Oncogene (2016) **35,** 158–172

www.nature.com/onc

ORIGINAL ARTICLE

miR-200 promotes the mesenchymal to epithelial transition by suppressing multiple members of the Zeb2 and Snail1 transcriptional repressor complexes

R Perdigão-Henriques[1,2,3], F Petrocca[1], G Altschuler[4], MP Thomas[1], MTN Le[1], SM Tan[1], W Hide[4,5] and J Lieberman[1,6]

The miR-200 family promotes the epithelial state by suppressing the Zeb1/Zeb2 epithelial gene transcriptional repressors. To identify other miR-200-regulated genes, we isolated mRNAs bound to transfected biotinylated miR-200c in mouse breast cancer cells. In all, 520 mRNAs were significantly enriched in miR-200c binding at least twofold. Putative miR-200-regulated genes included *Zeb2*, enriched 3.5-fold in the pull down. However, *Zeb2* knockdown does not fully recapitulate miR-200c overexpression, suggesting that regulating other miR-200 targets contributes to miR-200's enhancement of epithelial gene expression. Candidate genes were highly enriched for miR-200c seed pairing in their 3′UTR and coding sequence and for genes that were downregulated by miR-200c overexpression. Epidermal growth factor receptor and downstream MAPK signaling pathways were the most enriched pathways. Genes whose products mediate transforming growth factor (TGF)-β signaling were also significantly overrepresented, and miR-200 counteracted the suppressive effects of TGF-β and bone morphogenic protein 2 (BMP-2) on epithelial gene expression. miR-200c regulated the 3′UTRs of 12 of 14 putative miR-200c-binding mRNAs tested. The extent of mRNA binding to miR-200c strongly correlated with gene suppression. Twelve targets of miR-200c (*Crtap, Fhod1, Smad2, Map3k1, Tob1, Ywhag/14-3-3γ, Ywhab/14-3-3β, Smad5, Zfp36, Xbp1, Mapk12, Snail1*) were experimentally validated by identifying their 3′UTR miR-200 recognition elements. Smad2 and Smad5 form a complex with Zeb2 and Ywhab/14-3-3β and Ywhag/14-3-3γ form a complex with Snail1. These complexes that repress transcription assemble on epithelial gene promoters. miR-200 overexpression induced RNA polymerase II localization and reduced Zeb2 and Snail1 binding to epithelial gene promoters. Expression of miR-200-resistant *Smad5* modestly, but significantly, reduced epithelial gene induction by miR-200. miR-200 expression and *Zeb2* knockdown are known to inhibit cell invasion in *in vitro* assays. Knockdown of each of three novel miR-200 target genes identified here, *Smad5*, *Ywhag* and *Crtap*, also profoundly suppressed cell invasion. Thus, miR-200 suppresses TGF-β/BMP signaling, promotes epithelial gene expression and suppresses cell invasion by regulating a network of genes.

Oncogene (2016) **35,** 158–172; doi:10.1038/onc.2015.69; published online 23 March 2015

INTRODUCTION

The miR-200 family promotes the epithelial state of cells and suppresses mesenchymal properties during development and cellular differentiation. Transitions between epithelial and mesenchymal states (epithelial to mesenchymal (EMT) and mesenchymal to epithelial (MET)), which alter the motility of cells during development,[1] have an important role in tumor metastasis.[2] Epithelial cells express high levels of E-cadherin, occludin and claudins, while mesenchymal cells express high levels of N-cadherin, vimentin and fibronectin.[3] Some tumor cell lines, especially less differentiated cells, show plasticity in their display of epithelial and mesenchymal traits. Although the early steps of metastases—penetration through the basement membrane, invasion of the vasculature and extravasation—are facilitated by cells acquiring mesenchymal traits, the ability to colonize distant tissues and form macroscopic metastases may be facilitated by epithelial properties.[1,4–6] Accordingly, miR-200 expression has been associated with both increased[4,7–10] and decreased[11–13] malignancy and metastases, depending on the cancer model. High levels of circulating miR-200 are associated with poor prognosis in several human cancers, including ovarian, prostate, pancreatic and metastatic colorectal cancers.[14–17]

The miR-200 miRNAs are five homologous miRNAs (miR-141, miR-429, miR-200a, miR-200b and miR-200c), whose sequences are conserved between human and mouse. These miRNAs belong to two families whose seed regions (nucleotides 2–7 of the mature active strand, which form the most important region for target recognition) differ by a single nucleotide. The seed for miR-200b/c/429 sequence is 5′-AAUACU-3′, while for miR-200a/141 it is 5′-AACACU-3′. Because of its role as a master regulator of MET, miR-200 has been extensively studied. The Zeb epithelial gene transcriptional repressors contain multiple miR-200 binding sites in their 3′UTR (*Zeb1* has five and *Zeb2* has six conserved predicted binding sites) and their expression is profoundly downregulated by miR-200. Much of miR-200's effect on increasing epithelial gene expression and inhibiting cell motility and invasivity can be

[1]Cellular and Molecular Medicine Program, Boston Children's Hospital, Boston, MA, USA; [2]Animal Cell Technology Unit, Instituto de Tecnologia Química e Biológica (ITQB), Universidade Nova de Lisboa, Oeiras, Portugal; [3]Instituto de Biologia Experimental e Tecnológica (IBET), Oeiras, Portugal; [4]Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA; [5]Sheffield Institute for Translational Neuroscience, Department of Neuroscience, University of Sheffield, Sheffield, UK and [6]Department of Pediatrics, Harvard Medical School, Boston, MA, USA. Correspondence: Professor J Lieberman, Program in Cellular and Molecular Medicine, Children's Hospital Boston, 200 Longwood Avenue, Warren Alpert Building 255, Boston, MA 02115, USA.
E-mail: judy.lieberman@childrens.harvard.edu
Received 6 May 2014; revised 15 January 2015; accepted 20 January 2015; published online 23 March 2015

reproduced by knocking down *Zeb1* and/or *Zeb2*.[4,18–21] Many other gene targets have been described in mouse or man (Supplementary Table S1), including the transcription factors *Ets1*,[22] *Snail1* (*Snail*)[23] and *Snail2* (*Slug*).[24] miR-200 has also been shown to regulate cell-cell adhesion,[18,19,21] cancer stem cell self-renewal and differentiation,[25–27] cell division and apoptosis[28–31] and chemoresistance.[32,33]

miRNAs accomplish their biological functions by regulating networks of genes.[34] Transfection of a biotinylated (Bi-)miRNA mimic and streptavidin pull down (PD) of associated mRNAs provides an unbiased method to identify the genes regulated by a particular miRNA with high specificity.[34–39] To better understand how miR-200 functions, we carried out a Bi-miR-200c PD in the mouse breast cancer cell line 4TO7. 4TO7 is a mouse triple negative breast cancer (TNBC) cell line that is able to perform all the steps of metastasis when implanted in the mammary fat pad, except colonization of metastatic sites. 4TO7 does not express the miR-200 family, but ectopic expression of miR-200c, which accounts for 93% of mature miR-200b/c/429 family expression in human breast tumors,[40] leads to efficient colonization of lung and liver, macroscopic metastases and reduced survival.[4,8] Moreover, 4T1E, an isogenic clone derived from the same primary, spontaneously arising tumor that metastasizes, differs from 4TO7 by high expression of the miR-200 family. mRNAs for 520 genes, including *Zeb2* and other known miR-200 target genes, were significantly enriched at least twofold in the PD with miR-200c relative to a *C. elegans* control miRNA. The pulled-down genes were highly enriched for miR-200c seed-binding sequences and for genes knocked down by miR-200 overexpression. Of 18 genes chosen for experimental validation, 14 (78%) were confirmed as significantly enriched by PD and quantitative real-time PCR (qRT-PCR). Of those, 12 (86%) had functional miR-200 recognition elements in their 3′UTR. Genes that participate in epidermal growth factor receptor (EGFR) and transforming growth factor (TGF)-β/bone morphogenic protein (BMP) signaling were highly overrepresented in the miR-200 target gene set. In addition to *Zeb2* and *Snail1*, we found that miR-200 also suppressed expression of their cofactors, *Smad2/Smad5* and *Ywhab/Ywhag*, respectively.[41,42] Suppression of these multiple epithelial gene regulators enhanced the ability of miR-200 to promote MET. *Crtap*, which is required for type I collagen prolyl-hydroxylation, was a novel strong target of miR-200, whose knockdown (KD) strongly inhibited invasivity through Matrigel. Crtap is postulated to facilitate TGF-β binding to extracellular matrix.

RESULTS

miR-200c pull down

A previously published protocol,[34] which showed 90% specificity for identifying miR-34a-regulated genes, was used to capture miR-200-bound mRNAs. 4TO7 cells were transfected with control *C. elegans* miRNA mimic (Bi-cel-miR-67) or miR-200c mimic (Bi-miR-200c), both biotinylated at the 3′-end of the mature active strand. The next day cells were lysed and streptavidin-coated beads were used to isolate mRNAs bound to the biotinylated miRNAs. mRNA abundance in Bi-cel-miR-67 or Bi-miR-200c PD and input samples from two independent experiments was assessed on mRNA microarrays. An enrichment ratio *R* ((Bi-miR-200c PD/Bi-cel-miR-67 PD)/(Bi-miR-200c input/Bi-cel-miR-67 input)) was calculated as a measure of how much each transcript was enriched in the PD relative to its expression in the miRNA-transfected cells. We then calculated the average enrichment ratio and associated *P*-value for each microarray probe. In total, 520 transcripts had an enrichment ratio of >2 and $P<0.01$ (Figure 1a, Supplementary Table S2). To determine the sensitivity of the Bi-miR-200c PD, we first determined how many of the known mouse miR-200 family published targets were enriched (Supplementary Table S1). In all, 13 experimentally validated mouse targets of any miR-200 family member are expressed in 4TO7 cells (*Zeb1* and *Snail2* are not expressed in these cells) and 9 of these are known to be regulated by miR-200c. In all, 7 of the 13 (54%) and 5 of the 9 (56%) were identified in the PD. To estimate the false-positive rate of the Bi-miR-200c PD, we counted the number of transcripts that specifically bound to the control worm Bi-cel-miR-67, which has no mammalian ortholog, using the same criteria (enrichment ratio (Bi-cel-miR-67 PD/Bi-miR-200c PD)/(Bi-cel-miR-67 input/Bi-miR--200c input) >2 and $P<0.01$). Comparing 81 mRNAs enriched for binding to Bi-cel-miR-67 as a background with 520 transcripts enriched for binding to Bi-miR-200c, the estimated false-positive rate is 15.6%.

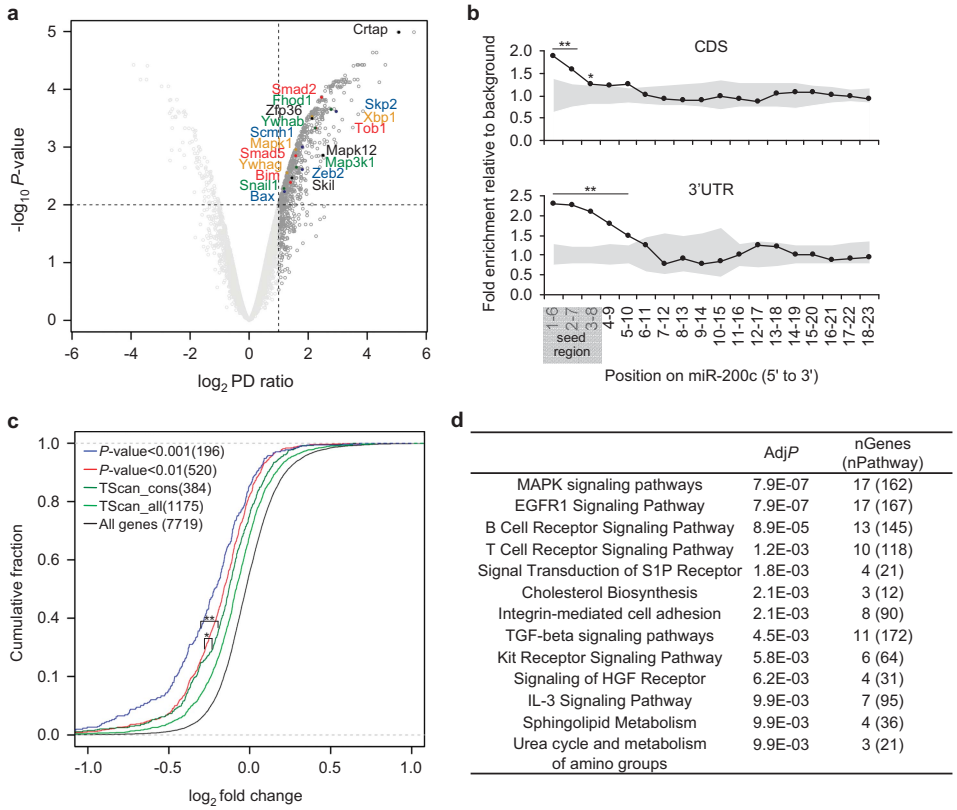
The putative miRNA target genes had properties expected of canonical miR-200c-regulated genes. Forty-three percent of the 520 enriched mRNAs were predicted miR-200c targets by TargetScan. These therefore contain a perfect match to the miR-200c seed in their 3′UTR. The coding sequence (CDS) and 3′UTRs (Figure 1b), but not the 5′UTRs (data not shown), of the 520 genes were also significantly enriched for hexamer seed matches ($P<0.0001$). Transcripts of the 520 pulled-down genes compared with all genes expressed in 4TO7 were significantly downregulated after Bi-miR-200c transfection ($P<2.2E{-}16$, Figure 1c). The extent of their downregulation was significantly greater than that of the 384 evolutionarily conserved, TargetScan predicted target genes ($P=0.023$) or the 1175 TargetScan predicted genes that included non-conserved binding sequences ($P=2.1E{-}15$). In all, 196 genes were enriched in the PD using a more stringent *P* cutoff of <0.001. This gene set was even more strongly downregulated than the larger set of 520 genes enriched by $P<0.01$ ($P=1.8E{-}3$) or the conserved TargetScan target gene set ($P=1.1E{-}6$).

miR-200 bound genes are enriched for genes involved in EGFR, MAPK and TGF-β signaling and metabolic pathways implicated in metastasis and oncogenic transformation

To understand better the biological processes regulated by miR-200, we first performed pathway enrichment analysis of the 520 enriched PD genes (Figure 1d, Supplementary Table S3). The most enriched pathways were MAPK and EGFR signaling, both with *P*-values of 8E–07, each containing 17 putative miR-200c targets, of which 6 overlapped. The MAPK and EGFR signaling was also significantly enriched pathways in the more stringently chosen gene list of 196 putative targets ($R>2$, $P<0.001$) (Supplementary Figure S1a, Supplementary Table S4). Other cell surface receptor signaling pathways were also enriched, such as the T- and B-cell receptor and Kit signaling pathways, which overlap with the EGFR pathway and use MAPK signaling. However, these other receptors are not expressed in breast tumor cells. TGF-β signaling genes were also significantly overrepresented ($P<0.005$) with 11 annotated genes. These included the *Zeb2*, *Smad2*, *Smad5* and *Tgif* transcription factors that help control epithelial gene expression (see below) and other oncogenic transcription factors, including *Fosb* and *Skil*. The sphingosine-1-phosphate receptor signaling pathway, implicated in tumor metastasis, as well as enzymes in sphingolipid and cholesterol biosynthesis pathways, were also overrepresented. The mRNAs for three enzymes involved in amino-acid metabolism, which is altered in cancer cells, were also significantly enriched.

miR-200-bound genes form a dense interaction network

The 520 putative miR-200c target genes were analyzed by Ingenuity Pathway Analysis software (Ingenuity Systems, www.ingenuity.com) to identify directly-interacting target gene products. In all, 222 gene products enriched in the PD were annotated to interact directly by protein–protein or protein–DNA/RNA interactions (Figure 2). Of note, the interactome contained a small module composed of genes that regulate transcriptional changes



	AdjP	nGenes (nPathway)
MAPK signaling pathways	7.9E-07	17 (162)
EGFR1 Signaling Pathway	7.9E-07	17 (167)
B Cell Receptor Signaling Pathway	8.9E-05	13 (145)
T Cell Receptor Signaling Pathway	1.2E-03	10 (118)
Signal Transduction of S1P Receptor	1.8E-03	4 (21)
Cholesterol Biosynthesis	2.1E-03	3 (12)
Integrin-mediated cell adhesion	2.1E-03	8 (90)
TGF-beta signaling pathways	4.5E-03	11 (172)
Kit Receptor Signaling Pathway	5.8E-03	6 (64)
Signaling of HGF Receptor	6.2E-03	4 (31)
IL-3 Signaling Pathway	9.9E-03	7 (95)
Sphingolipid Metabolism	9.9E-03	4 (36)
Urea cycle and metabolism of amino groups	9.9E-03	3 (21)

Figure 1. Bi-miR-200c PD identifies 520 putative miR-200c target genes. (**a**) Volcano plot for Bi-miR-200c PD showing the log2-transformed enrichment ratio for each probe detected above background in the microarray vs its corresponding log10-transformed *P*. Probes with a PD-enrichment ratio of $R > 2$ are shown in dark gray and the horizontal dashed line corresponds to $P = 0.01$. 520 genes fit the criteria $R > 2$ and $P < 0.01$. Probes detecting *Zeb2* (a known miR-200 target) and each of the 18 putative miR-200c targets chosen for validation are labeled (if there were multiple probes for the same gene, the one with the highest *R* was labeled). (**b**) Fold enrichment of hexamer sequences complementary to each 6-nucleotide segment of mature miR-200c in the CDS or 3′UTR of the 520 PD genes. Gray shading indicates hexamer enrichment ratios that did not differ by $P < 0.01$ from that in equal-sized gene sets drawn randomly from all expressed genes. *$P = 0.002$; **$P < 0.0001$. (**c**) Cumulative distribution plots of mRNA reduction in 4TO7 cells after transfection with miR-200c vs cel-miR-67 comparing the set of genes enriched in the Bi-miR-200c PD ($R > 2$ and $P < 0.01$ (red) or $P < 0.001$ (blue)) with all expressed genes detected above background in the microarray (black), and TargetScan (TScan) predictions of evolutionarily conserved (dark green) or all (light green) miR-200c target genes. Numbers in parentheses indicate the number of genes in each set (*, Kolmogorov–Smirnov test $P = 0.02$; **$P = 1.1E{-}6$). (**d**) Canonical pathways (Wikipathways) most enriched (adjusted *P* (Adj*P*) < 0.01) in the Bi-miR-200c PD ($R > 2$, $P < 0.01$). The PD genes in each enriched canonical pathway are listed in Supplementary Table S3.

during EMT. These included the master epithelial gene suppressors *Snail1* and *Zeb2*, the *Snail1* corepressors *Ywhag* (*14-3-3γ*) and *Ywhab* (*14-3-3β*),[42] the *Zeb2* corepressors *Smad2* and *Smad5*[41] and Smad corepressor, *Tgif1*. Relieving suppression of epithelial gene expression may also be enhanced by inhibition of *Anp32e*, which removes histone H2A.Z from promoters and enhancers to initiate transcription from paused RNA pol II.[43] *Snail1*, *Zeb2*, *Smad2* and *Smad5* are activated by the canonical TGF-β and BMP signaling pathways, suggesting that miR-200's induction of MET is accomplished in part by suppressing TGF-β and BMP signaling. *Map2k6* and *Mapk12* in the direct-interaction network also mediate Smad-independent TGF-β responses. These results suggest that miR-200 inhibits both Smad-dependent and -independent TGF-β and BMP signaling pathways. Another EMT-promoting transcription factor is *Xbp1*, which activates EMT in response to endoplasmic reticulum stress[44] and has a major role in the tumorigenicity, progression and recurrence of TNBC.[45] Genes known to regulate cell adhesion such as *Fhod1*,[46] *Wave1/Wasf1*[47] and *Map3k1*[48] were also in the miR-200 target gene interactome.

Experimental validation of the predictions of the Bi-miR-200c PD

To determine how reliable the predictions of miR-200-regulated genes were, we selected 18 putative target genes to test individually. These genes, which covered the full range of enrichment in the PD (2.3–33.7-fold, Figure 1a), were chosen because they were located at hubs in the interactome, were exceptionally enriched in the PD or might be related to miR-200's known functions. In particular, we were particularly interested in the five genes (*Ywhag*, *Ywhab*, *Snail1*, *Smad2* and *Smad5*) implicated with the known target *Zeb2* in suppression of epithelial gene expression. To validate these genes as miR-200c targets, we used an optimized version of the PD protocol[39] coupled with qRT-PCR to measure their enrichment for binding to Bi-miR-200c vs Bi-cel-miR-67 as a control. Bi-miR-200c PD in 4TO7 cells significantly enriched 14 of the 18 genes (Figure 3a). The four genes that did not confirm (*Skil*, *Bim/Bcl2l11*, *Mapk1* and *Bax*) had lower enrichment ratios (2.3–3.0). As a control, the housekeeping mRNAs *Sdha* and *Tuba1a* were not enriched in the Bi-miR-200c PD. To determine whether the confirmed pulled-down genes are direct miR-200 targets, we cloned their full 3′ UTRs into a dual-luciferase reporter vector and measured changes in luciferase activity in 4T1E cells after co-transfection with control or miR-200c mimic. miR-200c significantly repressed the 3′UTRs of 12 of the 14 genes by ~15–85% (Figure 3b), confirming that their 3′UTRs are directly regulated by miR-200c. The degree of repression of luciferase activity by miR-200c acting

on the 3'UTRs strongly correlated with their enrichment ratio in the PD, suggesting that those mRNAs that bound most strongly were the most highly regulated (Figure 3c). We next looked at whether miR-200c overexpression significantly reduced mRNA levels of the 14 genes in 4TO7 cells, which do not express endogenous miR-200, 48 h later. mRNAs for 13 of the 14 genes were significantly reduced by 25–85% (Figure 3d). To confirm this result in another cell line, we looked at whether miR-200c overexpression significantly reduced mRNA levels of the 12 of these genes expressed in NIH/3T3 cells 48 h later. mRNAs for all 11 target genes that confirmed in 4TO7 were significantly reduced by 15–80% (Supplementary Figure S1b). (The *Zfp36* gene that did not confirm in 4TO7 also was not suppressed in 3T3 cells.) The confirmed genes included the two genes (*Scmh1* and *Skp2*) whose 3′UTRs were not regulated by miR-200 overexpression. These might contain miR-200c miRNA recognition elements (MREs) in the CDS. However, this possibility was not examined experimentally. We also compared expression of the 14 target genes in 4TO7 and 4T1 cells. 4T1 cells are fully metastatic, express high levels of miR-200b/c/429 and have an epithelial-like morphology. In contrast, 4TO7 cells are poorly metastatic, do not express the miR-200 family and have a mesenchymal-like morphology.[4] Expression of the 13 of the 14 genes that were validated as targets was significantly down-regulated in 4T1 cells compared with 4TO7 cells (Figure 3e). For four genes (*Crtap*, *Smad5*, *Ywhag* and *Ywhab*), we also examined the effect of miR-200c overexpression on protein levels 48 h later (Figure 3f). All four proteins were significantly reduced after miR-200c overexpression.

To confirm that the 12 genes whose 3′UTRs were regulated by miR-200c were bona fide targets, we next wanted to determine whether their 3′UTRs contain miR-200c MREs. We used TargetScan,[49] PITA[50] and RNAhybrid[51] algorithms to identify potential MREs in their 3′UTRs (Supplementary Table S5). For all 12 genes, except *Smad2*, we identified one MRE, all but one of which had exact seed pairing and verified that it was functional by

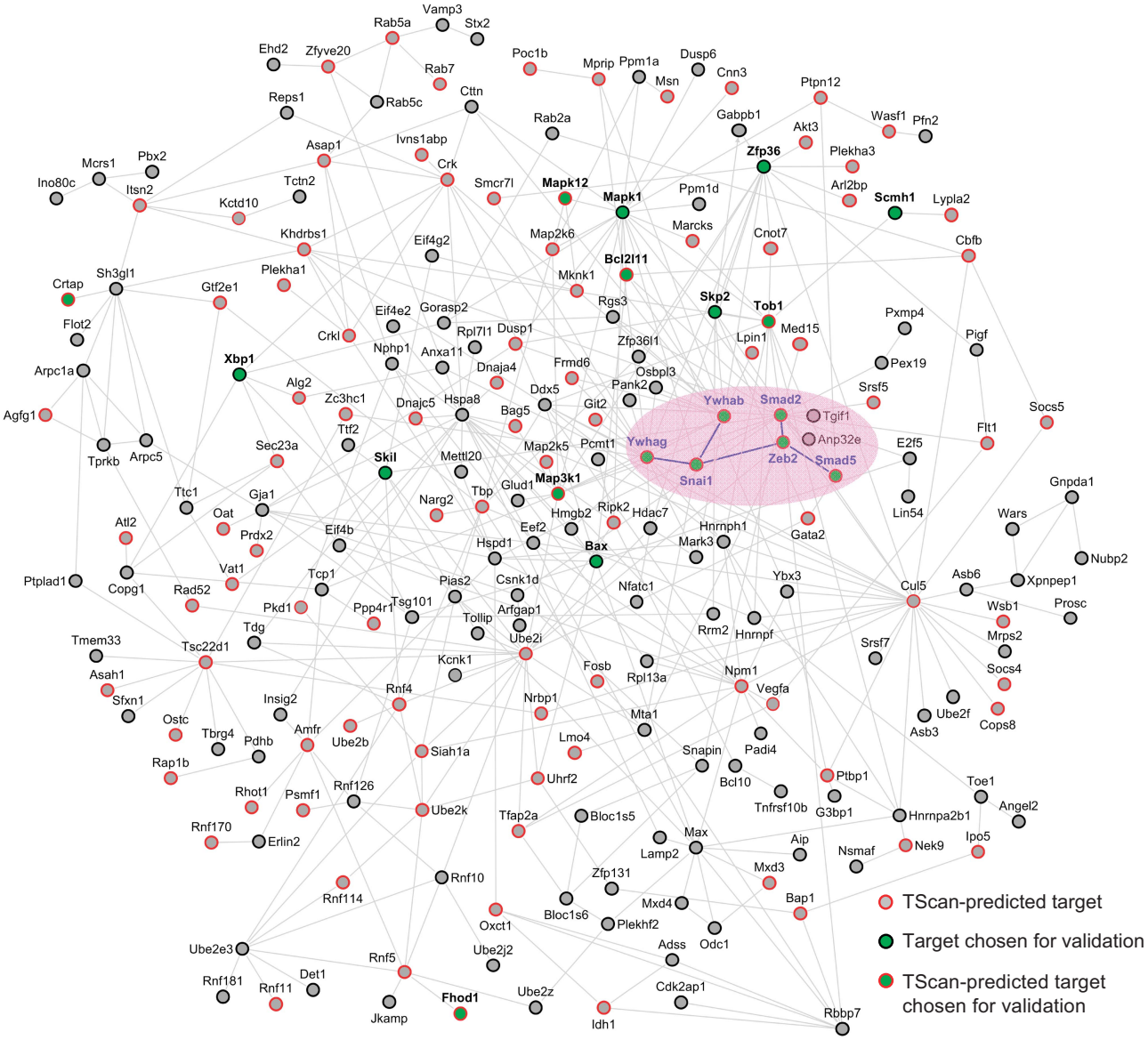
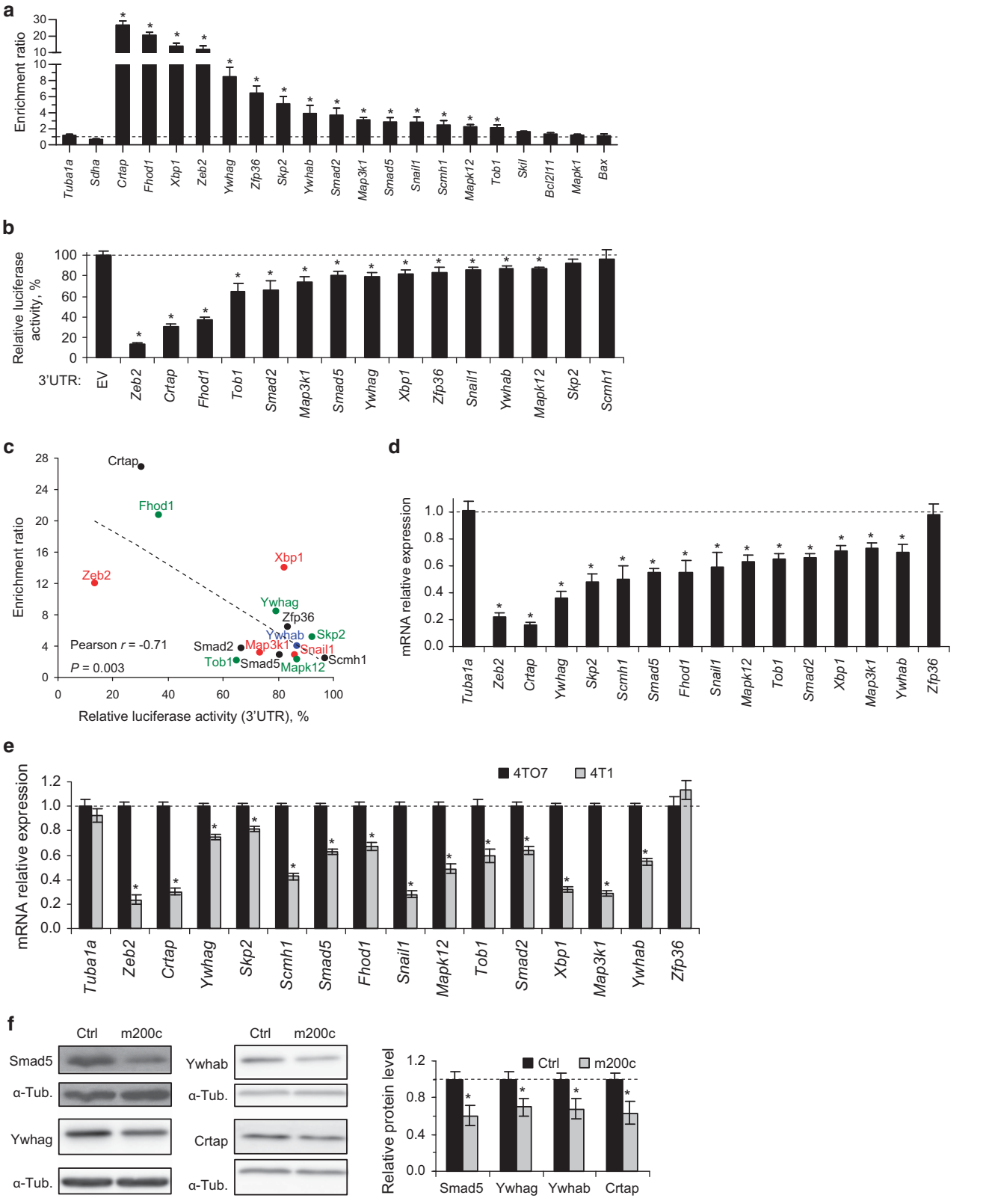
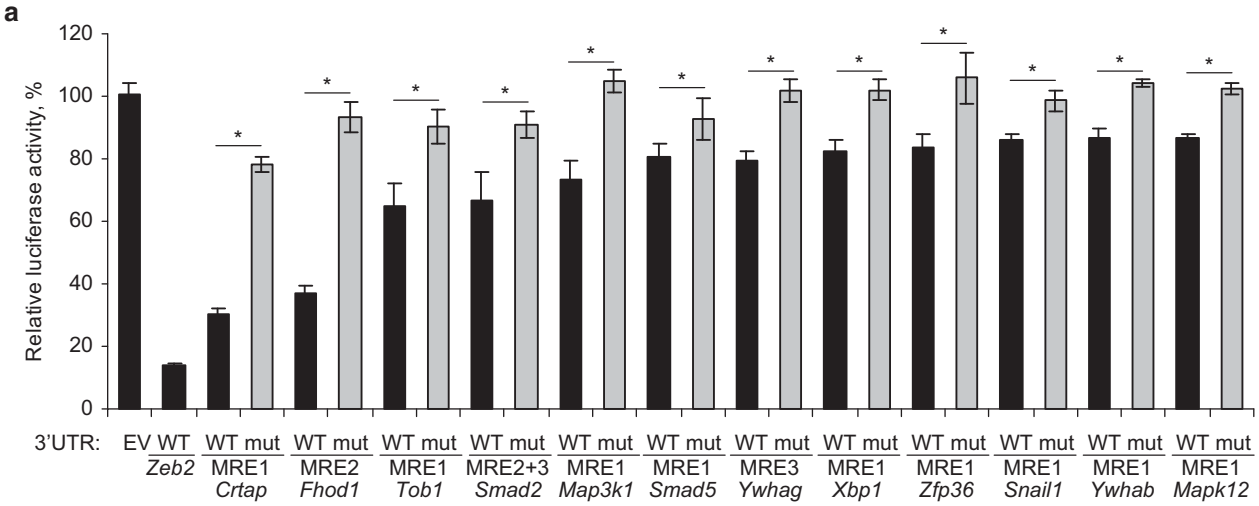


Figure 2. Interactome of directly-interacting pulled-down genes. 222 gene products enriched in the PD were annotated to interact directly. 45% of them (red border) are predicted by TargetScan (TScan) to be miR-200c targets in mice. 18 genes (green) were selected for experimental validation. Edges represent direct protein–protein or protein–DNA/RNA interactions. A module of genes involved in regulating transcription of epithelial genes is highlighted in pink.

showing that mutation of its seed sequence within the full 3′UTR restored luciferase reporter activity (Figure 4). For *Smad2*, we identified two MREs, neither of which had canonical seed pairing, but both had at least seven continuous exact matches to miR-200c overlapping with the seed; mutation of these pairing sequences in both MREs restored luciferase activity, but each individually only partially restored activity (data not shown). The results in Figures 3 and 4 taken together demonstrate that the miR-200c PD is highly





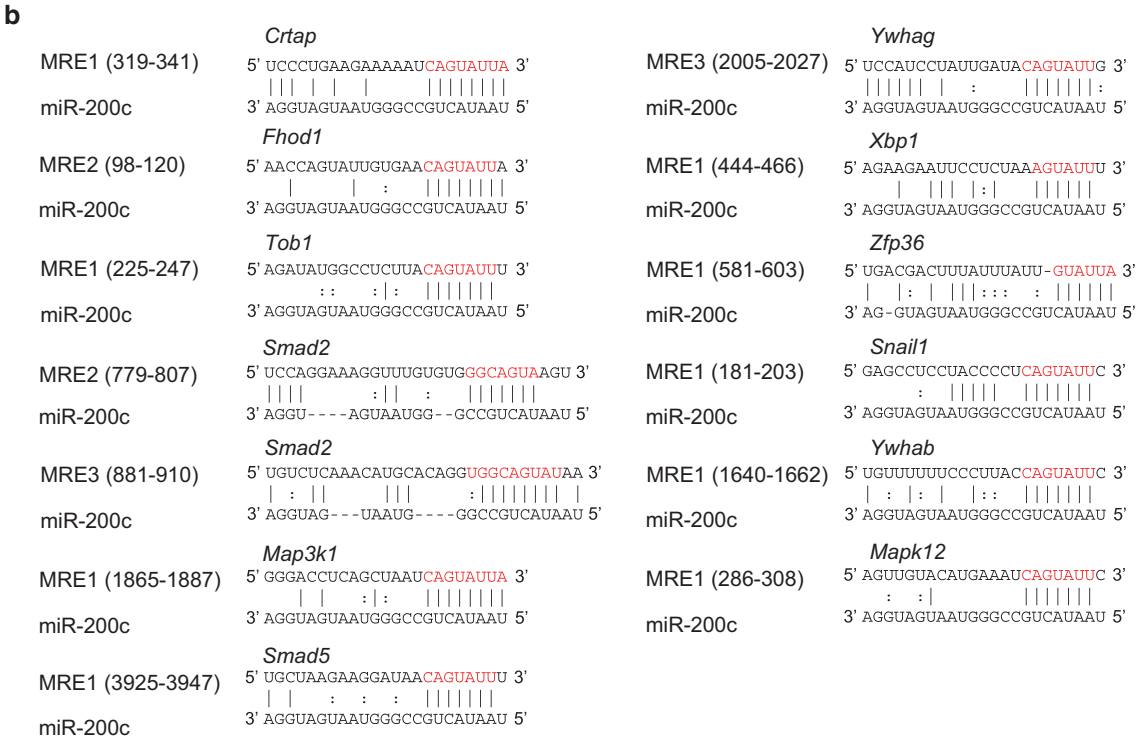


Figure 4. Identification of miR-200c MREs in the 3′UTRs of 12 new target genes. (**a**) 4T1E cells, co-transfected with control or miR-200c mimics and with a psiCHECK-2 luciferase reporter plasmid bearing the wild-type full 3′UTR of each gene or a mutated 3′UTR (mut) in which the miR-200c seed region of one or two MREs was mutated, were analyzed for luciferase activity 24 h later. Firefly luciferase activity was normalized to *Renilla* activity and results obtained for miR-200c mimic transfection were normalized to control mimic. Data represent mean ± s.d. of three independent experiments. *$P < 0.05$. (**b**) Functional miR-200c MREs in each 3′UTR are shown (numbers in parentheses indicate the position of the MRE in each 3′UTR). Point mutations in residues in red were introduced in each full 3′UTR in the mutant plasmid. Perfect matches are indicated by a line; G:U pairs, by a colon. Supplementary Table S5 shows all the predicted MREs that were tested.

Figure 3. Experimental validation of miR-200c target genes identified by the PD. (**a**) Bi-miR-200c enrichment ratio *R* in 4TO7 cells 24 h after transfection with Bi-miR-200c or Bi-cel-miR-67 mimics by qRT-PCR normalized to *Gapdh* for 18 genes chosen for validation. *R* = (miR-200c PD (gene mRNA/*Gapdh* mRNA))/(Cel-miR-67 PD (gene mRNA/*Gapdh* mRNA)). *Tuba1a* and *Sdha* are negative controls and *Zeb2* is a positive control. (**b**) miR-200c regulation of the 3′UTR of the 14 putative target genes confirmed to be pulled down by Bi-miR-200c. 4T1E cells, co-transfected with the dual luciferase reporter psiCHECK-2 bearing the full 3′UTR of each gene in the 3′ UTR of the Firefly reporter and with control or miR-200c mimics, were analyzed by luciferase assay 24 h later. Firefly luciferase activity was normalized to *Renilla* activity. Results obtained for miR-200c transfection were normalized to control miRNA transfection. (**c**) qRT-PCR PD enrichment ratio (**a**) correlates inversely with 3'UTR luciferase assay (**b**). (**d**) Effect of miR-200c overexpression on mRNA levels of the 14 confirmed pulled-down genes, assessed by qRT-PCR normalized to *Gapdh* 48 h after transfection of 4TO7 cells with miR-200c vs control miRNA mimics. (**e**) qRT-PCR comparison of miR-200 target gene mRNA expression, relative to *Gapdh*, in 4TO7 and 4T1 cells. (**f**) Immunoblot of total cell lysates extracted 48 h after transfecting 4TO7 cells with control (Ctrl) or miR-200c (m200c) mimics. α-Tubulin is a loading control. The bar graph shows the relative ratio of the signal (normalized to α-tubulin) in replicate experiments by densitometry, relative to cells transfected with control mimic. Bar graphs represent mean ± s.d. of three independent experiments. *$P < 0.05$.

specific for identifying miR-200c regulated genes. The low false positive rate in our data set suggests that miR-200c can potentially regulate hundreds of genes.

miR-200 targets Snail1 and Zeb2 transcription complexes

To confirm that miR-200 enhances epithelial gene expression in 4TO7 cells, we measured the effect of miR-200c overexpression on *Zeb2* and epithelial gene (*Cdh1*, the gene encoding E-cadherin; *Cldn3* and *Cldn7*, claudin genes) mRNAs by qRT-PCR (Figure 5a). As expected, *Zeb2* mRNA significantly declined, while the epithelial gene mRNAs increased 3–10-fold. To determine how much of the effect of miR-200 could be attributed to *Zeb2* KD, we compared the effect of miR-200 overexpression and *Zeb2* KD, both of which caused a comparable reduction in *Zeb2* mRNA. *Zeb2* KD only induced the epithelial genes by ~2-fold. Thus, miR-200 regulation of other genes likely contributes substantially to promote epithelial transition. In addition to *Zeb2*, miR-200c pulled down the E-cadherin transcription repressor *Snail1* and the corepressors *Smad2*/*Smad5* and *Ywhag*/*Ywhab* (Figures 2 and 3). Moreover, the mRNAs of all of these genes were significantly downregulated by miR-200 overexpression in 4TO7 cells. The other Smad and 14-3-3 genes were not pulled down with miR-200c and their mRNA levels were not reduced by miR-200c overexpression (Supplementary Table S2, Supplementary Figure S2a). To examine whether suppression of these other miR-200c target genes contributes to epithelial gene expression, we co-transfected 4TO7 cells with miR-200c or control miRNA mimics or siRNAs against *Zeb2, Snail1*, *Ywhag* or *Smad5* and a pGL3 reporter vector containing the Firefly luciferase gene under the control of the −178 to +92 fragment of the mouse E-cadherin promoter (Figure 5b). Depletion of endogenous *Snail1*, *Smad5* and *Ywhag* individually (verified by qRT-PCR, data not shown) enhanced transcription from the E-cadherin promoter about ~2.5-fold (almost as much as KD of *Zeb2*, which increased promoter activity ~3-fold) and more than the twofold increase in promoter activity caused by miR-200 overexpression.

Zeb2[52] and Snail1[53] are zinc finger transcription factors that bind to a consensus E-box hexanucleotide sequence (CACCTG) in the promoters of epithelial genes to repress their expression. *Cdh1* and *Cldn3* promoters have several E-box sequences (Figure 5c) that render them transcriptional targets of Zeb2 and Snail1. To determine whether miR-200 regulates the binding of endogenous Zeb2 and Snail1 to these promoters and transcription of epithelial genes, we performed Zeb2, Snail1 and RNA pol II chromatin-immunoprecipitation assays 1 day after 4TO7 cells were transfected with control or miR-200c mimics. miR-200 overexpression significantly enhanced RNA pol II occupancy of the *Cdh1* and *Cldn3* promoter and exonic regions, but not of the *Hbb1* (β-globin) promoter used as a control (Figure 5d), as expected, consistent with enhanced transcription of epithelial genes. In control cells, Zeb2 bound to the *Cdh1* promoter and Snail1 bound to both the *Cdh1* and *Cldn3* promoters, but neither bound above background to exonic regions of these genes (Figures 5e and f). miR-200 overexpression completely suppressed Zeb2 and Snail1 binding to these promoters. Thus miR-200c overexpression enhances transcription of epithelial genes, likely by reducing binding of both the Snail1 and Zeb2 transcriptional suppressor complexes.

Zeb2 is a well-validated strong target of miR-200c, which we confirmed in this mouse system by showing 80–90% reduction in luciferase reporter activity of the 3′UTR (Figure 3b) and mRNA level (Figure 3d) after miR-200 overexpression. We previously showed that Zeb2 protein becomes undetectable in 4TO7 cells upon miR-200c overexpression,[4] so it is not surprising that Zeb2 binding to epithelial gene promoters became undetectable. Changes in *Snail1* mRNA level and luciferase activity after miR-200c overexpression and Snail1 miR-200c binding were much less than for Zeb2 (Figures 3a and d), suggesting that it is a weaker target. In fact, miR-200c overexpression had no significant effect on Snail1 protein level (Figure 5g), presumably because other mechanisms compensated. Thus, reduced binding of Snail1 to the epithelial gene promoters (Figure 5f) was not due to a reduction in Snail1 protein. However, Snail1 binding to E-boxes and its repressor activity requires it to bind to Ywhag or Ywhab, since removing the Snail1 14-3-3 binding domain completely blocks its binding and repressor activity.[42] Ywhag and Ywhab protein levels declined with miR-200c overexpression (Figure 3f). Reduced 14-3-3 protein levels by miR-200c targeting likely contributed to the absence of detectable Snail1 binding to epithelial gene promoters, although other unknown mechanisms may have contributed.

miR-200 inhibits the EMT-promoting effects of TGF-β and BMP-2

Genes that mediate TGF-β signaling, which promotes mesenchymal properties, were significantly enriched in the miR-200 target network (Figure 1d). Both *Smad2 and Smad5*, the receptor Smads (R-Smad), which we validated as miR-200 targets (Figures 3 and 4 and Supplementary Figure S2a), were phosphorylated in response to TGF-β treatment of 4TO7 and 4T1 cells, as has been previously reported (Supplementary Figure S2b).[54] To investigate the effect of miR-200 on TGF-β signaling, we treated 4TO7 and 4T1 cells transfected with miR-200c or control miRNA with TGF-β or control vehicle. In both cell lines, TGF-β significantly reduced *Cdh1* expression as expected and also upregulated *Zeb2*. Cells overexpressing miR-200 significantly inhibited these TGF-β-mediated effects in both cell lines (Figure 6a). BMPs belong to the TGF-β superfamily, but mediate their signaling via R-Smads 1, 5 and 8. To study the effect of miR-200 on BMP signaling further, we treated 4TO7 cells transfected with miR-200c or control miRNA with BMP-2. BMP-2 treatment, as expected, significantly decreased the expression of two epithelial genes (*Cdh1* and *Cldn3*) and increased phosphorylation of the BMP-activated R-Smads (Figures 6b and c). miR-200 overexpression also significantly inhibited both these effects and also decreased Smad5 protein (Figures 6b and c). miR-200 did not affect *Smad1* and *Smad8* levels (Supplementary Figure S2a). Reduced Smad1/5/8 phosphorylation was recapitulated by knocking down *Smad5* (Figure 6c), suggesting that miR-200 represses BMP signaling by suppressing its signaling mediator, *Smad5*. Thus miR-200 inhibits both TGF-β and BMP signaling.

Smad5 gene regulation contributes to miR-200 enhancement of epithelial gene expression

The experiments performed so far have involved miR-200c overexpression, which can sometimes exaggerate the role or mechanism of endogenous miRNAs. To investigate the contribution of targeting individual gene targets to miR-200-mediated MET, we first assessed the effect on epithelial gene expression of knocking down some of the epithelial gene transcriptional repressor and corepressor genes in 4TO7 cells. KD of *Zeb2* or *Snail1* increased E-cadherin and claudin-3 mRNA and protein expression as expected (Figures 6d and e), and knocking down both was additive. *Smad5* or *Ywhag* KD did not affect epithelial gene expression, probably because other family members, known to have overlapping functions, may have substituted. The effect of miR-200c overexpression and *Zeb2*, *Snail1*, *Smad5*, *Ywhag* and *Crtap* KD on E-cadherin expression and cell morphology of 4TO7 cells was also analyzed by fluorescence microscopy (Supplementary Figure S3). Overexpression of miR-200c in 4TO7 cells increased E-cadherin protein and changed 4TO7 morphology from spindle-shaped to cobblestone-forming epithelial-like cells. Although there was no clear microscopic effect of *Snail1* KD, KD of *Zeb2* increased E-cadherin protein expression and epithelial morphology, and knocking down *Zeb2* together with *Snail1* was additive. KD of *Smad5*, *Ywhag* or *Crtap* individually did not affect

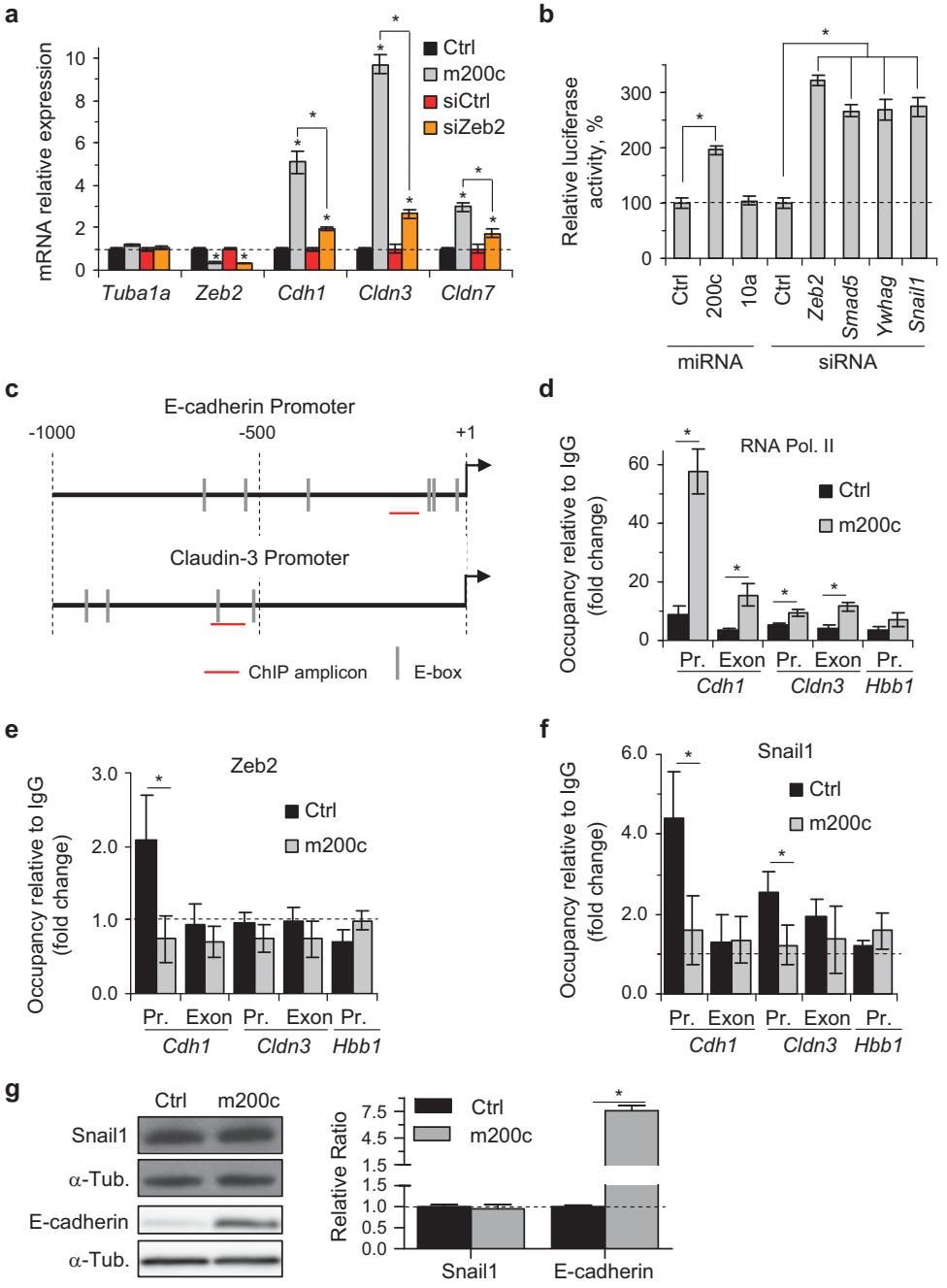


Figure 5. miR-200 induces epithelial gene expression by repressing Zeb2 and Snail1 complexes. (**a**) miR-200c overexpression enhances epithelial gene expression more than *Zeb2* KD. Gene expression in 4TO7 cells was analyzed by qRT-PCR, relative to *Gapdh*, 48 h after transfection with the indicated siRNA or miRNA mimics, relative to control siRNA or miRNA transfection. (**b**) Effect of miR-200c overexpression and *Zeb2*, *Smad5*, *Ywhag* and *Snail1* KD on the transcriptional activity of a luciferase reporter plasmid bearing the − 178 to +92 fragment of the *Cdh1* promoter. 4TO7 cells were co-transfected with the Firefly luciferase promoter construct and a control *Renilla* reporter vector and 24 h later with the indicated siRNA or miRNA mimics. Relative luciferase activity measured 48 h after the last transfection was normalized to that in cells transfected with the respective control siRNA or miRNA. (**c**) Schematic of putative Zeb2 and Snail1 binding sites (E-boxes) in the 1000-bp genomic sequence upstream of the transcription start sites (nucleotide +1) of the mouse *Cdh1* (E-cadherin) and *Cldn3* (claudin-3) genes. Sequences amplified by qPCR are indicated in red. (**d**–**f**) Binding of RNA polymerase II, Zeb2 or Snail1 to sequences in the *Cdh1* or *Cldn3* promoter (Pr) or exon 24 h after transfecting 4TO7 cells with control (Ctrl) or miR-200c (m200c) mimics was analyzed by chromatin immunoprecipitation (ChIP). Graphs show the binding to the specific antibody compared with binding to the same amount of IgG as measured by qPCR. The *Hbb1* (β-globin) promoter region was used as a control. (**g**) Immunoblot analysis of total protein extracted 48 h after 4TO7 cells were transfected with control (Ctrl) or miR-200c (m200c) mimics. α-Tubulin is a loading control. The bar graph shows the relative ratio of the signal (normalized to α-tubulin) in replicate experiments by densitometry, relative to cells transfected with control mimic. Bar graphs show mean ± s.d. of three independent experiments. $^*P < 0.05$.

E-cadherin protein expression, consistent with qRT-PCR and immunoblotting analyses.

To probe the contribution of Ywhag and Smad5 suppression in miR-200c-mediated epithelial gene activation, we co-transfected 4TO7 cells with miR-200c or control miRNA and miR-200c-resistant expression plasmids for these genes. Ywhag and Smad5 protein increased in these cells even in the presence of exogenous miR-200c (Supplementary Figure S2c). BMP-2, which increased phosphorylation of Smad1/5/8 (Figure 6c), reduced epithelial gene expression and the induction of epithelial genes by miR-200 (Figure 6b). Expression of miR-200-resistant *Smad5* slightly, but significantly, reduced miR-200c-mediated E-cadherin and claudin-3 upregulation in cells treated with the Smad5 activator (BMP-2), but not in cells not treated with BMP-2. Thus, miR-200 suppression of endogenous Smad5 contributes modestly to promoting the MET. However, overexpression of miR-200-resistant *Ywhag* had no significant effect in this assay.

The miR-200c target *Crtap* promotes cell invasion

Crtap was one of the strongest miR-200c targets, based on its enrichment in the PD (~34-fold) and reduction in luciferase activity and mRNA expression (Figure 3). Not much is known about Crtap. Crtap is an enzyme cofactor that facilitates prolyl-hydroxylation of type I collagen, and its homozygous mutation leads to osteogenesis imperfecta, a severe defect in bone mineralization. Of note, a recent study suggests that prolyl-hydroxylation of type I collagen is required for collagen binding to decorin, a proteoglycan that captures TGF-β in the extracellular matrix.[55] KD of *Crtap* had no effect on epithelial gene expression (Figures 6d and e). One of the hallmarks of mesenchymal cells and TGF-β stimulated cells is their motility and ability to invade basement membranes. However, Crtap has no known role in cell motility. To determine whether suppression of *Crtap* or the epithelial gene suppressor targets of miR-200c promote invasion, we knocked them down in 4TO7 cells and measured the effect on cell invasion across a collagen-containing Matrigel-coated membrane (Figure 6f) (KD of these genes had no effect on cell viability; Supplementary Figure S2d). KD of *Zeb2* reduced invasivity comparably to transfection of miR-200c. KD of *Smad5*, *Ywhag* or *Snail1* inhibited invasivity even more strongly than either miR-200 overexpression or *Zeb2* KD. However, KD of *Crtap* had the most potent inhibitory effect on invasion. These results suggest that Crtap and other miR-200-regulated gene products have unexpected roles in cell motility/invasion.

CRTAP and miR-200 expression is inversely correlated in NCI-60 tumors and primary human breast cancers

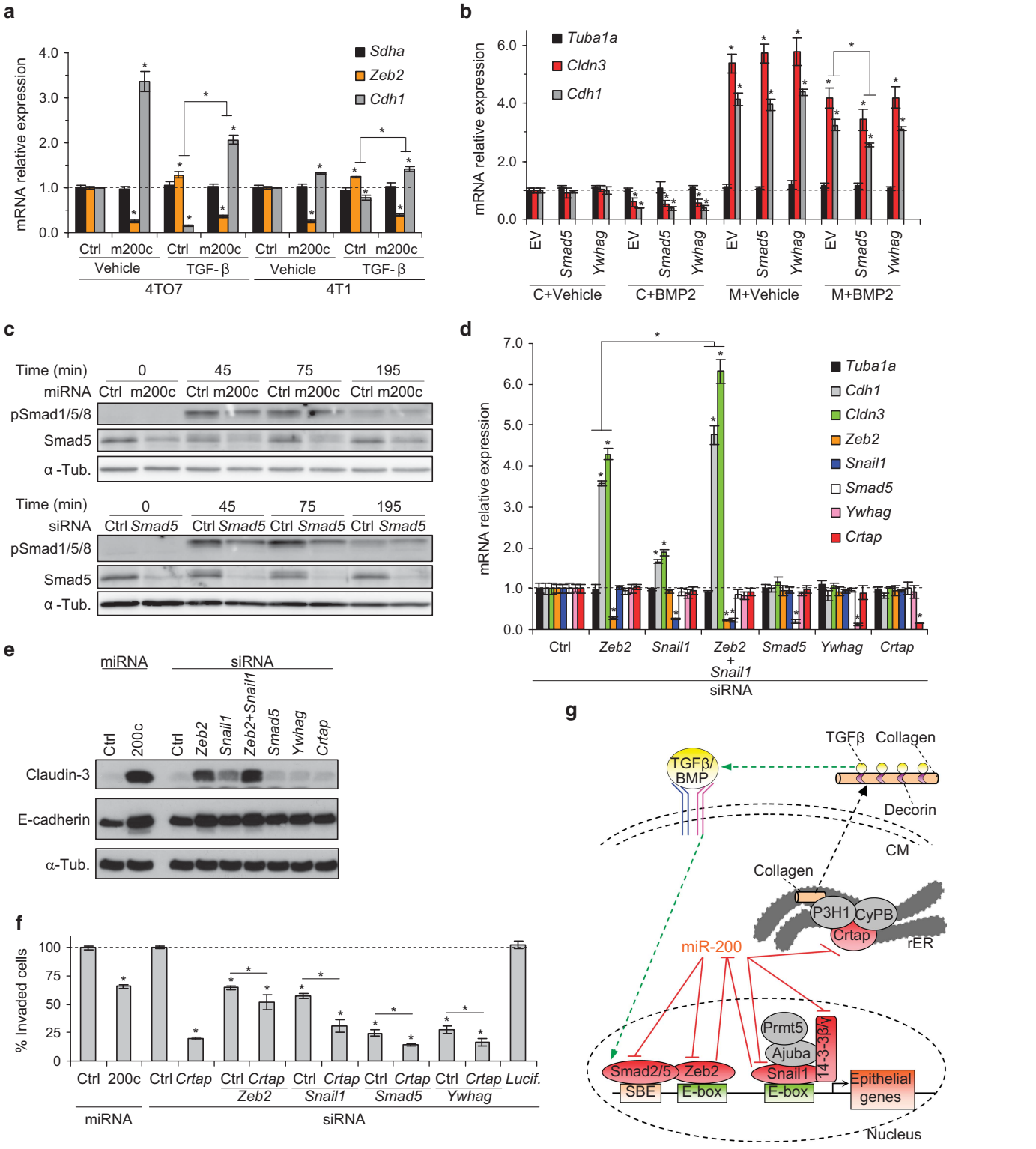
To further study the potential relationship between *Crtap* and miR-200c in cancer, we analyzed the correlation between *CRTAP* and miR-200c expression in the NCI-60 panel of human cancer cell lines. The miR-200 MRE we identified in the mouse gene is evolutionarily conserved in mammals, and human *CRTAP* is a high confidence TargetScan-predicted miR-200c target, which contains a perfect 8-mer seed binding site (Supplementary Figure S4a). *CRTAP* expression was almost as strongly inversely correlated with miR-200c expression in the NCI-60 panel of human cancer cell lines (Pearson $r = -0.44$, $P < 0.05$) as *ZEB2* expression ($r = -0.50$, $P < 0.05$). The housekeeping gene *SDHA* used as a control was not significantly correlated. We also analyzed data from a study that performed genome-wide mRNA and miRNA profiling in 101 human primary breast cancer samples.[56] *CRTAP* and *ZEB2* expression was also significantly inversely correlated with miR-200c expression in that cohort (Supplementary Figure S4b). These findings suggest that *CRTAP* is a physiologically relevant miR-200 target in human cancer.

DISCUSSION

In this study, we used a biochemical method with high specificity to identify without bias candidate genes that bind to Bi-miR-200c as putative miR-200c target genes. In total, 520 gene mRNAs were enriched at least twofold in binding to miR-200c with $P < 0.01$ and 196 were similarly enriched with greater confidence ($P < 0.001$).

Figure 6. *Snail1*, *Smad5*, *Ywhag* and *Crtap* are physiologically important miR-200c targets. (**a**) Effect of miR-200c overexpression on TGF-β signaling. qRT-PCR data were normalized to *Gapdh.* 4TO7 and 4T1 cells were transfected with control (Ctrl) or miR-200c (m200c) mimics in the presence of TGF-β or vehicle. RNA was harvested 48 h after transfection. For each cell line, results were normalized to transfection with control miRNA and treatment with vehicle. Asterisks above bars indicate significant differences in relative mRNA expression compared with control miRNA-transfected, untreated samples, while asterisks above brackets indicate significant differences between paired control miRNA and miR-200-transfected samples. (**b**) Effect of overexpression of miR-200-resistant *Smad5* and *Ywhag* or empty vector control (EV) on miR-200c-mediated epithelial gene mRNA upregulation in cells treated with vehicle or BMP-2. qRT-PCR data were normalized to *Gapdh*. 4TO7 cells were co-transfected with the indicated pcDNA3.1 expression vectors and control (C) or miR-200c (M) mimics in the presence of BMP-2 or vehicle. RNA was harvested 48 h after transfection. All results were normalized to control conditions in which cells were co-transfected with pcDNA3.1 EV and control miRNA and treated with vehicle. Asterisks above bars indicate significant differences in relative mRNA expression compared with the vehicle-treated sample co-transfected with EV and control mimic, while an asterisk above a bracket indicates significant differences between paired EV and expression plasmid-transfected samples. (**c**) Effect of miR-200c overexpression (top) or *Smad5* KD (bottom) on BMP signaling. Immunoblot of total protein extracted at indicated times after BMP-2 stimulation of serum-starved 4TO7 cells that were transfected with the indicated miRNA or siRNA 48 h earlier. α-Tubulin is a loading control. (**d**) Effect of *Zeb2*, *Snail1*, *Smad5*, *Ywhag* and *Crtap* KD on epithelial gene expression. mRNA was analyzed by qRT-PCR relative to *Gapdh* 72 h after transfection of 4TO7 cells with indicated target gene siRNAs. Asterisks above bars indicate significant differences in relative mRNA expression compared with control siRNA-transfected samples, while asterisk above brackets indicates significant difference when *Snail1* KD is combined with *Zeb2* KD. (**e**) Immunoblot of total protein extracted 72 h after 4TO7 cells were co-transfected with control (Ctrl) or miR-200c (200c) mimics and indicated target gene siRNAs or control (Ctrl) siRNAs. α-Tubulin is a loading control. (**f**) Invasion through Matrigel of 4TO7 cells co-transfected with miRNA mimics or 50 nM target gene siRNAs and an siRNA against *Crtap* or control (Ctrl) siRNA. In (**d**–**f**) 50 nM of each siRNA was used. Ctrl siRNA was added if needed to achieve a final total siRNA concentration of 100 nM. Results were normalized to control siRNA or miRNA transfection. An siRNA targeting Firefly luciferase (Lucif.) was used as a control. Bar graphs show mean ± s.e.m. of three independent experiments. (**g**) Schematic model of how miR-200c induces epithelial gene expression. miR-200c inhibits the expression of multiple members of the Zeb2 and Snail1 epithelial gene suppressor complexes that are activated by BMP or TGF-β signaling. Both of these complexes also suppress miR-200 transcription in a negative feedback loop.[78,79] Additionally, miR-200 knocks down *Crtap*, which is required for type I collagen prolyl-hydroxylation by prolyl 3-hydroxylase (P3H1) in complex with cyclophilin B (CyPB) in the endoplasmic reticulum. This post-translational modification is required for collagen to bind to decorin, which sequesters TGF-β in the extracellular matrix. Thus, miR-200 KD of *Crtap* may change TGF-β capture by the extracellular matrix and alter TGF-β-stimulated EMT and invasivity. CM, cell membrane; rER, rough endoplasmic reticulum; SBE, SMAD-binding element. Bar graphs show mean ± s.d. of three independent experiments, unless otherwise specified. *$P < 0.05$.

Gene expression of these biochemically identified miR-200 putative target genes was much more significantly downregulated ($P = E{-}6$) after miR-200c overexpression than genes predicted to be miR-200c-regulated genes by the best available prediction algorithm (evolutionarily conserved TargetScan targets). The most significantly enriched pathways were EGFR and other cell surface receptor signaling pathways and the MAPK pathways activated secondarily. Importantly, genes involved in TGF-β receptor signaling, which promotes mesenchymal transition and mesenchymal properties, were significantly enriched, and BMP and TGF signaling effects were inhibited by miR-200 overexpression. These target genes included genes encoding the R-Smad transcription factors, *Smad2* (activated by TGF-β signaling) and *Smad5* (activated by BMP signaling), which we confirmed were miR-200c targets, and *Tgif*, a corepressor of Smad transcription, which we did not experimentally validate. Ectopic expression of



miR-200c-resistant *Smad5* slightly but significantly reduced epithelial gene suppression by BMP-2 when miR-200 was ectopically expressed. These results suggest that inhibition of TGF-β/BMP signaling genes by miR-200 contributes to maintain epithelial gene expression. miR-200 expression is known to be repressed by BMP-2[57] and TGF-β.[58] Thus, our data suggest that miR-200c reinforces the epithelial state not only by inhibiting epithelial gene suppressors, but also by inhibiting TGF-β/BMP signaling.

The miR-200 family is well known to promote epithelial transition through suppression of the epithelial gene transcriptional repressors, *Zeb1/2* and *Snail1/2* (Snail and Slug).[4,18–21,23,24] Here, we confirmed that *Zeb2* and *Snail1* were miR-200c target genes (*Zeb1* and *Snail2* were not expressed in these cells). However, although we previously showed that exogenous miR-200c expression strongly suppressed Zeb protein levels in these cells,[4] Snail1 protein levels did not change. Snail1 protein was also not downregulated in 4TO7 cells by miR-200c overexpression in a previous study.[8] Although *Snail1* mRNA was consistently downregulated by miR-200 and an MRE was identified, verifying that it is a bona fide target of miR-200, it was a weak target. Snail1 is a very labile protein with a short half-life (~25 min) regulated by phosphorylation and ubiquitylation.[59] Post-translational modifications of Snail1 protein likely counteracted the weak effect of miR-200 on *Snail1* mRNA. miR-200c orchestrates the expression of epithelial genes not only by suppressing key transcriptional repressors, but also by targeting multiple gene products that act as cofactors in the Zeb and Snail1 repressor complexes, namely Smad2 and Smad5 (with Zeb) and 14-3-3β and 14-3-3γ (with Snail1) (Figure 6g).[42] In fact miR-200c overexpression had a much stronger effect on epithelial gene expression than *Zeb2* KD, and KD of any of these genes strongly induced the E-cadherin promoter. As a consequence, miR-200 overexpression greatly reduced Zeb2 and Snail1 binding specifically to epithelial gene promoters and enhanced RNA pol II binding to these promoters and transcription of epithelial genes. Snail1 binding to promoters likely depends on the availability of its corepressors. A previous study showed that KD of the Snail1 corepressor Ajuba[60] also impairs Snail1 binding without changing Snail1 protein level. KD of other genes in the miR-200c PD, especially *Cbf-A* (*Nfyb*),[61] whose product is a component of the CBF-A/KAP-1/FTS-1 complex that is a master transcriptional regulator of mesenchymal genes, and the NURD complex component *Mta1*,[62] which enhances transcription of *Snail1/2* and their recruitment to epithelial gene promoters, may also contribute to promote the epithelial transition. Thus, the ability of miR-200c to foster MET is reinforced by its suppression of multiple components of epithelial gene repressor complexes, its suppression of TGF-β/BMP signaling and likely direct suppression of mesenchymal gene transcriptional activators as well.

We experimentally examined the accuracy of the PD predictions for 18 genes that were not known miR-200c targets when we began this study. In all, 14 of 18 (78%) identified as enriched by the PD and microarray analysis were found to be enriched by at least twofold by more accurate qRT-PCR analysis. The four genes that were not confirmed had a fold enrichment that was close to our arbitrary cutoff of twofold enrichment. We estimated the false positive rate of the PD method used here as 16%, which compares favorably with estimated false positive rates of the best bioinformatics prediction algorithms (which vary between 22 and 31%[63]) or Ago-immunoprecipitation (~40–70%[64]). It is similar to the estimated false positive rate of Ago HITS-CLIP (~13–27%[64]). Our method, which does not use cross-linking, has the advantages that it is simpler and easier to do and requires much fewer cells. The current estimate is similar to an estimate we made for the specificity of a miR-34a PD, where only 1 of 11 (9%) of a random list of putative miR-34a-regulated targets was not verified.[34] That paper required a higher enrichment in the PD (>2.5-fold) than we used here, which could explain the higher specificity. Increasing the cutoff for the microarray data would undoubtedly improve the specificity of the PD method for identifying bona fide target genes, but at the cost of reduced sensitivity. Performing more replicates, and/or changing from microarray analysis of bound mRNAs to RNA sequencing (as we recently described[39]) would likely increase the specificity of the PD. The 3′UTRs of 12 of the 14 confirmed enriched genes (86%) had miR-200-regulated 3′UTRs and we were able to identify 3′-UTR MREs for all of them. Four of these hits were meanwhile independently identified as miR-200-regulated genes by other groups, further supporting our results (*Snail1*,[23] *Smad2*,[23] *Fhod1*[65] and *Ywhag*[66]). Most of these new 3′UTR MREs (11/13) had exact seed matches, suggesting that recognition of target genes by miR-200c closely follows canonical seed pairing. The remaining two confirmed gene targets may either be false positives or have MREs in the CDS, since recent studies suggest that a large fraction of MREs are in coding regions[64,67,68] and the target gene list was significantly enriched for mRNAs that contain seed pairing sequences in the CDS. The extent of enrichment in the PD correlated well with the extent of miR-200 suppression by luciferase activity (Figure 3c) or mRNA downregulation after miR-200c overexpression (data not shown). These results suggest that strong enrichment in the PD can be used to identify the set of genes most highly regulated by a miRNA, which may be the more important targets.

The high level of specificity of the PD, coupled with the large number of enriched genes in one cell line that we examined, suggests that miR-200c is capable of regulating hundreds of genes. How many genes are suppressed under physiological conditions is uncertain, since our study relied on miR-200c overexpression where miR-200c levels are not rate-limiting. Moreover, how much the set of regulated genes varies among cell types is also unclear for miR-200c or, for that matter, for any individual miRNA. Another open question is how much the target genes for miR-200c are shared with miR-200b and miR-429, which have identical seed sequences or with miR-200a and miR-141, which, although they are considered family members, actually have seed sequences that differ by one nucleotide. The Bi-miRNA PD method could be used to address these questions.

Crtap, which was enriched in the Bi-miR-200c PD by ~30-fold by microarray binding to each of two probes (Supplementary Table S2) and by qRT-PCR (Figure 3a) and whose both mRNA and protein were substantially reduced by miR-200c overexpression (by ~6-fold and ~2-fold, respectively; Figures 3d and f), was found to have an unanticipated role in cell invasivity through Matrigel. *CRTAP* expression in human tumor cell lines and a human breast cancer cohort was inversely correlated with miR-200c expression, suggesting that *CRTAP* is also an important miR-200 target in humans. Crtap, which is hypomorphic in some patients with osteogenesis imperfecta, is an endoplasmic reticulum cofactor for prolyl 3-hydroxylation of collagen. Tumor cell secretion of collagen with reduced prolyl 3-hydroxylation may promote tumor invasion. Loss of *Crtap* may reduce collagen binding of small leucine-rich proteoglycans, such as decorin,[55] that sequester mature TGF-β within the extracellular matrix. In fact, many of the symptoms of osteogenesis imperfecta in $Crtap^{-/-}$ mice can be ameliorated by injection of anti-TGF-β. Changes in TGF-β capture by the extracellular matrix could alter TGF-β-stimulated EMT and invasivity. Another possibility is that modifications of prolines on other unknown secreted substrate(s) are responsible for enhanced invasivity/metastasis of miR-200 overexpressing cells. Further study of the role of Crtap in cancer and metastasis is warranted. Another highly enriched gene in the PD, which was previously identified as a miR-200c target gene,[65] which we showed is a strongly regulated miR-200c target gene, was *Fhod1*. Fhod1 is a formin that bundles and stabilizes actin filaments in lamellipodia and filopodia at the leading edge of moving cells and has been implicated in regulating cytokinesis during mitosis in an

E-cadherin-dependent pathway. Fhod1 is upregulated in cancer cells and during EMT and has been linked to enhanced invasion and metastasis in some models. In this study, we did not look at the functional consequences of *Fhod1* suppression on cancer properties. Because miR-200 expression enhances metastasis in some cancer models[4,7–10] and reduces it in others,[11–13] it will be worthwhile to study more the effect of miR-200 on *Crtap* and *Fhod1* in cancer cell migration and metastasis.

MATERIALS AND METHODS

Cell culture

4TO7 and 4T1 cells were kindly provided by Fred Miller (Wayne State University).[69] 4T1E cells were previously generated in our laboratory.[70] 4TO7, 4T1, 4T1E and NIH/3T3 cell lines were grown in DMEM (Gibco, Rockville, MD, USA) supplemented with 10% fetal bovine serum, 1 mM L-glutamine and penicillin/streptomycin (Gibco).

siRNA and miRNA mimics transfection

4TO7, 4T1 or NIH/3T3 cells were transfected with 50 nM negative control miRNA mimic (Dharmacon, Lafayette, CO, USA, CN-001000-01-10), miR-200c miRNA mimic (Dharmacon, C-310569-07-0010), control siRNA (Dharmacon, D-001210-04-05) or siRNAs against *Zeb2* (Dharmacon, M-059671-01-0005), *Snail1* (Dharmacon, M-062765-00-0005), *Ywhag* (Dharmacon, M-059307-00-0005), *Smad5* (Dharmacon, M-057015-00-0005), *Crtap* (Dharmacon, M-049986-00-0005) or Firefly luciferase (sequences listed in Supplementary Table S6) using Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA). Cells were incubated with miRNA or siRNA-containing lipid complexes, according to the manufacturer's protocol, for 4 h before culture supernatants were removed and replaced with complete growth medium. Cells were harvested 48 or 72 h post transfection for mRNA and protein analyses.

RNA isolation and quantitative RT-PCR

Total RNA was isolated using Trizol reagent (Ambion, Austin, TX, USA), treated with DNase I (Promega, Madison, WI, USA) and reverse transcribed using random hexamers and Thermoscript RT kit (Invitrogen) as per the manufacturer's protocols. qRT-PCR was performed in triplicate samples using Ssofast Evagreen qPCR assay (Bio-Rad, Hercules, CA, USA) on a Bio-Rad CFX96 C1000 Thermal Cycler. mRNA levels were normalized to the housekeeping gene *Gapdh*. miRNA was quantified in triplicate using the TaqMan MicroRNA Assay (Applied Biosystems, Foster City, CA, USA) as per the manufacturer's instructions and normalized to snoRNA234. Primer sequences are listed in Supplementary Table S6.

Immunoblot

Whole-cell lysates were prepared using RIPA buffer (150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris pH 8.0) supplemented with Complete Mini-protease Inhibitor Cocktail (Roche, Indianapolis, IN, USA). For phosphoprotein analysis, lysis buffer was further supplemented with 10 mM NaF, 2 mM Na_3VO_4 and 2.5 mM sodium pyrophosphate. Protein concentration was determined using the BCA Protein Assay (Pierce, Rockford, IL, USA). Samples were resolved on SDS-Page gels and transferred using a Transblot semi-dry transfer apparatus (Bio-Rad). Blots were probed with antibodies to Smad5 (sc-7443) and Ywhab (sc-59419) from Santa Cruz Biotechnology, Santa Cruz, CA, USA; Ywhag (EMD Millipore, Billerica, MA, USA, 05-639); Claudin-3 (Life Technologies, Carlsbad, CA, USA, 341700); Crtap (kind gift of Prof. Brendan Lee, Baylor College of Medicine); Snail1 (3895), phospho-Smad1/5/8 (9511), Smad2 (5339) and phospho-Smad2 (3101) from Cell Signaling, Danvers, MA, USA; E-cadherin (BD Biosciences, San Jose, CA, USA, 61081); and α-tubulin (Sigma, St Louis, MO, USA, T5168). Blots were quantified by densitometry.

Bi-miR-200c pull down

4TO7 cells (2.4×10^6) were transfected as described above with 50 nM Bi-miR-200c or Bi-cel-miR-67 miRNA mimics biotinylated at the 3′-end of the active strand (Dharmacon). Biotin PDs were performed 24 h after transfection as previously described.[34] The level of mRNA in the Bi-miR-200c or Bi-cel-miR-67 control PD was then quantified by mRNA microarray (Illumina, San Diego, CA, USA) and analyzed as described below. For technical validation of the PD using qRT-PCR, cells were transfected as described above and an optimized version of the PD protocol was used.[39] In this case, mRNA levels in the PD and input samples were first normalized to *Gapdh* and then the enrichment ratio of the control-normalized PD mRNA to the control-normalized input levels was calculated.

Microarray analysis

Total RNA (from two independent experiments) was amplified, labeled and hybridized to Illumina MouseRef-8 v2.0 expression beadchip (Illumina) mouse whole-genome expression arrays. The quality of the RNA was assessed before performing the microarray and the quality of the microarray data was assessed using the affyPLM and Affy software (Bioconductor, www.bioconductor.org). The microarray data were normalized using RMA[71] to reduce interarray variation. Probes were mapped to Entrez Gene identifications (IDs) using systematically updated annotations from AILUN (Array Information Library Universal Navigator).[72] The PD enrichment ratio was defined as (Bi-miR-200c PD/Bi-cel-miR-67 PD)/(Bi-miR-200c input/Bi-cel-miR-67 input) and calculated for each probe. A detection *P* cutoff of 0.05 was used to exclude probes in any of the control arrays. For each probe, a significance value associated with the enrichment ratio was obtained using a *t*-test that compared the PD to input ratio for the Bi-miR-200c experiments with the control Bi-cel-miR-67 experiments. Probes with a *P* of < 0.01, and an enrichment ratio of > 2 were considered 'hits' for the informatic analysis of the PD data.

Gene downregulation after Bi-miR-200c overexpression

The fold change in probe expression caused by Bi-miR-200c overexpression was calculated by the ratio of input RNA from Bi-miR-200c-transfected 4TO7 cells to the ratio of input RNA from Bi-cel-miR-67-transfected cells. A cumulative distribution function plot was used to compare the distribution of Bi-miR-200c overexpression fold changes for PD gene lists defined by different cutoffs. The Kolmogorov–Smirnov test was used for statistical comparisons between gene sets.

Analysis of miR-200c target genes by target prediction algorithms

To determine whether a gene was also a predicted target of miR-200c, the presence of miR-200c binding sites was analyzed using TargetScan 6.2 (www.targetscan.org/),[49,73,74] PITA (genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html)[50] or RNAhybrid (bibiserv.techfak.uni-bielefeld.de/rnahybrid).[51]

Hexamer analysis

The mature mmu-miR-200c sequence was downloaded from miRBase; all RefSeq mouse mRNAs were obtained from NCBI. Hexamers complementary to each position of the miR-200c sequence were found in the complete gene set and their frequency was length normalized (matches/kb). Significance of enrichment was calculated by Monte Carlo sampling without replacement from the complete expressed gene set as described.[34]

Pathway enrichment analysis and network visualization

Canonical pathway gene sets were compiled from Wikipathways[75] and genes not expressed in 4TO7 cells (as determined by using Bi-cel-miR-67-transfected 4TO7 microarray data using a detection *P* cutoff of 0.05) were removed. The hypergeometric test was used to assess the enrichment of these biological gene sets in the list of Bi-miR-200c PD genes. The interactome of directly-interacting pulled-down genes was generated using the Ingenuity Pathway Analysis software (Ingenuity Systems, www.ingenuity.com). Edges represent direct functional interactions (protein–protein or protein–DNA/RNA interactions) between gene products.

miRNA and mRNA correlation in NCI-60 cell lines and primary human breast cancers

The relationship between *CRTAP*, *SDHA*, *ZEB2* and miR-200c expression levels in the NCI-60 panel of cancer cell lines was analyzed from genome-wide mRNA and miRNA profiling data using Cellminer (http://discover.nci.nih.gov/cellminer).[76] Matched miRNA and mRNA profiling in human primary breast tumor samples was reported in GSE19783 deposited in GEO (www.ncbi.nlm.nih.gov/geo/). Multiple probes mapping to the same gene were aggregated using the median expression. Pearson correlation

with a one-sided test for negative correlation was used to investigate the correlation between miR-200c and *CRTAP*, *ZEB2* and *SDHA* expression levels.

Dual-luciferase reporter assay

The full 3′UTR of each gene was amplified by PCR from genomic DNA isolated from 4TO7 cells and cloned into the psiCHECK-2 vector (Promega) immediately upstream of the *Renilla* luciferase gene. These constructs were used to generate constructs containing mutations in putative miR-200c MREs within the full 3′UTR of indicated genes. The sequences of all constructs were verified by sequencing. 4T1E cells were co-transfected with 50 nM miRNA mimics together with psiCHECK-2 vector containing the 3′UTR of indicated genes. Cells were lysed 24 h after transfection in Passive Lysis Buffer (Promega) and Firefly and *Renilla* luciferase activities were measured using the Dual Luciferase Assay System (Promega) on a Synergy2 plate reader (Biotek, Winooski, VT, USA). *Renilla* luciferase activity levels were normalized to Firefly luciferase and results from three independent experiments were compared to activity in cells expressing empty psiCHECK-2 vector. Sequence of primers used for cloning 3′UTRs and to generate mutated 3′UTRs is listed in Supplementary Table S6.

E-cadherin promoter reporter assay

The − 178/+92 fragment of the mouse E-cadherin promoter was amplified by PCR from genomic DNA isolated from 4TO7 cells using oligonucleotides linked to *Kpn*I and *Xho*I restriction sites and cloned into the same restriction sites in the pGL3 vector (Invitrogen) fused to the Firefly luciferase reporter gene. To determine the activity of the E-cadherin promoter, 4TO7 cells (2×10^4 cells) were co-transfected in 24-well plates with 50 ng of the reporter and 50 ng of TK-*Renilla* plasmid (Promega). Cells were transfected with 50 nM siRNA or miRNA 24 h later and 48 h after that cells were then lysed in Passive Lysis Buffer (Promega) and Firefly and *Renilla* luciferase activities were measured consecutively using the Dual Luciferase Assay System (Promega) on a Synergy2 plate reader (Biotek). Firefly luciferase activity was normalized to *Renilla* luciferase activity and relative to activity measured for control siRNA or miRNA transfection.

Chromatin immunoprecipitation assay

4TO7 cells (12.5×10^6) were transfected with 50 nM miRNA as above and 24 h later cross-linked with 1% formaldehyde for 10 min at room temperature. Crosslinking was stopped by adding glycine to a final concentration of 0.125 M. Cells were washed with cold PBS and kept on ice for 10 min in 25 mM HEPES (pH 7.8), 1.5 mM $MgCl_2$, 10 mM KCl, 0.1% NP-40, 1 mM DTT and Complete Mini-protease Inhibitor Cocktail (Roche). Following homogenization, nuclei were pelleted and resuspended in sonication buffer (50 mM HEPES (pH 7.9), 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na deoxycholate, 0.1% SDS, and Complete Mini-protease Inhibitor Cocktail (Roche)) and sonicated on ice to an average length of 200–1500 bp. Chromatin from ~5×10^6 cells was used for each immunoprecipitation, which was carried out overnight in the presence of 25 µl Protein G Dynabeads (Life Technologies, 10004D) using 2 µg of mouse monoclonal antibody against RNA Polymerase II (Santa Cruz Biotechnology, sc-56767), 5 µg of rabbit antiserum against Zeb2 (kind gift of Prof. Danny Huylebroeck, University of Leuven), 1 µg of rabbit antiserum against Snail1+Snail2 (Abcam, Cambridge, MA, USA, ab85931; we confirmed by qRT-PCR (data not shown) that *Snail2* is not expressed in 4TO7 cells) or equal amounts of rabbit IgG (Cell Signaling, 2729S) or mouse IgG (Santa Cruz Biotechnology, sc-2025). Samples were then extensively washed with sonication buffer, wash buffer A (50 mM Hepes pH 7.9, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na deoxycholate, 0.1% SDS and Complete Mini-protease Inhibitor Cocktail (Roche)), wash buffer B (20 mM Tris, pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% Na deoxycholate and Complete Mini-protease Inhibitor Cocktail (Roche)) and Tris-EDTA buffer to remove unbound DNA. Immunoprecipitated DNA was then eluted using elution buffer (50 mM Tris, pH 8.0, 1 mM EDTA, 1% SDS, 50 mM $NaHCO_3$) and treated with proteinase K (Ambion, AM2546) and RNAse (Sigma, R4642) overnight at 65 °C. DNA was then purified using a PCR purification kit (Qiagen, Valencia, CA, USA) and analyzed by qRT-PCR using the primers listed in Supplementary Table S6.

Transwell invasion assays

4TO7 cells (4×10^5) were transfected in 6-well plates with 50 nM siRNA or miRNA mimics as described above and 48 h later were incubated in serum-free medium for 2 h. Cells were then harvested with TrypLE Express (Invitrogen) and counted using a hemocytometer. A BD BioCoat Tumor Invasion System (BD Biosciences, 354165) containing BD Falcon Fluoroblock 24-Multiwell inserts (8-µm pore size; PET membrane) was used to measure invasion following the manufacturer's instructions. Briefly, 1.5×10^5 cells were added in serum-free medium to each insert and complete medium containing 10% fetal bovine serum was added into each basal chamber. After 20 h, invaded cells were quantified by labeling with calcein AM (BD Biosciences), and measuring the fluorescence underneath the membrane using a Synergy2 plate reader (Biotek).

Smad5 and *Ywhag* expression plasmids

Smad5 and *Ywhag* CDS were cloned into the multiple cloning site of pcDNA3.1(+) (Invitrogen) by RT-PCR using mRNA from 4TO7 cells. Primers used for PCR amplification are listed in Supplementary Table S6.

TGF-β treatment

4TO7 cells (4×10^5) were transfected in 6-well plates with 50 nM miRNA mimics as described above. Culture supernatants were removed after 4 h and replaced with complete medium containing 5 ng/ml TGF-β (Cell Signaling, 5231) or the same volume of vehicle (20 mM sodium citrate, pH 3.0). After 24 h, medium was replaced with fresh medium containing TGF-β or vehicle. Cells were harvested 24 h after that for mRNA analysis.

Plasmid DNA transfection and BMP-2 treatment

4TO7 cells (4×10^5) were co-transfected in 6-well plates as described above with 6 µg empty vector, *Ywhag* or *Smad5* expression vectors together with 50 nM miRNA mimics. Culture supernatants were removed after 4 h and replaced with complete medium containing 25 ng/ml BMP-2 (Cell Signaling, 4697) or the same volume of vehicle (10 mM HCl, 0.1% BSA). After 24 h, medium was replaced with fresh medium containing BMP-2 or vehicle. Immunoblot was performed on cell lysates harvested 24 h after that.

Cell viability assay

4TO7 cells (2×10^4), transfected in 24-well plates with 50 nM siRNA or miRNA mimics, were analyzed 48 h later for cell viability by CellTiter-Glo assay (Promega) following the manufacturer's protocol. An siRNA targeting *Plk1* (Dharmacon, M-040566-01-0005) was used as a control. Signals were normalized to control siRNA or miRNA treatment.

E-cadherin immunostaining

4TO7 cells, transfected with siRNA or miRNA mimics as described above, were plated 24 h later (10^4 cells/well) on a Lab-Tek chamber slide (Thermo Fisher, Waltham, MA, USA, 177445) in 0.2 ml and fixed 48 h later in 4% formaldehyde in PBS. The slides were washed twice with PBS, then incubated in blocking buffer (PBS with 0.5% BSA and 0.05% Triton-X) for 30 min, then for 1 h in blocking buffer plus E-cadherin antibody (BD Biosciences, 61081) at a final dilution of 1:100. After washing, slides were placed in blocking buffer with Alexa-Fluor-488-conjugated donkey anti-mouse secondary antibody (Life Technologies, A-21202) for 1 h and then washed and stained with 4 µg/ml DAPI in PBS and mounted using polyvinyl alcohol (Sigma, P-8136) aqueous mounting medium. Cells were imaged using an Axiovert 200 M microscope (Pan Apochromat, 1.4 NA; Carl Zeiss, Thornwood, NY, USA) and analyzed with SlideBook 4.2 (Intelligent Imaging Innovations Inc., Denver, CO, USA).

Statistical analysis

In all experiments, except where indicated, paired Student's *t*-tests (two-tailed) were used to analyze the difference between two groups of data.

Note added in proof

After the submission of this manuscript, a report[77] was published, which used Ago-HITS-CLIP to identify candidate miR-200a/b targets in MDA-MB-231 human breast cancer cells. They found that miR-200a/b controls actin cytoskeleton dynamics by regulating gene products involved in Rho-ROCK signaling, invadopodia formation, matrix metalloproteinase activity and focal adhesions. They also identified cell adhesion and TGF-β signaling as among the most enriched pathways for miR-200a/b targets in human cells in agreement with our findings for miR-200c targets in mouse.

Integrin-mediated cell adhesion (Figure 1d) and focal adhesion (adjusted $P = 2.1E{-}02$, data not shown) networks, which share 29 and 48 genes, respectively, with the actin cytoskeleton regulation network were among the most-enriched pathways we identify here for miR-200c targets in mouse. About 25% of the target genes we identified for miR-200c in mice were also found by Bracken *et al.* for miR-200a/b in humans. Of note, 50% of the 12 genes we validated as direct miR-200c targets in mouse by luciferase assay were also identified by Bracken *et al.* to be miR-200a/b targets (*Crtap*, *Map3k1*, *Smad5*, *Ywhag*, *Zfp36* and *Ywhab*), further supporting our findings. However, we did not identify control of actin cytoskeleton dynamics on its own as a significantly enriched process in miR-200c-regulated mouse cells. This difference in conclusions between our study, which studied a mouse TNBC that has both mesenchymal and epithelial properties, and theirs, which examined a mesenchymal human TNBC cell line, could reflect differences in technique, differences between the relevant miR-200 family members (miR-200c vs miR-200a/b), differences in mouse and human targets, or differences in gene expression between these cell lines. In fact, about half of the mouse miR-200c target genes we found that were not identified as targets of miR-200a/b in human cells do not have TargetScan-predicted miR-200 binding sites that are conserved between mice and humans. Additionally, the Bracken *et al.* study showed that epithelial human TNBC cell lines, which are not motile, do not express many of the actin motility-related targets identified in the mesenchymal cell line they studied. These differences suggest that some of the target genes and functions of a miRNA may be shared in different cellular contexts, while others may be context specific.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

R Perdigão-Henriques acknowledges the Portuguese Ministry of Science and Technology (FCT) for Ph.D. fellowship SFRH/BD/37188/2007. We thank Alex Amiet and Devin Leake (Dharmacon) for providing the biotinylated miRNAs, Linfeng Huang (Lieberman laboratory) for assistance with ChIP assays and Francisco Navarro (Lieberman laboratory) for helpful suggestions.

REFERENCES

1 Baum B, Settleman J, Quinlan MP. Transitions between epithelial and mesenchymal states in development and disease. *Semin Cell Dev Biol* 2008; **19**: 294–308.
2 Thiery JP. Epithelial-mesenchymal transitions in tumour progression. *Nat Rev Cancer* 2002; **2**: 442–454.
3 Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. *J Clin Invest* 2009; **119**: 1420–1428.
4 Dykxhoorn DM, Wu Y, Xie H, Yu F, Lal A, Petrocca F *et al.* miR-200 enhances mouse breast cancer cell colonization to form distant metastases. *PLoS ONE* 2009; **4**: e7181.
5 Wells A, Yates C, Shepard CR. E-cadherin as an indicator of mesenchymal to epithelial reverting transitions during the metastatic seeding of disseminated carcinomas. *Clin Exp Metastasis* 2008; **25**: 621–628.
6 Cheung KJ, Gabrielson E, Werb Z, Ewald AJ. Collective invasion in breast cancer requires a conserved basal epithelial program. *Cell* 2013; **155**: 1639–1651.
7 Gravgaard KH, Lyng MB, Laenkholm AV, Sokilde R, Nielsen BS, Litman T *et al.* The miRNA-200 family and miRNA-9 exhibit differential expression in primary versus corresponding metastatic tissue in breast cancer. *Breast Cancer Res Treat* 2012; **134**: 207–217.
8 Korpal M, Ell BJ, Buffa FM, Ibrahim T, Blanco MA, Celia-Terrassa T *et al.* Direct targeting of Sec23a by miR-200s influences cancer cell secretome and promotes metastatic colonization. *Nat Med* 2011; **17**: 1101–1108.
9 Dai Y, Xia W, Song T, Su X, Li J, Li S *et al.* MicroRNA-200b Is Overexpressed in Endometrial Adenocarcinomas and Enhances MMP2 Activity by Downregulating TIMP2 in Human Endometrial Cancer Cell Line HEC-1A Cells. *Nucleic Acid Ther* 2013; **23**: 29–34.
10 Li J, Du L, Yang Y, Wang C, Liu H, Wang L *et al.* MiR-429 is an independent prognostic factor in colorectal cancer and exerts its anti-apoptotic function by targeting SOX2. *Cancer Lett* 2013; **329**: 84–90.
11 Pichler M, Ress AL, Winter E, Stiegelbauer V, Karbiener M, Schwarzenbacher D *et al.* MiR-200a regulates epithelial to mesenchymal transition-related gene expression and determines prognosis in colorectal cancer patients. *Br J Cancer* 2014; **110**: 1614–1621.
12 Sun L, Yao Y, Liu B, Lin Z, Lin L, Yang M *et al.* MiR-200b and miR-15b regulate chemotherapy-induced epithelial-mesenchymal transition in human tongue cancer cells by targeting BMI1. *Oncogene* 2012; **31**: 432–445.
13 Ye F, Tang H, Liu Q, Xie X, Wu M, Liu X *et al.* miR-200b as a prognostic factor in breast cancer targets multiple members of RAB family. *J Transl Med* 2014; **12**: 17–27.
14 Li A, Omura N, Hong SM, Vincent A, Walter K, Griffith M *et al.* Pancreatic cancers epigenetically silence SIP1 and hypomethylate and overexpress miR-200a/200b in association with elevated circulating miR-200a and miR-200b levels. *Cancer Res* 2010; **70**: 5226–5237.
15 Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL *et al.* Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci USA* 2008; **105**: 10513–10518.
16 Taylor DD, Gercel-Taylor C. MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. *Gynecol Oncol* 2008; **110**: 13–21.
17 Toiyama Y, Hur K, Tanaka K, Inoue Y, Kusunoki M, Boland CR *et al.* Serum miR-200c is a novel prognostic and metastasis-predictive biomarker in patients with colorectal cancer. *Ann Surg* 2014; **259**: 735–743.
18 Burk U, Schubert J, Wellner U, Schmalhofer O, Vincan E, Spaderna S *et al.* A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO Rep* 2008; **9**: 582–589.
19 Korpal M, Lee ES, Hu G, Kang Y. The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *J Biol Chem* 2008; **283**: 14910–14914.
20 Park SM, Gaur AB, Lengyel E, Peter ME. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev* 2008; **22**: 894–907.
21 Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G *et al.* The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol* 2008; **10**: 593–601.
22 Chan YC, Khanna S, Roy S, Sen CK. miR-200b targets Ets-1 and is down-regulated by hypoxia to induce angiogenic response of endothelial cells. *J Biol Chem* 2011; **286**: 2047–2056.
23 Shin JO, Lee JM, Cho KW, Kwak S, Kwon HJ, Lee MJ *et al.* MiR-200b is involved in Tgf-beta signaling to regulate mammalian palate development. *Histochem Cell Biol* 2012; **137**: 67–78.
24 Liu YN, Yin JJ, Abou-Kheir W, Hynes PG, Casey OM, Fang L *et al.* MiR-1 and miR-200 inhibit EMT via Slug-dependent and tumorigenesis via Slug-independent mechanisms. *Oncogene* 2013; **32**: 296–306.
25 Shimono Y, Zabala M, Cho RW, Lobo N, Dalerba P, Qian D *et al.* Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells. *Cell* 2009; **138**: 592–603.
26 Iliopoulos D, Lindahl-Allen M, Polytarchou C, Hirsch HA, Tsichlis PN, Struhl K *et al.* Loss of miR-200 inhibition of Suz12 leads to polycomb-mediated repression required for the formation and maintenance of cancer stem cells. *Mol Cell* 2010; **39**: 761–772.
27 Samavarchi-Tehrani P, Golipour A, David L, Sung HK, Beyer TA, Datti A *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* 2010; **7**: 64–77.
28 Uhlmann S, Zhang JD, Schwager A, Mannsperger H, Riazalhosseini Y, Burmester S *et al.* miR-200bc/429 cluster targets PLCgamma1 and differentially regulates proliferation and EGF-driven invasion than miR-200a/141 in breast cancer. *Oncogene* 2010; **29**: 4297–4306.
29 Xia W, Li J, Chen L, Huang B, Li S, Yang G *et al.* MicroRNA-200b regulates cyclin D1 expression and promotes S-phase entry by targeting RND3 in HeLa cells. *Mol Cell Biochem* 2010; **344**: 261–266.
30 Yao CX, Wei QX, Zhang YY, Wang WP, Xue LX, Yang F *et al.* miR-200b targets GATA-4 during cell growth and differentiation. *RNA Biol* 2013; **10**: 465–480.
31 Schickel R, Park SM, Murmann AE, Peter ME. mir-200c regulates induction of apoptosis through CD95 by targeting FAP-1. *Mol Cell* 2010; **38**: 908–915.
32 Feng B, Wang R, Song HZ, Chen LB. MicroRNA-200b reverses chemoresistance of docetaxel-resistant human lung adenocarcinoma cells by targeting E2F3. *Cancer* 2012; **118**: 3365–3376.
33 Zhu W, Xu H, Zhu D, Zhi H, Wang T, Wang J *et al.* miR-200bc/429 cluster modulates multidrug resistance of human cancer cell lines by targeting BCL2 and XIAP. *Cancer Chemother Pharmacol* 2012; **69**: 723–731.
34 Lal A, Thomas MP, Altschuler G, Navarro F, O'Day E, Li XL *et al.* Capture of microRNA-bound mRNAs identifies the tumor suppressor miR-34a as a regulator of growth factor signaling. *PLoS Genet* 2011; **7**: e1002363.

35 Orom UA, Nielsen FC, Lund AH. MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. *Mol Cell* 2008; **30**: 460–471.
36 Krishnan K, Steptoe AL, Martin HC, Wani S, Nones K, Waddell N *et al.* MicroRNA-182-5p targets a network of genes involved in DNA repair. *RNA* 2013; **19**: 230–242.
37 Kang H, Davis-Dusenbery BN, Nguyen PH, Lal A, Lieberman J, Van Aelst L *et al.* Bone morphogenetic protein 4 promotes vascular smooth muscle contractility by activating microRNA-21 (miR-21), which down-regulates expression of family of dedicator of cytokinesis (DOCK) proteins. *J Biol Chem* 2012; **287**: 3976–3986.
38 Cloonan N, Wani S, Xu Q, Gu J, Lea K, Heater S *et al.* MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol* 2011; **12**: R126.
39 Tan SM, Kirchner R, Jin J, Hofmann O, McReynolds L, Hide W *et al.* Sequencing of captive target transcripts identifies the network of regulated genes and functions of primate-specific miR-522. *Cell Rep* 2014; **8**: 1225–1239.
40 Chaffer CL, Marjanovic ND, Lee T, Bell G, Kleer CG, Reinhardt F *et al.* Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. *Cell* 2013; **154**: 61–74.
41 Verschueren K, Remacle JE, Collart C, Kraft H, Baker BS, Tylzanowski P *et al.* SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad proteins and binds to 5'-CACCT sequences in candidate target genes. *J Biol Chem* 1999; **274**: 20489–20498.
42 Hou Z, Peng H, White DE, Wang P, Lieberman PM, Halazonetis T *et al.* 14-3-3 binding sites in the snail protein are essential for snail-mediated transcriptional repression and epithelial-mesenchymal differentiation. *Cancer Res* 2010; **70**: 4385–4393.
43 Obri A, Ouararhni K, Papin C, Diebold ML, Padmanabhan K, Marek M *et al.* ANP32E is a histone chaperone that removes H2A.Z from chromatin. *Nature* 2014; **505**: 648–653.
44 Tanjore H, Cheng DS, Degryse AL, Zoz DF, Abdolrasulnia R, Lawson WE *et al.* Alveolar epithelial cells undergo epithelial-to-mesenchymal transition in response to endoplasmic reticulum stress. *J Biol Chem* 2011; **286**: 30972–30980.
45 Chen X, Iliopoulos D, Zhang Q, Tang Q, Greenblatt MB, Hatziapostolou M *et al.* XBP1 promotes triple-negative breast cancer by controlling the HIF1alpha pathway. *Nature* 2014; **508**: 103–107.
46 Koka S, Neudauer CL, Li X, Lewis RE, McCarthy JB, Westendorf JJ *et al.* The formin-homology-domain-containing protein FHOD1 enhances cell migration. *J Cell Sci* 2003; **116**: 1745–1755.
47 Yamazaki D, Fujiwara T, Suetsugu S, Takenawa T. A novel function of WAVE in lamellipodia: WAVE1 is required for stabilization of lamellipodial protrusions during cell spreading. *Genes Cells* 2005; **10**: 381–392.
48 Cuevas BD, Abell AN, Witowsky JA, Yujiri T, Johnson NL, Kesavan K *et al.* MEKK1 regulates calpain-dependent proteolysis of focal adhesion proteins for rear-end detachment of migrating fibroblasts. *EMBO J* 2003; **22**: 3346–3355.
49 Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005; **120**: 15–20.
50 Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. *Nat Genet* 2007; **39**: 1278–1284.
51 Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. Fast and effective prediction of microRNA/target duplexes. *RNA* 2004; **10**: 1507–1517.
52 Comijn J, Berx G, Vermassen P, Verschueren K, van Grunsven L, Bruyneel E *et al.* The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. *Mol Cell* 2001; **7**: 1267–1278.
53 Batlle E, Sancho E, Franci C, Dominguez D, Monfar M, Baulida J *et al.* The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nat Cell Biol* 2000; **2**: 84–89.
54 Liu IM, Schilling SH, Knouse KA, Choy L, Derynck R, Wang XF *et al.* TGFbeta-stimulated Smad1/5 phosphorylation requires the ALK5 L45 loop and mediates the pro-migratory TGFbeta switch. *EMBO J* 2009; **28**: 88–98.
55 Grafe I, Yang T, Alexander S, Homan EP, Lietman C, Jiang MM *et al.* Excessive transforming growth factor-β signaling is a common mechanism in osteogenesis imperfecta. *Nat Med* 2014; **20**: 670–675.
56 Enerly E, Steinfeld I, Kleivi K, Leivonen SK, Aure MR, Russnes HG *et al.* miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. *PLoS ONE* 2011; **6**: e16915.
57 Itoh T, Nozawa Y, Akao Y. MicroRNA-141 and -200a are involved in bone morphogenetic protein-2-induced mouse pre-osteoblast differentiation by targeting distal-less homeobox 5. *J Biol Chem* 2009; **284**: 19272–19279.
58 Gregory PA, Bracken CP, Smith E, Bert AG, Wright JA, Roslan S *et al.* An autocrine TGF-beta/ZEB/miR-200 signaling network regulates establishment and maintenance of epithelial-mesenchymal transition. *Mol Biol Cell* 2011; **22**: 1686–1698.
59 Zhou BP, Deng J, Xia W, Xu J, Li YM, Gunduz M *et al.* Dual regulation of Snail by GSK-3beta-mediated phosphorylation in control of epithelial-mesenchymal transition. *Nat Cell Biol* 2004; **6**: 931–940.
60 Langer EM, Feng Y, Zhaoyuan H, Rauscher FJ, Kroll KL, Longmore GD *et al.* AjubaLIM proteins are snail/slug corepressors required for neural crest development in Xenopus. *Dev Cell* 2008; **14**: 424–436.
61 Venkov CD, Link AJ, Jennings JL, Plieth D, Inoue T, Nagai K *et al.* A proximal activator of transcription in epithelial-mesenchymal transition. *J Clin Invest* 2007; **117**: 482–491.
62 Mishra SK, Talukder AH, Gururaj AE, Yang Z, Singh RR, Mahoney MG *et al.* Upstream determinants of estrogen receptor-alpha regulation of metastatic tumor antigen 3 pathway. *J Biol Chem* 2004; **279**: 32709–32715.
63 Ahmadi H, Ahmadi A, Azimzadeh-Jamalkandi S, Shoorehdeli MA, Salehzadeh-Yazdi A, Bidkhori G *et al.* HomoTarget: a new algorithm for prediction of microRNA targets in Homo sapiens. *Genomics* 2013; **101**: 94–100.
64 Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 2009; **460**: 479–486.
65 Jurmeister S, Baumann M, Balwierz A, Keklikoglou I, Ward A, Uhlmann S *et al.* MicroRNA-200c represses migration and invasion of breast cancer cells by targeting actin-regulatory proteins FHOD1 and PPM1F. *Mol Cell Biol* 2012; **32**: 633–651.
66 Capobianco V, Nardelli C, Ferrigno M, Iaffaldano L, Pilone V, Forestieri P *et al.* miRNA and protein expression profiles of visceral adipose tissue reveal miR-141/YWHAG and miR-520e/RAB11A as two potential miRNA/protein target pairs associated with severe obesity. *J Proteome Res* 2012; **11**: 3358–3369.
67 Helwak A, Kudla G, Dudnakova T, Tollervey D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 2013; **153**: 654–665.
68 Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 2010; **141**: 129–141.
69 Aslakson CJ, Miller FR. Selective events in the metastatic process defined by analysis of the sequential dissemination of subpopulations of a mouse mammary tumor. *Cancer Res* 1992; **52**: 1399–1405.
70 Petrocca F, Altschuler G, Tan SM, Mendillo ML, Yan H, Jerry DJ *et al.* A genome-wide siRNA screen identifies proteasome addiction as a vulnerability of basal-like triple-negative breast cancer cells. *Cancer Cell* 2013; **24**: 182–196.
71 Chang TC, Wentzel EA, Kent OA, Ramachandran K, Mullendore M, Lee KH *et al.* Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Mol Cell* 2007; **26**: 745–752.
72 Chen R, Li L, Butte AJ. AILUN: reannotating gene expression data automatically. *Nat Methods* 2007; **4**: 879.
73 Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 2007; **27**: 91–105.
74 Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP *et al.* Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. *Nat Struct Mol Biol* 2011; **18**: 1139–1146.
75 Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C *et al.* WikiPathways: pathway editing for the people. *PLoS Biol* 2008; **6**: e184.
76 Reinhold WC, Sunshine M, Liu H, Varma S, Kohn KW, Morris J *et al.* CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res* 2012; **72**: 3499–3511.
77 Bracken CP, Li X, Wright JA, Lawrence DM, Pillman KA, Salmanidis M *et al.* Genome-wide identification of miR-200 targets reveals a regulatory network controlling cell invasion. *EMBO J* 2014; **33**: 2040–2056.
78 Garibaldi F, Cicchini C, Conigliaro A, Santangelo L, Cozzolino AM, Grassi G *et al.* An epistatic mini-circuitry between the transcription factors Snail and HNF4α controls liver stem cell and hepatocyte features exhorting opposite regulation on stemness-inhibiting microRNAs. *Cell Death Differ* 2012; **19**: 937–946.
79 Bracken CP, Gregory PA, Kolesnikoff N, Bert AG, Wang J, Shannon MF *et al.* A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. *Cancer Res* 2008; **68**: 7846–7854.

Supplementary Information accompanies this paper on the Oncogene website (http://www.nature.com/onc)