

RESEARCH

Open Access



Integrative genome-scale analyses reveal post-transcriptional signatures of early human small intestinal development in a directed differentiation organoid model

Yu-Han Hung¹, Meghan Capeling², Jonathan W. Villanueva¹, Matt Kanke¹, Michael T. Shanahan¹, Sha Huang², Rebecca Cubitt¹, Vera D. Rinaldi¹, John C. Schimenti¹, Jason R. Spence^{2,3} and Praveen Sethupathy^{1*}

Abstract

Background MicroRNAs (miRNAs) are important post-transcriptional gene regulators controlling cellular lineage specification and differentiation during embryonic development, including the gastrointestinal system. However, miRNA-mediated regulatory mechanisms involved in early embryonic development of human small intestine (SI) remains underexplored. To explore candidate roles for miRNAs in prenatal SI lineage specification in humans, we used a multi-omic analysis strategy in a directed differentiation model that programs human pluripotent stem cells toward the SI lineage.

Results We leveraged small RNA-seq to define the changing miRNA landscape, and integrated chromatin run-on sequencing (ChRO-seq) and RNA-seq to define genes subject to significant post-transcriptional regulation across the different stages of differentiation. Small RNA-seq profiling revealed temporal dynamics of miRNA signatures across different developmental events of the model, including definitive endoderm formation, SI lineage specification and SI regional patterning. Our multi-omic, integrative analyses showed further that the elevation of miR-182 and reduction of miR-375 are key events during SI lineage specification. We demonstrated that loss of miR-182 leads to an increase in the foregut master marker *SOX2*. We also used single-cell analyses in murine adult intestinal crypts to support a life-long role for miR-375 in the regulation of *Zfp36l2*. Finally, we uncovered opposing roles of SMAD4 and WNT signaling in regulating miR-375 expression during SI lineage specification. Beyond the mechanisms highlighted in this study, we also present a web-based application for exploration of post-transcriptional regulation and miRNA-mediated control in the context of early human SI development.

Conclusion The present study uncovers a novel facet of miRNAs in regulating prenatal SI development. We leveraged multi-omic, systems biology approaches to discover candidate miRNA regulators associated with early SI developmental events in a human organoid model. In this study, we highlighted miRNA-mediated post-

*Correspondence:
Praveen Sethupathy
pr46@cornell.edu

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

transcriptional regulation relevant to the event of SI lineage specification. The candidate miRNA regulators that we identified for the other stages of SI development also warrant detailed characterization in the future.

Keywords Micro-RNA, Post-transcriptional regulation, Intestine, Development, Functional genomics

Introduction

MicroRNAs (miRNAs) are small, non-coding RNA molecules (~22 nucleotides) that function as post-transcriptional gene suppressors in various biological processes including development [1, 2]. The biogenesis of mature miRNAs occurs through a multi-step process. Longer primary miRNA transcripts are usually transcribed by RNA polymerase II and subsequently processed by enzymes into mature, functional miRNAs. Mature miRNAs confer gene silencing through complementary base pairing with mRNA target sites (typically in 3' untranslated region of transcripts), leading to destabilization and degradation of the target mRNAs [3]. As demonstrated by several seminal studies, miRNAs play an essential role in controlling dynamics of gene expression during mammalian development. Murine embryonic lethality was observed when disrupting miRNA maturation [4, 5] or function [6] in vivo. In addition, different fetal tissues were shown to have distinct miRNA signatures and temporal dynamics in both mouse and humans [7], suggesting that miRNAs participate in lineage specification during pre-natal development.

This study sought to identify candidate miRNA regulators of early small intestinal (SI) development in humans. While specific miRNAs have been linked to the regulation of embryonic development of various organs such as the heart [8–11], brain [12–15], pancreatic islets [16–18] and skeletal muscle [19–21], the role of miRNAs in SI development has not been as closely studied. Mice with intestinal epithelial-specific knockout of Dicer (a key miRNA processing enzyme) were reported to exhibit disrupted gut epithelial architecture and function [11], conveying the importance of miRNAs to proper SI development. However, the roles of specific miRNAs in the context of SI development are unexplored in humans. Directed differentiation from human pluripotent stem cells (hPSCs) to human intestinal organoids (HIOs) represents a state-of-art model to study human SI development [22–24]. To generate HIOs, hPSCs are differentiated into definitive endoderm patterned into mid/hindgut, which subsequently leads to the formation of 3-dimensional (3D) spheroids fated to the SI lineage and eventually mature HIOs. These SI spheroids can be regionally patterned (duodenum vs. ileum) by controlling the length of time that they are exposed to FGF and WNT signaling [23]. We have previously applied chromatin run-on sequencing (ChRO-seq) [25] in this model system to successfully profile enhancer dynamics and define transcriptional programs relevant to SI developmental events in human

[26]. Here we aimed to define post-transcriptional programs that underlie human SI development and identify candidate master miRNA regulators in this context.

To that end, we developed a multi-omic analysis strategy that integrates ChRO-seq, RNA-seq and small RNA-seq (smRNA-seq) generated from the hPSC-HIO model. Specifically, we first used smRNA-seq to characterize the dynamics of the miRNA landscape across different cellular stage transitions of SI differentiation. Next, we integrated ChRO-seq (which measures nascent gene transcription) and RNA-seq (which measures steady-state gene expression) to tease apart the contribution of transcriptional and post-transcriptional (PT) regulation for each gene individually. We made quantitative measurements of PT regulation and then identified the miRNAs that likely mediate the most PT regulation during endoderm formation, SI lineage specification, and SI regional patterning. Our analyses pointed to the gain of miR-182 (and the consequent loss of *SOX2*) and the loss of miR-375 (and the consequent gain of *ARL4C* and *ZFP36L2*) as key events during SI lineage specification. Further analyses showed that whereas miR-182 levels are likely controlled by PT mechanisms, miR-375 is regulated robustly by transcription during human SI lineage specification. Follow-up functional studies in the hPSC-HIO model confirmed that miR-182 suppresses the foregut driver gene *SOX2*. In addition, single cell RNA-seq of adult miR-375 knockout mice supported a continued, post-developmental role of miR-375 in the regulation of *ARL4C* and *ZFP36L2*. In this study, by using an integrative multi-omics approach, we offer the first look into PT regulation during early human SI development and provide novel insights into the function of miRNAs in this context. We also provide the research community with a searchable web server (<https://sethupathy-lab.shinyapps.io/ProHumID/>) to explore the omics data we generated from the hPSC-HIO model.

Results

SmRNA-seq shows a changing miRNA landscape during human SI lineage commitment

To study early SI development in humans, we carried out directed differentiation from human embryonic stem cells (hESCs) toward the SI lineage as previously described [22, 23, 27] (Fig. 1A). Cells from the stage of hESC, definitive endoderm (DE), duodenal (Duo; proximal SI) spheroids, and ileal (Ile; distal SI) spheroids were collected to study molecular changes (Fig. 1A). The important markers of cell type identity and regional

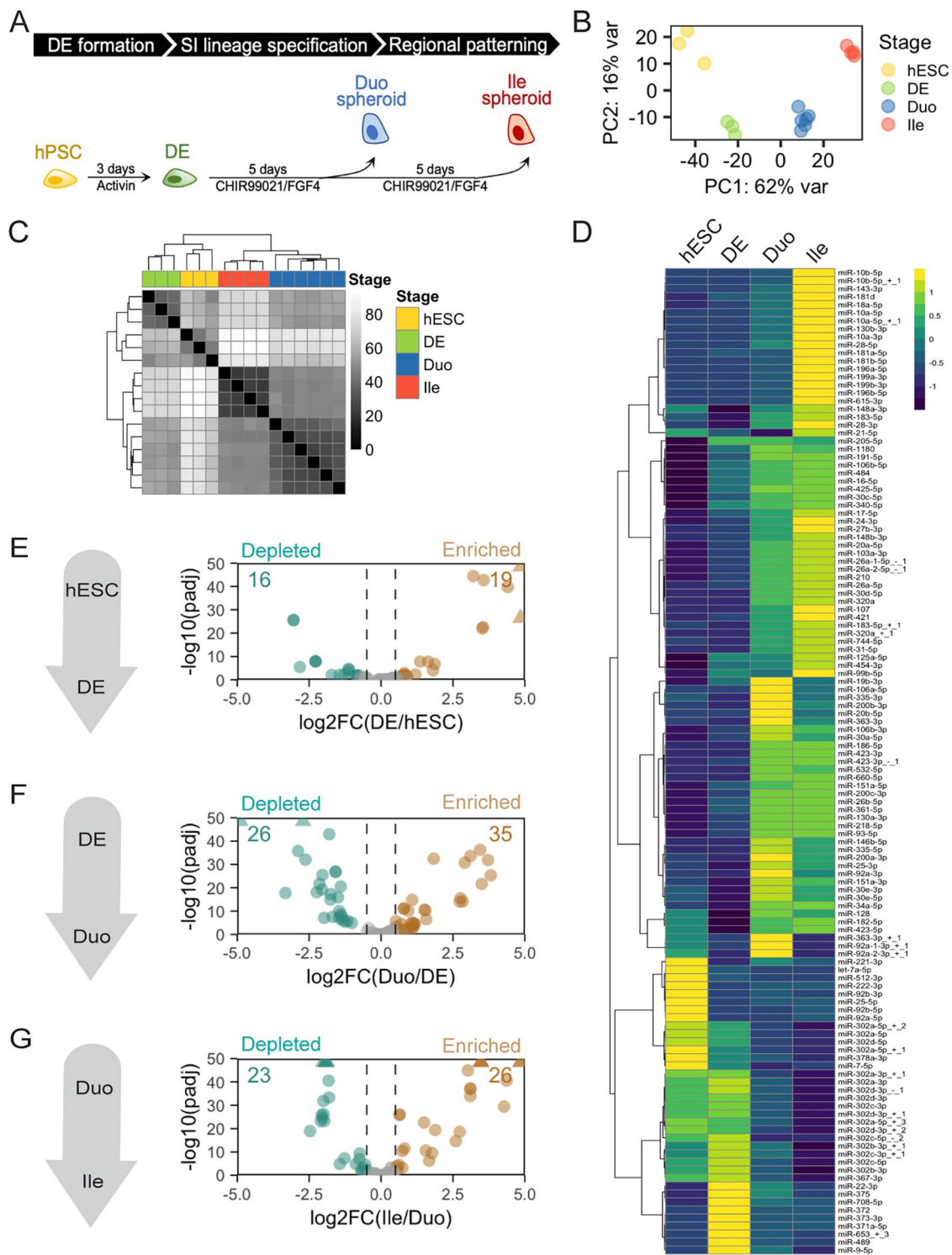


Fig. 1 smRNA-seq revealed changing miRNA signatures during early SI development in a human organoid model. **(A)** The stage transitions of the hPSC-HIO model recapitulate early events of human SI development: DE formation, SI lineage formation and SI regional patterning. **(B)** PCA of smRNA-seq miRNA profiles of hESC, DE, Duo and Ile spheroids. **(C)** Hierarchical clustering analysis (Euclidean distance) of miRNA profiles of hESC, DE, Duo and Ile spheroids. **(D)** Heatmap showing changing expression of miRNAs across hESC, DE, Duo and Ile spheroids. Only showing those miRNAs with RPMMM > 1000 in at least one stage. **(E-G)** Volcano plots showing differentially expressed miRNAs in the event of DE formation **(E)**, SI lineage specification **(F)**, and SI regional patterning **(G)**. Numbers in bronze and pine are for up- and down-regulated miRNAs in the indicated comparison, respectively (RPMMM > 1000 in at least in one stage, log₂FC > 0.5, adjusted P < 0.05 by Wald test; DESeq2). RPMMM, reads per million mapped to miRNAs. smRNA-seq: human embryonic stem cell (hESC), n = 3; definitive endoderm (DE), n = 3; Duo spheroid (Duo), n = 6; Ile spheroid (Ile), n = 4

patterning of these directed differentiation samples have been validated in our previous study [26]. In the present study, to characterize the dynamics of miRNA expression, we performed smRNA-seq across these stages. Principal component analysis (PCA) and unsupervised hierarchical clustering demonstrated distinct miRNA profiles at each stage (Fig. 1B and C). We observed that miRNAs (RPMMM > 1000 in at least one stage) exhibit changing expression patterns over the course of directed differentiation (Fig. 1D). Consistent with previous microarray and smRNA-seq studies comparing hESC and DE [28–30], we found several expected miRNAs enriched in hESCs (e.g., miR-512b, miR-222, miR-7) or in DE (e.g., miR-708, miR-375, miR-489, miR-9) (Fig. 1D). In addition, previous reports showed that the miR-302/367 family is critical for maintaining stem cell pluripotency [31–33] and that miR-302 isoforms may have functional roles during differentiation of hPSCs to DE [28]. Our sequencing data indeed confirmed enrichment of miR-302 family members and isoforms in the hESC and/or DE stages (Fig. 1D).

By performing differential expression analysis (adjusted $P < 0.05$ and absolute \log_2 foldchange > 0.5 and by DESeq2 Wald test, RPMMM > 1000 at least in one stage of a two-group comparison), we identified 35 miRNAs altered during DE formation (16 down and 19 up; Fig. 1E, **Supplementary Data 3**), 61 miRNAs altered during SI lineage specification (26 down and 35 up; Fig. 1F, **Supplementary Data 3**), and 49 miRNAs altered during distal SI patterning (23 down and 26 up; Fig. 1G, **Supplementary Data 3**). Notably, 3 miRNAs (miR-798, miR-375 and miR-27b) were found to be significantly altered in all the stage transitions during directed differentiation process (**Supplementary Fig. 2**) and may indicate their important roles throughout the SI developmental process. Overall, the dramatically changing miRNA landscape during directed differentiation towards SI lineage is suggestive of active miRNA-mediated gene regulation during early SI development.

Integration of ChRO-seq and RNA-seq defines post-transcriptionally regulated genes during major stages of early human SI development

In theory, the depletion and enrichment of miRNAs during a stage transition should attenuate and enhance, respectively, post-transcriptional suppression of target genes. Therefore, we sought to identify genes that are subject to significant post-transcriptional (PT) regulation during each stage transition of the hPSC-HIO model by generating and integrating ChRO-seq and RNA-seq data. ChRO-seq provides high-resolution mapping of active RNA polymerases on chromatin and directly quantifies nascent transcriptional activity of genes prior to PT regulation (e.g., miRNA) (Fig. 2A).

RNA-seq, on the other hand, determines steady-state gene expression levels, which is the result of the net effect of nascent transcriptional activity and PT regulation of genes (Fig. 2A). Therefore, discordance between the ChRO-seq and RNA-seq readouts for a given gene indicates that the gene is likely subject to PT regulation (Fig. 2A). We performed two-factor DESeq2 analysis to identify genes that exhibit significant, discordant changes at the level of transcription (ChRO-seq) versus steady-state expression (RNA-seq) in a given stage transition (Fig. 2B; **Methods**). For each stage transition, we identified genes for which PT suppression is lost (ChRO-seq TPM > 20, RNA-seq base mean > 100, adjusted $P < 0.05$ and $\log_2FC > 0.5$ in the interaction term comparing RNA-seq vs. ChRO-seq changes during a given stage transition by Wald test DESeq2, Fig. 2C–E) and these genes were denoted hereafter as “stable genes” (see gene lists in **Supplementary Data 4**). We also identified genes for which PT suppression is gained (ChRO-seq TPM > 20, RNA-seq base mean > 100, adjusted $P < 0.05$ and $\log_2FC < -0.5$ in the interaction term comparing RNA-seq vs. ChRO-seq changes during a given stage transition by Wald test DESeq2; Fig. 2C–E) and these genes were denoted hereafter as “unstable genes” (see gene lists in **Supplementary Data 4**). This analysis revealed a total of 825 stable and 612 unstable genes during DE formation from hESC (Fig. 2C); 1972 stable and 1756 unstable genes during SI lineage specification from DE to Duo (Fig. 2D); and 505 stable and 609 unstable genes during distal SI patterning from Duo to Ile (Fig. 2E). We further pruned these gene lists by applying a specific stringent criterion: very little change in nascent transcriptional activity (adjusted $P > 0.1$ in DESeq2 Wald test assessing ChRO-seq change during a stage transition) but significant alterations in steady-state expression levels (adjusted $P < 0.05$ and $\log_2FC > 0.5$ or < -0.5 in DESeq2 Wald test assessing RNA-seq change during a stage transition). This stringent set signifies genes (hereafter as “stringent stable” or “stringent unstable”) that are regulated primarily through post-transcriptional mechanisms during one or more stage transitions (Fig. 2B, see gene lists in **Supplementary Data 4**). This analysis revealed 236 stringent stable and 208 stringent unstable genes during DE formation (Fig. 2F); 292 stringent stable and 324 stringent unstable genes during SI lineage specification (Fig. 2G); and 139 stringent stable and 139 stringent unstable genes during distal SI patterning (Fig. 2H). For the rest of the study, we focused mostly on the initial event of SI lineage specification: the transition from DE to Duo, during which we observed the most remarkable changes in the miRNA landscape (Fig. 1B) as well as the greatest number of stable/unstable genes (which is indicative of the largest amount of PT regulation) (Fig. 2D).

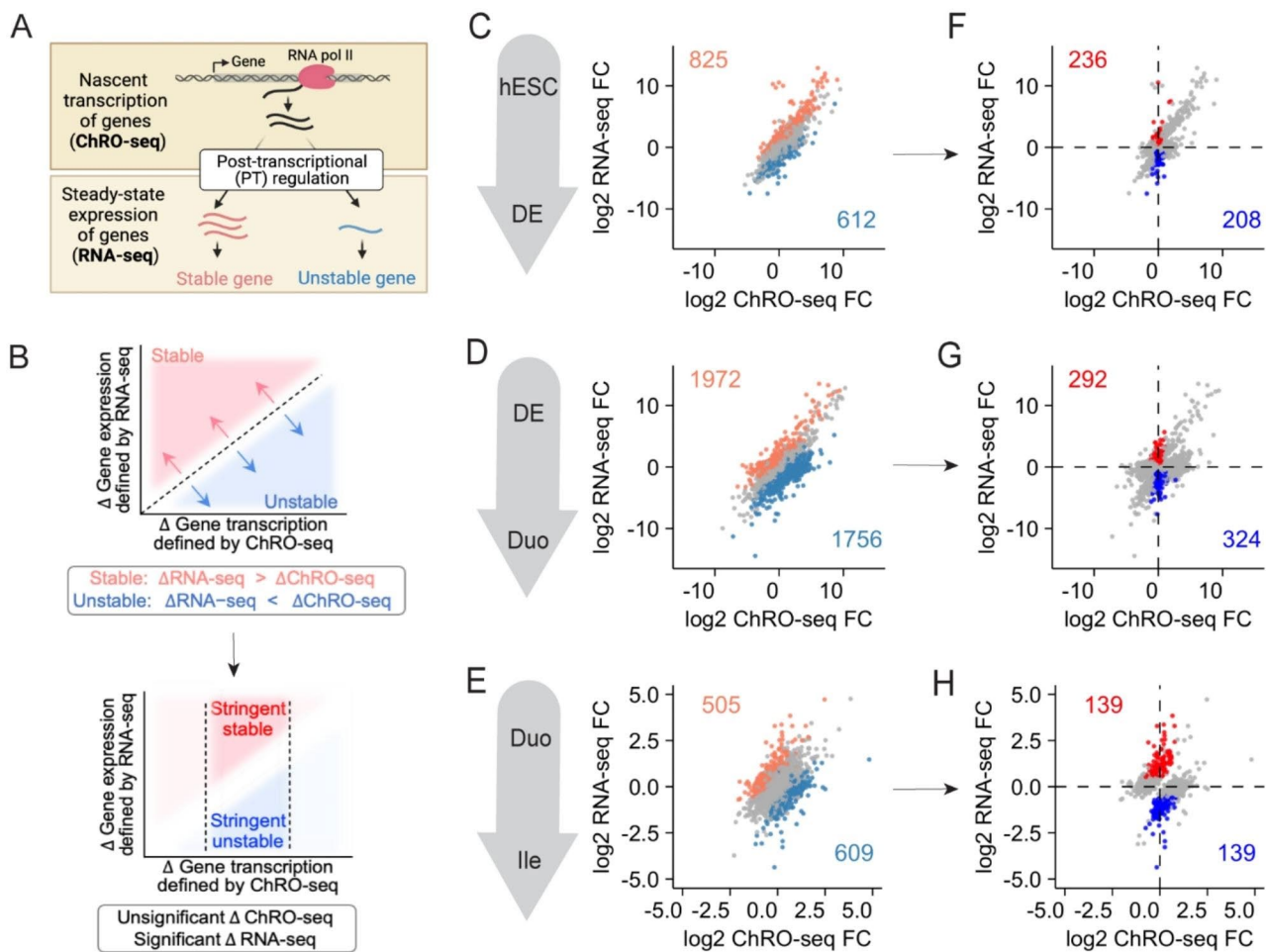


Fig. 2 Integrative analysis of ChRO-seq and RNA-seq defined genes that were post-transcriptionally (PT) regulated during early SI development in a human organoid model. **(A)** Illustration showing the strategy of integrating ChRO-seq and RNA-seq to identify genes that are subject to PT regulation (PT stable or unstable) during a stage transition. **(B)** Top: Genes with discordant RNA-seq and ChRO-seq changes were detected by two-factor DESeq2 analysis (adjusted $P < 0.05$ and absolute $\log_2FC > 0.5$ in the interaction term comparing RNA-seq vs. ChRO-seq changes during a given stage transition by Wald test DESeq2; ChRO-seq TPM > 20 ; RNA-seq normalized counts > 100). They were defined as stable ($\log_2FC > 0.5$) and unstable ($\log_2FC < -0.5$) genes. Bottom: Additional filtering was applied to stable and unstable genes (adjusted $P > 0.1$ in DESeq2 Wald test assessing ChRO-seq changes; adjusted $P < 0.05$ and absolute $\log_2FC > 0.5$ in DESeq2 Wald test assessing RNA-seq changes in a given stage transition), yielding genes that were “primarily” subject to PT regulation during a stage transition. They were defined as stringent stable and unstable genes. **(C-E)** Two-factor DESeq2 analysis defined PT regulated genes during the event of DE formation **(C)**, SI lineage specification **(D)** and SI regional patterning **(E)**. In **(C-E)**, genes in coral and teal are defined as stable and unstable genes, respectively. **(F-H)** Additional filtering revealed genes that were primarily subject to PT regulation during the event of DE formation **(F)**, SI lineage specification **(G)** and SI regional patterning **(H)**. In **(F-H)**, genes in blue and red are defined as stringent stable and unstable genes, respectively. ChRO-seq: human embryonic stem cell (hESC), $n = 3$; definitive endoderm (DE), $n = 4$; Duo spheroid (Duo), $n = 3$; Ile spheroid (Ile), $n = 3$. RNA-seq: hESC, $n = 2$; DE, $n = 3$; Duo, $n = 6$; Ile, $n = 4$

Elevated activity of miR-182 leads to post-transcriptional suppression of SOX2 during SI lineage specification

A miRNA can serve as a master regulator of gene expression in various contexts by targeting key regulators atop the regulatory hierarchy, such as transcription factors (TFs). The stage transition from DE to Duo spheroids is driven by the induction of CDX2, which is required for SI lineage specification and is therefore known as the master TF of intestinal identity [34, 35]. Concomitant with acquisition of SI lineage specification is the suppression of DE associated TFs (e.g., SOX17, GATA6,

EOMES, LHX1) [36–40], and the suppression of SOX2, which is essential for foregut identity [41, 42]. During SI lineage specification (DE to Duo), we observed concordant RNA-seq and ChRO-seq changes in *CDX2*, *SOX17* and *GATA6*, indicating that the steady-state expression changes in these genes are driven primarily by the changes in nascent transcriptional activity (Fig. 3A). In contrast, the ChRO-seq and RNA-seq changes are discordant for *EOMES*, *LHX1* and *SOX2*; specifically, there are no remarkable alterations in nascent transcription at these gene loci, and yet the steady-state mRNA levels are

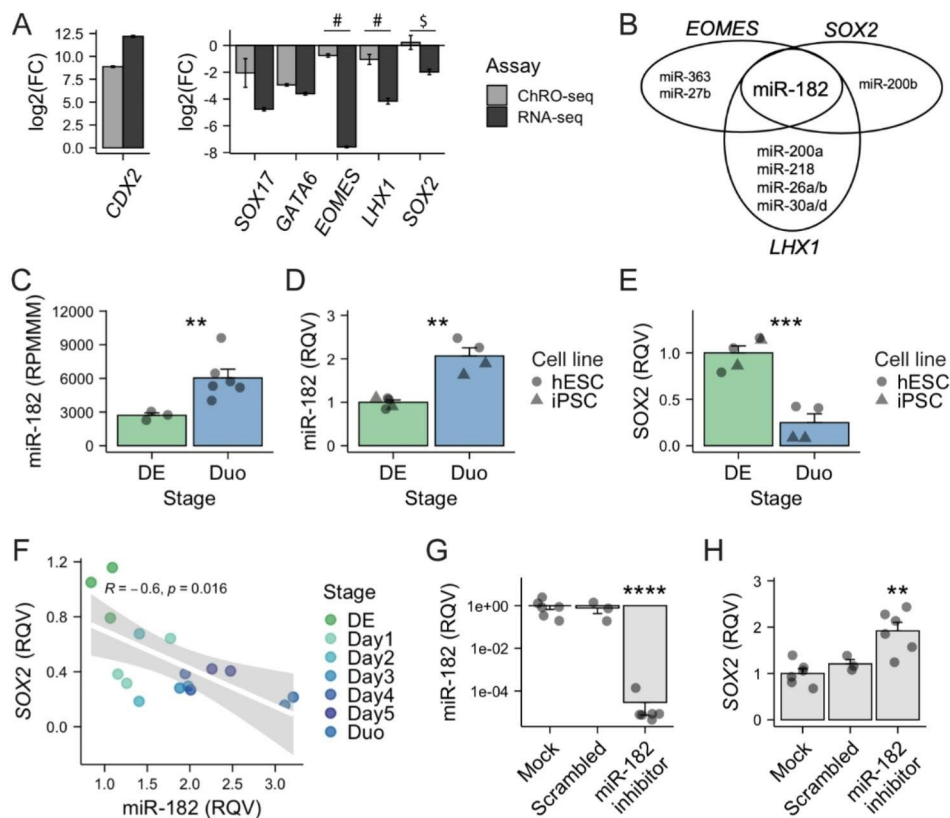


Fig. 3 Elevated activity of miR-182 promotes SI lineage formation in part through post-transcriptionally suppressing *SOX2*. **(A)** Nascent transcriptional (ChRO-seq) and steady-state expressional (RNA-seq) changes of transcription factors critical for SI lineage specification (transition from DE to Duo). # defined as unstable genes in Fig. 2D. \$ defined as stringent unstable genes in Fig. 2G. **(B)** miRNAs that are enriched during SI lineage specification (adjusted $P < 0.05$ and absolute \log_2 foldchange > 0.5 and by DESeq2 Wald test, RPMMM > 1000 at least in one stage of a two-group comparison) and have conserved target sites on *EOMES*, *LHX1* or *SOX2* (TargetScan). **(C)** smRNA-seq showing a significant upregulation of miR-182 during SI lineage specification. ** Adjusted $P < 0.01$ by DESeq2 Wald test in smRNA-seq. DE, $n = 3$; Duo, $n = 6$. **(D)** qPCR showing a significant upregulation of miR-182 during SI lineage specification in both hESC and iPSC cell line. ** $P < 0.01$ by two-tailed t-test. $N = 2-3$ samples per cell line per stage. **(E)** qPCR showing a significant downregulation of *SOX2* during SI lineage specification in both hESC and iPSC cell line. *** $P < 0.001$ by two-tailed t-test. $N = 2-3$ samples per cell line per stage. **(F)** Reverse Spearman correlation between miR-182 and *SOX2* during 5-day window of DE cells receiving SI induction cocktail. $N = 2-3$ samples per day. **(G)** qPCR of miR-182 levels in iPSC directed differentiated cells receiving treatment of mock, scrambled control or miR-182 inhibitors. **** $P < 0.0001$ by two-tailed t-test. **(H)** qPCR of *SOX2* in iPSC directed differentiated cells receiving treatment of mock, scrambled control or miR-182 inhibitors. ** $P < 0.01$ by two-tailed t-test. $N = 3-6$ per condition across 2 independent experiments. DE, definitive endoderm. Duo, duodenal spheroid. RPMMM, reads per million mapped to miRNAs. RQV, relative quantitative value

significantly reduced. This finding is indicative of extensive post-transcriptional suppression (Fig. 3A). Indeed, *EOMES* and *LHX1* were identified as unstable genes (Fig. 2C) and *SOX2* as a stringent unstable gene (Fig. 2F) during SI lineage specification in our earlier analysis.

To explore candidate miRNAs that mediate the post-transcriptional suppression of *EOMES*, *LHX1* and *SOX2*, we identified miRNAs that are elevated during this stage transition (Fig. 1E) and also have highly conserved target sites in these genes (TargetScan). This analysis revealed only one miRNA: miR-182 (Fig. 3B). In addition, among these three TFs, *SOX2* is the only one that satisfies the criteria for “stringent unstable gene” (Fig. 2G). We next sought to support miR-182-mediated suppression of *SOX2* during SI lineage specification. First, we validated the up-regulation of miR-182 and suppression of *SOX2*

during the transition from DE to Duo spheroid in two independent experiments using two different hPSC lines (Fig. 3C-E). Second, we demonstrated a strong negative correlation between miR-182 and *SOX2* over the course of the 5-day SI fate specification process (Spearman correlation $R = -0.6$, $p = 0.016$; Fig. 3F).

Next, to test miR-182 regulation of *SOX2* during SI lineage specification, we performed miR-182 loss-of-function studies during the transition from DE to Duo spheroids. Specifically, transfection of miR-182 inhibitors (500 nM) was performed in DE cells on day 3 of SI induction and the cells were collected on day 4, the earliest time point at which Duo spheroids are formed in the model. As determined by qPCR, the treatment of miR-182 inhibitors led to a robust suppression of miR-182 compared to the mock and scrambled control groups

($P < 0.0001$ by two-tailed t-test; Fig. 3G). Consistent with our multi-omics prediction suggesting miR-182 regulation of *SOX2*, the results showed a significant elevation of *SOX2* upon treatment of miR-182 inhibitors compared to both mock and scrambled control groups ($P < 0.01$ by two-tailed t-test; Fig. 3H). *SOX2* has been previously reported to suppress *CDX2* through an induction of *SOX21* [43]. However, treatment with miR-182 inhibitors for only 24-hr did not lead to significant changes in *CDX2* or *SOX21* expression in this experiment (Supplementary Fig. 3). In sum, the data indicates that the up-regulation of miR-182 is one component of a larger regulatory network during SI lineage induction that controls *SOX2* levels and activity.

Stable genes during SI lineage specification are strongly associated with the depletion of miR-375

To determine which if any miRNAs are likely most responsible for post-transcriptional regulation during SI lineage specification, we leveraged our previously developed tool called miRhub [44] that determines enrichment of miRNA target sites in a given gene set. We found that stringent unstable genes (defined in Fig. 2G) during the stage transition from DE to Duo are significantly

enriched for predicted target sites of six miRNAs that are significantly up-regulated during the same transition (miR-24, miR-34a, miR-423, miR-26a, miR-30d, miR-744) ($P < 0.05$ by miRhub analysis; Supplementary Fig. 4B). We also found that stringent stable genes (Fig. 2G) during the same stage transition are significantly enriched for predicted target sites of only two miRNAs that exhibit reduced expression (miR-375 and miR-302a, $P < 0.05$ by miRhub analysis) (Fig. 4A). Similar analyses were also performed for the other stage transitions in the hPSC-HIO model (Supplementary Fig. 1).

Among the stable genes during the transition from DE to Duo, 119 have validated or predicted target sites for miR-375 (TargetScan; Fig. 4B; see gene list in Supplementary Data 1). Less than half of these ($n = 46$) have at least one miR-375 target site that is conserved across multiple species (TargetScan; Fig. 4B). Two of these genes namely *ARL4C* and *ZFP36L2*, have particularly high affinity predicted binding sites for miR-375 (Fig. 4B). Specifically, *ARL4C* harbors an 8-mer site (Watson–Crick match to miRNA positions 2–8 followed by ‘A’), which is thought to be a high affinity match for miR-375 (based on TargetScan prediction). *ZFP36L2* has two highly conserved, closely spaced (33 bases apart) miR-375 target

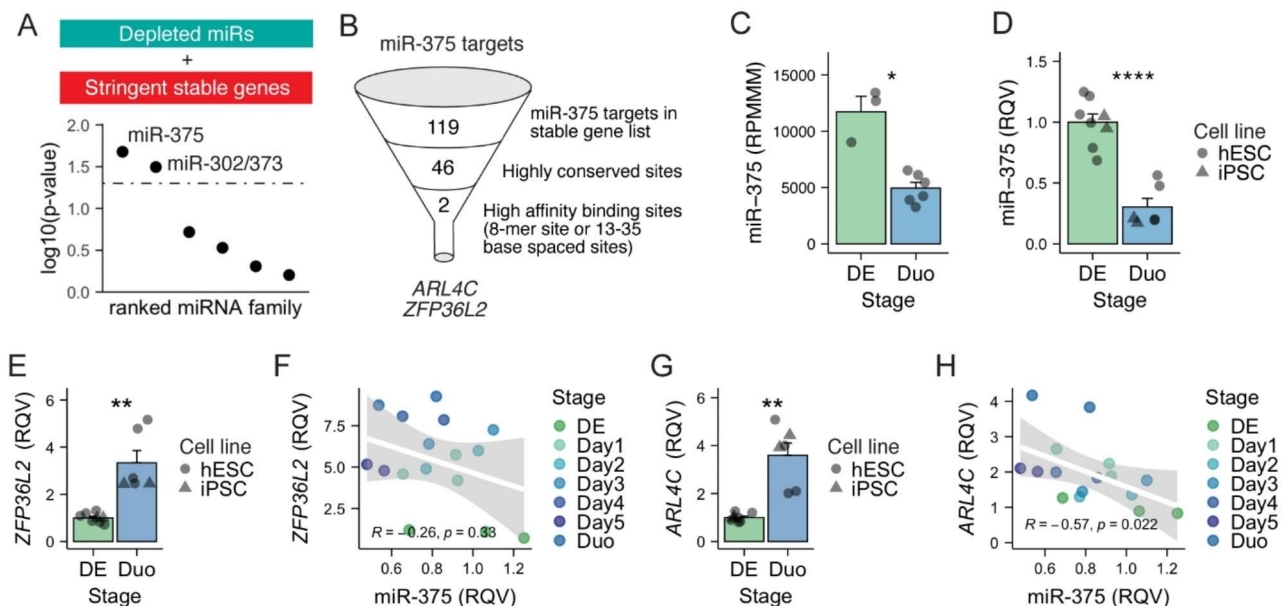


Fig. 4 Post-transcriptionally stable genes during SI lineage specification are strongly associated with the depletion of miR-375. **(A)** MiRhub analysis suggested that genes that stabilized during SI lineage specification were enriched in target sites for two depleted miRNAs, miR-375 and miR-302 ($P < 0.05$ by miRhub analysis; see Methods). **(B)** miR-375 target genes ($n = 119$; TargetScan) among stable genes in SI lineage specification. **(C)** smRNA-seq showing a significant reduction of miR-375 during SI lineage specification. Adjusted $P < 0.05$ by DESeq2 Wald test. DE, $n = 3$; Duo, $n = 6$. **(D)** qPCR showing a significant reduction of miR-375 during SI lineage specification in both hESC and iPSC cell line. **** $P < 0.0001$ by two-tailed t test. $N = 2-4$ samples per cell line per stage. **(E)** qPCR showing a significant downregulation of *ZFP36L2* during SI lineage specification in both hESC and iPSC cell line. ** $P < 0.01$ by two-tailed t test. $N = 2-4$ samples per cell line per stage. **(F)** Modest reverse Spearman correlation between miR-375 and *ZFP36L2* during 5-day window of DE cells receiving SI induction cocktail. $N = 2-3$ samples per day. **(G)** qPCR showing a significant downregulation of *ARL4C* during SI lineage specification in both hESC and iPSC cell line. $P < 0.01$ by two-tailed t test. $N = 2-4$ samples per cell line per stage. **(H)** Significant reverse Spearman correlation between miR-375 and *ARL4C* during 5-day window of DE cells receiving SI induction cocktail. $N = 2-3$ samples per day. DE, definitive endoderm. Duo, duodenal spheroid. RPMMM, reads per million mapped to miRNAs. RQV, relative quantitative value

sites, a scenario that typically leads to a synergistic suppression effect [45, 46]. To validate our observations in the sequencing studies, we performed qPCR experiments and demonstrated that the reduction of miR-375 (Fig. 4C-D) as well as the up-regulation of *ZFP36L2* and *ARL4C* during the transition from DE to Duo (Fig. 4E and G) are reproducible across independent directed differentiation experiments with two different hPSC lines. In addition, our data showed that miR-375 exhibits a modest negative correlation with *ZFP36L2* (Spearman correlation $R = -0.26$, $p = 0.33$; Fig. 4F) and a strong negative correlation with *ARL4C* (Spearman correlation $R = -0.57$, $p = 0.022$; Fig. 4H) during the 5-day SI fate specification process.

We next sought to interrogate miR-375-mediated regulation of *ZFP36L2* and *ARL4C*. It is important to note that *ZFP36L2* and *ARL4C* harbor conserved target sites not only for miR-375 but also for miR-302, both of which are depleted during SI lineage specification and identified as strong candidate miRNA regulators of genes that lose PT suppression during SI lineage specification (Fig. 4A). To assess the regulatory impact of miR-375 apart from miR-302, we turned to adult mouse SI tissues in which miR-302 is absent but miR-375 remains robustly expressed [47].

Single cell transcriptomics reveals a post-developmental role for miR-375 in regulating *Zfp36l2* and *Arl4c* in adult mouse SI intestinal stem cells

[43] First we generated mice with whole body knock-out of miR-375 using CRISPR/Cas9 technology (375KO mice). We then performed single cell RNA-seq of crypts isolated from 375KO and wild-type mice. This experiment captured a total of 5798 cells from SI epithelial tissues ($n = 2591$ cells from wild-type; $n = 3207$ cells from 375KO) after quality control filtering (Fig. 5A-B; **Supplementary Fig. 5**). All major intestinal epithelial cell types were recovered from both wild-type and 375KO tissues (**Supplementary Fig. 5**). We observed that *Zfp36l2* is robustly expressed across different cell types present in adult SI epithelium and especially enriched in progenitor and stem cell populations (Fig. 5C-D). We performed differential expression analysis in 375KO vs. wild-type for each cell type, which demonstrated that the absence of miR-375 leads to the significantly upregulation of *Zfp36l2* in intestinal stem cells (Adjusted P value = 0.000279, $\log_2FC = 0.346$ by differential expression analysis using MAST method; Fig. 5E-F). Although *Arl4c* is very lowly detected in our single cell RNA-seq data, we performed pseudo-bulk differential expression analysis and found that *Arl4c* is elevated in 375KO relative to wild-type (1.42 fold increase, $P = 0.37$ by two-tailed t-test; **Supplementary Fig. 6**). Collectively, this study provides in vivo

evidence to support a life-long role for miR-375 in regulating *Zfp36l2*.

SMAD4 and WNT signaling exert opposite effects on regulating miR-375 expression

Because our data suggests that up-regulation of miR-182 and reduction of miR-375 are critical events during SI lineage specification (Figs. 3 and 4), we were motivated to understand how the expression of these miRNAs are regulated. We found that the upregulation of miR-182 during SI lineage specification is likely driven by stabilization of miR-182 transcripts, as transcriptional activity of the miR-182 locus was unchanged (adjusted $P = 0.23$, $\log_2FC = 0.78$ by DESeq2 Wald test in ChRO-seq) whereas smRNA-seq showed a robust increase in mature miR-182 levels during the DE-to-Duo transition (adjusted $P = 0.003$, $\log_2FC = 3.47$ by DESeq2 Wald test in ChRO-seq; Fig. 6A). On the other hand, we found that the depletion of mature miR-375 levels (adjusted $P = 8.65E-08$, $\log_2FC = -5.79$ by DESeq2 Wald test in smRNA-seq) is concordant with the reduction in nascent transcription (adjusted $P = 0.004$, $\log_2FC = -3.27$ by DESeq2 Wald test in ChRO-seq), indicating that miR-375 is regulated during the DE-to-Duo transition largely at the transcriptional level (Fig. 6A). Indeed, our ChRO-seq data showed overall high transcriptional activity and several active transcriptional regulatory elements (TREs) around the miR-375 locus only in the DE stage, but not in other stages including Duo spheroids [26] (Fig. 6B). Specifically, four out of the six TREs adjacent to the miR-375 locus are uniquely active in DE (TRE#2, #3, #4 and #6 in Fig. 6B). We also showed that the summed transcriptional activity of TREs that were overlapped with the putative promoter region of miR-375 defined in [48] (TRE#5–6) is highly correlated with the transcriptional activity of the nearby enhancer TREs (TRE#1–4) across all four stages of the directed differentiation model (Spearman correlation $R = 0.86$, $P = 6e-04$; Fig. 6C). Taken together, these observations demonstrate that miR-375 transcription is in large part regulated by the activity of nearby enhancers and that the significant down-regulation of mature miR-375 during SI lineage specification is likely due to reduced transcriptional activity at the corresponding genomic locus.

To explore the candidate pathway(s)/regulator(s) that control miR-375 transcription during human SI development, we first identified 474 genes (RNA-seq data) that are positively correlated (Spearman correlation $R^2 > 0.8$) with miR-375 expression across hESC, DE, Duo spheroid and Duo HIO (Day 28 HIOs) (Fig. 6D). Pathway analysis of these genes revealed significant enrichment of TGF β signaling ($P = 0.0013$ by Fisher's exact test in KEGG 2019 human database; Fig. 6E) and transcriptional networks controlled by SMAD4 ($P = 2.046544e-09$ by Fisher's

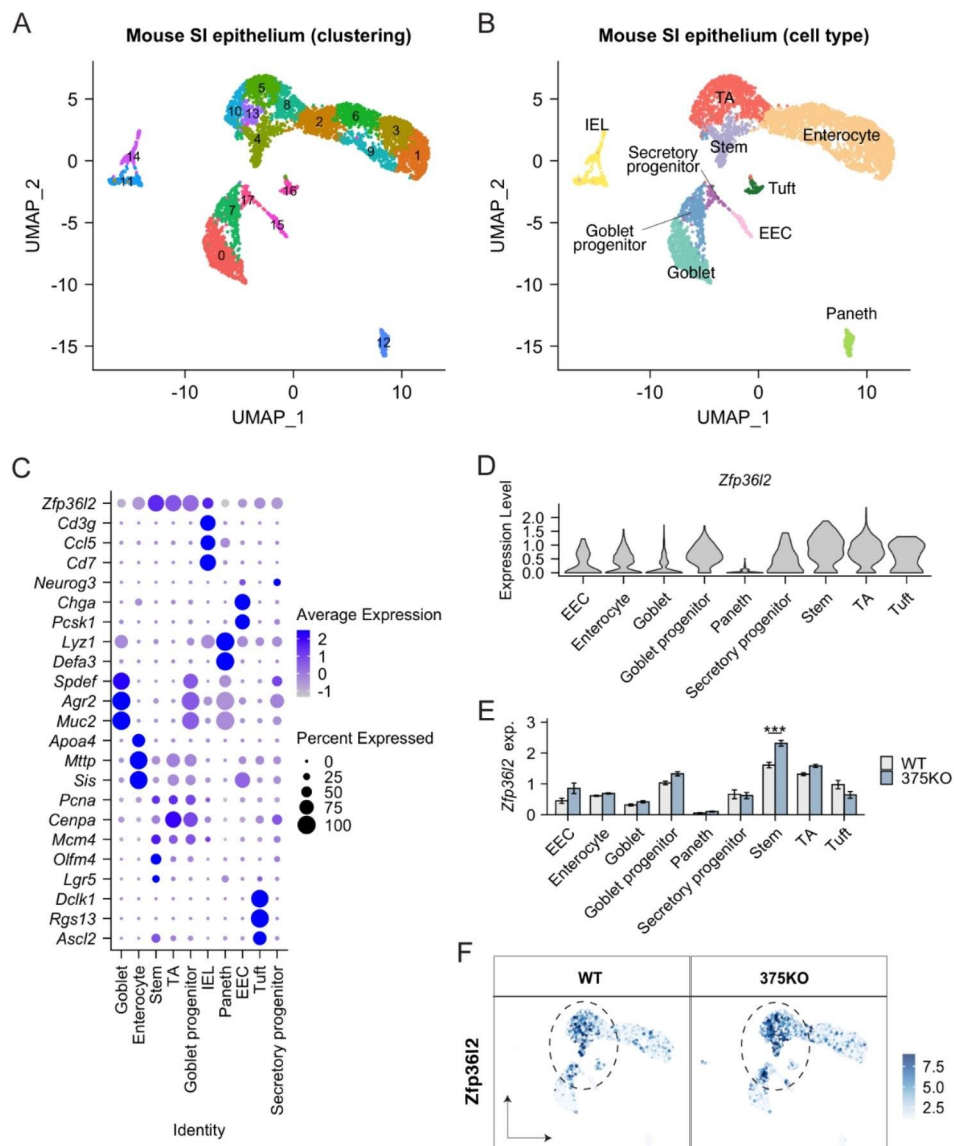


Fig. 5 Single cell RNA-seq revealed a long-lasting, conserved role for miR-375 in regulating *Zfp36l2* in intestinal stem cell population of adult mice. **(A)** UMAP of initial clustering analysis with intestinal epithelial cells of adult WT and whole body miR-375 knockout (375KO) mice. N=2 mice per genotype; 2937 cells from WT and 3275 cell from 375KO after doublet removal and quality control filtering. **(B)** UMAP of cell type annotation with intestinal epithelial cells of adult WT and 375KO mice. **(C)** Dot plot of known cell type markers and *Zfp36l2* across all cell types defined in the single cell RNA-seq dataset. **(D)** Violin plot showing *Zfp36l2* most enriched in stem cell population. **(E)** Significant upregulation of *Zip36l2* in stem cell population of 375KO compared to WT mice. *** Adjusted P value = 0.000279 and log₂FC = 0.346 by differential expression analysis using MAST method [95]. **(F)** UMAP overlay with the normal expression of *Zfp36l2* in WT and 375KO mice. UMAP, Uniform Manifold Approximation and Projection. IEL, Intraepithelial lymphocytes; EEC, enteroendocrine cell; TA, transient amplifying cell

exact test in ChEA Consensus TFs database; Fig. 6F). Consistent with these findings, we observed that DE-specific TRE#3 harbors several high-confidence SMAD4 motifs (JASPAR database), which have been validated by SMAD4 ChIP-seq in hESC-derived DE [49](Fig. 6G). These findings together suggest key roles for SMAD4 in promoting transcription at the miR-375 locus.

During SI lineage specification, the levels of miR-375 are significantly reduced upon treatment with a cocktail that augments Wnt signaling. In addition, we noticed that

miR-375 expression continued to decrease in the subsequent stage transition where the cells were patterned to gain distal SI regional identity by prolonged Wnt activation (Supplementary Fig. 2). These observations motivated us to test whether Wnt signaling itself serves as a negative regulator of miR-375 expression in the SI. To this end, we leveraged CRISPR-engineered mouse enteroids in which a mutation was introduced to *Apc* (*Apc*^{Q883}) or *Ctnnb1* (*Ctnnb1*^{S33F}), leading to constitutive activation of Wnt signaling and proliferation phenotypes

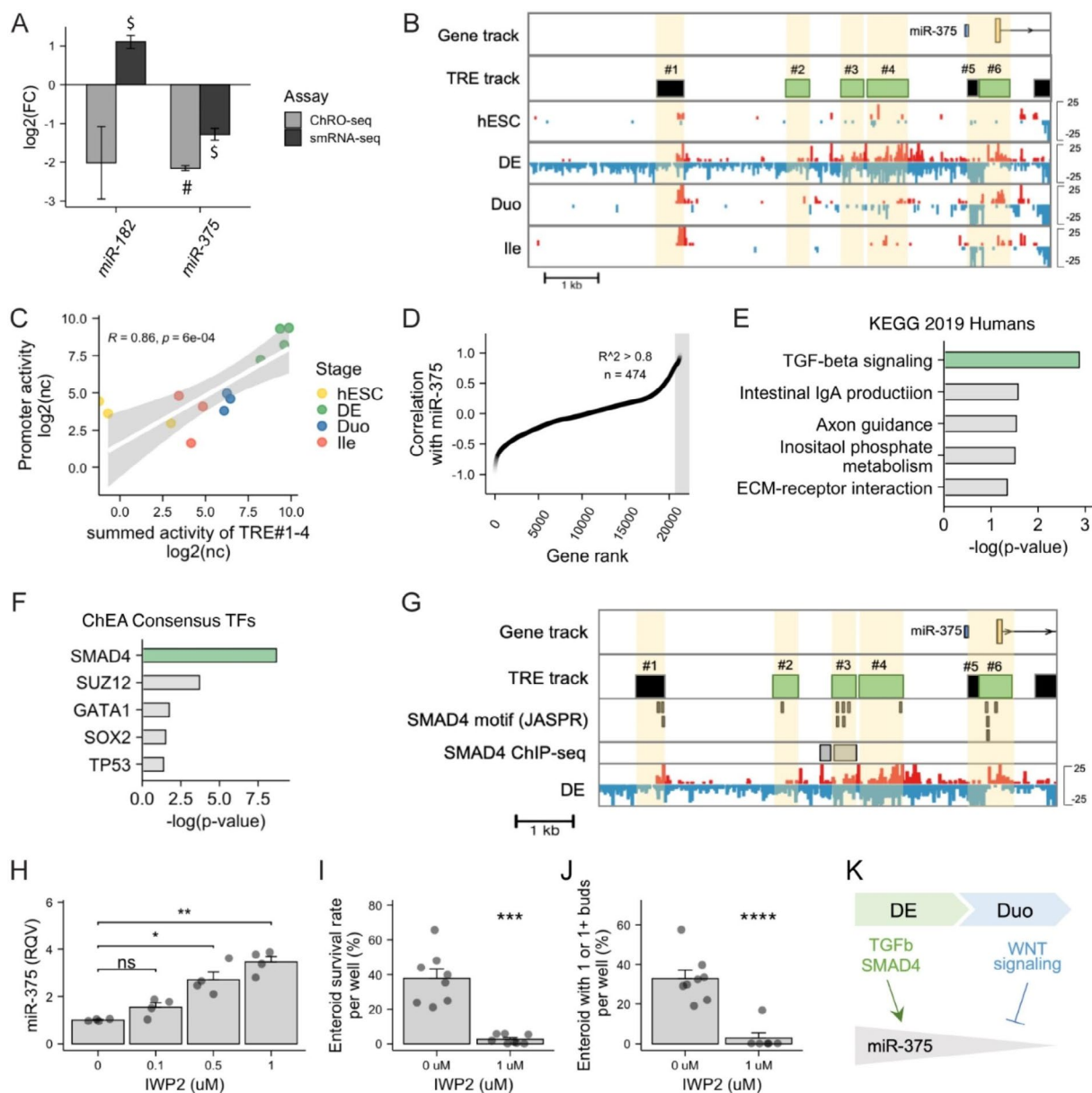


Fig. 6 SMAD4 and WNT signaling respectively serve as a positive and negative regulator of miR-375 expression in SI epithelial cells. **(A)** Nascent transcriptional (ChRO-seq) and steady-state expression (smRNA-seq) changes of miR-182 and miR-375 during SI lineage specification (transition from DE to Duo). \$ Adjusted P value < 0.05 by DESeq2 Wald test in small RNA-seq; # Adjusted P value < 0.05 by DESeq2 Wald test in ChRO-seq. **(B)** Normalized ChRO-seq signal and active TREs detected around miR-375 locus across different stages. Reads in red and blue are signals from positive and negative strand, respectively. DE-specific TREs (in green) were defined previously in Hung et al., (2021). Scale (± 25 reads/kb/ 10^6) is fixed across all stages. **(C)** Positive Spearman correlation between ChRO-seq activity of putative miR-375 promoter (TRE#5–6) and the summed activity of nearby enhancers (TRE#1–4) across all stages. **(D)** Genes ranked based on their correlation with miR-375 expression across stages of hESC, DE, Duo spheroid and Duo HIO. Genes with positive correlation with miR-375 expression ($n = 474$; Spearman correlation $R > 0.8$) are highlighted. **(E)** Enrichment of TGF beta signaling (KEGG 2019 Human) in genes that have positive correlation with miR-375 (Spearman correlation $R > 0.8$). **(F)** Enrichment of SMAD4 (ENCODE & ChEA consensus) in genes that have positive correlation with miR-375 (Spearman correlation $R > 0.8$). **(G)** DE-specific TREs that are nearby miR-375 locus contain sequences of SMAD4 motif (MA1153.1; JASPAR) and overlap with SMAD4 sites defined by the ChIP-seq study (Kim et al., 2011). **(H)** qPCR showing an upregulation of miR-375 in mouse enteroids treated with IWP2, a Wnt signaling inhibitor, compared to mock. $N = 4$ wells per condition. * $P < 0.05$, *** $P < 0.01$ by two-tailed t-test. **(I)** Survival rate of mouse enteroids with and without IWP2 treatment. 400 enteroids were seeded per well of 96-well plate on day 0 and survival rate per well was assessed on day 5. *** $P < 0.001$ by two-tailed t-test. $N = 5-8$ wells per condition. **(J)** Enteroid budding phenotype of mouse enteroids with and without IWP2 treatment. Percentage of enteroids with 1 or 1+ buds per well was determined on day 5 after treatment. **** $P < 0.0001$ by two-tailed t-test. $N = 5-8$ wells per condition. **(K)** Proposed model of TGFb/SMAD4 and Wnt signaling in regulating miR-375 expression. TRE, transcriptional regulatory element. hESC, human embryonic stem cell. DE, definitive endoderm. Duo, duodenal spheroid. Ile, ileal spheroid

as previously described [50, 51]. By performing qPCR analysis, we showed a very robust reduction of miR-375 expression in these Wnt-activated mouse enteroids compared to the control group (~10 fold downregulation; **Supplementary Fig. 7A**). Consistent with this finding, we observed a significant reduction of miR-375 in SI polyps developed in mouse models with hyperactivation of Wnt signaling (*Apc^{Min/+}* or *Apc^{q1405x/+}*) [50], compared normal epithelial tissues of control mice (**Supplementary Fig. 7B**). In addition, we perturbed Wnt signaling by treating WT mouse enteroids with Wnt signaling inhibitor, IWP2. The results showed that miR-375 expression was significantly elevated by IWP2 in a dose-dependent manner (Fig. 6H) and that the proliferative phenotypes (survival rate and budding phenotype) of enteroids were significantly compromised upon the treatment of 1 μ M IWP2 compared to mock control ($P < 0.05$ by two-tailed t-test; Fig. 6I–J). Taken together, these data strongly support Wnt signaling as a negative regulator of miR-375 expression during SI lineage specification (Fig. 6K). miR-375 likely exhibits a long-lasting role in regulating stem cell/progenitors in mature, adult SI beyond the pre-natal developmental stages.

Discussion

The prenatal development of the SI requires a well-orchestrated transcriptional program to achieve cellular identity acquisition and proper morphological patterning [52, 53]. The present study represents the first investigation to our knowledge of miRNA-mediated post-transcriptional mechanisms in very early events of human SI development. An important SI developmental event included in this study is SI lineage specification from the endodermal layer, which occurs during post-conceptual week 2 of human development. Studying such early developmental events in humans remains extremely challenging due to limited access to primary human fetal tissues from appropriate embryonic stages. A major strength of this study is the use of the directed differentiation model that generates fetal-like human intestinal organoids (HIOs) from human pluripotent stem cells (hPSCs). By leveraging this state-of-art human model of SI development, we provide novel insights into post-transcriptional regulation and miRNA activity during human SI development.

Our smRNA-seq analysis demonstrated changing miRNA signatures across different cellular stage transitions of the hPSC-HIO model. The up- or down-regulation of miRNAs during a cellular stage transition can in theory lead to changes in expression of target genes. Leveraging RNA-seq data alone to test this hypothesis is fraught with limitations, as changes detected in steady-state gene expression levels by RNA-seq could be due to transcriptional and/or post-transcriptional (PT)

processes. ChRO-seq was recently developed not only to detect active regulatory elements but also provide sensitive measurements of gene transcriptional activity [25, 26]. In this study, by integrating RNA-seq and ChRO-seq data, we teased apart transcriptional from PT regulation of genes across the entire genome and identified potential miRNA regulators relevant to human SI development.

We showed how a miRNA may exert gene regulatory impact through targeting an important TF during early SI development. Specifically, we found that the increase in miR-182 leads to the suppression of *SOX2*. Although miR-182 has been shown to target and regulate *SOX2* previously in cochlea development and glioblastoma [54, 55], our study demonstrated the functional relevance of this regulatory connection in the context of SI development. *SOX2* is known as a master TF of foregut identity [41, 42]; therefore, the increased expression and activity of miR-182 during SI lineage establishment may serve as a critical mechanism to favor midgut vs. foregut lineages.

We also demonstrated that the depletion of miR-375 during SI lineage specification significantly alleviates post-transcriptional suppression and thereby upregulates its target genes, including *ARL4C* and *ZFP36L2*, which could play an important role during this developmental event. *ARL4C*, which belongs to the superfamily of GTP-binding proteins, is reported to control the formation of 3D tubular structure in the rat intestinal IEC6 cell line [56]. *ZFP36L2* is a previously validated miR-375 target in pancreatic carcinoma [57] and known to code for an RNA binding protein that can destabilize RNA transcripts containing AU-rich elements [58]. Notably, *SOX2* is identified as one of the PT unstable genes during SI lineage formation in our analyses and known to harbor AU-rich elements. Whether *SOX2* is a direct target of *ZFP36L2* for mRNA decay during this developmental event warrants further investigation. Both *ZFP36L2* and *ARL4C* also harbor conserved predicted target sites for miR-302, which our analysis suggests may work in coordination with miR-375 to regulate SI lineage specification. To delineate the regulatory impact of miR-375 apart from that of miR-302, we studied adult mouse SI tissue in which miR-302 is absent but miR-375 is robustly expressed. As shown by our scRNA-seq study of adult miR-375KO mice, the miR-375:*Zfp36l2* and miR-375:*Arl4c* regulatory interactions are conserved across species and maintained in adult intestinal stem cells beyond early SI development. This long-lasting regulatory impact of miR-375 could in part contribute to changes in intestinal function and tissue homeostasis observed previously [59, 60].

Our study also revealed that while miR-182 expression may be regulated primarily by post-transcriptional mechanisms, miR-375 levels are changed largely by nascent transcriptional activity. Our analyses suggested SMAD4 and TGF β signaling as positive regulators of miR-375

expression during human SI development, supporting only one previous study in which it was shown that TGF β /Smad pathway promotes miR-375 expression in pancreatic islets [61]. By mining our ChRO-seq data, we showed that transcription of miR-375 is likely driven by nearby enhancers, one of which harbors a previously validated SMAD4 binding site [49]. In addition, several lines of evidence suggest that miR-375 and WNT signaling have a reciprocal inhibitory relationship. First, many of the miR-375 targets identified in our analyses are potentiators of WNT signaling (e.g., *ARLAC*, *ZFP36L2*, *TCF12*, *TCF4*). Second, miR-375 expression is suppressed during SI lineage specification upon Wnt activating cocktail and continues to decrease in the subsequent event of distal SI regional patterning with prolonged Wnt activation. Third, we performed perturbation experiments in enteroids to demonstrate that WNT signaling is a robust negative regulator of miR-375. These data together suggest a long-lasting functional role of miR-375 in regulating stem cell/progenitors of SI not only during pre-natal developmental stages but also in adult, mature tissues.

Many studies have uncovered individual miRNAs that control certain aspects of adult SI biology, including SI cell type allocation and function [62–65], epithelial homeostasis [59, 66–68], and disease pathogenesis [69–71]. The present study uncovers a novel facet of miRNAs in regulating prenatal SI development. We leveraged multi-omic, systems biology approaches to discover candidate miRNA regulators associated with the events of DE formation, SI lineage specification, and SI regional patterning in a human organoid model. While this work mainly focused on post-transcriptional regulation relevant to SI lineage specification, the candidate miRNA regulators that we identified for the other stages of SI development also warrant detailed characterization in the future.

Methods

Directed differentiation of human pluripotent stem cells The human embryonic stem cell (hESC) line H9 (NIH registry #0062) was obtained from the WiCell Research Institute. The induced pluripotent stem cell line (iPSC) 72.3 was obtained from Cincinnati Children's Hospital [72]. hPSCs were maintained and passaged in mTeSR1 medium (STEMCELL Technologies, Vancouver, BC, Canada) in cell culture plates coated with Matrigel (BD Biosciences, San Jose, CA). hPSCs were differentiated toward human duodenal or ileal fates as previously described [22, 23, 27]. Briefly, the definitive endoderm was generated by treatment of activin A (100 ng/ml) for 3 consecutive days in Roswell Park Memorial Institute 1640 (RPMI-1640) media supplemented with 0% (v/v; day 1), 0.2% (v/v; day 2) and 2.0% (v/v; day 3) Hyclone defined fetal bovine serum (dFBS). Midgut patterning that gen-

erates intestinal spheroids was carried out in RPMI-1640 supplemented with 2% HyClone dFBS, FGF4 (500 ng/mL) and CHIR99021 (2 μ M) with daily media changes. The intestinal spheroids that represent progenitor cells of human fetal duodenum and ileum were collected after 5 and 10 days of FGF4/CHIR99021 cocktail, respectively.

Whole body miR-375 knockout mouse model Mouse miR-375 gene was deleted using the CRISPR/Cas9 system with a guide RNA (gRNA) targeting the 946–965 bp region of the mouse miR-375 gene (ENSMUSG00000065616). 173 FVB x B6J F1 hybrid 1-cell embryos were co-injected with 2.5 μ g of gRNA and 7.5 μ g of Cas9 mRNA. 107 embryos at two-cell stage were transferred to pseudo-pregnant recipient female mice at ~20 embryos/recipient. 20 founder pups were born and validated for deletion of the miR-375 gene by targeted genomic sequencing. Mice null for miR-375 were backcrossed at least three generations with B6(Cg)-Tyr-2 J (B6 albino) wildtype mice. Synthesis of the gRNA, mouse embryo co-injections and viable embryo emplacement were performed by the Cornell Stem Cell and Transgenic Core Facility at Cornell University and supported in part by Empire State Stem Cell Fund (contract number C024174). Genotyping was performed on genomic DNA extracted from tail tissues, using forward primer 5'-GAAGCTCATCCACCAGACACC-3' and reverse primer 5'-GAGCTATGCCTGGCAGCTCA-3' amplifying a 307 bp product in wild-type mice and a 260 bp product in miR-375 knockout mice. All animal procedures were performed with the approval and authorization of the Institutional Animal Care and Use Committee at Cornell University (IACUC protocol: 2017-0086).

Apc mutant mouse models: generation and genotype validation of *Apc*^{Min/+} and *Apc*^{q1405x/+} mouse line were

Described previously in [50]. These mouse lines with a heterozygous mutation at *Apc* locus have been shown to form polyps in the small intestine due to a hyperactivation of Wnt signaling [50]. In this study, small intestinal polyps from *Apc*^{Min/+} and *Apc*^{q1405x/+} mice were identified using a dissection microscope, individually collected from cold PBS-flushed small intestine, and subsequently flash frozen at -80 °C. Collected villi scrapings of Wild type mice from Rosa26-CAGs-LSL-rtTA3 knock-in or Rosa26-CAGs-rtTA3 knock-in colonies were used as control group. All animal procedures were performed with the approval and authorization of the Institutional Animal Care and Use Committee at Weill Cornell Medicine (IACUC protocol: 2014-0038).

Isolation of mouse jejunal crypts system with a guide RNA/jejunal crypts were isolated from male C57BL/6 or B6 albino mice as previously described [73]. Small intes-

tine was excised from mice and divided into three equal segments. The middle region was considered jejunum. Subsequent to luminal flushing with ice cold PBS, the tissue was longitudinally cut and subjected to incubation in 3 mM EDTA in ice cold PBS with 1% (v/v) Primocin™ (InvivoGen, San Diego, CA; ant-pm-1) for 15 min at 4 °C. The mucosa of the intestinal pieces was gently scrapped of mucus, shaken in ice cold PBS with 1% (v/v) primocin for 2 min, and incubated in fresh 3 mM EDTA in ice cold PBS with 1% (v/v) primocin for 40 min at 4 °C. After 2 to 6 min of gentle manual shaking in ice cold PBS with 1% (v/v) primocin, the intestinal pieces were inspected microscopically for detached intestinal crypts, diluted 1:2 with ice cold PBS with 1% (v/v) primocin, filtered with a 70 µm cell strainer, and collected by pelleting with centrifugation at 110 g for 10 min at 4 °C. The isolated crypts were used for enteroid culture or single cell RNA-seq library preparation.

Mouse enteroid culture Mouse enteroid culture was performed as previously described [59, 66]. The jejunal crypts (day 0) isolated from male 3- to 5-month-old B6 albino mice were grown in Reduced Growth Factor Matrigel (Corning, Corning, NY; 356,231) and advanced DMEM/F12 (Gibco; 12634-028) containing GlutaMAX (Gibco; 35050-061), Primocin (Invivogen, San Diego, CA; ant-pm-05), HEPES (Gibco; cat. 15630-080), N2 supplement (Gibco; cat. 17502-048), 50 ng/µL EGF (R&D Systems, Minneapolis, MN; 2028-EG), 100 ng/µL Noggin (PeproTech, Rocky Hill, NJ; 250–38), 250 ng/µL murine R-spondin (R&D Systems, Minneapolis, MN; 3474-RS-050), and 10 µM Y27632 (Enzo Life Sciences, East Farmingdale, NY; ALX270-333). For Wnt suppression studies, IWP2 (Tocris Bioscience, Bristol, UK; 3533), the compound that inhibits Wnt processing and secretion, was added on day 0 and day 3. 400 intestinal crypts were seeded per well of 96-well plate on day 0. Enteroids on day 5 were harvested for total RNA isolation and morphological assessments.

CRISPR/Cas9 editing of mouse enteroids The proximal 15 cm of the small intestine was harvested from 6-week-old C57BL/6 mice, followed by crypt isolation as previously described [74]. Crypts were plated in Reduced Growth Factor Matrigel (Corning, Corning, NY; 356,231) and grown in basal media containing 50 ng/µL EGF (Invitrogen Waltham, MA; PMG8043), 100 ng/µL Noggin (PeproTech, Rocky Hill, NJ; 250–38) and 500 ng/µL murine R-spondin (R&D Systems, Minneapolis, MN; 3474-RS-050) for 3–4 weeks [50]. To generate enteroids with constitutive Wnt activation, enteroids were dissociated and transfected with using CRISPR base editing to introduce a *Apc* mutation (*Apc*^{Q883*}) or *Ctnnb1* mutation (*Ctnnb1*^{S33F}) to the system, detailed in [50, 51]. Exogenous R-spondin1

were withdrawn 2 days after transfection to select for mutant enteroids.

Real-time quantitative PCR: Total RNA was isolated using the Total Purification kit (Norgen Biotek, Thorold, ON, Canada) per manufacturer's instructions. Reverse transcription of miRNA was performed using TaqMan microRNA Reverse Transcription kit (Thermo Fisher Scientific, Waltham MA). miRNA expression was measured using TaqMan assays (Thermo Fisher Scientific, Waltham MA) with TaqMan Universal PCR Master Mix. The microRNA TaqMan assays used in this study are miR-375 (assay ID: 000564) and miR-182 (assay ID: 002334). miRNA expression was normalized to U6 (assay ID: 001973) or RNU48 (assay ID: 001006). Reverse transcription of RNA was performed using High Capacity RNA-to-cDNA kit (Thermo Fisher Scientific, Waltham, MA). Gene expression was measured using TaqMan assays (Thermo Fisher Scientific, Waltham MA) with TaqMan Gene Expression Master Mix. The Gene expression TaqMan assays used in this study are *SOX2* (Hs04234836_s1), *ARL4C* (Hs00255039_s1), *ZFP36L2* (Hs00272828_m1), *CDX2* (Hs01078080_m1) and *SOX17* (Hs00751752_s1), *CDH1* (Hs01023895_m1) and *SOX21* (Hs01072517_s1). RNA expression levels were normalized to RPS9 (Hs02339424_g1). Reaction was quantified on a BioRad CFX96 Touch Real Time PCR Detection System (Bio-Rad Laboratories, Richmond, CA, United States).

Transfection studies in directed differentiation model We used human induced pluripotent stem cell line 72.3 for transfection studies. To test candidate miRNA regulators, DE-patterned cells were transfected by miRCURY LNA inhibitors (Qiagen, Hilden, Germany) against specific miRNAs during hindgut fate induction (FGF4/CHIR99021 cocktail). Transfection of LNA inhibitors were performed using Lipofectamine 3000 (ThermoFisher Scientific, Waltham MA; L3000-008) per manufacturer's instructions. Specifically, on day 3 of FGF4/CHIR99021 cocktail, DE cells were treated with 500 nM LNA miR-182 inhibitor (Qiagen, Hilden, Germany; #4101243-101) or scrambled control (Power Negative Control A; YI00199006-DDA). After 24 h treatment, cell monolayer was collected for RNA isolation followed by qPCR measurements. The efficacy of directed differentiation was confirmed by qPCR measurements of *CDX2* (intestinal marker gene) and *SOX17* (DE marker gene) in mock treated cells. The efficacy of LNA transfection was assessed by qPCR of target miRNA.

Small RNA sequencing Samples for small RNA sequencing were processed through the procedures described previously [75, 76]. Briefly, total RNA was isolated using the Total Purification kit (Norgen Biotek, Thorold, ON, Canada). RNA was quantified with the Nanodrop 2000

(Thermo Fisher Scientific, Waltham, MA), and RNA integrity was determined by the Agilent 2100 Bioanalyzer or 4200 TapeStation (Agilent Technologies, Santa Clara, CA). Libraries were prepared using the CleanTag Small RNA Library Prep kit (TriLink Biotechnologies, San Diego, CA). Sequencing (single-end 50x) was performed on the HiSeq2000 platform (Illumina, San Diego, CA) at the Genome Sequencing Facility of the Greehey Children's Cancer Research Institute (University of Texas Health Science Center, San Antonio, TX). Mapping statistics of small RNA-seq samples can be found in **Supplementary Data 1**. Raw sequencing and processed data can be accessed through GEO record GSE207467.

RNA sequencing Total RNA was isolated using the Total Purification kit (Norgen Biotek, Thorold, ON, Canada). RNA purity was quantified with the Nanodrop 2000 (Thermo Fisher Scientific, Waltham, MA), and RNA integrity was quantified with the Agilent 2100 Bioanalyzer or 4200 TapeStation (Agilent Technologies, Santa Clara, CA). Libraries were generated using the NEBNext Ultra II Directional Library Prep Kit (New England Biolabs, Ipswich, MA, USA) and subjected to sequencing (single-end 92x) on the NextSeq500 platform (Illumina) at the Cornell University Biotechnology Resource Center. At least 80 M reads per sample were acquired. Raw and processed sequencing data is available through GEO record GSE142316.

Chromatin run-on sequencing Chromatin isolation for chromatin run-on sequencing (ChRO-seq) was performed as previously described [25, 26]. To perform run-on reaction with the isolated chromatin, samples were mixed with an equal volume of 2X run-on reaction mix [10 mM Tris-HCl pH 8.0, 5 mM MgCl₂, 1 mM DTT, 300 mM KCl, 400 μM ATP, 0.8 μM CTP, 400 μM GTP, 400 μM UTP, 40 μM Biotin-11-CTP (Perkin Elmer, Waltham, MA; NEL542001EA), 100 ng yeast tRNA (VWR, 80054–306), 0.8 units/μL SUPERase In RNase Inhibitor, 1% (w/v) Sarkosyl]. The run-on reaction was incubated in an Eppendorf Thermomixer at 37°C for 5 min (700 rpm) and stopped by adding Trizol LS (Life Technologies, Carlsbad, CA; 10296–010). RNA samples were precipitated by GlycoBlue (Ambion, Austin, TX; AM9515) and resuspended in diethylpyrocarbonate (DEPC)-treated water. To perform base hydrolysis reaction, RNA samples in DEPC water were heat denatured at 65°C for 40 s and incubated in 0.2 N NaOH on ice for 4 min. Base hydrolysis reaction was stopped by neutralizing with Tris-HCl pH 6.8. Nascent RNA was enriched using streptavidin beads (NEB, Ipswich, MA; S1421) followed by RNA extraction using Trizol. RNA samples were then processed through the following steps: (i) 3' adapter ligation with T4 RNA Ligase 1 (NEB, Ipswich, MA; M0204), (ii) RNA binding with strep-

tavidin beads (NEB, Ipswich, MA; S1421) followed by RNA extraction with Trizol, (iii) 5' de-capping with RNA 5' pyrophosphohydrolase (NEB, Ipswich, MA; M0356), (iv) 5' end phosphorylation using T4 polynucleotide kinase (NEB, Ipswich, MA; M0201), (v) 5' adapter ligation with T4 RNA Ligase 1 (NEB, Ipswich, MA; M0204). The 5' adaptor contained a 6-nucleotide unique molecular identifier (UMI) to allow for bioinformatic detection and elimination of PCR duplicates. Streptavidin bead binding followed by Trizol RNA extraction was performed again before final library construction. To generate ChRO-seq library, cDNA was generated through a reverse-transcription reaction using Superscript III Reverse Transcriptase (Life Technologies, Carlsbad, CA; 18,080–044) and amplified using Q5 High-Fidelity DNA Polymerase (NEB, Ipswich, MA; M0491). Finally, ChRO-seq libraries were sequenced (5' single end; single-end 75x) using the NextSeq500 high-throughput sequencing system (Illumina, San Diego, CA) at the Cornell University Biotechnology Resource Center. Raw and processed sequencing data is available through GEO record GSE142316.

trimmed from the

Single cell RNA sequencing Samples for single cell RNA sequencing were processed through the procedures described previously [77]. Briefly, mouse jejunal crypts from male B6 albino wildtype and miR-375 knockout mice were isolated and resuspended in an ice-cold PBS containing 0.04% (w/v) bovine serum albumin and pelleted at 1000 x g at 4°C for 5 min. The crypts were subsequently digested with 0.3 U/mL Dispase I (MilliporeSigma Burlington, MA; D4818) in HBSS at 37°C for 12 min with gentle agitation. The dispase I activity was stopped by adding fetal bovine serum to a final concentration of 10% (v/v) and DNaseI to a final concentration of 50 μg/mL. The sample was filtered through 40 μm strainer, pelleted at 500 x g at 4°C for 5 min, washed with cold HBSS, filtered again, and then resuspended in an ice-cold PBS containing 0.04% (w/v) bovine serum albumin. Prior to submission, single cell suspensions were triturated by pipetting and evaluated for total viable cell number by using trypan blue staining with a TC20 automated cell counter (Bio-Rad Laboratories, Richmond, CA). Single cell RNA sequencing of these samples was performed at the Cornell University Biotechnology Resource Center. Libraries were prepared using the 10X Genomics Chromium preparation kit (3' v2 chemistry) (10X Genomics, Pleasanton, CA). Sequencing quality of single cell RNA-seq samples can be found in **Supplementary Data 2**. Single cell RNA-seq data of wildtype mice was included in our previous study [73] (GSE178826) and re-analyzed in the present study. Single cell RNA-seq data of miR-375 knockout mice is available through GEO record GSE208224.

Small RNA-seq analyses SmRNA-seq reads were trimmed, mapped, and quantified to the hg19 genome using miRquant 2.0 [78], our previously described smRNA-seq analysis pipeline. Briefly, reads were trimmed using Cutadapt v1.12. miRquant 2.0 requires a minimum 10 nucleotide overlap between the adapter sequence and the 3'-end of the sequencing read with less than 10% errors in the alignment. Those reads that did not meet this criterion, or were <14 nucleotides in length following trimming, were discarded. Alignment of reads were performed using Bowtie v1.1.0 [79]. Perfectly aligned reads represented miRNA loci and imperfectly mapped reads (from miRNA isoforms) were re-aligned to these loci using SHRiMP v2.2.2 [80]. If a read aligned equally well to multiple genomic loci, those reads were proportionally assigned to each of the loci. Aligned reads were quantified and normalized using reads per million mapped to miRNAs (RPM). Principal components analysis (PCA) of miRNAs was performed using raw counts with rlog transformation. To identify differentially expressed miRNAs, smRNA-seq raw counts were analyzed through DESeq2 v1.30.1 [81]. MiRNAs that were significantly altered between two conditions were defined by criteria of RPM >1000 at least in one condition, log₂ fold-change >0.5 (or < -0.5) and adjusted P < 0.05 by Wald test (DESeq2).

RNA-seq analyses RNA-seq reads were mapped to human genome hg38 using STAR v2.5.3a [82] and transcript quantification was performed using Salmon v0.6.0 [83] with GENCODE v33 transcript annotations. The expression levels of genes were normalized using DESeq2 v1.30.1 [81]. To identify differentially expressed genes, RNA-seq raw counts were analyzed through DESeq2 v1.30.1 [81]. Genes that were significantly altered between two conditions were defined by criteria of base mean >100, log₂ foldchange >0.5 (or < -0.5) and adjusted P < 0.05 by Wald test (DESeq2). Pathway analyses of genes were performed using Enrichr [84–86].

ChRO-seq analyses ChRO-seq reads were mapped to human genome hg38 using an established pipeline [87]. Briefly, PCR deduplication was performed by collapsing UMIs using PRINSEQ lite v0.20.2 [88]. Adapters were trimmed from the 3' end of remaining reads using Cutadapt v1.16 with a maximum 10% error rate, minimum 2 bp overlap, and minimum 20 quality score. Processed reads with a minimum length of 15 bp were mapped to the hg38 genome modified with the addition of a single copy of the human Pol I ribosomal RNA complete repeating unit (GenBank: U13369.1) using Burrows-Wheeler Aligner (BWA) v0.7.13 [89]. The location of the RNA polymerase active site was represented by a single base that denotes the 5' end of the nascent RNA, which corresponds to the

position on the 3' end of each sequenced read. The ChRO-seq reads were normalized by the length of gene bodies to transcripts per million (TPM). To visualize ChRO-seq signal, data were converted to bigwig format using bedtools and UCSC bedGraphToBigWig. Bigwig files within a sample category were then merged and normalized to a total signal of 1 × 10⁶ reads. Genomic loci snapshots were generated using R package Gviz [90].

Quantification of gene transcription activity Gene transcription activity was determined using ChRO-seq data. Gene body information was extracted from GENCODE v33 (hg38). Genes with gene body smaller than 1 kb were excluded from this analysis. To quantify transcription activity of gene loci (gene length > 1 kb), stranded ChRO-seq reads within gene coordinates were counted. Reads within 150 b downstream of transcription start site (TSS) were excluded to avoid bias of RNA polymerase pausing at the promoters (Supplementary Fig. 1). Since mature miRNA species are transcribed from longer primary miRNAs transcripts, we used +/-5 kb flanking region of mature miRNA coordinate to determine transcription activity of miRNA species. Reads from transcriptional regulatory elements (TREs) that are present within gene bodies were also excluded to avoid bias of intragenic TRE activity. The TREs that are present in specific stage of directed differentiation experiments were defined previously in Hung et al. [26] using dREG [87]. BEDTools was used to remove reads mapping to these TREs within gene bodies. The remaining reads were quantified using the R package bigwig (<https://github.com/andrelmartins/bigwig>). In differential transcription analysis, ChRO-seq raw counts of gene loci were analyzed through DESeq2 v1.30.1 [81].

Define post-transcriptionally regulated genes The development of this analysis pipeline was modified from [91]. To define post-transcriptional regulation of genes, ChRO-seq and RNA-seq were analyzed through the two-factor analysis of DESeq2 v1.30.1 [81]. Specifically, in the two-factor analysis, one factor referred to cell identities during a stage transition (e.g., duodenal spheroids versus definitive endoderm) and the other factor referred to type of assays (RNA-seq versus ChRO-seq). In this full model that contains these two factors and the interaction term between the two, the interaction term of the DESeq2 result was extracted to query the differences between assay types (RNA-seq, ChRO-seq) during a stage transition. Genes that were subject to significant post-transcriptionally regulation (become post-transcriptionally stable or unstable during a stage transition) were defined by criteria of log₂ foldchange >0.5 (or < -0.5) and adjusted P < 0.05 in the interaction term by Wald test (DESeq2). To identify genes that had little changes at the transcriptional

level (ChRO-seq) but significant changes at the steady state level (RNA-seq) during a stage transition, the post-transcriptionally regulated genes were filtered with additional criteria of adjusted $P > 0.1$ in ChRO-seq differential expression analysis, adjusted $P < 0.05$ in RNA-seq differential expression analysis, and \log_2 foldchange > 0.5 (or < -0.5) in RNA-seq differential expression analysis. In this study, only genes with ChRO-seq TPM > 20 and RNA-seq base mean > 100 were presented.

miRhub analysis The bioinformatic tool miRhub [92] was used to identify candidate master miRNA regulators of genes that were subject to post-transcriptional regulation in this study. miRhub uses the TargetScan algorithm to predict target sites for miRNAs of interest in the 3' untranslated regions (UTRs) of a given gene set and then determines by Monte Carlo simulation for each miRNA whether the number and strength of predicted targets is significantly greater than expected by chance. The Monte Carlo simulation in a miRhub analysis was performed with 1000 iterations, with a set of randomly selected human genes each time, to generate a background distribution of the predicted targeting scores for each miRNA. These score distributions were then used to calculate an empirical p-value of the targeting score for each miRNA in a given gene set. In the present study, miRhub analyses were performed with predated miRNA target sites that were present in humans and also preserved in one additional species (chicken, dog, mouse, rat) (using miRhub cons1 mode). Predicted miRNA target sites (species and sequence conservation) were obtained from TargetScan [93].

Single cell RNA-seq analysis Single cell sequencing reads from age and sex matched wildtype and miR-375 knockout mice (12-month-old; male) were aligned to the mouse genome (10x genomics pre-built reference version mm10-1.2.0) and quantified using Cell Ranger (v3.0.1) to obtain cell by gene count matrices. Ambient RNA removal was carried out using SoupX (v1.5.2). Subsequent filtering, data integration, clustering, and visualization of the data was completed using Seurat (v4.0.5). To maintain data quality, cells with less than 1500 detected genes or greater than 25% of the reads mapping to mitochondrial genes were discarded. 702 potential doublets, which were detected using scDblFinder (v1.4.0), were discarded, resulting in 11,697 singlets for downstream analyses. To control sequencing depth, counts were normalized using sctransform [94] and glmGamPoi (v1.2.0). Samples were merged using the Seurat integration anchor workflow based on the 3000 most variable genes. Clustering and identification of nearest neighbors relied on 30 principal component analysis (PCA) dimensions. Cells were clustered at a resolution of 0.7, resulting in 17 cell clusters in

the dataset. Clustering visualization was performed using uniform manifold approximation and projection (UMAP). Highly enriched markers for each cluster were determined using the Seurat function FindAllMarkers (log fold change > 0.5 ; using MAST method [95]), which compares the gene expression within a cluster with all other clusters. Identity of each of the cluster was assigned using addModuleScore function comparing highly enriched markers with known markers of intestinal epithelial cell types (curated from Habor et al., [96]) and classical immune cell markers (specifically *Gzma*, *Gzmb*, *Itgae*, *Ccl5*, *Cd7*, *Cd69*, *Cd3g* and *Cd8a*). Differential expression of genes between wildtype and miR-375 knockout samples was determined using function FindMarkers (using MAST method).

Statistics Statistical analyses were performed using R (4.0.4). In sequencing studies, statistical significance was determined using DESeq2, where P values were calculated by Wald test and adjusted using Benjamin and Hochberg (BH) method. Statistical significance of quantitative PCR was determined using unpaired two sample t-test unless otherwise noted. The P values in pathway enrichment analysis by Enrichr were calculated using Fisher's exact test. Values of $p < 0.05$ (or adjusted $p < 0.05$ if in sequencing studies) were considered statistically significant. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09743-1>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6

Acknowledgements

We thank Dr. Zhao Li and the Greehey Children's Cancer Research Institute at University of Texas Health Science Center at San Antonio for small RNA library preparation and sequencing; Dr. Peter Schweitzer and the Genomics Facility at Cornell University. Some of illustrations were made in BioRender.com.

Author contributions

Conceptualization and investigation, Y.-H.H., and P.S.; Formal analyses and bioinformatics, Y.-H.H.; Experiments, Y.-H.H., M.C., S.H., M.T.S. and J.V.; Data curation, Y.-H.H., M.K., M.T.S., and J.V.; Mouse colony, R.C. and V.D.R.; Supporting and Resources, J.C.S.; Writing (original draft), Y.-H.H. and P.S.; Review and editing, Y.-H.H., M.C., J.R.S. and P.S.; Supervision, J.R.S. and P.S.

Funding

Y.-H. H. is supported by the Empire State Stem Cell Fund (C30293GG); P.S. is supported by ADA Pathway to Stop Diabetes Research Accelerator Award (1-16-ACE-47), Cornell Inter-campus Multi-Investigator Seed Grant, and an R21 from NIH-National Institute of Child Health & Human Development (NICHD) (R21HD104922). J.C.S. is supported by Small RNA Pathways in Mammalian Gametogenesis by NIH-NICHD (5P50HD076210-05). J.R.S. is supported in part

by grant CZF2019-002440 from the Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation, in part by the Intestinal Stem Cell Consortium (U01DK103141), a collaborative research project funded by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) and the National Institute of Allergy and Infectious Diseases (NIAID), and in part by the University of Michigan Center for Gastrointestinal Research (UMCGR) (NIDDK 5P30DK034933).

Data Availability

The small RNA-seq data generated from this study is available through GEO accession GSE207467. ChRO-seq and bulk RNA-seq data have been generated previously [26] and available through GEO accession GSE142316. For the single cell RNA-seq analyses, the data of wild type and whole body miR-375 knockout mice is available through GSE178826 [73] and GSE208224, respectively. We established an interactive web server called ProHumID (<https://sethupathy-lab.shinyapps.io/ProHumID/>) for users to explore sequencing data generated from this study.

Declarations

Ethical approval

All animal procedures were performed with the approval and authorization of the Institutional Animal Care (IACUC protocol: 2014-0038 and 2017-0086). All experimental procedures were conducted following the guidelines of Institutional Animal Care and Use Committee at Cornell University. We confirm this study was conducted in accordance with the ARRIVE guidelines (<https://arriveguidelines.org>).

Consent for publication

Not applicable.

Conflict of interest

The authors disclose no conflicts.

Author details

¹Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA

²Department of Cell and Developmental Biology, University of Michigan, Ann Arbor, MI, USA

³Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, MI, USA

Received: 4 July 2023 / Accepted: 14 October 2023

Published online: 26 October 2023

References

- Bartel DP. Metazoan MicroRNAs. *Cell*. 2018;173(1):20–51.
- Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*. 2009;19(1):92–105.
- Lim LP, Lau NC, Garrett-Engle P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*. 2005;433(7027):769–73.
- Bernstein E, Kim SY, Carmell MA, Murchison EP, Alcorn H, Li MZ, Mills AA, Elledge SJ, Anderson KV, Hannon GJ. Dicer is essential for mouse development. *Nat Genet*. 2003;35(3):215–7.
- Wienholds E, Koudijs MJ, van Eeden FJ, Cuppen E, Plasterk RH. The microRNA-producing enzyme Dicer1 is essential for zebrafish development. *Nat Genet*. 2003;35(3):217–8.
- La Rocca G, King B, Shui B, Li X, Zhang M, Akat KM, Ogorodowski P, Mastroleo C, Chen K, Cavalieri V, et al. Inducible and reversible inhibition of miRNA-mediated gene repression in vivo. *Elife*. 2021. 10.
- Rahmanian S, Murad R, Breschi A, Zeng W, Mackiewicz M, Williams B, Davis CA, Roberts B, Meadows S, Moore D et al. Dynamics of microRNA expression during mouse prenatal development. *Genome Res* 2019.
- Wei Y, Peng S, Wu M, Sachidanandam R, Tu Z, Zhang S, Falce C, Sobie EA, Lebeche D, Zhao Y. Multifaceted roles of miR-1s in repressing the fetal gene program in the heart. *Cell Res*. 2014;24(3):278–92.
- Espinoza-Lewis RA, Wang DZ. MicroRNAs in heart development. *Curr Top Dev Biol*. 2012;100:279–317.
- Cordes KR, Srivastava D. MicroRNA regulation of cardiovascular development. *Circ Res*. 2009;104(6):724–32.
- Tao G, Martin JF. MicroRNAs get to the heart of development. *Elife*. 2013;2:e01710.
- Shibata M, Nakao H, Kiyonari H, Abe T, Aizawa S. MicroRNA-9 regulates neurogenesis in mouse telencephalon by targeting multiple transcription factors. *J Neurosci*. 2011;31(9):3407–22.
- Prodromidou K, Vlachos IS, Gaitanou M, Kouroupi G, Hatzigeorgiou AG, Matsas R. MicroRNA-934 is a novel primate-specific small non-coding RNA with neurogenic function during early development. *Elife* 2020, 9.
- Nowakowski TJ, Rani N, Golkaram M, Zhou HR, Alvarado B, Huch K, West JA, Leyrat A, Pollen AA, Kriegstein AR, et al. Regulation of cell-type-specific transcriptomes by microRNA networks during human brain development. *Nat Neurosci*. 2018;21(12):1784–92.
- Nowakowski TJ, Fotaki V, Pollock A, Sun T, Pratt T, Price DJ. MicroRNA-92b regulates the development of intermediate cortical progenitors in embryonic mouse brain. *Proc Natl Acad Sci U S A*. 2013;110(17):7056–61.
- Poy MN, Hausser J, Trajkovski M, Braun M, Collins S, Rorsman P, Zavolan M, Stoffel M. miR-375 maintains normal pancreatic alpha- and beta-cell mass. *Proc Natl Acad Sci U S A*. 2009;106(14):5813–8.
- Lopez-Beas J, Capilla-Gonzalez V, Aguilera Y, Mellado N, Lachaud CC, Martin F, Smani T, Soria B, Hmadcha A. miR-7 modulates hESC differentiation into insulin-producing Beta-like cells and contributes to cell maturation. *Mol Ther Nucleic Acids*. 2018;12:463–77.
- Kloosterman WP, Lagendijk AK, Ketting RF, Moulton JD, Plasterk RH. Targeted inhibition of miRNA maturation with morpholinos reveals a role for miR-375 in pancreatic islet development. *PLoS Biol*. 2007;5(8):e203.
- Williams AH, Liu N, van Rooij E, Olson EN. MicroRNA control of muscle development and Disease. *Curr Opin Cell Biol*. 2009;21(3):461–9.
- Horak M, Novak J, Bienertova-Vasku J. Muscle-specific microRNAs in skeletal muscle development. *Dev Biol*. 2016;410(1):1–13.
- Kim HK, Lee YS, Sivaprasad U, Malhotra A, Dutta A. Muscle-specific microRNA miR-206 promotes muscle differentiation. *J Cell Biol*. 2006;174(5):677–87.
- Spence JR, Mayhew CN, Rankin SA, Kuhar MF, Vallance JE, Tolle K, Hoskins EE, Kalinichenko VV, Wells SJ, Zorn AM, et al. Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. *Nature*. 2011;470(7332):105–9.
- Tsai YH, Nattiv R, Dedhia PH, Nagy MS, Chin AM, Thomson M, Klein OD, Spence JR. In vitro patterning of pluripotent stem cell-derived intestine recapitulates in vivo human development. *Development*. 2017;144(6):1045–55.
- Watson CL, Mahe MM, Munera J, Howell JC, Sundaram N, Poling HM, Schweitzer JJ, Vallance JE, Mayhew CN, Sun Y, et al. An in vivo model of human small intestine using pluripotent stem cells. *Nat Med*. 2014;20(11):1310–4.
- Chu T, Rice EJ, Booth GT, Salamanca HH, Wang Z, Core LJ, Longo SL, Corona RJ, Chin LS, Lis JT, et al. Chromatin run-on and sequencing maps the transcriptional regulatory landscape of Glioblastoma Multiforme. *Nat Genet*. 2018;50(11):1553–64.
- Hung YH, Huang S, Dame MK, Yu Q, Yu QC, Zeng YA, Camp JG, Spence JR, Sethupathy P. Chromatin regulatory dynamics of early human small intestinal development using a directed differentiation model. *Nucleic Acids Res*. 2021;49(2):726–44.
- McCracken KW, Howell JC, Wells JM, Spence JR. Generating human intestinal tissue from pluripotent stem cells in vitro. *Nat Protoc*. 2011;6(12):1920–8.
- Hinton A, Hunter SE, Afrikanova I, Jones GA, Lopez AD, Fogel GB, Hayek A, King CC. sRNA-seq analysis of human embryonic stem cells and definitive endoderm reveals differentially expressed microRNAs and novel lsmiRNAs with distinct targets. *Stem Cells*. 2014;32(9):2360–72.
- Tzur G, Levy A, Meiri E, Barad O, Spector Y, Bentwich Z, Mizrahi L, Katzenellenbogen M, Ben-Shushan E, Reubinoff BE, et al. MicroRNA expression patterns and function in endodermal differentiation of human embryonic stem cells. *PLoS ONE*. 2008;3(11):e3726.
- Hinton A, Afrikanova I, Wilson M, King CC, Maurer B, Yeo GW, Hayek A, Pasquelli AE. A distinct microRNA signature for definitive endoderm derived from human embryonic stem cells. *Stem Cells Dev*. 2010;19(6):797–807.
- Anokye-Danso F, Trivedi CM, Juhr D, Gupta M, Cui Z, Tian Y, Zhang Y, Yang W, Gruber PJ, Epstein JA, et al. Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell*. 2011;8(4):376–88.

32. Lin SL, Chang DC, Lin CH, Ying SY, Leu D, Wu DT. Regulation of somatic cell reprogramming through inducible mir-302 expression. *Nucleic Acids Res.* 2011;39(3):1054–65.
33. Miyoshi N, Ishii H, Nagano H, Haraguchi N, Dewi DL, Kano Y, Nishikawa S, Tanemura M, Mimori K, Tanaka F, et al. Reprogramming of mouse and human cells to pluripotency using mature microRNAs. *Cell Stem Cell.* 2011;8(6):633–8.
34. Gao N, White P, Kaestner KH. Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. *Dev Cell.* 2009;16(4):588–99.
35. Kumar N, Tsai YH, Chen L, Zhou A, Banerjee KK, Saxena M, Huang S, Toke NH, Xing J, Shivdasani RA et al. The lineage-specific transcription factor CDX2 navigates dynamic chromatin to control distinct stages of intestine development. *Development* 2019, 146(5).
36. Tam PP, Khoo PL, Wong N, Tsang TE, Behringer RR. Regionalization of cell fates and cell movement in the endoderm of the mouse gastrula and the impact of loss of Lhx1(Lim1) function. *Dev Biol.* 2004;274(1):171–87.
37. Arnold SJ, Hofmann UK, Bikoff EK, Robertson EJ. Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. *Development.* 2008;135(3):501–11.
38. Costello I, Pimeisl IM, Drager S, Bikoff EK, Robertson EJ, Arnold SJ. The T-box transcription factor eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. *Nat Cell Biol.* 2011;13(9):1084–91.
39. Viotti M, Nowotschin S, Hadjantonakis AK. SOX17 links gut endoderm morphogenesis and germ layer segregation. *Nat Cell Biol.* 2014;16(12):1146–56.
40. Heslop JA, Pournasir B, Liu JT, Duncan SA. GATA6 defines endoderm fate by controlling chromatin accessibility during differentiation of human-induced pluripotent stem cells. *Cell Rep.* 2021;35(7):109145.
41. Trisno SL, Philo KED, McCracken KW, Cata EM, Ruiz-Torres S, Rankin SA, Han L, Nasr T, Chaturvedi P, Rothenberg ME, et al. Esophageal organoids from human pluripotent stem cells delineate Sox2 functions during esophageal specification. *Cell Stem Cell.* 2018;23(4):501–515e507.
42. Raghoebir L, Bakker ER, Mills JC, Swagemakers S, Kempen MB, Munck AB, Driegen S, Meijer D, Grosveld F, Tibboel D, et al. SOX2 redirects the developmental fate of the intestinal epithelium toward a premature gastric phenotype. *J Mol Cell Biol.* 2012;4(6):377–85.
43. Kuzmichev AN, Kim SK, D'Alessio AC, Chenoweth JG, Wittko IM, Campanati L, McKay RD. Sox2 acts through Sox21 to regulate transcription in pluripotent and differentiated cells. *Curr Biol.* 2012;22(18):1705–10.
44. Selitsky SR, Dinh TA, Toth CL, Kurtz CL, Honda M, Struck BR, Kaneko S, Vickers KC, Lemon SM, Sethupathy P. Transcriptomic analysis of Chronic Hepatitis B and C and Liver Cancer reveals MicroRNA-Mediated Control of Cholesterol Synthesis Programs. *mBio.* 2015;6(6):e01500–01515.
45. Saetrom P, Heale BS, Snove O Jr, Aagaard L, Alluin J, Rossi JJ. Distance constraints between microRNA target sites dictate efficacy and cooperativity. *Nucleic Acids Res.* 2007;35(7):2333–42.
46. Grimson A, Farh KK, Johnston WK, Garrett-Engle P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell.* 2007;27(1):91–105.
47. Shanahan MT, Kanke M, Oyesola OO, Hung YH, Koch-Laskowski K, Singh AP, Peck BCE, Biraud M, Sheahan B, Cortes JE, et al. Multiomic analysis defines the first microRNA atlas across all small intestinal epithelial lineages and reveals novel markers of almost all major cell types. *Am J Physiol Gastrointest Liver Physiol.* 2021;321(6):G668–81.
48. Sethupathy P. Illuminating microRNA transcription from the Epigenome. *Curr Genomics.* 2013;14(1):68–77.
49. Kim SW, Yoon SJ, Chuong E, Oyulu C, Wills AE, Gupta R, Baker J. Chromatin and transcriptional signatures for nodal signaling during endoderm formation in hESCs. *Dev Biol.* 2011;357(2):492–504.
50. Schatoff EM, Goswami S, Zafra MP, Foronda M, Shusterman M, Leach BI, Katti A, Diaz BJ, Dow LE. Distinct Colorectal Cancer-Associated APC mutations dictate response to tankyrase inhibition. *Cancer Discov.* 2019;9(10):1358–71.
51. Schatoff EM, Zafra MP, Dow LE. Base editing the mammalian genome. *Methods.* 2019;164–165:100–8.
52. Wells JM, Spence JR. How to make an intestine. *Development.* 2014;141(4):752–60.
53. Chin AM, Hill DR, Aurora M, Spence JR. Morphogenesis and maturation of the embryonic and postnatal intestine. *Semin Cell Dev Biol.* 2017;66:81–93.
54. Weston MD, Tarang S, Pierce ML, Pyakurel U, Rocha-Sanchez SM, McGee J, Walsh EJ, Soukup GA. A mouse model of miR-96, miR-182 and miR-183 misexpression implicates miRNAs in cochlear cell fate and homeostasis. *Sci Rep.* 2018;8(1):3569.
55. Kouri FM, Hurley LA, Daniel WL, Day ES, Hua Y, Hao L, Peng CY, Merkel TJ, Queisser MA, Ritner C, et al. miR-182 integrates apoptosis, growth, and differentiation programs in glioblastoma. *Genes Dev.* 2015;29(7):732–45.
56. Matsumoto S, Fujii S, Sato A, Ibuka S, Kagawa Y, Ishii M, Kikuchi A. A combination of wnt and growth factor signaling induces Arl4c expression to form epithelial tubular structures. *EMBO J.* 2014;33(7):702–18.
57. Yonemori K, Seki N, Kurahara H, Osako Y, Idichi T, Arai T, Koshizuka K, Kita Y, Maemura K, Natsugoe S. ZFP36L2 promotes cancer cell aggressiveness and is regulated by antitumor microRNA-375 in pancreatic ductal adenocarcinoma. *Cancer Sci.* 2017;108(1):124–35.
58. Dumdie JN, Cho K, Ramaiah M, Skarbrevik D, Mora-Castilla S, Stumpo DJ, Lykke-Andersen J, Laurent LC, Blackshear PJ, Wilkinson MF, et al. Chromatin modification and global transcriptional silencing in the oocyte mediated by the mRNA decay activator ZFP36L2. *Dev Cell.* 2018;44(3):392–402e397.
59. Peck BC, Mah AT, Pitman WA, Ding S, Lund PK, Sethupathy P. Functional transcriptomics in diverse intestinal epithelial cell types reveals robust MicroRNA sensitivity in intestinal stem cells to Microbial Status. *J Biol Chem.* 2017;292(7):2586–600.
60. Biton M, Levin A, Slyper M, Alkalay I, Horwitz E, Mor H, Kredon-Russo S, Avnitsagi T, Cojocaru G, Zreik F, et al. Epithelial microRNAs regulate gut mucosal immunity via epithelium-T cell crosstalk. *Nat Immunol.* 2011;12(3):239–46.
61. Gao Y, Zhang R, Dai S, Zhang X, Li X, Bai C. Role of TGF-beta/Smad Pathway in the transcription of pancreas-specific genes during Beta cell differentiation. *Front Cell Dev Biol.* 2019;7:351.
62. Rawat M, Nighot M, Al-Sadi R, Gupta Y, Viszwapriya D, Yochum G, Koltun W, Ma TY. IL1B increases intestinal tight Junction permeability by Up-regulation of MIR200C-3p, which degrades occludin mRNA. *Gastroenterology.* 2020;159(4):1375–89.
63. Goga A, Yagabasan B, Herrmanns K, Godbersen S, Silva PN, Denzler R, Zund M, Furter M, Schwank G, Sunagawa S, et al. miR-802 regulates paneth cell function and enterocyte differentiation in the mouse small intestine. *Nat Commun.* 2021;12(1):3339.
64. Kwon MS, Chung HK, Xiao L, Yu TX, Wang SR, Piao JJ, Rao JN, Gorospe M, Wang JY. MicroRNA-195 regulates Tuft cell function in the intestinal epithelium by altering translation of DCLK1. *Am J Physiol Cell Physiol.* 2021;320(6):C1042–54.
65. Knudsen LA, Petersen N, Schwartz TW, Egerod KL. The MicroRNA Repertoire in Enteroendocrine cells: identification of miR-375 as a potential Regulator of the Enteroendocrine lineage. *Endocrinology.* 2015;156(11):3971–83.
66. Singh AP, Hung YH, Shanahan MT, Kanke M, Bonfini A, Dame MK, Biraud M, Peck BCE, Oyesola OO, Freund JM, et al. Enteroendocrine Progenitor Cell-enriched miR-7 regulates intestinal epithelial proliferation in an xiap-dependent manner. *Cell Mol Gastroenterol Hepatol.* 2020;9(3):447–64.
67. Li Y, Wen S, Yao X, Liu W, Shen J, Deng W, Tang J, Li C, Liu K. MicroRNA-378 protects against intestinal ischemia/reperfusion injury via a mechanism involving the inhibition of intestinal mucosal cell apoptosis. *Cell Death Dis.* 2017;8(10):e3127.
68. Strubberg AM, Veronese Paniagua DA, Zhao T, Dublin L, Pritchard T, Bayguinov PO, Fitzpatrick JAJ, Madison BB. The zinc finger transcription factor PLAGL2 enhances stem cell fate and activates expression of ASCL2 in intestinal epithelial cells. *Stem Cell Reports.* 2018;11(2):410–24.
69. Jiang L, Hermeking H. miR-34a and miR-34b/c suppress intestinal tumorigenesis. *Cancer Res.* 2017;77(10):2746–58.
70. Zhang B, Tian Y, Jiang P, Jiang Y, Li C, Liu T, Zhou R, Yang N, Zhou X, Liu Z. MicroRNA-122a regulates zonulin by Targeting EGFR in Intestinal Epithelial Dysfunction. *Cell Physiol Biochem.* 2017;42(2):848–58.
71. Tian Y, Xu J, Li Y, Zhao R, Du S, Lv C, Wu W, Liu R, Sheng X, Song Y, et al. MicroRNA-31 reduces Inflammatory Signaling and promotes regeneration in Colon epithelium, and delivery of Mimics in Microspheres reduces Colitis in mice. *Gastroenterology.* 2019;156(8):2281–2296e2286.
72. McCracken KW, Cata EM, Crawford CM, Sinagoga KL, Schumacher M, Rockich BE, Tsai YH, Mayhew CN, Spence JR, Zavros Y, et al. Modelling human development and Disease in pluripotent stem-cell-derived gastric organoids. *Nature.* 2014;516(7531):400–4.
73. Shanahan M, Kanke M, Oyesola OO, Hung YH, Koch-Laskowski K, Singh AP, Peck BCE, Biraud M, Sheahan B, Cortes JE et al. Multi-omic analysis defines the first microRNA atlas across all small intestinal epithelial lineages and reveals novel markers of almost all major cell types. *Am J Physiol Gastrointest Liver Physiol* 2021.
74. O'Rourke KP, Ackerman S, Dow LE, Lowe SW. Isolation, culture, and maintenance of mouse intestinal stem cells. *Bio Protoc* 2016, 6(4).

75. Dinh TA, Jewell ML, Kanke M, Francisco A, Sritharan R, Turnham RE, Lee S, Kasstenhuber ER, Wauthier E, Guy CD, et al. MicroRNA-375 suppresses the Growth and Invasion of Fibrolamellar Carcinoma. *Cell Mol Gastroenterol Hepatol*. 2019;7(4):803–17.
76. Hung YH, Kanke M, Kurtz CL, Cubitt R, Bunaciu RP, Miao J, Zhou L, Graham JL, Hussain M, Havel PJ et al. Acute suppression of insulin resistance-associated hepatic miR-29 in vivo improves glycemic control in adult mice. *Physiol Genomics* 2019.
77. Shanahan MT, Matt Kanke APS, Jonathan W, Villanueva, Adrian J, McNairn, Oyebola O, Oyesola A, Bonfini Y-H, Hung B, Sheahan JC, Bloom RL, Cubitt, Ennessa G, Curry WA, Pitman VD, Rinaldi CM, Dekaney S, Ding, Bailey CE, Peck JC, Schimenti LE, Dow, Nicolas Buchon, Elia D. Tait-Wojno, Praveen Sethupathy: Single Cell Analysis Reveals Multi-faceted miR-375 Regulation of the Intestinal Crypt. *bioRxiv* 2020.
78. Kanke M, Baran-Gale J, Villanueva J, Sethupathy P. miRquant 2.0: an expanded Tool for Accurate Annotation and quantification of MicroRNAs and their isomiRs from small RNA-Sequencing data. *J Integr Bioinform*. 2016;13(5):307.
79. Langmead B. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinformatics* 2010, Chap. 11:Unit 11 17.
80. Rumble SM, Lacroute P, Dalca AV, Fiume M, Sidow A, Brudno M. SHRiMP: accurate mapping of short color-space reads. *PLoS Comput Biol*. 2009;5(5):e1000386.
81. Love MI, Huber W, Anders S. Moderated estimation of Fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
82. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
83. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14(4):417–9.
84. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, Lachmann A, Wojciechowicz ML, Kropiwnicki E, Jagodnik KM, et al. Gene Set Knowledge Discovery with Enrichr. *Curr Protoc*. 2021;1(3):e90.
85. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44(W1):W90–97.
86. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR. Ma'ayan A: Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14:128.
87. Chu T, Wang Z, Chou SP, Danko CG. Discovering Transcriptional Regulatory Elements from Run-On and sequencing data using the web-based dREG gateway. *Curr Protoc Bioinformatics*. 2019;66(1):e70.
88. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011;27(6):863–4.
89. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589–95.
90. Hahne F, Ivanek R. Visualizing genomic data using Gviz and Bioconductor. *Methods Mol Biol*. 2016;1418:335–51.
91. Patel RK, West JD, Jiang Y, Fogarty EA, Grimson A. Robust partitioning of microRNA targets from downstream regulatory changes. *Nucleic Acids Res*. 2020;48(17):9724–46.
92. Baran-Gale J, Fannin EE, Kurtz CL, Sethupathy P. Beta cell 5'-shifted isomiRs are candidate regulatory hubs in type 2 Diabetes. *PLoS ONE*. 2013;8(9):e73240.
93. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 2015, 4.
94. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol*. 2019;20(1):296.
95. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ, Prlic M, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol*. 2015;16:278.
96. Haber AL, Biton M, Rogel N, Herbst RH, Shekhar K, Smillie C, Burgin G, Delorey TM, Howitt MR, Katz Y, et al. A single-cell survey of the small intestinal epithelium. *Nature*. 2017;551(7680):333–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.