Supplementary information: Genome-wide identification of microRNA targets in human ES cells reveals a role for miR-302 in modulating BMP response

Inna Lipchina\textsuperscript{1}, Yechiel Elkabetz\textsuperscript{2}, Markus Hafner\textsuperscript{3}, Robert Sheridan\textsuperscript{4}, Aleksandra Mihailovic\textsuperscript{3}, Thomas Tuschi\textsuperscript{3}, Chris Sander\textsuperscript{4}, Lorenz Studer\textsuperscript{1}, Doron Betel\textsuperscript{4,5}

Correspondence to mirstemcell@cbio.mskcc.org reaches all authors

\textsuperscript{1}Developmental Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY
\textsuperscript{2}Department of Cell and Developmental Biology, Sackler School of Medicine, Tel Aviv University, Israel
\textsuperscript{3}Howard Hughes Medical Institute, Laboratory of RNA Molecular Biology, The Rockefeller University, New York, NY
\textsuperscript{4}Computational Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY
\textsuperscript{5}Current address: Department of Medicine and Institute of Computational Biomedicine, Weill Cornell Medical College, New York, NY

September 12, 2011

Supplementary Figures

Supplemental Figure 1: MicroRNA expression profile determined by hybridization array is consistent with sequence-based expression profile

Supplemental Figure 2: Efficacy of coding region CCRs in target downregulation

Supplemental Figure 3: Sequence similarity among the highly expressed microRNAs results in ambiguous target assignment in PAR-CLIP data

Supplemental Figure 4: Validation of miR-302/367 inhibition and overexpression by luciferase assays

Supplemental Figure 5: Antagomir treatment downregulates both miR-302 and miR-367 activity independently

Supplemental Figure 6: Identification of the high confidence miR-302/367 targets

Supplemental Figure 7: miR-302/367 inhibition leads to reduction of BMP activity during neural induction

Supplemental Figure 8: miR-302/367 inhibition decreases the efficiency of mesoderm and endoderm induction

Supplemental Figure 9: Luciferase assay validation of DAZAP2, SLAIN1, TOB2 regulation by miR-302

Supplementary Files

Supplementary Table 1: 454 and Solexa microRNA sequencing data

Supplementary Table 2: Upregulated targets from miR-302/367 antagomir experiments

Supplementary File 1: PAR-CLIP target sites (5’UTR, CD, 3’UTR). Excel file containing the miRNA binding sites found in CCRs that matched annotated regions.
Description of Supplementary Files

Supplementary File 1: PAR-CLIP target sites (5'UTR, CD, 3'UTR).

This Excel file contains three tables:

- CLIP_3UTR_Hits
- CLIP_5UTR_Hits
- CLIP_CD_Hits

Each table includes a description of the crosslink centered regions (CCRs) that were found in annotated gene regions (3'UTR, 5'UTR and CD) and that contain target sites to the top 29 expressed seed families in hESC.

CCR records begin with a description line followed by the specific microRNA targets. The description line contains the following fields:

- [Gene Name]
- [RefSeq ID]
- [CCR id]
- [number of crosslinked reads]
- [number of reads]
- [miRNA seed family]

Note that there could be more than one seed family with target sites in the CCR. The microRNA target lines contain the predicted target sites found in the CCR. Each line includes the following fields:

- [miRNA name]
- [miRNA alignment]
- [alignment string]
- [mRNA alignment]
- [mirSVR score (Betel et al., 2010)]
- [conservation score (Siepel et al., 2005)]
- [position in UTR or CD]

For example:

**LEFTY2**

<table>
<thead>
<tr>
<th>NM_003240</th>
<th>slc26a14_chr1 10 43</th>
<th>mir-302a/302b/302c/302d/372/520a-3p/520b/520c-3p/520d-3p/302e</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-miR-302a</td>
<td>aguGGU--UUUGUA-CCUUCGUGAAu</td>
<td></td>
</tr>
<tr>
<td>hsa-miR-302b</td>
<td>gaugAUUUGUA-CCUUCGUGAAu</td>
<td></td>
</tr>
</tbody>
</table>
Supplemental Figure S1: MicroRNA expression profile determined by hybridization array is consistent with sequence-based expression profile. MicroRNA expression levels measured by Agilent array are consistent with the deep sequencing profile as determined by 454 sequencing. In particular, the normalized array intensities are highly correlated with the sequencing coverage for the abundantly expressed microRNA clusters in undifferentiated hESCs.
Supplemental Figure S2: CCRs in coding regions are generally less effective in target downregulation than CCRs in 3’UTRs. Similar to figure 2d,e in the manuscript, we performed a CDF analysis of genes with miR-302/367 CCR sites in their coding regions (184 genes) and similarly for genes with miR-302/367 CCRs in their 3’UTRs (654 genes). These gene sets were mutually exclusive such that genes with miR-302/367 CCRs in both coding region and 3’UTR were excluded from the analysis. We found that the extent of target upregulation following miR-302/367 inhibition for genes with coding region targets (green) is significantly smaller than genes with 3’UTR target sites (blue). KS p-value=0.0334 for the difference between coding-region set and background and the p-value=0.000111 for the difference between coding-region and 3’UTR.
Supplemental Figure S3: Sequence similarity among the highly expressed microRNAs results in ambiguous target assignment. Most target prediction methods are heavily biased towards perfect complementarity between the microRNA seed sequence (positions 2-8) and the target mRNA. As a result, microRNAs with similar seed sequences that are few bases shifted are often predicted to target the same sites. Among the highly expressed microRNAs in undifferentiated hESC are miR-302, miR-17/20 and miR-130 which are highly similar in their seed sequences (marked in red). Consequently, there is a large overlap among their respective targets identified by the PAR-CLIP experiment.
Supplemental Figure S4: Validation of miR-302/367 inhibition and overexpression by luciferase assays. Cells were transfected with psi-check2 plasmid containing four miR-302a target sites and assayed for luciferase activity after 24 hours by dual-luciferase assay. (a) Inhibition of miR-302a by antagomirs led to increased luciferase activity relative to the corresponding controls of random antagomir and transfection vehicle (Lipofectamine) indicating loss of miR-302a-mediated downregulation (p-value < 0.05 compared to control antagomir). (b) Conversely, overexpression of miR-302/367 (mimics) resulted in reduction of luciferase activity relative to controls indicating increased downregulation of the target.
Supplemental Figure S5: Antagomir treatment downregulates both miR-302 and miR-367 activity. CDF analysis of miR-367 and miR-302 targets following antagomir treatment indicates clear upregulation of miR-367 targets independently of miR-302 targets (gene sets are mutually exclusive). Furthermore, the expression levels of miR-106 targets are not significantly different from the background gene set even though miR-106 seed sequence is one base shifted relative to miR-302 (see Supplemental Fig. 3) indicating that the antagomir treatment is highly specific to the mir-302 cluster.
Supplemental Figure S6: Identification of the high confidence miR-302/367 targets. CDF plots demonstrating the enrichment of both gene upregulation and number of crosslinked reads in the set of high-confidence targets. The list of 146 high-confidence targets was defined as described in the text (Fig. 3a) and compared to the respective background datasets. (a) Among the set of 622 miR-302/367 targets that are upregulated by more than 0.2 log change upon miR-302/367 inhibition (blue circle in Fig. 3a), the 146 high-confidence targets (green line) are significantly more upregulated than the remaining 476 targets (p-value $\leq 3.72e^{-7}$, KS-test). (b) Similarly, among 734 genes identified in the PAR-CLIP experiment (green circle in Fig. 3a) the set of high-confidence 146 miR-302/367 targets (green line) are enriched for crosslinked sequence reads (p-value $\leq 1.62e^{-5}$, KS-test) relative to the remaining 588 targets.
Supplemental Figure S7: miR-302/367 inhibition leads to reduction of BMP activity during neural induction. (a) qRT-PCR for early neural marker SOX1 after 7 days of differentiation in knockout serum replacement (KSR)-based medium containing inhibitors of BMP (Noggin) and TGF-β/Activin/Nodal (SB-431542) signaling pathways. In all pathway inhibition experiments the miR-302/367 antagomir treatment led to increased SOX1 levels relative to control antagomir (p-values 0.036 for SB, 0.159 for Noggin, 0.005 for Noggin SB, Wilcoxon rank sum test, n=7). In contrast, without the inhibitor treatments there is no difference between the two antagomir treatments (average ratio of 1).

In the analysis of PAX6 expression by intracellular FACS, hESCs were transfected with control antagomirs (b) or miR-302/367 antagomir (c) and cultured in media containing the inhibitor of TGF-β/Activin/Nodal signaling pathway (SB-431542). On day 11 of differentiation, cells were fixed and analyzed for PAX6 expression by intracellular flow cytometry. (d) In all 4 replicates the percentage of PAX6 positive cells was larger in the miR-302 antagomir vs. control antagomir (p-value = 0.004). As a positive control (e) cells were grown in the presence of both TGF-β/Activin/Nodal and BMP inhibitors (SB-431542 and Noggin). (f) unstained negative control.
Supplemental Figure S8: miR-302/367 inhibition decreases the efficiency of mesoderm and endoderm induction. Cells were transfected with miR-302/367 antagonirs and cultured in endoderm (a) or mesoderm (b) differentiation conditions. Induction was measured by qRT-PCR of SOX17 for endoderm differentiation and BRACHYURY for mesoderm differentiation. In both cases loss of miR-302/367 led to reduced differentiation possibly through increased inhibition of TGF-β signaling (** p-value < 0.01, *** p-value < 0.001 compared to control antagonir).
Supplemental Figure S9: Validation of DAZAP2, SLAIN1, TOB2 siRNA knockdown and regulation by miR-302 by luciferase assay. (a) siRNA screen of miR-302/367 targets for BMP signaling regulation. BMP-specific marker ID1 levels were assayed by qRT-PCR following siRNA treatment of the 11 genes as well as two positive controls SMAD6/7 (known inhibitors of BMP signaling) and non-targeting siRNA control. Of the 11 targets knockdown of DAZAP2, SLAIN1 and TOB2 had robust and consistent upregulation of ID1 levels when compared to the control siRNA suggesting that these are negative regulators of BMP signaling. (b) siRNA functional validation by qRT-PCR 48 hours after knockdown, normalized to negative control siRNA. (c) Sequences of miR-302 target sites identified by PAR-CLIP and corresponding mutated sites. Luciferase reporter constructs containing wild type or mutant 3′ UTR of DAZAP2, SLAIN1, and TOB2 were transfected into (d) hESC together with miR-302/367 antagonirs, or (e) HEK293T cells with miR-302/367 mimics. Three DAZAP2 miR-302 target sites were mutated either individually or together. Data is presented as fold change of luciferase signal between miR-302/367 antagonist or mimic treated cells over the control, normalized to empty pMIR-REPORT plasmid (WT = wild type 3′ UTR; M = mutated 3′ UTR; * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001)