden, C., Mastriano, S., Wang, N.Y. & Lu, J. microRNA Expression Profiling: Technologies, Insights, and Prospects. *Adv Exp Med Biol* **888**, 409-421 (2015).
36. Gennarino, V.A. et al. MicroRNA target prediction by expression analysis of host genes. *Genome Res* **19**, 481-490 (2009).
37. Su, N.F., Qian, M.P. & Deng, M.H. Integrative Approaches for microRNA Target Prediction: Combining Sequence Information and the Paired mRNA and miRNA Expression Profiles. *Curr Bioinform* **8**, 37-45 (2013).
38. Cho, S. et al. miRGator v3.0: a microRNA portal for deep sequencing, expression profiling and mRNA targeting. *Nucleic Acids Res* **41**, D252-D257 (2013).
39. Giles, C.B., Girija-Devi, R., Dozmorov, M.G. & Wren, J.D. mirCoX: a database of miRNA-mRNA expression correlations derived from RNA-seq meta-analysis. *Bmc Bioinformatics* **14** (2013).
40. Vaz, C. et al. Analysis of microRNA transcriptome by deep sequencing of small RNA libraries of peripheral blood. *Bmc Genomics* **11** (2010).
41. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* **10**, 1096-1098 (2013).
42. Macosko, E.Z. et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214 (2015).
43. Yuan, J.Z. & Sims, P.A. An Automated Microwell Platform for Large-Scale Single Cell RNA-Seq. *Sci Rep-Uk* **6** (2016).

44. Gierahn, T.M. et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods* **14**, 395-+ (2017).
45. Mukherji, S. et al. MicroRNAs can generate thresholds in target gene expression. *Nat Genet* **43**, 854-U860 (2011).
46. Liberzon, A. et al. The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst* **1**, 417-425 (2015).
47. Lamb, J. et al. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929-1935 (2006).
48. Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *P Natl Acad Sci USA* **102**, 15545-15550 (2005).
49. Bosson, A.D., Zamudio, J.R. & Sharp, P.A. Endogenous miRNA and Target Concentrations Determine Susceptibility to Potential ceRNA Competition. *Mol Cell* **56**, 347-359 (2014).
50. Denzler, R. et al. Impact of MicroRNA Levels, Target-Site Complementarity, and Cooperativity on Competing Endogenous RNA-Regulated Gene Expression. *Mol Cell* **64**, 565-579 (2016).
51. Arvey, A., Larsson, E., Sander, C., Leslie, C.S. & Marks, D.S. Target mRNA abundance dilutes microRNA and siRNA activity. *Mol Syst Biol* **6** (2010).
52. Schmiedel, J.M. et al. MicroRNA control of protein expression noise. *Science* **348**, 128-132 (2015).
53. Horman, S.R. et al. Akt-mediated phosphorylation of argonaute 2 downregulates cleavage and upregulates translational repression of MicroRNA targets. *Mol Cell* **50**, 356-367 (2013).
54. Goodarzi, H. et al. Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement. *Cell* **161**, 790-802 (2015).
55. Kirchner, S. & Ignatova, Z. Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nat Rev Genet* **16**, 98-112 (2015).
56. Martens-Uzunova, E.S., Olvedy, M. & Jenster, G. Beyond microRNA - Novel RNAs derived from small non-coding RNA and their implication in cancer. *Cancer Lett* **340**, 201-211 (2013).
57. Lu, J. et al. MicroRNA-mediated control of cell fate in megakaryocyte-erythrocyte progenitors. *Dev Cell* **14**, 843-853 (2008).
58. Guo, Y.W. et al. Characterization of the mammalian miRNA turnover landscape. *Nucleic Acids Res* **43**, 2326-2341 (2015).

## End Notes

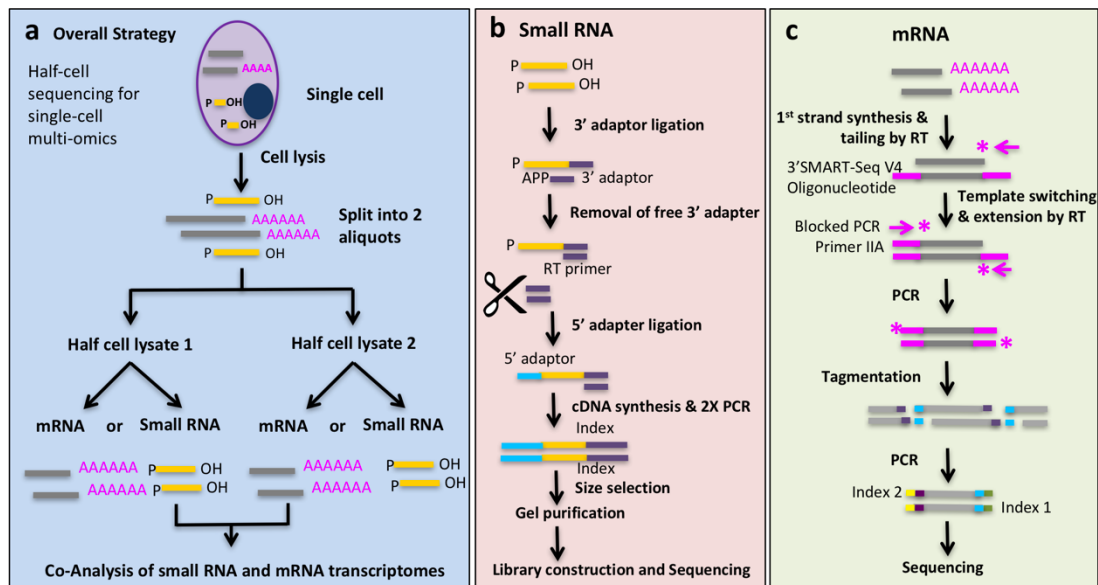
This work was supported in part by NIH grants R01CA149109 (to J.L.), R01GM116855 (to Y.D. and J.L.), R21CA177393 (to R.F.), U54CA193461 (to R.F.), U54CA209992 (Sub-Project ID: 7297 to R.F.), Connecticut RMRF grant 15-RMB-YALE-06 (to J.L.), and National Science Foundation CAREER Award CBET-1351443 (to R.F.). Services provided by the NIDDK-supported Yale Cooperative Center of Excellence in Hematology (U54DK106857) assisted this study.

RF and JL conceived the study and supervised the work. NW established the single and half-cell small RNA profiling method. NW and ZC generated half-cell mRNA profiles. NW, JZ and YL generated single/half-cell miRNA profiles, performed single/half-cell qRT-PCR, and performed experiments on drug treatment. BD, MK and ZC established, obtained and analysed data for high-throughput single-cell RNAseq. JX, YL and DK contributed to the drug treatment experiments. MZ, CR and JC contributed to cell acquisition and subsequent experiments. JL and YD analysed data from single cell profiles. NW, RF and JL wrote the manuscript.

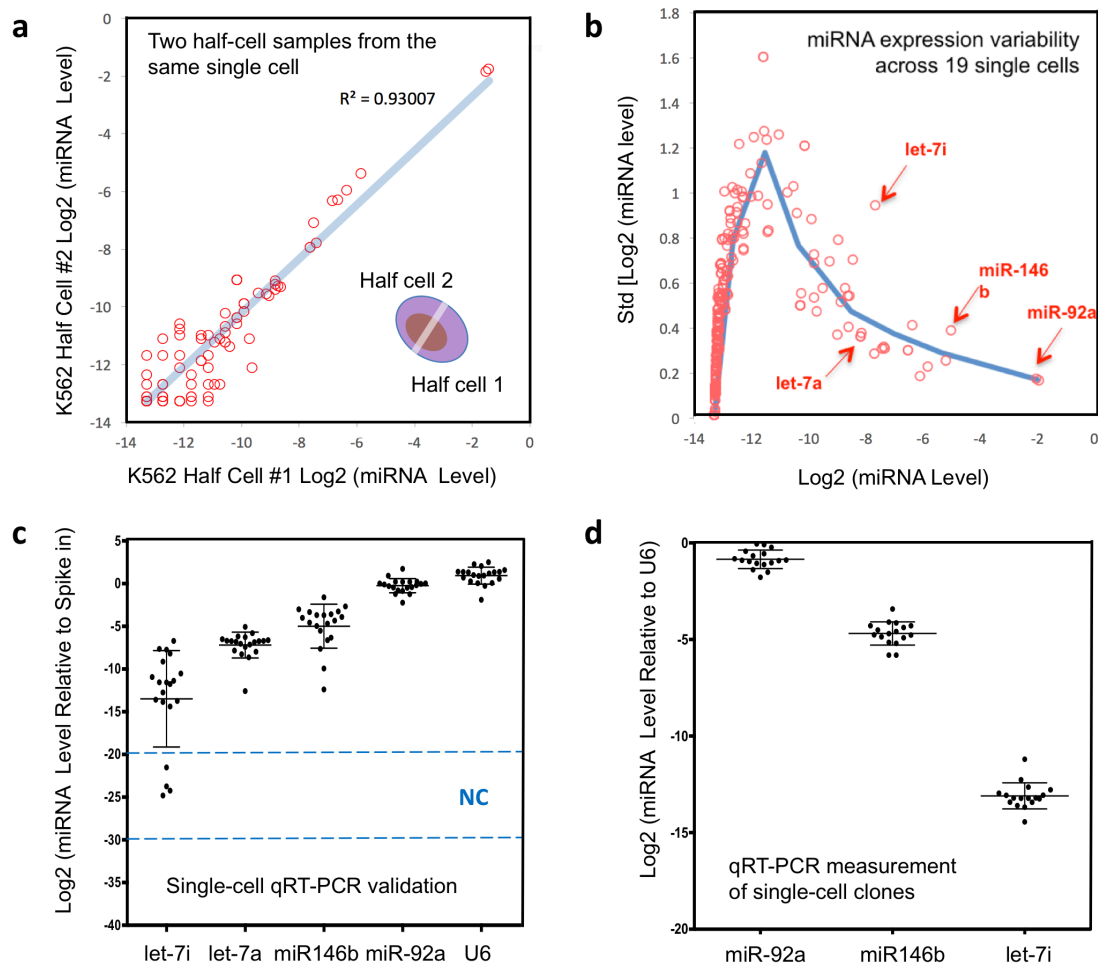
### **Conflict of Interest**

Rong Fan is on the Scientific Advisory Boards of IsoPlexis, Bio-Techne, and Singleron Biotechnologies.

## Figures and Figure Legends

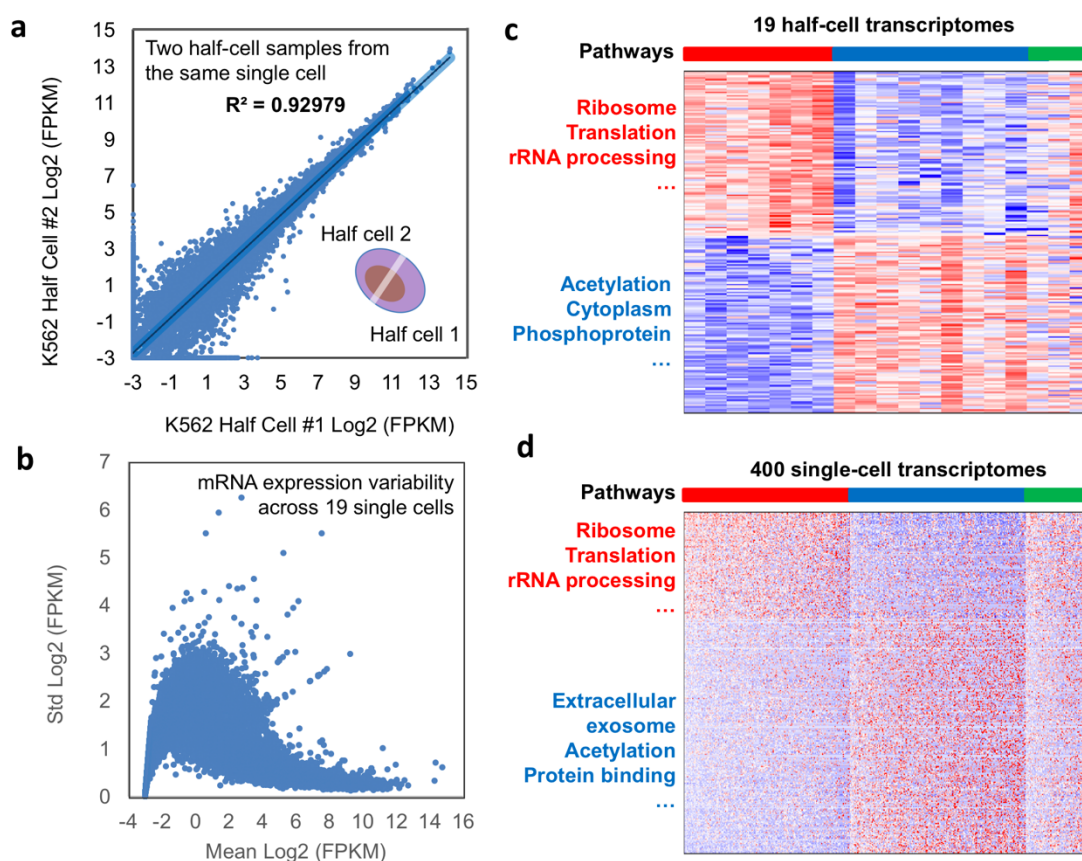


**Figure 1. Experimental workflow.** (a) Overall strategy for profiling miRNA and mRNA from the same single cells using half-cell genomics. It involves cell lysis, half-cell split, followed by small RNA & large RNA library preparation. (b) For miRNA library preparation, a pre-adenylated (APP) 3' adaptor was used to ligate to the 3' end of miRNA molecules, followed by digestion of unreacted 3' adaptor, ligation with 5' adaptor, RT and PCR amplification. (c) For mRNA library preparation, first-strand cDNA synthesis was primed by the 3' SMART-Seq CDS Primer IIA. Template switching at the 5' end of transcript was performed using the SMART-Seq v4 oligonucleotides. After PCR amplification, cDNA was fragmented using Illumina's Tagmentation process.



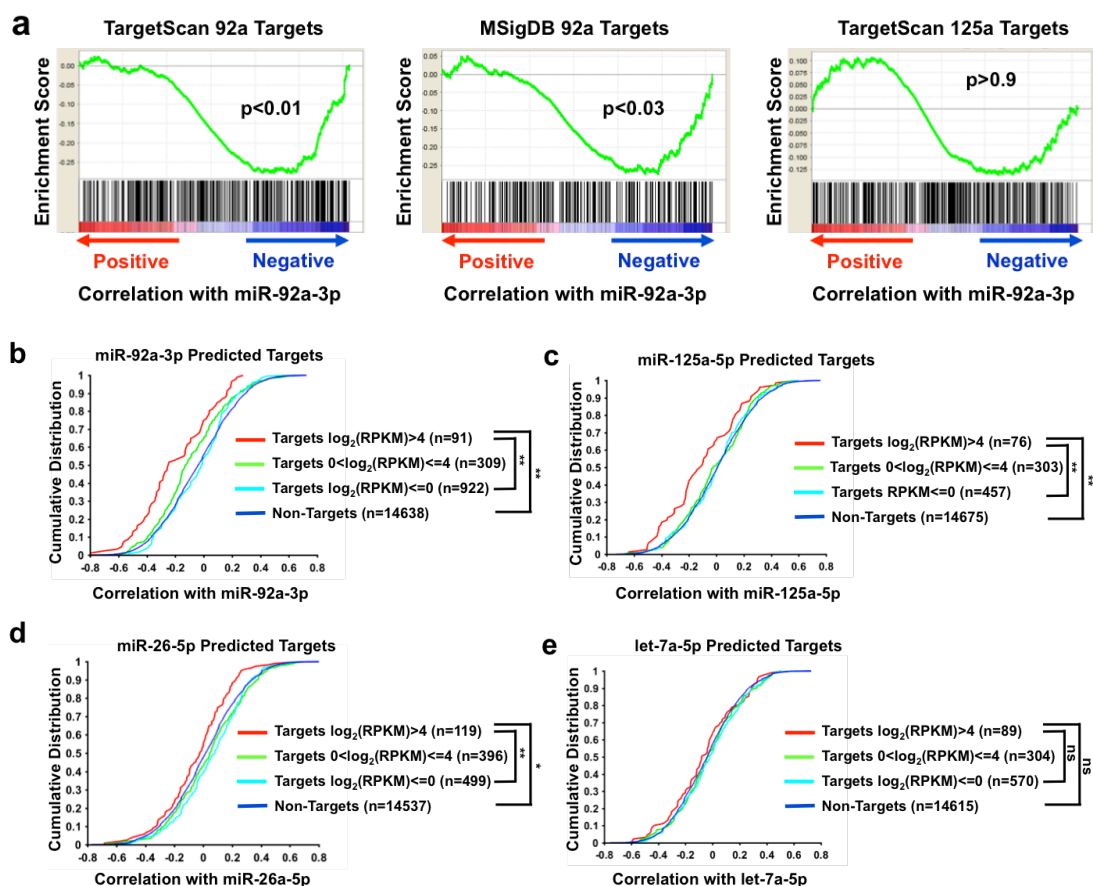
**Figure 2. Profiling of miRNAs from half-cell lysate.** (a) A single K562 cell was lysed, and two halves of the lysate were separately subjected to small RNA sequencing. Scatter plot of normalized and log<sub>2</sub>-transformed miRNA expression levels (miRNA expression levels represent the fraction among total miRNA, see Methods) is shown. Each circle represents one annotated miRNA. The correlation coefficient  $R^2=0.93$ . (b) The standard deviations (on y-axis) of miRNA expression across 19 successfully profiled half cells were plotted against the mean expression on x-axis, using normalized and log<sub>2</sub>-transformed miRNA expression data as described in (a). Each circle represents one annotated miRNA. The blue line shows the trend of the distribution based on locally averaged values. Specific miRNAs are highlighted with red arrows. (c) Validation experiment was carried out by quantifying miRNA expression in 20 single K562 cells using qRT-PCR. Relative levels of let-7i-5p, let-7a-5p, miR-146b-5p, miR-92a-3p and U6 were determined relative to a spike-in control (synthetic miR-371). Each dot represents one K562 cell. Each data point reflects the average measurements from two technical replicates. Error bars present standard deviation. The levels of miRNA are shown in log<sub>2</sub> scale. The dashed lines and NC (negative control) indicates the detection range of water samples. (d) Single K562 cells were grown to derive 17 clones, and the relative levels of miR-92a-3p, miR-146b-

5p and let-7i-5p were determined in clonal populations by qRT-PCR. Roughly 40,000 cells were used from each clone. Each dot represents one K562 clone. Each data point reflects the average from two technical replicates. Error bars present standard deviation. U6 was measured as a control for normalization. The levels of miRNAs are shown in log scale.

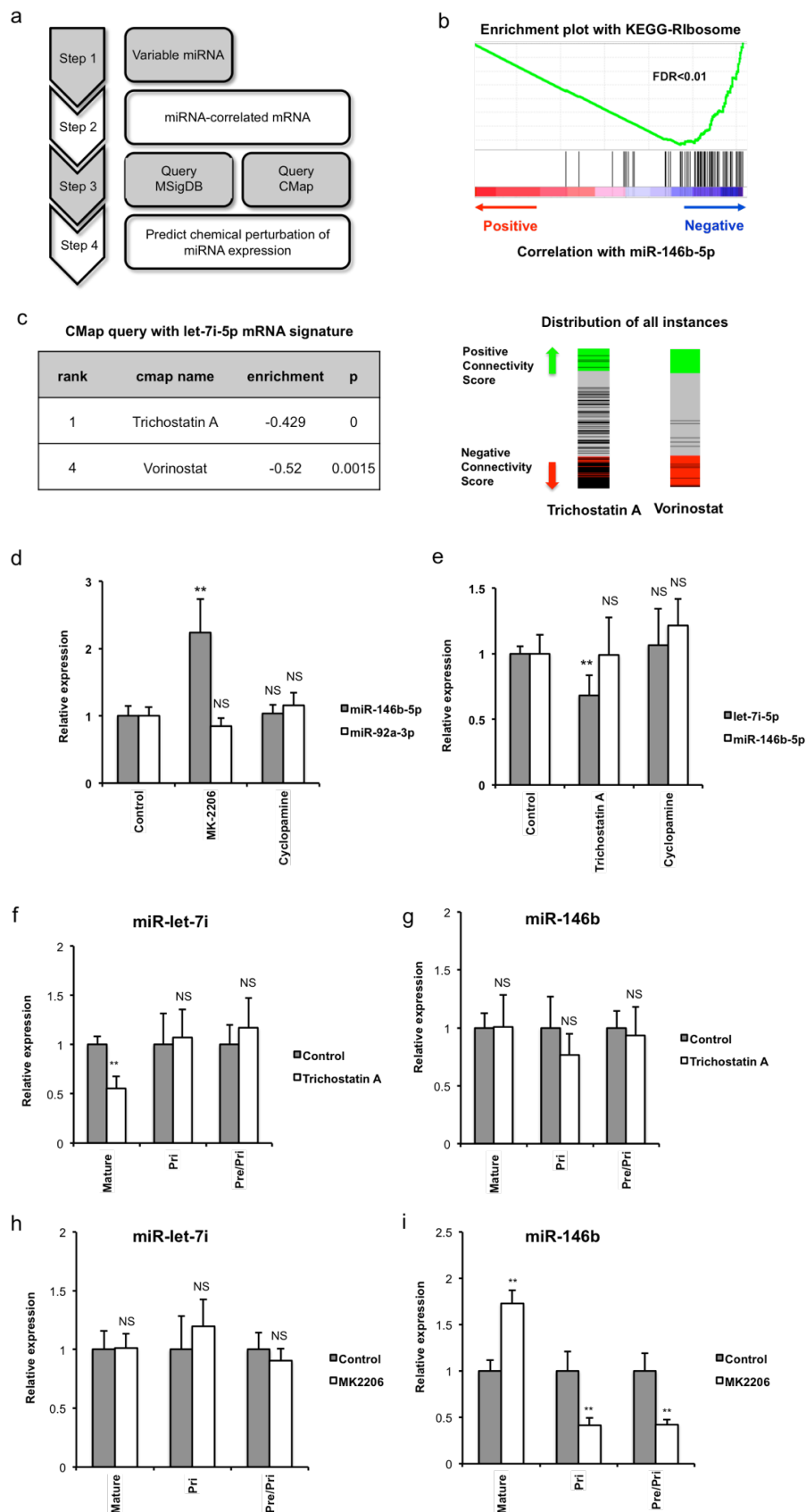


**Figure 3. Global profiling of mRNAs from half cells.** (a) A single K562 cell was lysed, and two halves of the lysate were separately subjected to RNAseq. Scatter plot of normalized and log<sub>2</sub>-transformed mRNA expression levels (in FPKM units) is shown. Each circle represents one annotated mRNA. The correlation coefficient  $R^2$  is 0.930. (b) The standard deviations of mRNA expression across 19 successfully profiled half cells were plotted against the mean expression, using log<sub>2</sub>-transformed mRNA expression data. Each circle represents one annotated mRNA. (c) Consensus clusters were identified in the 19 K562 half-cell mRNA expression data. A heatmap is shown for differentially expressed genes between the red and blue clusters, with enriched pathways annotated by the DAVID pathway analysis. Each row represents one annotated gene whereas each column represents a single cell. (d) K562 cells were profiled using a massively parallel single-cell 3'-end RNAseq technology. A heatmap is shown for differentially expressed genes between the red and blue consensus clusters, with enriched pathways annotated by the DAVID pathway analysis. Each row represents one annotated gene whereas each column represents a single cell.





**Figure 4. Relationships between miRNA and miRNA targets in paired half-cell data.** (a) mRNAs were ranked according to their correlation with miR-92a-3p using data from the 19 pairs of half-cell miRNA and mRNA profiles. The resultant rank list was queried with predicted targets of miR-92a-3p obtained from TargetScan or MSigDB (left and middle panels) or with targets of miR-125a-5p from TargetScan (right panel). Gene Set Enrichment Analysis plots are shown with p values indicated. (b-e) mRNAs were ranked according to their correlation with (b) miR-92a-3p, (c) miR-125a-5p, (d) miR-26a-5p or (e) let-7a-5p. Cumulative distribution functions were plotted with non-targets and with predicted targets of the corresponding miRNA (based on TargetScan). Predicted targets were further categorized based on the mean RPKM values across the 19 K562 half-cell samples. The number of genes in each category is indicated in parentheses. P values were calculated based on the Kolmogorov-Smirnov test between the indicated pairs of conditions. \*\*: P < 0.005; \*: P < 0.05; ns: not significant.



**Figure 5. Uncovering regulators of miRNA expression through paired half-cell data.** (a) Schematics of the approach to infer chemical perturbation of miRNA expression through paired half-cell mRNA and miRNA profiles. Regulators of miRNA were predicted based on Gene Set Enrichment Analysis (GSEA) Molecular Signature Database (MSigDB) and Connectivity Map (CMap). (b) mRNAs were ranked according to Pearson correlation coefficients with miR-146b-5p. The resultant rank list was used to query the MSigDB database. An enrichment plot is shown for the gene set KEGG-Ribosome in which genes in this gene set were enriched for negative correlation with the miRNA. False discovery rate (FDR) is indicated. (c) mRNAs were ranked according to Pearson correlation coefficients with let-7i-5p. mRNAs with correlation coefficient  $R > 0.45$  were defined as the positive signature whereas mRNAs with correlation coefficient  $R < -0.45$  were used as the negative signature. mRNA signatures of let-7i-5p were used to query the CMap database. Two HDAC inhibitors were ranked at the top of the list (left panel). Right panel shows the enrichment plots of Trichostatin A and Vorinostat treatment instances within the CMap database. (d-i) K562 cells were treated with 10 $\mu$ M AKT inhibitor MK-2206, 10 $\mu$ M Cyclopamine, 1 $\mu$ M HDAC inhibitor Trichostatin A or vehicle control (DMSO) for 24 hours. (d, e) The expression levels of miR-146b-5p, miR-92a-3p or let-7i-5p were determined with qRT-PCR. N=6 biological replicates. Error bars: standard deviation. A representative experiment out of two is shown. (f-i) Mature miR-146b-5p and let-7i-5p levels, as well as the expression levels of the corresponding primary miRNA transcript were determined by qRT-PCR. In addition, primer sets that detect both primary and precursor miRNAs were also used (pri/pre). N=6 biological replicates. Error bars: standard deviation. A representative experiment out of two is shown. \*:  $p < 0.01$ ; \*\*  $p > 0.05$ ; ns: not significant, as determined by student's t-test.

## **Supplementary Information**

### **Single-cell microRNA/mRNA co-sequencing reveals non-genetic heterogeneity and novel regulatory mechanisms**

**Nayi Wang<sup>1,2,3</sup>, Ji Zheng<sup>2,3,4</sup>, Zhuo Chen<sup>1,2,3</sup>, Yang Liu<sup>2,3</sup>, Burak Dura<sup>1</sup>, Minsuk Kwak<sup>1</sup>, Juliana Xavier-Ferrucio<sup>3,5</sup>, Yi-Chien Lu<sup>3,5</sup>, Miaomiao Zhang<sup>2,4,6</sup>, Christine Roden<sup>2,3</sup>, Jijun Cheng<sup>2,3</sup>, Diane Krause<sup>3,5</sup>, Ye Ding<sup>7</sup>, Rong Fan<sup>1,3\*</sup> & Jun Lu<sup>2,3,8,9\*</sup>**

<sup>1</sup>Department of Biomedical Engineering, Yale University, New Haven, CT 06520, USA

<sup>2</sup>Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06510, USA

<sup>3</sup>Yale Stem Cell Center, Yale Cancer Center, New Haven, CT, 06520, USA

<sup>4</sup>Southwest Hospital, Third Military Medical University, Chongqing, 400038, China

<sup>5</sup>Department of Laboratory Medicine, Yale University School of Medicine, New Haven, Connecticut 06510, USA

<sup>6</sup>School of Medicine, Jiangsu University, Zhenjiang, Jiangsu, 212013, China

<sup>7</sup>Wadsworth Center, New York State Department of Health, Albany, New York 12208, USA

<sup>8</sup>Yale Center for RNA Science and Medicine, New Haven, CT, 06520, USA

<sup>9</sup>Yale Cooperative Center of Excellence in Hematology, New Haven, CT, 06520, USA

\* Correspondence should be addressed to:

Rong Fan, 55 Prospect Street, Room 213, New Haven, CT 06520.

Telephone: 203-432-9905; Fax: 203-432-6441;

Email: rong.fan@yale.edu

or

Jun Lu, 10 Amistad Street, Room 237C, New Haven, CT 06520.

Telephone: 203-737-3426; Fax: 203-785-4305;

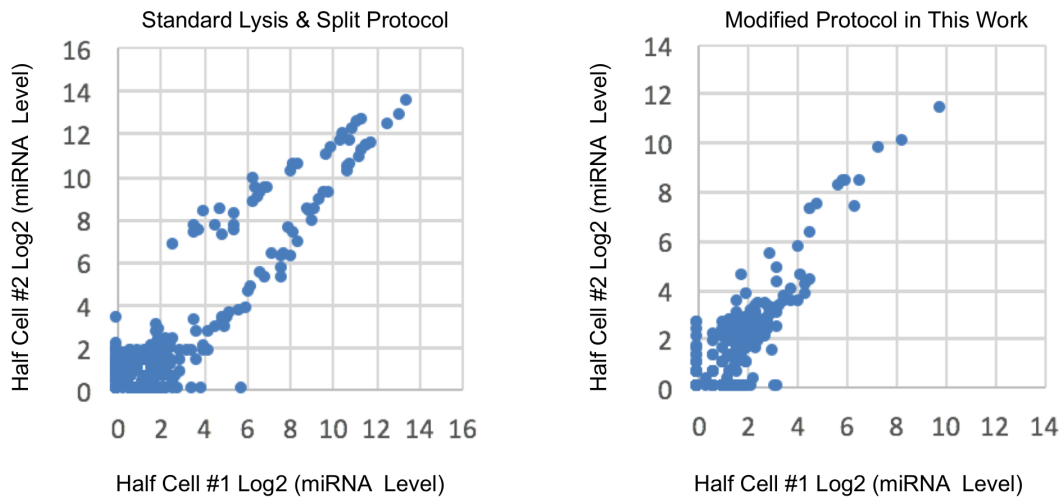
Email: jun.lu@yale.edu

Enclosed are :

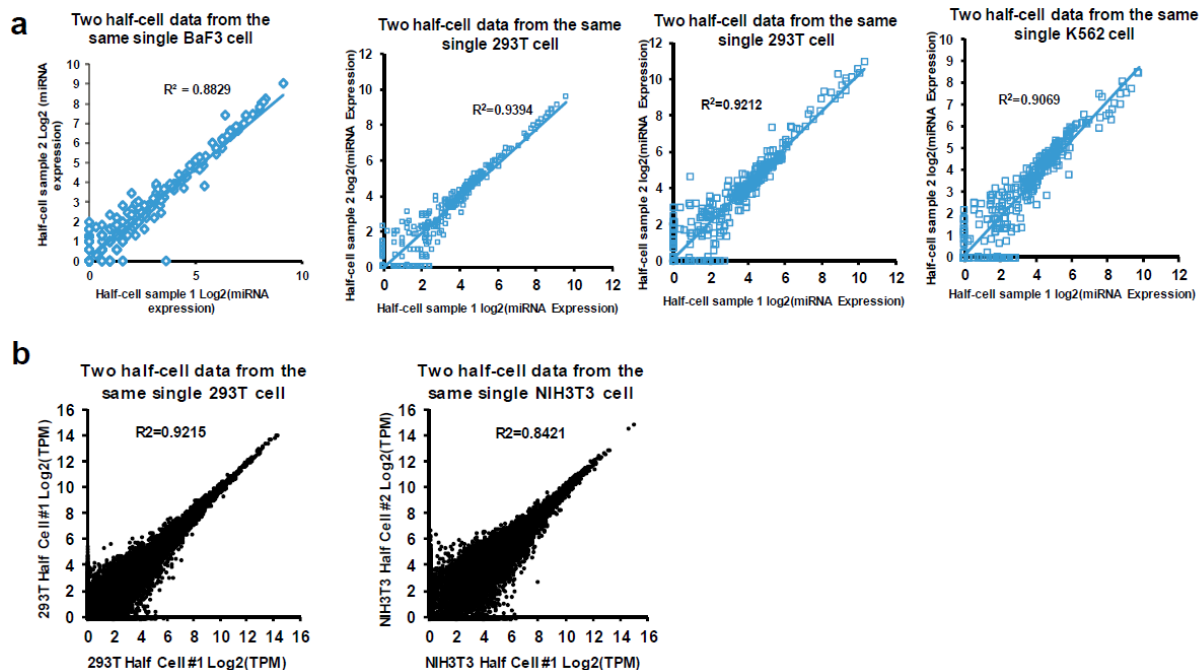
11 Supplementary Figures and

6 Supplementary Tables.

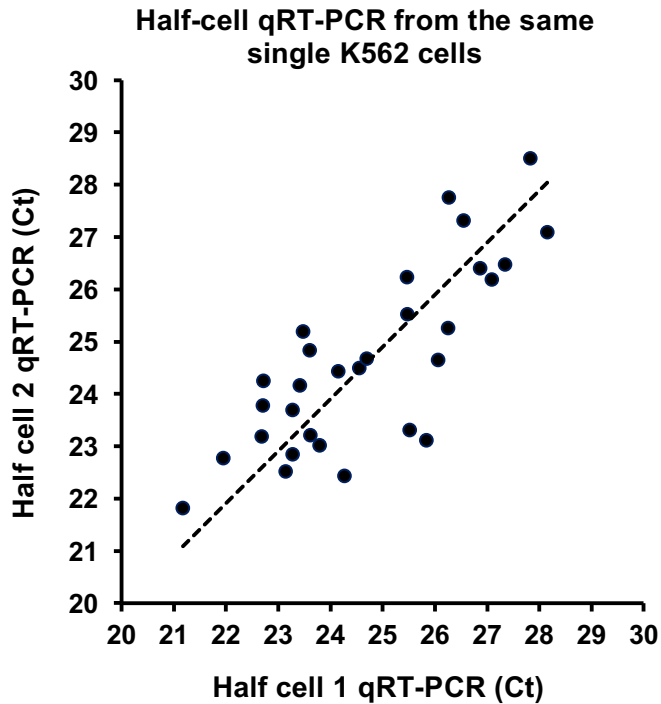
## Supplementary Figures



**Supplementary Figure 1. New protocol to improve the splitting of single-cell lysate to profile miRNAs.** Direct splitting of single-cell lysate prepared using current methods did not yield even distribution of miRNAs in two half-cell samples. This is presumably due to the binding of miRNAs in cytoplasmic proteins that are still tethered to partially digested cellular materials. The new protocol that combines multiple freeze/thaw and heat treatment led to significant improvement in single-cell lysate splitting. Left: profiling of microRNAs from two half-cell samples split from the same single cell. Cell lysis was performed using standard methods. Right: the same profiling except that the protocol for the cell lysis and pre-treatment method has been modified for single cell miRNA profiling. The log<sub>2</sub> transformed miRNA expression levels (see Methods) are plotted, with each dot representing a single miRNA.

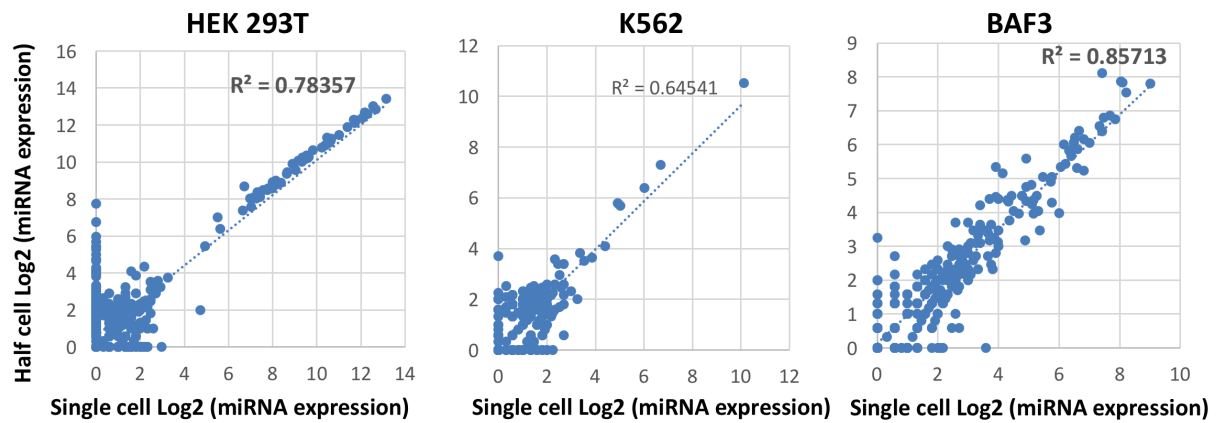


**Supplementary Figure 2. Profiling of miRNA expression in two half-cell samples derived from the same single cells.** (a) The indicated single cells were lysed, and two halves of the lysate from the same single 293T cell were split (using the improved approach) and independently subjected to the processing shown in Figure 1b and then miRNAs were measured on the Luminex Multiplex Bead Array platform. Scatter plots of normalized and log<sub>2</sub>-transformed miRNA expression levels (see Methods) are shown, with each panel showing data from one of the indicated cells. Each dot represents one annotated miRNA, with  $R^2$  indicated. (b) Similar to (a), the indicated single cells were lysed, and two halves of the lysate from the same single cell were split and subjected to RNAseq analyses. Scatter plots of log<sub>2</sub>-transformed mRNA expression levels are shown. Each dot represents one annotated mRNA, with  $R^2$  indicated.

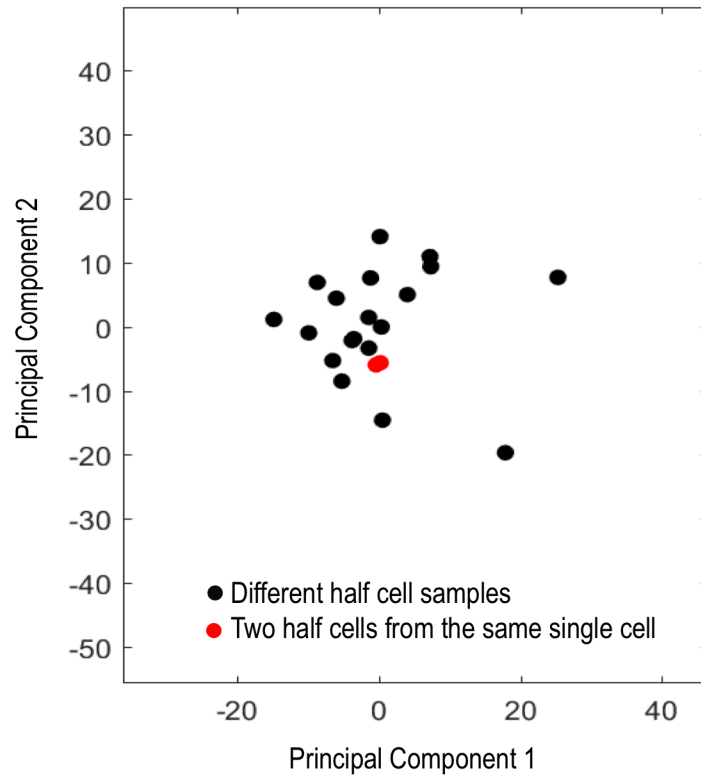


**Supplementary Figure 3. qRT-PCR analysis of miRNA expression in two half-cell samples derived from the same single cells.** Single K562 cells were lysed. Lysate was split with each half of the lysate analyzed for miR-146b-5p expression using qRT-PCR. Each dot reflects a single cell. A total of 30 cells were analyzed. The Scatter plot shows the Ct values (see Methods on single cell qRT-PCR) of the two halves of the same single cells.



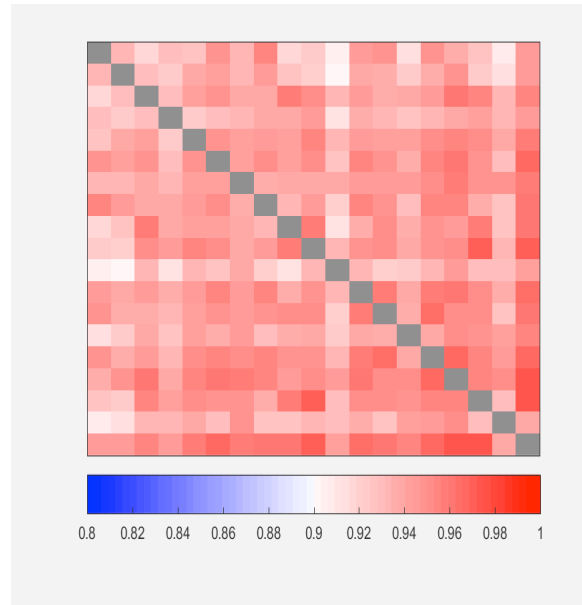


**Supplementary Figure 4. Single-cell vs half-cell miRNA profiles.** Scatter plots showing the comparison of miRNA profile between a single cell and a half cell cross three different types of cell lines. Cell types and  $R^2$  values are indicated. Log2-transformed miRNA expression data are shown (see Methods), with each dot presenting a single miRNA.



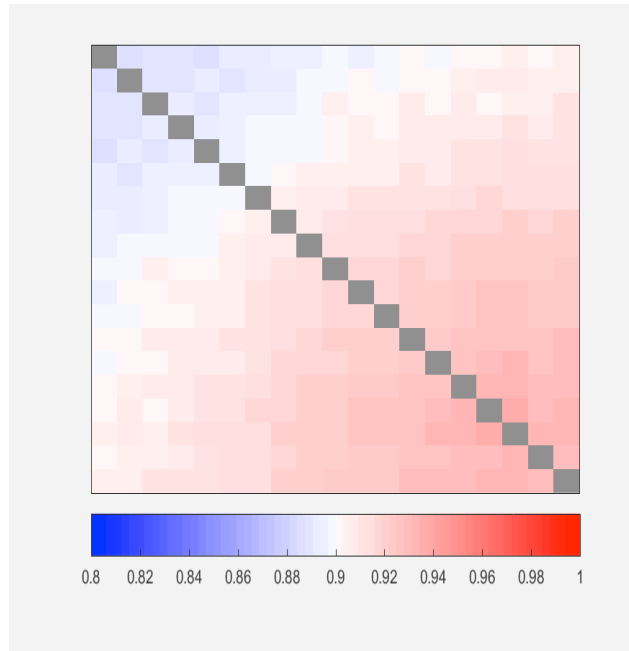
**Supplementary Figure 5. Principal component analysis of half-cell miRNA profiles.**

miRNA expression profiles from 19 different half-cells (black) and two half-cell miRNA profiles from the same single cell (red) were analyzed with principle component analysis, with each dot representing a half cell. Note that the two red dots are very close by.



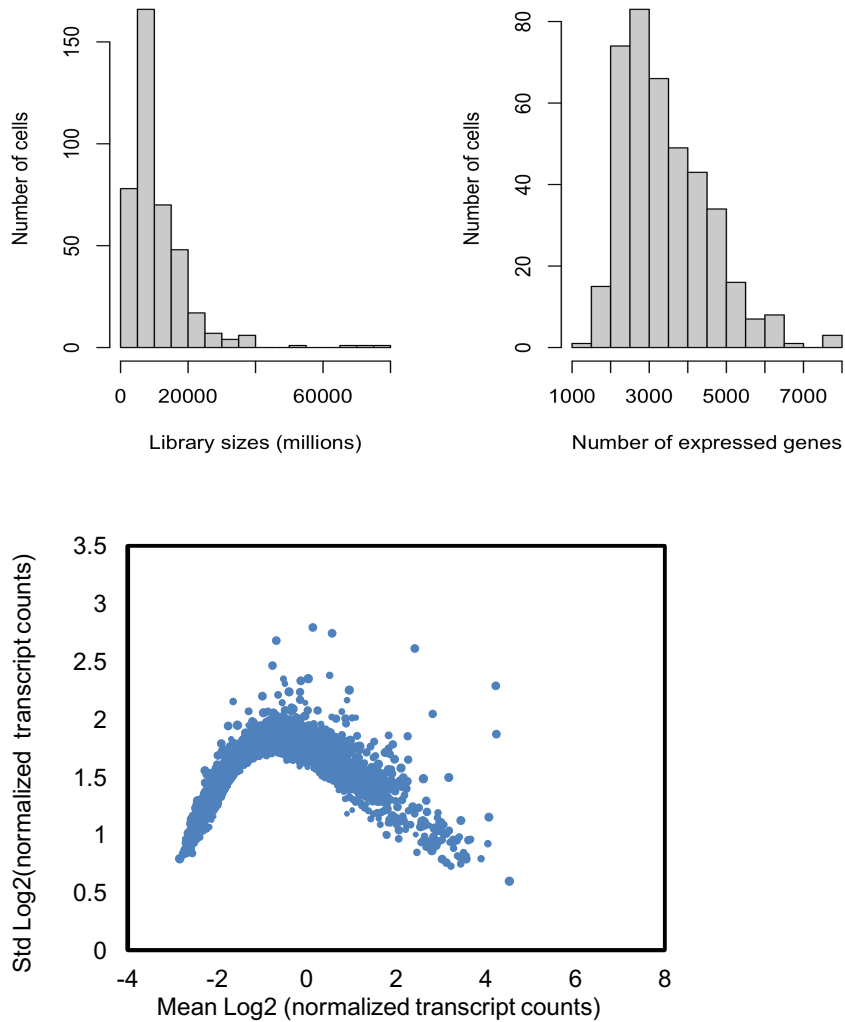
**Supplementary Figure 6. Pair-wise correlation across 19 half-cell miRNA profiles.**

Paired half-cell mRNA and miRNA sequencing was successfully performed for 19 single K562 cells. The heatmap shows the pair-wise correlation matrix of 19 K562 half-cell microRNA data subjected to hierarchical clustering. Color key is shown below the plot as the Pearson Correlation Coefficient  $R$ . A grey color was assigned to designate self-correlations. Overall, there is a good correlation ranging from  $R=0.902$  to  $0.975$  among the group of 19 single cells. The order of cells is the same as in Supplementary Figure 7.



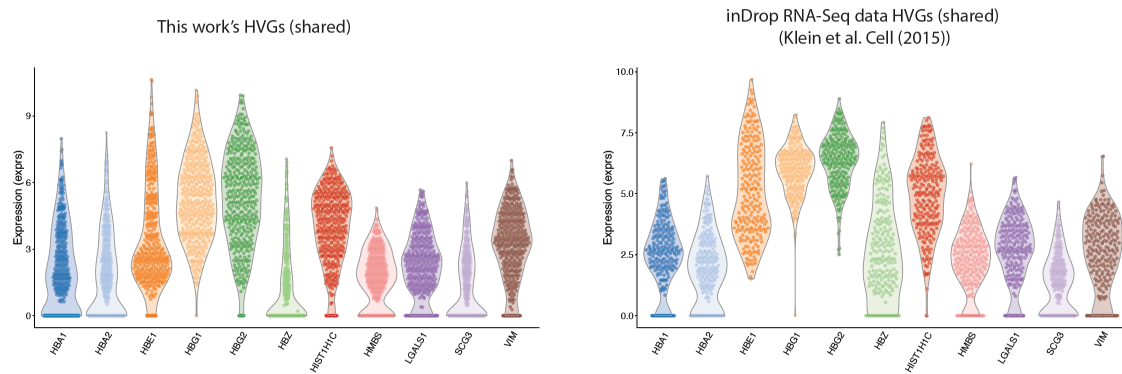
**Supplementary Figure 7. Pair-wise correlation across 19 half-cell mRNA profiles.**

Paired half-cell mRNA and miRNA sequencing was successfully performed for 19 single K562 cells. The heatmap shows the pair-wise correlation matrix of half-cell mRNA data subjected to hierarchical clustering. Color key is shown below the plot as the Pearson Correlation Coefficient R. A grey color was assigned to designate self-correlations. The R value ranges from R=0.884 to 0.937 among the group of 19 single cells. The order of cells is the same as in Supplementary Figure 6.



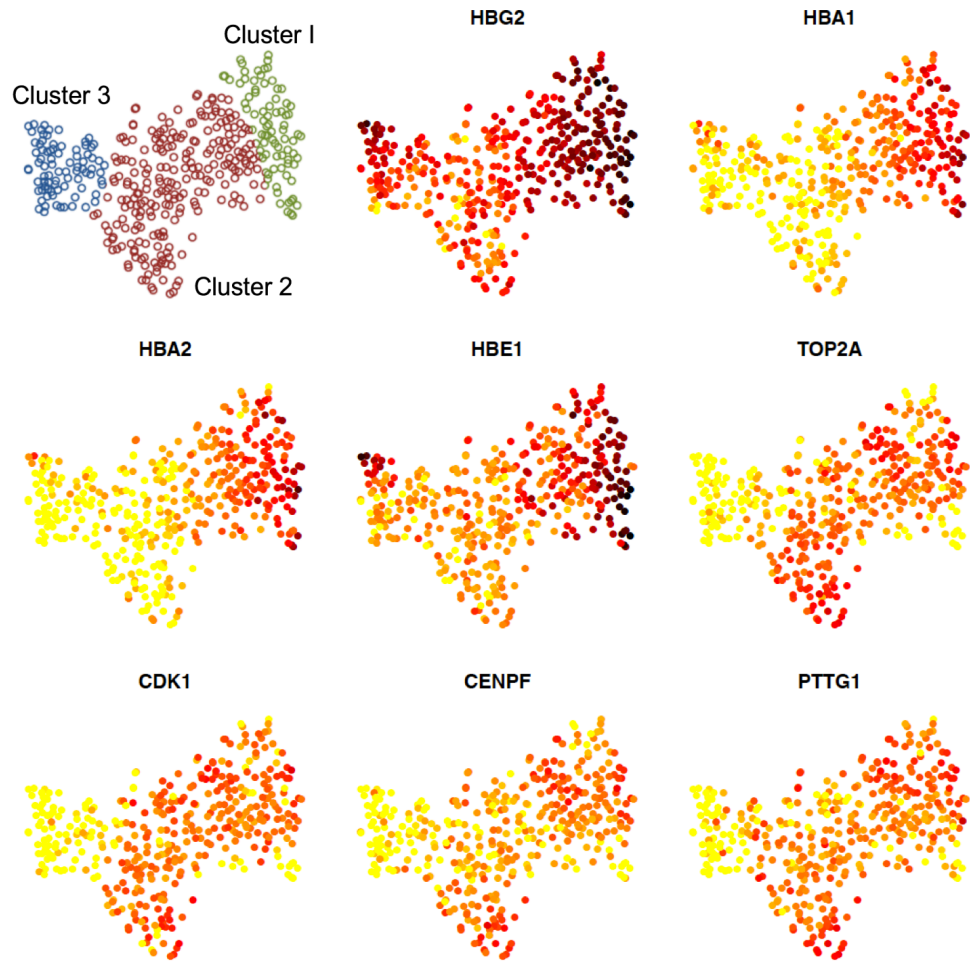
**Supplementary Figure 8. Library quality check for 400 single K562 transcriptome sequencing data.** Single K562 cells were analyzed using our in-house massively parallel single-cell 3'-end RNAseq technology. The quality of the libraries was examined through histograms of single cell RNAseq data as a function of the library size (reflecting total number of mapped reads, top left) or of the number of genes detected (top right). These results are in agreement with literature results (Klein AM, et al., Cell, 161(5), 1187-1120. 2015). The relationship between mean log2 normalized transcript counts and variation is

shown in the bottom figure with each dot representing a single detected gene.  
Normalization procedure is described in Methods.

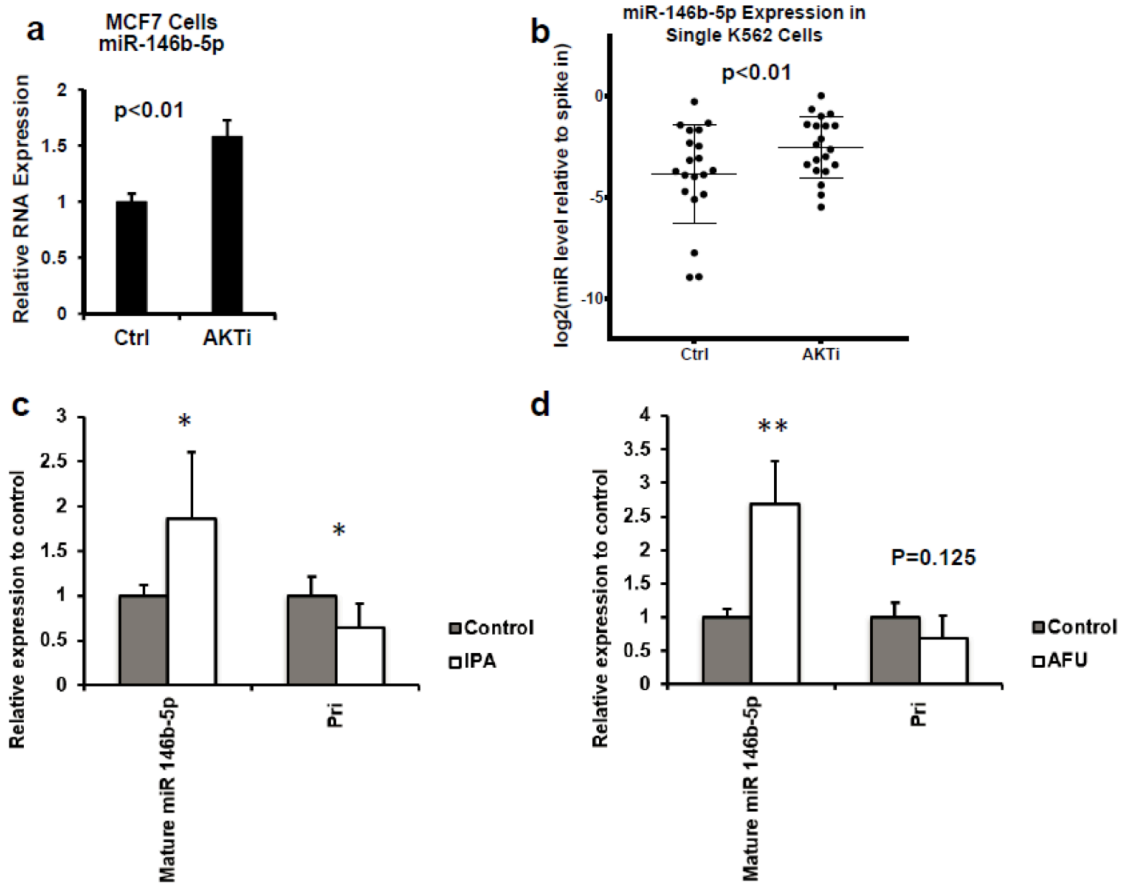


**Supplementary Figure 9. Comparison of K562 single-cell mRNA-seq data between this work and the data generated by InDrop.** Single K562 cells were analyzed using our in-house massively parallel single-cell 3'-end RNAseq technology (this work). The top 20 highly variable genes (HVGs) were determined from our dataset (this work, 400 single cells) and that from the data generated with InDrop published by Klein et al. (right panel, 238 single cells). The two panels show scattered and violin plots for 11 HVGs that were shared between the two datasets. Despite higher number of single cells (N=400) in this work compared to InDrop data (N=238) and the resultant lower depth per cell, the expression levels of these HVGs are consistent between this work and the InDrop K562 data.





**Supplementary Figure 10. tSNE Analysis of 400 single K562 cell mRNA sequencing data.** Single K562 cells were analyzed using our in-house massively parallel single-cell 3'-end RNAseq technology. tSNE analysis was performed to visualize the transcriptomic data from 400 single cells. The result indicates the existence of three clusters, consistent with the heatmap in Figure 3d. Also shown is the distribution of selected highly variable genes in order to visualize the expression of specific genes in major clusters.



**Supplementary Figure 11: The regulation of miR-146b expression by AKT. (a)**

MCF7 cells were treated with vehicle control (Ctrl, DMSO) or AKT inhibitor (AKTi, MK2206) for 24 hours, and the expression of miR-146b-5p was determined by qRT-PCR. N=3 biological replicates. Data are from a representative experiment out of two performed. P<0.01, student's t-test. **(b)** K562 cells were treated with vehicle control (DMSO) or AKTi MM2206 for 24 hours. Single K562 cells were analyzed for miR-146b expression using qRT-PCR. Each dot stands for one cell. Standard deviation in the Ctrl group is 2.41 and in the AKTi group is 1.50. Error bars stand for standard deviation. P<0.01. **(c, d)** K562 cells were treated with two additional AKT inhibitors IPA (c) and AFU (d). The expression of mature miR-146b-5p or primary miR-146b were determined

by qRT-PCR. N=6 biological replicates. Data are from a representative experiment out of two performed. Error bars stand for standard deviation. \*P<0.05; \*\*P<0.01, student's t-test.

## Supplementary Tables

(all supplementary tables are uploaded separately)

**Supplementary Table 1.** List of genes enriched in the first & second cluster for 19 K562 half-cell. Table is sorted by p-values (low to high).

Provided as an excel file: [Supplementary Table 1.xlsx](#)

**Supplementary Table 2.** List of genes enriched in the first & second cluster for 400 K562 single-cell. Table is sorted by p-values (low to high) .

Provided as an excel file: [Supplementary Table 2.xlsx](#)

**Supplementary Table 3.** Gene set enrichment analysis (GSEA) with miR-146b-5p & let-7i-5p

**Supplementary Table 4.** CMAP query with let-7i-5p mRNA signatures

Provided as an excel file: [Supplementary Table 4.xlsx](#)

**Supplementary Table 5.** Summary of small RNAs in 19K562 half-cell mappable to other genomic regions.

Provided as an excel file: [Supplementary Table 5.xlsx](#)

**Supplementary Table 6.** List oligonucleotides and primer sequences, with their respective modifications.

Provided as an excel file: [Supplementary Table 6.xlsx](#)