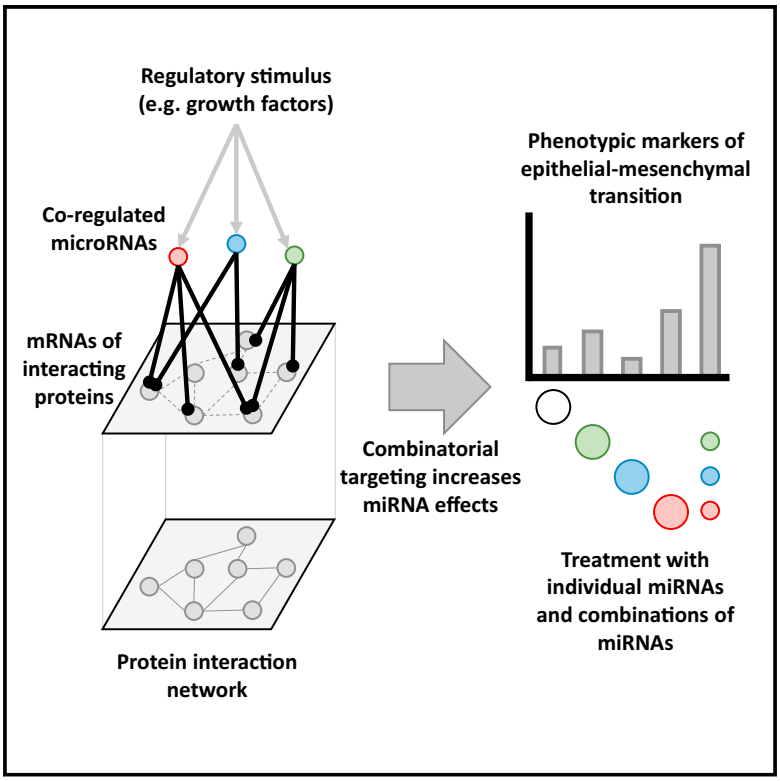
Article

Cell Systems

Combinatorial Targeting by MicroRNAs Co-ordinates Post-transcriptional Control of EMT

Graphical Abstract



Highlights

- miRNAs regulate EMT and the reverse MET as part of complex regulatory circuits
- miRNA overexpression can drive changes but often with off-target effects
- We identify miRNAs co-regulated in EMT that share functionally interacting targets
- Combinations of these miRNAs can alter phenotype at much lower abundances

Authors

Joseph Cursons, Katherine A. Pillman, Kaitlin G. Scheer, ..., Gregory J. Goodall, Cameron P. Bracken, Melissa J. Davis

Correspondence

cameron.bracken@sa.gov.au (C.P.B.), davis.m@wehi.edu.au (M.J.D.)

In Brief

Micro-RNAs (miRNAs) are important modulators of gene expression, providing an additional layer of post-transcriptional regulation. In epithelial-mesenchymal transition (EMT), interactions between miRNAs and key transcription factors allow for switching between phenotypes. We identify a set of miRNAs co-regulated during breast cancer EMT and demonstrate that treatment with combinations of pro-epithelial miRNAs at very low concentrations can drive highly specific phenotypic change with fewer off-target effects than individual miRNA overexpression. Our results suggest co-operative interactions should be considered when studying miRNA function.

Cursons et al., 2018, Cell Systems 7, 77–91
July 25, 2018
https://doi.org/10.1016/j.cels.2018.05.019

CellPress

Cell Systems
Article

Combinatorial Targeting by MicroRNAs Co-ordinates Post-transcriptional Control of EMT

Joseph Cursons,[1,2] Katherine A. Pillman,[3,4] Kaitlin G. Scheer,[3] Philip A. Gregory,[3,4] Momeneh Foroutan,[1,5] Soroor Hediyeh-Zadeh,[1] John Toubia,[3] Edmund J. Crampin,[6,7,8] Gregory J. Goodall,[3,4] Cameron P. Bracken,[3,4,*] and Melissa J. Davis[1,2,9,10,*]

[1]Bioinformatics Division, Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, VIC 3052, Australia
[2]Department of Medical Biology, University of Melbourne, Parkville, VIC 3010, Australia
[3]Centre for Cancer Biology, an Alliance of SA Pathology and University of South Australia, North Terrace, Adelaide, SA 5000, Australia
[4]Department of Medicine, University of Adelaide, Adelaide, SA 5005, Australia
[5]The University of Melbourne Department of Surgery, St. Vincent's Hospital, Fitzroy, VIC 3065, Australia
[6]Systems Biology Laboratory, University of Melbourne, Parkville, VIC 3010, Australia
[7]ARC Centre of Excellence in Convergent Bio-Nano Science, Department of Biomedical Engineering, University of Melbourne, Parkville, VIC 3010, Australia
[8]School of Mathematics and Statistics, University of Melbourne, Parkville, VIC 3010, Australia
[9]Department of Biochemistry, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Parkville, VIC 3010, Australia
[10]Lead Contact
*Correspondence: cameron.bracken@sa.gov.au (C.P.B.), davis.m@wehi.edu.au (M.J.D.)
https://doi.org/10.1016/j.cels.2018.05.019

SUMMARY

MicroRNAs (miRNAs) are important post-transcriptional regulators of gene expression, functioning in part by facilitating the degradation of target mRNAs. They have an established role in controlling epithelial-mesenchymal transition (EMT), a reversible phenotypic program underlying normal and pathological processes. Many studies demonstrate the role of individual miRNAs using overexpression at levels greatly exceeding physiological abundance. This can influence transcripts with relatively poor targeting and may in part explain why over 130 different miRNAs are directly implicated as EMT regulators. Analyzing a human mammary cell model of EMT we found evidence that a set of miRNAs, including the miR-200 and miR-182/183 family members, co-operate in post-transcriptional regulation, both reinforcing and buffering transcriptional output. Investigating this, we demonstrate that combinatorial treatment altered cellular phenotype with miRNA concentrations much closer to endogenous levels and with less off-target effects. This suggests that co-operative targeting by miRNAs is important for their physiological function and future work classifying miRNAs should consider such combinatorial effects.

INTRODUCTION

Epithelial-mesenchymal transition (EMT) is a reversible phenotypic switch that has gained significant attention due to its role in both normal physiological (e.g., gastrulation and wound healing) and pathological (e.g., fibrosis, metastasis, and chemoresistance) processes (Ye and Weinberg, 2015). Many aspects of EMT regulation are well known: it is driven by co-ordinated changes in gene expression, controlled by both transcription factors (particularly the ZEB, SNAIL, and TWIST families) and miRNAs (such as the miR-200 family) linked through multiple co-regulatory relationships (Bracken et al., 2016; Lu et al., 2013; Friard et al., 2010; Gosline et al., 2016; Re et al., 2009). Many studies investigating the role of specific miRNAs have employed overexpression or ectopic exposure at concentrations greatly exceeding physiological abundances. Given that miRNA binding affinities are dose dependent, this overexpression likely has a multitude of off-target effects, influencing gene transcripts that are not functional targets at endogenous miRNA levels (Witwer and Halushka, 2016). EMT-promoting stimuli (such as transforming growth factor β [TGF-β]) cause changes in the expression of multiple miRNAs, but induced abundances are typically far lower than those achieved by transfection with miRNA mimics. This suggests a discrepancy between the function of miRNAs with more modest endogenous expression levels and the effect of individual miRNAs when expressed at levels far beyond their physiological concentration. We hypothesized that the co-operative actions of multiple miRNAs may thus contribute to their function in controlling EMT.

The role of some miRNAs in EMT is well established; however, over 130 different miRNAs have been directly implicated in regulating EMT (Table S1), raising questions about the effects of common experimental manipulations (e.g., miRNA overexpression) and the extent to which these relationships are biologically meaningful or a reasonable reflection of endogenous function. As miRNAs can directly regulate a multitude of targets, there has been a focus on their characterization following manipulation; however, it appears that the primary effects of miRNA perturbation are mediated by transcription factors (Gosline et al., 2016), and thus by solely characterizing direct targets, the field has failed to capture the impact of miRNAs upon regulatory networks. Furthermore, it was recently reported that many predicted miRNA targets are functionally insensitive at

endogenous miRNA levels and many miRNA target sites identified by computational algorithms may be conserved due to other evolutionary drivers (Pinzón et al., 2017).

Significant questions remain regarding the role of miRNAs in EMT regulation: which miRNAs are truly capable of regulating the EMT/mesenchymal-epithelial transition (MET) process *in vivo* and during cancer progression; to what extent do miRNAs regulate EMT at physiological concentrations; what are the endogenous roles of individual miRNAs given that EMT-promoting stimuli drive co-ordinated up- and downregulation for dozens of miRNAs? To address these questions, we combined computational and experimental methods to identify co-regulated miRNAs that have the potential for functional co-operation during EMT. We explore the contribution of endogenous miRNAs to the regulation of EMT/MET and demonstrate that the combinatorial activities of co-regulated miRNAs is central to their natural function. To an extent, these co-operative functions can be achieved by the massive overexpression of individual miRNA components but this fails to fully re-capitulate the endogenous function and is associated with off-target effects.

We demonstrate that miRNAs provide a secondary regulatory layer after transcription, amplifying transcriptional effects on relevant EMT-associated processes (such as cell adhesion and extracellular matrix organization) while simultaneously buffering transcriptional effects on non-EMT genes. We also show that, at physiologically relevant expression levels, EMT is regulated by multiple miRNAs acting in a combinatorial manner. We observe a dominant role for miR-200c-3p, which is augmented by miR-141-3p, miR-182-5p, and miR-183-5p; these miRNAs are not only simultaneously suppressed during TGF-β-induced EMT and co-regulated across a large cohort of breast cancer samples, but computational analyses also suggest functional co-operation, co-targeting genes that are in close proximity within protein interaction networks. Furthermore, as miRNA function is proportional to abundance, we find that transfection with multiple miRNAs at sub-nanomolar concentrations is sufficient to regulate EMT without off-target effects that are evident at high concentrations. While this work seeks to better understand co-ordinated regulation of the EMT process, we anticipate that our findings of the importance of co-operative miRNA function at non-supra-physiological levels, and the global role of miRNAs in supporting and modulating transcriptional output, should be applicable to all miRNA systems.

RESULTS

The HMLE Cell Model of EMT Shares a Transcriptional Profile with Claudin-Low Metaplastic Tumors

Following prolonged (24-day) exposure to TGF-β, human mammary epithelial (HMLE) cells establish a stable sub-line (MesHMLE) with morphological (Figure 1A), protein (Figure 1B), and mRNA transcript (Figure 1C) changes indicative of EMT. Subsequent overexpression of miR-200c (MesHMLE + miR-200c) is sufficient to reverse morphological changes (Figure 1A) consistent with MET. However, molecular changes suggest MET is partial, with *VIM* (vimentin) remaining high despite increased *CDH1* (E-cadherin) and reduced *ZEB1* expression (Figures 1B and S1A).

We have previously examined the relative molecular phenotypes of cells on a 2D epithelial and mesenchymal landscape (Foroutan et al., 2017). To characterize our model system, gene set scores for the HMLE cell system (Figure 1D) were compared with TCGA breast cancer samples (The Cancer Genome Atlas Network, 2012) and a number of established breast cancer cell lines (Daemen et al., 2013; Heiser et al., 2012). The TGF-β-driven HMLE to MesHMLE transition is associated with a large reduction in epithelial score and a smaller increase in the mesenchymal score (Figure 1D). The subsequent MesHMLE + miR-200c cells have a similar mesenchymal score to the original HMLE cells, but only a partial restoration of the epithelial phenotype. This is consistent (Figures 1A–1C) and supports the notion of partial MET driven by exogenous miR-200c exposure. MesHMLE cells also show a similar molecular phenotype to basal B claudin-low cell lines (Blick et al., 2008) (Figure S1B) and have scores similar to a rare subset of TCGA breast cancer samples that typically have a metaplastic histological annotation (Table S2A; Figure S1C). These cancers are characterized by lower expression of epithelial markers and higher expression of mesenchymal markers than other samples (Figure S1D).

Post-transcriptional Gene Regulation Simultaneously Reinforces and Buffers Transcriptional Changes

During EMT, complex changes in gene expression are co-ordinated through transcriptional and post-transcriptional regulation, largely governed by transcription factors (TFs) and miRNAs, respectively. To assess their contributions we used exon-intron split analysis (EISA) (Gaidatzis et al., 2015): as most miRNA-mediated post-transcriptional regulation occurs after transcripts are exported from the nucleus to the cytoplasm, intronic read differences between two conditions (Δintron) are indicative of altered gene transcription, whereas differences between exon and intron read changes (Δexon − Δintron) suggest altered post-transcriptional regulation. To ensure that the Δintron and Δexon − Δintron data were robust to alignment and quantification methods, an alternative pipeline was run in parallel; the resulting data were largely consistent (Figure S2).

The EISA metrics show that transcriptional and post-transcriptional regulation can either work in unison or opposition. Genes that are both transcriptionally and post-transcriptionally upregulated or downregulated we annotate as "co-ordinately increased" (CI) or "co-ordinately decreased" (CD). Genes where the transcriptional increase is counteracted by post-transcriptional downregulation (or vice versa) we label as "increased, buffered" (IB) and "decreased, buffered" (DB) (Figure 2A; Table 1). Comparing HMLE and MesHMLE cells we expect post-transcriptionally downregulated genes to be targeted by miRNAs associated with a mesenchymal state. Conversely, the loss of repression from miRNAs predominantly expressed in the epithelial state would drive post-transcriptional upregulation.

There was a marked difference in the enrichment of gene ontology terms, with genes subject to co-ordinated regulation highly enriched for processes central to EMT (Figure 2B). For example, terms including "extracellular matrix" and "cell migration" were enriched for the CI gene set, while CD genes were involved in processes such as "cell junction organization" and "epithelial cell development." In contrast, genes with buffered

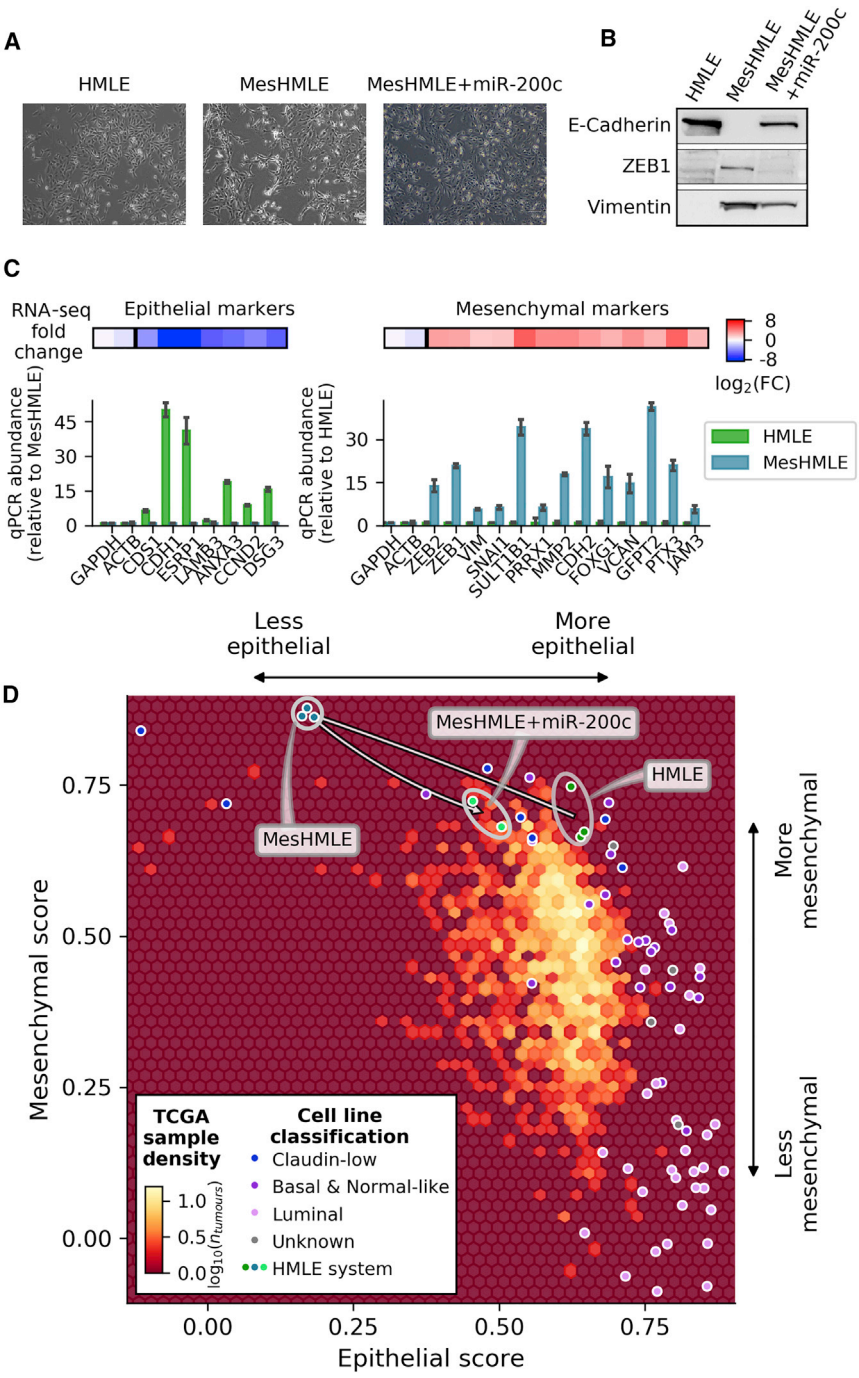


Figure 1. Characterizing Our Human Cell-Line Model of EMT

(A) The morphology of HMLE, MesHMLE (HMLE +24 days TGF-β), and MesHMLE cells exposed to 20 nM of miR-200c mimic (MesHMLE + miR-200c). Scale bar, 100 μm.

(B) Western blots showing protein expression for selected epithelial (E-cadherin, encoded by *CDH1*) and mesenchymal (ZEB1, vimentin) markers.

(C) Gene expression changes from RNA sequencing (RNA-seq) and qPCR transcript abundance data. *GAPDH* was used for qPCR normalization and *ACTB* is an internal control. Error bars represent SD.

(D) A density plot of epithelial and mesenchymal score for TCGA breast cancer samples (hexbin), overlaid with a scatterplot of epithelial and mesenchymal scores for individual breast cancer cell lines. The epithelial and mesenchymal gene sets derived by Tan et al. (2014) have distinct cell line and tissue signatures to account for the more complex composition of tissue, however, factors such as sample purity may influence the transcriptional profile.

expression had far less enrichment, indicating diverse genes with no strong functional similarity. Terms significant for the buffered sets were associated with relatively small numbers of genes and had no obvious EMT link.

During EMT miRNAs Co-regulate Functionally Related Transcripts

Within gene regulatory networks "feedback motifs" between miRNAs and TFs are relatively common, and specific relationships, such as that between miR-200 and the ZEB transcription factors (Bracken et al., 2015, 2016), are known to underpin EMT. The transcriptomic changes we observe during HMLE cell EMT (Figure 2) speak to a broader degree of transcriptional and post-transcriptional co-regulation, with miRNAs acting to both reinforce TF activity and buffer genes with altered transcription that are not essential for EMT.

Thus we sought to characterize the roles of endogenous miRNAs in co-ordinating EMT gene regulation. We ranked miRNAs by their relative abundance and magnitude of differential expression (Figure 3A). In general, miRNAs that decreased have been described as pro-epithelial and associated with MET, while miRNAs that increased in abundance are generally pro-mesenchymal and associated with EMT (Table S1). An exception is miR-204 which has been described as pro-epithelial in a number of studies, but undergoes a large increase in abundance with the HMLE-to-MesHMLE transition (Figure 3A). Epithelial- and mesenchymal-specific expression of the EMT-responsive miRNAs were also seen in TCGA breast cancer data, with a particularly strong positive correlation with epithelial score and negative correlation with mesenchymal score found for the most "pro-epithelial" miRNAs (Figure 3B).

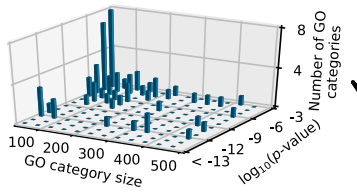
To determine the contribution of each miRNA from Figure 3A in directly controlling gene expression, we calculated the relative enrichment of high-confidence predicted targets within each of the EISA-partitioned gene sets (Figure 3C). Genes post-transcriptionally upregulated during EMT (the CI and DB gene sets)

A

Transcription factor type	Transcriptional regulation	Post-transcriptional regulation	Transcriptional effect	Post-transcriptional effect	Category of EISA result
Transcriptional Activator	TF ⇧ Gene ⇧	mRNA AAA ⇧ miR ⇩	⇧	⇧	Coordinated Increase
	TF ⇩ Gene ⇧	mRNA AAA ⇧ miR ⇩	⇩	⇧	Decreased, Buffered
	TF ⇧ Gene ⇩	mRNA AAA ⇩ miR ⇧	⇧	⇩	Increased, Buffered
	TF ⇩ Gene ⇩	mRNA AAA ⇩ miR ⇧	⇩	⇩	Coordinated Decrease
Transcriptional Inhibitor	TF ⇩ Gene ⇧	mRNA AAA ⇧ miR ⇩	⇧	⇧	Coordinated Increase
	TF ⇧ Gene ⇩	mRNA AAA ⇧ miR ⇩	⇩	⇧	Decreased, Buffered
	TF ⇩ Gene ⇧	mRNA AAA ⇩ miR ⇧	⇧	⇩	Increased, Buffered
	TF ⇧ Gene ⇩	mRNA AAA ⇩ miR ⇧	⇩	⇩	Coordinated Decrease

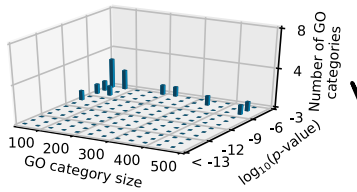
B

Coordinated increase (n_{genes} = 463)



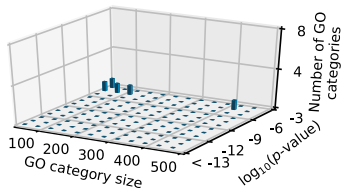
p-value	GO list size	GO annotation
< 4.10E-13	353	extracellular matrix
< 4.10E-13	357	extracellular matrix organization
< 4.10E-13	476	cell migration
< 4.10E-13	107	growth factor binding
< 4.10E-13	250	proteinaceous extracellular matrix
< 4.10E-13	316	extracellular structure organization
8.15E-11	472	cell adhesion
2.31E-10	460	biological adhesion

Decreased buffered (n_{genes} = 223)



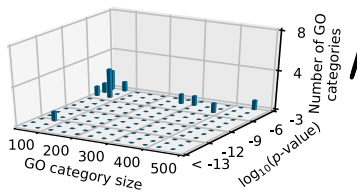
p-value	GO list size	GO annotation
3.83E-08	42	negative regulation of Ras protein signal transduction
3.51E-06	50	negative regulation of cell projection organization
4.35E-06	92	negative regulation of neuron differentiation
6.95E-06	498	regulation of GTPase activity

No change (n_{genes} = 222)



p-value	GO list size	GO annotation
< 4.10E-13	81	anaphase-promoting complex-dependent catabolic process
6.53E-06	46	regulation of synaptic transmission, glutamatergic
6.99E-06	84	mitochondrial translational termination
6.99E-06	84	mitochondrial translational initiation
7.01E-06	409	mitotic cell cycle
9.34E-06	85	mitochondrial translational elongation

Increased buffered (n_{genes} = 250)



p-value	GO list size	GO annotation
3.28E-12	88	epidermis development
3.57E-10	238	cell-cell junction
4.96E-09	92	cell junction organization
4.12E-08	247	tissue development
1.15E-07	50	cell-cell adherens junction
2.05E-07	51	epithelial cell development
3.79E-07	64	anchoring junction
6.13E-07	53	granulocyte chemotaxis

Coordinated decrease (n_{genes} = 451)

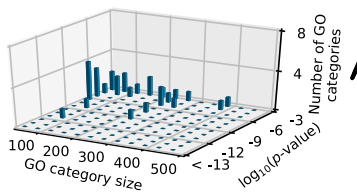


Figure 2. Post-transcriptional Regulation Can Reinforce or Buffer Transcriptional Changes Associated with Specific Biological Processes

(A) The categories into which genes were classified using exon-intron split analysis (EISA).

(B) For each gene set the number of significant gene ontology (GO) annotations is shown, grouped across GO category list size and the degree of statistical support ($\log_{10}$(p value)). For visualization only significant (adjusted p value < 1×10^{-3}) groups with a minimum GO list size of 40 are plotted (3D histograms), while the largest and most significant GO categories are listed. More comprehensive lists are given in Table S3.

Table 1. Gene Sets Partitioned by mRNA Abundance Changes

	Gene Set Label				
Metric Partitioned	Co-ordinated Increase	Decreased, Buffered	No Change	Increased, Buffered	Co-ordinated Decrease
Δintron	top 15%	bottom 15%	−0.1 -> 0.1 (30th–47th %ile)	top 15%	bottom 15%
Δexon – Δintron	top 15%	top 15%	−0.1 -> 0.1 (46th–65th %ile)	bottom 15%	bottom 15%
Differential transcript abundance	–	–	−0.3 -> 0.31 (31st–66th %ile)	–	–

Gene sets were partitioned using thresholds against mRNA differential expression and exon-intron split analysis (EISA) metrics.

are enriched for pro-epithelial miRNA target sites (Figure 3C, upper panel), consistent with a loss of repression across these genes as epithelial miRNA expression decreases during EMT. The greatest relative enrichment of target sites is consistently seen within the DB gene set, suggesting particularly important roles for miRNAs in buffering unrelated transcription. Targets for epithelial miRNAs are under-represented among post-transcriptionally downregulated genes (the IB and CD gene sets, upper panel). The absence of clear trends for predicted targets of EMT upregulated miRNAs (Figure 3C, lower panel), coupled with less-concordant changes for these miRNAs in TCGA data (Figure 3B) suggests a dominant role for miRNAs associated with a more differentiated epithelial state.

As multiple miRNAs appear co-regulated between epithelial and mesenchymal states, both in our HMLE model and in TCGA, we further investigated co-ordinated miRNA function through co-operative regulation. Correlated expression was examined between pairs of miRNAs across breast cancer samples (Figure 3D, top right), revealing extensive co-regulation among epithelial and mesenchymal miRNAs (Figure 3D, red shading of the upper left and lower sections) and, in general, a slightly weaker inverse association when comparing epithelial and mesenchymal miRNAs (Figure 3D, light blue shading of upper right panel). Most positively correlated miRNA pairs, such as miR-200c/miR-141 and miR-182/miR-183, are encoded at genomic loci within 20 kb of each other (Figure 3D, bottom left, green shading), and correlations likely reflect their polycistronic clustering. However, moderate positive correlation in the expression of non-polycistronic miRNAs, such as the miR-200 and miR-182/183 clusters, suggest co-regulation. The miRNAs co-expressed our HMLE system also show similar associations across a broader set of primary cell lines from the microRNAome (McCall et al., 2017) (Figures S3A and S3B).

We hypothesized that miRNAs amplify their function through co-regulation and co-operative targeting of functionally related transcripts. This might be evident as direct co-targeting of a transcript by co-expressed miRNAs, or by targeting of transcripts that encode proteins with functional relationships (e.g., direct protein-protein interactions [PPIs]; Figure S4). To assess this we constructed a regulatory network containing miRNAs with large changes in abundance during HMLE cell EMT (Figure 3A) and predicted target mRNA transcripts, with known protein interactions between targets. The resulting network (Figure S5) contains 13,851 nodes (20 miRNAs and 13,831 mRNAs) and 147,487 edges (42,908 predicted miRNA:mRNA interactions without filtering, and 104,578 protein interactions). It is difficult to distinguish structure within such complex networks so we used network topology metrics to quantify miRNA co-regulation (Figure S4). In particular, we assessed miRNA direct co-targeting by testing the null hypothesis that target overlap frequency is independent between miRNAs (Figure 3E, top right). Between many pro-epithelial miRNAs downregulated during EMT there was particularly strong evidence against this (indicated by red squares; adjusted p value $< 1 \times 10^{-3}$), suggesting enriched co-targeting of individual genes. This was particularly evident between miR-200b-3p, miR-200c-3p, miR-141-3p, and miR-182-5p (and expected for miR-200b-3p and miR-200c-3p, which share a seed sequence).

We also quantified the connectivity of high-confidence predicted miRNA target genes in the context of PPI networks, examining the density of bipartite graphs induced from paired miRNA target lists to quantify co-targeting around protein interactions and complexes (Figure 3E, bottom left). This shows relatively high co-targeting between pro-epithelial miRNAs, suggesting another dimension of co-operation between these miRNAs. With the exception of miR-381-3p, which has a very large number of targets, this predicted co-regulation is largely absent for pro-mesenchymal miRNAs, again consistent with the notion that epithelial-associated miRNAs have a dominant role in orchestrating epithelial-mesenchymal plasticity (EMP). We also examined correlations between the abundance of miRNAs expressed across a number of primary cell lines (Figure S3A), which included several miRNAs studied here (Figure S3B). We identified miRNAs with a strong cross-correlation (i.e., within the top 1 percentile of Figure S3A) and performed the co-targeting analysis across an expanded set of miRNA pairs (Figure S3C), revealing a number of cross-correlated miRNA pairs (beyond the miRNAs studied here) with evidence for co-targeting (Figures S3B and S3D).

These results show highly co-regulated expression between epithelial miRNAs, especially the miR-200c/141 and miR-182/183 clusters (Figure 3A, 3C, and 3D), as well as evidence of co-operative function through both direct overlap of target transcripts, and through regulation of functionally interacting targets (Figure 3E).

Multiple miRNAs Act in a Combinatorial Manner to Promote MET

We tested the functional significance of these predicted co-regulators by re-expressing combinations of pro-epithelial miRNAs in MesHMLE cells and investigating the resulting MET. Based upon titrations of miR-200c mimics we selected a concentration of 0.2 nM (far lower than the 5–20 nM typically used) for subsequent miRNA transfection into MesHMLE cells. At this abundance, only miR-200c was sufficient to induce epithelial marker genes (Figure 4A). While miR-141, miR-182, and miR-183

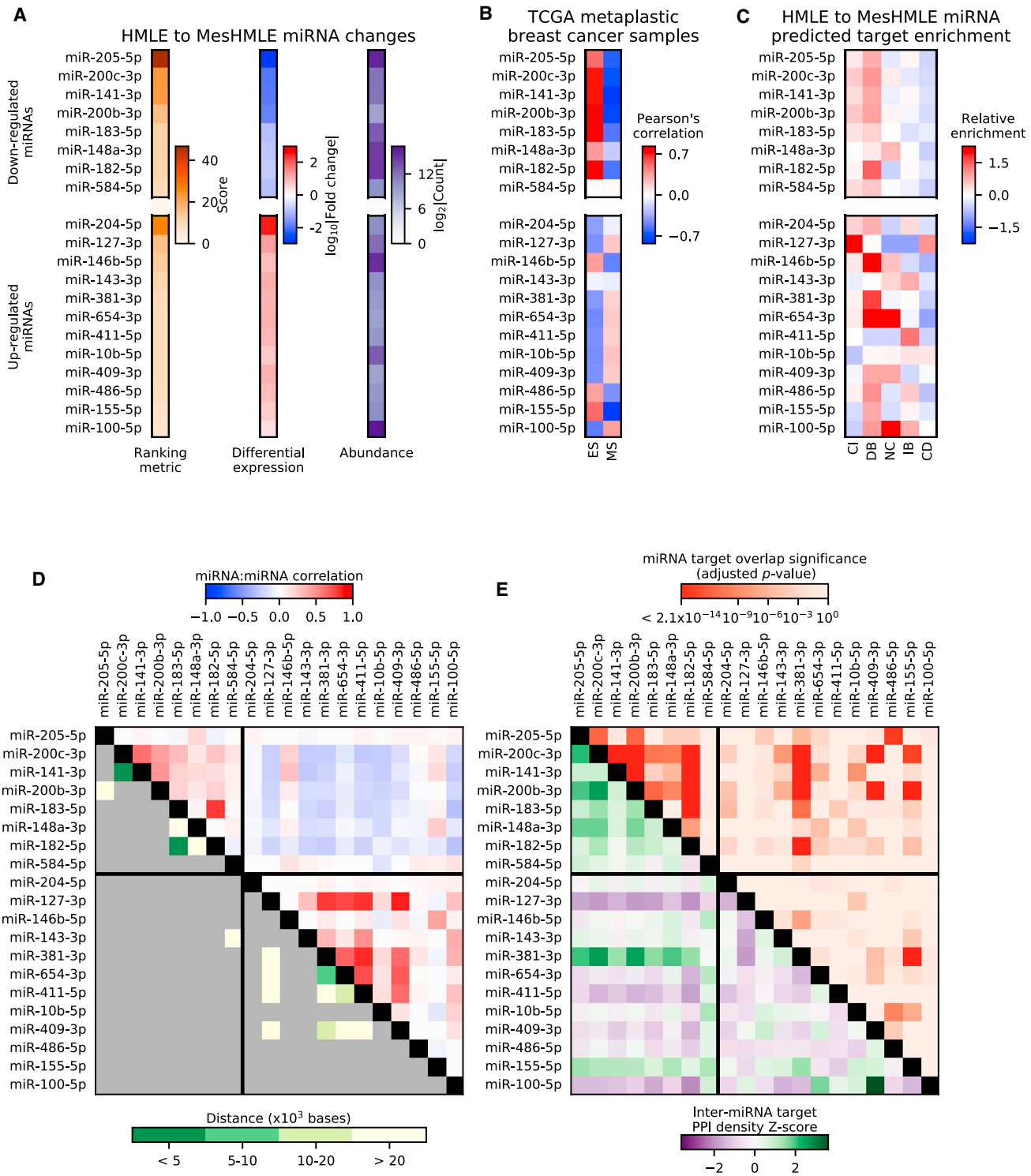


Figure 3. miRNAs Undergoing Large Changes in Abundance during HMLE Cell EMT Show Concordant Changes within TCGA Data and Overlap in their Predicted Target Genes

(A) miRNAs were selected by relative abundance (summed across both states) and differential expression between HMLE and MesHMLE cells (details in the STAR Methods); the top 20 miRNAs are shown, split by the direction of their change in abundance.

(B) Correlations between miRNA abundance and epithelial score or mesenchymal score across metaplastic TCGA breast cancer samples (Table S2B; $n_{tumours}$ = 16).

(legend continued on next page)

individually had no effect on epithelial gene expression they acted in a co-operative manner with miR-200c, indicating the importance of combinatorial miRNA activity at abundances at least an order of magnitude lower (and closer to natural levels of expression) than those at which miRNAs are typically manipulated.

Conversely, treatment of epithelial HMLE cells with combinations of antagomiRs did not appear to drive EMT to the same extent. AntagomiRs against all four miRNAs led to minor repression of *ESRP1* ($p \leq 1.84 \times 10^{-2}$; all combined antagomiRs against control) and minor upregulation of *SNAI1*, but had little effect on the mRNA abundance of other EMT markers tested (Figure S7A). It is important to note that *ZEB1* and *ZEB2* transcripts are present at almost undetectable levels in HMLE cells (Figure S7B) and relative to the MesHMLE state, these genes are depleted for activating histone marks and enriched for repressive histone marks (Figure S7C). This suggests that these TFs are downregulated through epigenetic mechanisms, and miRNAs may not be the dominant suppressors under stable epithelial conditions. Consistent with this, *SNAI1*, which had a minor increase in mRNA abundance with combined antagomiR treatment (Figure S7A), is not as strongly regulated by histone modifications (Figure S7C).

The relative contribution of miR-183 or miR-182 to epithelial gene expression is dependent upon the gene in question: upregulation of *CDH1* depends upon miR-200c and miR-182, whereas *CDS1* is more reliant on miR-200c and miR-183 (Figure 4B). Further, although miR-200c drives the regulation of epithelial genes, the miR-183/-182 family primarily suppresses *VIM* (Figure 4C), probably through an indirect mechanism, as neither TargetScan nor DIANA-microT predict *VIM* targeting by either miRNA. A more prominent suppression of *VIM* and *ZEB1* proteins was observed for the miR-183/-182 and miR-200 families, respectively (Figures 4C and 4D), under these conditions.

A large increase in miR-204 abundance was observed with the HMLE-to-MesHMLE transition (Figure 3A), contradicting a number of studies that have implicated miR-204 as pro-epithelial (Table S1). We expressed miR-204 alone at varying concentrations, and together at low levels with miR-200c (Figure 4E). In this context, miR-204 alone (even at 20 nM) had no effect on *CDH1*, *CDH2*, and *ESRP1* mRNA abundance, and only minor repressive effects upon *ZEB2* and *DSG3* (consistent with an individually pro-epithelial role, at least when overexpressed). When co-expressed with miR-200c, however, miR-204 antagonized epithelial gene induction, suggesting that it may promote a mesenchymal phenotype. This emphasizes the importance of testing miRNA function at physiological levels and in the context of other miRNAs, and may explain contradictory reports of this apparently pro-mesenchymal miRNA promoting epithelial characteristics.

Ectopic miRNA Expression Effects Genes that Are Not Functional Targets at Endogenous miRNA Levels

Supra-physiological miRNA concentrations (5–20 nM) may influence mRNA transcripts that are not functionally regulated by that miRNA at endogenous levels (Mayya and Duchaine, 2015; Witwer and Halushka, 2016). Further, computational databases may over-predict the number of transcripts functionally regulated by miRNAs at endogenous expression levels (Pinzón et al., 2017). To test this we compared EISA data for predicted miR-200c targets during TGF-β-driven EMT against miR-200c-driven MET, which involved exposure to ectopic miRNA at a concentration (20 nM) well in excess of physiological levels (Figure 5).

As noted earlier, non-zero values for Δintron are indicative of altered transcription, while Δexon – Δintron indicates altered post-transcriptional regulation. For reference, distributions for all mRNAs are shown (Figure 5, top row) with density contours superimposed upon plots of predicted target mRNAs with varying confidence (see percentile). The rotation and skew of these distributions reflect the earlier observation that, at a systems level, many genes experience transcriptional changes that are supported by concordant regulatory changes (Figure 2).

During the HMLE-to-MesHMLE transition there was an approximately 80-fold decrease in miR-200c driven by endogenous factors downstream of TGF-β signaling; accordingly, there is a bulk shift in target mRNAs toward positive Δexon – Δintron values, particularly when compared with the distribution of all genes. However, even for the strongest predicted targets, where there is good overlap between databases, only a small fraction are represented. Conversely, with miR-200c-induced MET, over 70% of predicted targets show evidence of strong post-transcriptional suppression. For miRNA targets predicted with less confidence (lower percentile bins) there is a reduced enrichment for post-transcriptional upregulation in EMT, while the corresponding bins for miR-200c overexpression still show over 60% of targets with post-transcriptional downregulation. These results highlight substantial differences between miR-200c overexpression effects and endogenous changes (Figure 5, middle column) in comparison with the effects associated with miR-200c overexpression (Figure 5, right column). We strongly suspect that the attribution of function to some miRNAs has been confounded by miRNA overexpression effects upon targets that would otherwise not be modulated at endogenous levels; this may have contributed to an over-estimation of the impact of many individual miRNAs on EMT (Table S1).

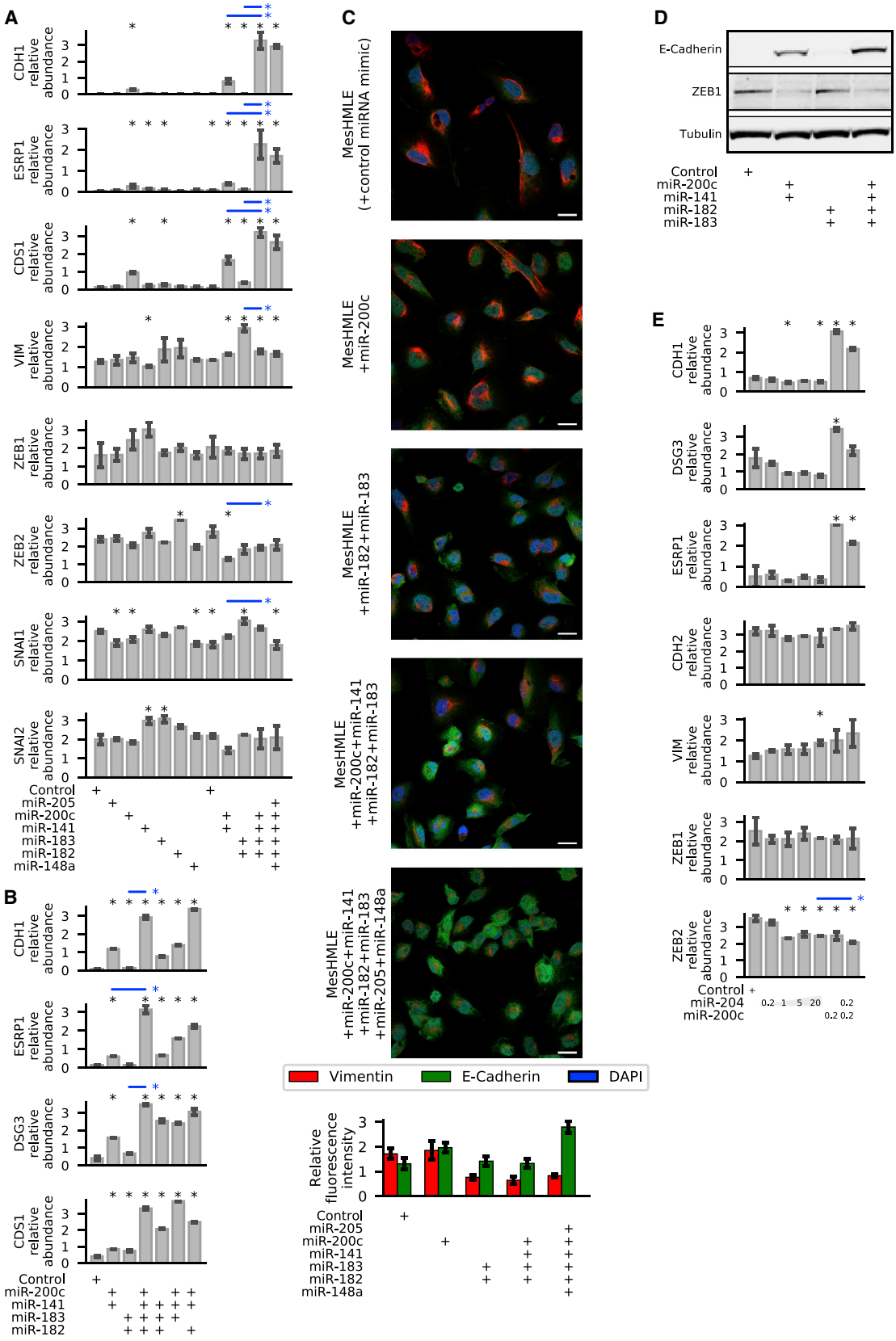
Combinations of miRNAs Can Induce MET with Increased Specificity

As discussed above, overexpression of miR-200c drives the post-transcriptional degradation of several hundred predicted targets that are not under strong control by miR-200c during TGF-β-induced EMT. This suggests that miR-200c

(C) Relative enrichments of high-confidence predicted targets (top 5% of TargetScan and DIANA-microT targets) for each miRNA within the EISA-partitioned gene sets (as in Figure 2A).

(D) Pearson's correlation for each pair of miRNAs across all TCGA breast cancer samples (at top right) as a measure of miRNA co-regulation across a larger set of tissue samples, together with the genomic distance between miRNAs located on the same chromosome (at bottom left).

(E) Several metrics were applied to measure miRNA target set interactions (Figure S4) Overlap in miRNA targets is shown in red, top right, while the density of protein interactions between the targets of each miRNA pair is shown on the bottom left.



(legend on next page)

overexpression may have "off-target" effects that would not occur endogenously, nor when EMT is driven by miRNA combinations at lower levels. Increasing levels of miR-200c increased the expression of multiple epithelial genes in a dose-dependent manner (*CDH1, ESRP1,* and *DSG3*; Figure 6A) but not mesenchymal genes (with the prominent exception of *ZEB2*; Figures 6B and 6C). We identified overexpression off-target effects by selecting genes that were: (1) altered in the RNA sequencing data with ectopic miR-200c transfection (MesHMLE to MesHMLE + miR-200c) but not TGF-β-induced EMT, (2) not previously implicated in EMT, and (3) not predicted targets of either the miR-200 or miR-182/-183 families. A number of these off-target genes had dose-dependent effects with miR-200c transfection, being either up- (*FLJ44056* and *TRIM58*) or down-regulated (*CAMK2N2, GTF2I,* and *GBGT1*). Consistent with our previous result (Figure 4), co-transfection with low levels of miR-200 and miR-182/-183 family members had a co-operative effect on epithelial gene expression (Figure 6A), but importantly, despite still driving MET-like gene expression, these low-level miRNA combinations no longer influenced off-target mRNAs (Figure 6D).

Finally, we investigated the phenotype of MesHMLE cells following transfection of individual miRNAs or combinations of miRNAs using a proliferation and migration assay. The relative density of cells transfected with miR-200c, miR-141, miR-182, and miR-183 was reduced compared with all other conditions at both the 48- and 72-hr time points (Figure 6E). Furthermore, MesHMLE cells treated with this combination showed reduced migratory capacity (Figure 6F). These results show that low-abundance pro-epithelial miRNAs exert a combinatorial effect and induce a more-robust MET than that achieved through treatment with extremely high concentrations of individual pro-epithelial miRNAs.

DISCUSSION

Results presented here provide evidence for an alternative model of miRNA function, where small, simultaneous changes in the abundance of several miRNAs can exert strong functional effects through distributed targeting across molecular networks, with buffering of unrelated expression. Working with an established model of TGF-β-induced EMT in breast cells (Figure 1), we used the EISA method to separate transcriptional and post-transcriptional changes. A small number of interesting genes were lost with EISA due to very small intronic regions (e.g., the well-studied miR-200c target *CFL2*; Bracken et al., 2014), or very low abundance under certain conditions (e.g., *ZEB2* with the miR-200c-induced MET); however, it still provides a powerful approach to examine system-wide changes in the regulation of mRNA transcripts. Subsequent gene ontology enrichment analyses using EISA-partitioned data identified several EMT-associated gene sets with increased abundance from promotion of transcription that was enhanced by a loss of repression from epithelial-promoting miRNAs, and, similarly, a set of mesenchymal-associated miRNAs that supported the effects of repressive TFs, which exerted their effects during EMT (Figure 2A; distribution skew/rotation in Figure 5). A number of TFs within the CI and decreased gene sets are important regulators of EMT, in particular *ZEB1/ZEB2* (Gregory et al., 2008; Lamouille et al., 2014; Perdigao-Henriques et al., 2016), RUNX2 (Chimge et al., 2011; Tan et al., 2016), and GRHL2 (Cieply et al., 2013; Mooney et al., 2017), reflecting the high degree of interaction between transcriptional and post-transcriptional control.

Not only do miRNAs provide an additional regulatory layer to enhance transcriptional regulation, but they also prominently buffer effects on "bystander" genes that are not needed in EMT but that nevertheless are transcriptionally affected by regulators of EMT. This can be seen both in Figure 3C, where "buffered" genes (lane 2, epithelial panel) have the greatest enrichment of miRNA target sites, and in Figure 5, where targets in the buffered quadrants are especially regulated after miR-200c expression. Given that these genes are buffered by miRNAs (and thus, do not appear among the genes with strong differential expression), such relationships may be under-represented in studies aiming to characterize the strongest targets and, as such, this important (and general) aspect of miRNA function remains under-appreciated. miRNAs therefore play an important role in transcriptomic homeostasis extending beyond the well-studied miRNA:TF reciprocal inhibition motif (Gregory et al., 2008; Tam and Weinberg, 2013; Bracken et al., 2016), which provides an important regulatory module within biological networks (Lu et al., 2013; Jolly et al., 2016).

Another key finding is the importance of combinatorial miRNA function, which is often masked in single miRNA overexpression studies (Witwer and Halushka, 2016). We observe this clearly when comparing EISA profiles for HMLE cells that are exposed to TGF-β, where miR-200c-3p levels are naturally reduced 80-fold, against the profiles of cells after exogenous miR-200c overexpression, where hundreds of weak target genes become miR-200c responsive (Figure 5). For miR-200c, perhaps the most well-known enforcer of an epithelial phenotype, we note that individually it has minimal impact driving MET when expressed at low levels (Figures 4 and 6). Instead, we have identified miRNAs that are co-regulated with miR-200c and predicted to co-target common protein complexes (Figure 3E), and we observe a dramatic additive effect when simultaneously modulating these miRNAs to promote MET (Figure 4). Although the EMT-promoting effects of multiple low-level miRNAs are similar to that of single miRNA overexpression, combinatorial miRNAs at modest expression levels minimize off-target effects. This

Figure 4. miRNAs Act in a Combinatorial Manner to Promote an Epithelial Phenotype

All qPCR abundances shown relative to *GAPDH*. Black asterisks indicate a significant difference ($p < 0.01$) from control samples, blue asterisks indicate a significant difference between annotated samples.

(A and B) Marker gene expression in MesHMLE cells 3 days after transfection with 0.2 nM of indicated miRNAs.

(C) Expression of E-cadherin and vimentin in MesHMLE cells detected by immunofluorescence 3 days after transfection with 0.2 nM of the indicated miRNAs. Scale bars represent 20 μm. Quantified fluorescence intensity (inset) was normalized against DAPI intensity (for individual cells). Scale bars represent 20 μm.

(D) Expression of E-cadherin and ZEB1 by western blotting following treatment with specified miRNA combinations at 0.2 nM.

(E) Expression of marker genes with varying concentrations of miR-204 and/or 0.2 nM miR-200c.

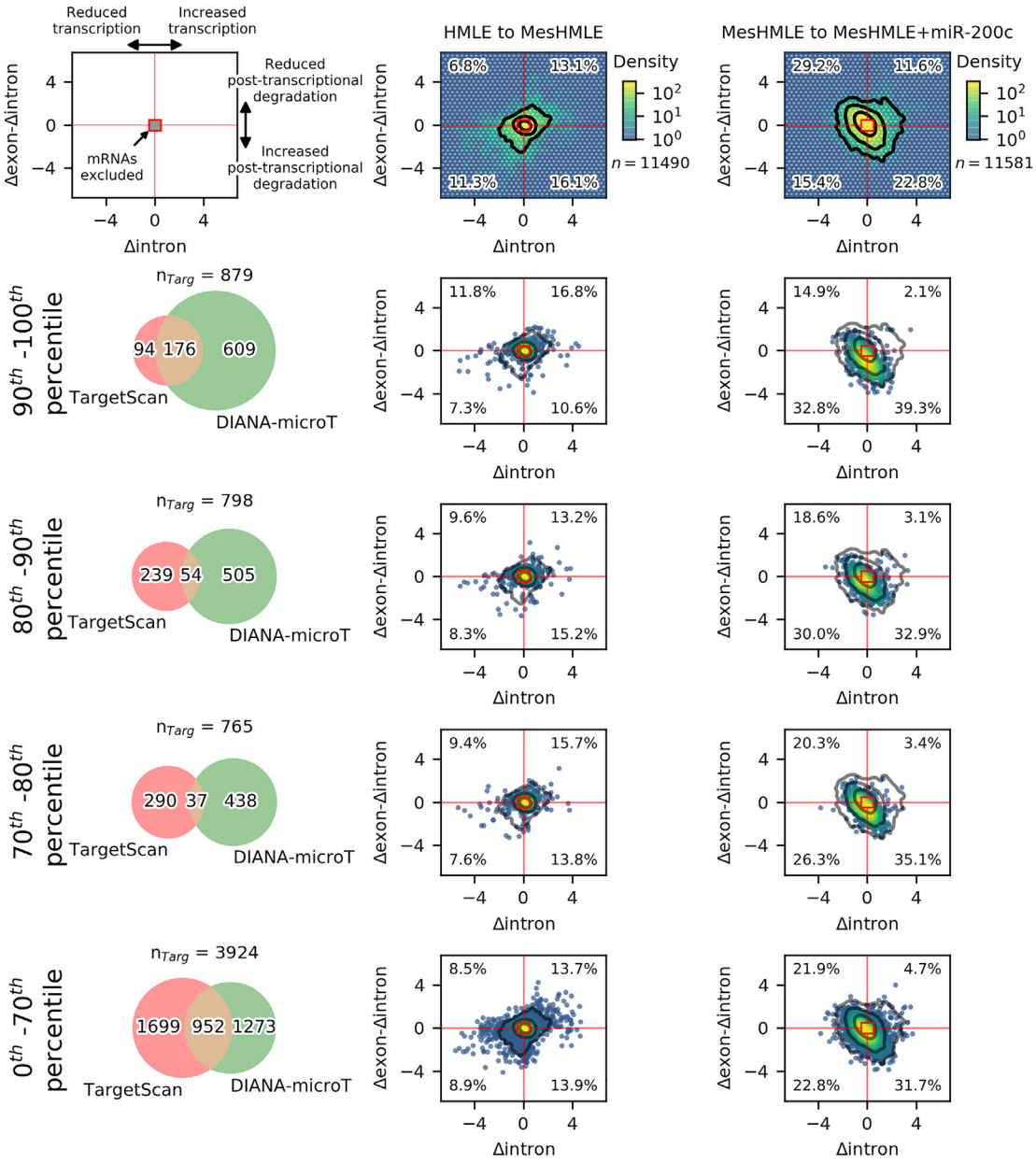


Figure 5. Differences in the Transcriptional and Post-transcriptional Regulation of mRNA Transcripts

The number of miRNA targets from each database and the overlap is shown (Venn diagrams on the left). The distribution of Δintron and Δexon – Δintron values are shown for mRNA transcripts during the HMLE-to-MesHMLE and MesHMLE-to-MesHMLE + miR-200c transitions. The density of all mRNA transcripts is shown (top row), and contour values are superimposed over the distribution of mRNA transcripts within the specified percentile range of database scores. Within each EISA data plot, red horizontal and vertical lines indicate the origin of each plot and the relative proportion of mRNA transcripts within each resulting quadrant is shown. In the scatterplots of percentile-binned targets, markers are colored by a kernel estimate of sample density. Note that there are differences in the number of miR targets across percentile bins of each database, and DIANA-microT in particular predicts a very large number of miR-200c-3p targets within the highest-scoring group (Figure S6).

may have clinical relevance, as this work suggests combinatorial manipulation could both achieve greater specificity than single miRNA expression and require lower levels of miRNA perturbation, which might be more easily achievable *in vivo*. While we demonstrate these combinatorial relationships between members of the miR-200 and miR-182/-183 families (Figures 4 and 6) in the induction of an epithelial phenotype, it is likely that more miRNAs participate in this role and that this research describes an important type of combinatorial function that is likely applicable to all endogenous aspects of miRNAs.

The relative contribution of miRNAs to the control of EMP is dependent not only upon miRNA abundance but also the expression level of key miRNA-targeted transcripts and the activity of other regulatory mechanisms. This is well demonstrated

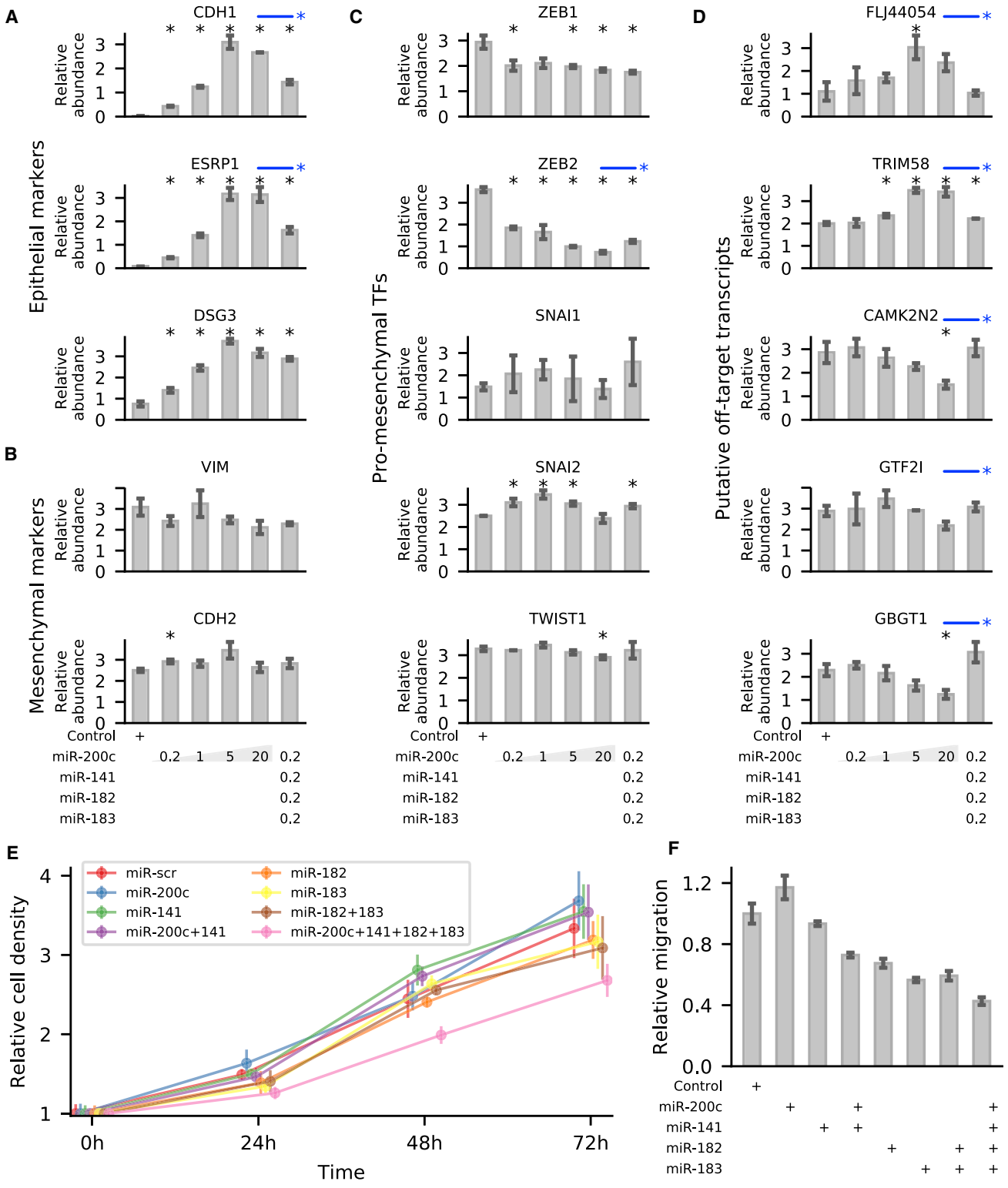


Figure 6. Transfection of Sub-nanomolar Concentrations of Co-operative Epithelial-Enforcing miRNAs Can Drive MET with Reduced Off-Target Effects, Reduced Proliferation and Reduced Migration

All qPCR abundances shown relative to *GAPDH*. Black asterisks indicate a significant difference ($p < 0.01$) from control samples, blue asterisks indicate a significant difference between annotated samples. Genes are divided into (A) known pro-epithelial markers, (B) known pro-mesenchymal markers,

(legend continued on next page)

by the inability of antagomiR repression against miR-141, miR-200c, miR-182, and miR-183 to effectively drive EMT in HMLE cells (Figure S7A); *ZEB1* and *ZEB2* are already expressed at very low levels in HMLE cells (Figure S7B) with an associated depletion of activating histone marks and an enrichment of the repressive H3K27me3 mark (Figure S7C). Accordingly, further repression is unnecessary, and the miRNAs tested here do not appear to play an active role in repressing *ZEB1* and *ZEB2* under these conditions. Consistent with this, the well-studied *ZEB1/2* target gene *CDH1* (Comijn et al., 2001) does not undergo significant changes; however, *SNAI1*, which has an earlier response with induction of EMT (Dave et al., 2011), is not subject to the same degree of histone-mediated regulation (Figure S7C) and shows minor increases with antagomiR treatment against all target miRNAs (Figure S7A). This provides strong evidence that MET is not simply a direct reversal of EMT and is consistent with previous reports that TGF-β-induced EMT involves transitions through several states which are mediated by the switching of multiple double-negative feedback loops (Zhang et al., 2014; Tian et al., 2013). This lack of symmetry in gene regulation is likely mediated by histone modifications and other epigenetic mechanisms that can reinforce phenotypic change.

Combinatorial activities of miRNAs have been reported previously and recently reviewed (Bracken et al., 2016). The use of antisense inhibitors to miR-21, miR-23a, and miR-27a for example showed synergistic effects reducing the proliferation of cells in culture and the growth of xenograft tumors in mice to a greater extent than the inhibition of single miRNAs alone (Frampton et al., 2014). Combinatorial activities of miRNAs co-regulated from polycistronic clusters have also been characterized. The miR-192–miR-194–miR-215 cluster, for example, can co-ordinately suppress tumor progression in renal cell carcinoma (Khella et al., 2013). Similarly, each member of the let-7c-miR-99b-miR-125b cluster directly targets interleukin-6 receptor (IL-6R) and other components of the IL-6-signal transducer and activator of transcription 3 signaling pathway to decrease mammosphere growth, invasion, and the metastatic spread of tumors in xenograft mouse models (Lin et al., 2016). Computational studies also suggest that clustered miRNAs co-regulate genes in shared protein interaction networks and that the closer the proximity of proteins in the network, the more likely they are to be targeted by miRNAs from the same cluster (Yuan et al., 2009), but these predicted effects were not experimentally validated. In agreement with the findings we present here, miR-200 family members have been shown to be co-expressed with the miR-96∼miR-183 cluster in lung cancer, with both miRNA families promoting an epithelial phenotype and preventing cancer invasion and metastasis (Kundu et al., 2016). It should be noted, however, that many of these studies performed transfections with miRNA mimics/precursors at levels greatly exceeding physiological abundances. The concentrations used here are much closer to endogenous levels and we suspect that observed combinatorial activities may be increasingly important at lower abundance levels.

Several of the miRNAs altered during the HMLE-to-MesHMLE transition (Figure 3A) have been associated with mammary cell differentiation, such that the downregulated miR-200 family members are expressed in luminal cells, while upregulated miR-143, miR-100, and miR-204 have been linked to mammary stem cells/basal cells (Pal et al., 2015). It is possible that the molecular changes we observed during this TGF-β-induced EMT correspond to a loss of cellular differentiation, while the phenotypic changes induced by our combinatorial miRNA treatment (Figures 6E and 6F) reflect activation of a terminal differentiation program, rather than a pure induction of MET achieved through miR-200c alone. Indeed, the co-operation of these miRNAs in both lung (Kundu et al., 2016) and breast tissue suggests a more-general role in epithelialization.

In conclusion, we have shown that combinatorial miRNA regulation provides a more efficacious approach to modulating cell phenotype with reduced off-target effects when compared with overexpression of individual miRNAs. The control of cell phenotype by miRNAs appears to be mediated through both the reinforcement of transcriptional changes and the buffering of transcriptional noise, a feature of miRNA functionality that has thus far been under-appreciated. Using an established cell line model of EMT we have demonstrated that distributed, combinatorial targeting by multiple miRNAs can produce a more pronounced phenotypic effect than individual miRNAs used at higher concentrations. Together, these results support the hypothesis that EMT is co-ordinated through distributed co-targeting by multiple miRNAs, suggesting that low-level co-manipulation of miRNAs may provide an effective strategy to minimize off-target effects while maintaining biological efficacy.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Tissue Culture Cell Lines
  - Selection of Target miRNAs
- METHOD DETAILS
  - Tissue Culture
  - Isolation of RNA and Real-Time PCR
  - Western Blotting
  - Immunofluorescence
  - Cellular Proliferation Assay
  - Transwell Migration Assay
  - RNA-sequencing
  - Histone Modification ChIP-seq
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Real-Time Quantitative PCR
  - Immunofluorescence
  - Cellular Proliferation Assay

(C) pro-mesenchymal EMT-promoting transcription factors, and (D) putative off-target genes that change in the MesHMLE-to-MesHMLE + miR-200 transition but not the TGF-β-induced HMLE-to-MesHMLE transition, are not previously associated with EMT, and do not have TargetScan-predicted sites for the miR-200 family, miR-182 or miR-183. Error bars represent SD. Phenotypic effects of single miRNAs and miRNA combinations were also assessed using (E) a cellular proliferation assay and (F) a trans-well migration assay. Assays were performed in triplicate and error bars represent SD.

SUPPLEMENTAL INFORMATION

Supplemental Information includes eight figures and five tables and can be found with this article online at https://doi.org/10.1016/j.cels.2018.05.019.

ACKNOWLEDGMENTS

The authors would like to acknowledge Prof. E. Thompson from Queensland University of Technology for his input and discussion, and Daniel Esposito from the Walter and Eliza Hall Institute for the combined PPI databases. Results published here are, in part, based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/. This work was supported in part by: National Breast Cancer Foundation funding of the EMPathy Breast Cancer Network (CG-10-04), a National Collaborative Research Program; the Australian Research Council Center of Excellence in Convergent Bio-Nano Science and Technology (project number CE140100036); and Australian National Health and Medical Research Council project grants GNT1068773 (to G.J.G. and P.A.G.), GNT1034633 and GNT1069128 (to G.J.G. and C.P.B.). M.F. is supported by a Melbourne International Fee Remission Scholarship (MIFRS) and Melbourne International Research Scholarship (MIRS). P.A.G. is supported by a Cancer Council SA Beat Cancer Fellowship. G.J.G. is supported by fellowships from the Australian National Health and Medical Research Council (GNT1026191 and GNT1118170). C.P.B. is supported by a Florey Fellowship from the Royal Adelaide Hospital Research Foundation. M.J.D. is funded by the National Breast Cancer Foundation ECF-14-043.

AUTHOR CONTRIBUTIONS

J.C., P.A.G., E.J.C., G.J.G., C.P.B., and M.J.D. designed the study. K.S., P.A.G., and C.P.B. generated the data. J.C., K.A.P., K.S., M.F., S.H.Z., J.T., C.P.B., and M.J.D. performed the data analysis. J.C., C.P.B., and M.J.D. wrote the paper with input from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 24, 2017
Revised: February 22, 2018
Accepted: May 25, 2018
Published: July 11, 2018

REFERENCES

Aken, B.L., Achuthan, P., Akanni, W., Amode, M.R., Bernsdorff, F., Bhai, J., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., et al. (2017). Ensembl 2017. Nucleic Acids Res. *45*, D635–D642.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29.

Attema, J.L., Bert, A.G., Lim, Y.Y., Kolesnikoff, N., Lawrence, D.M., Pillman, K.A., Smith, E., Drew, P.A., Khew-Goodall, Y., Shannon, F., et al. (2013). Identification of an enhancer that increases miR-200b∼200a∼429 gene expression in breast cancer cells. PLoS One *8*, e75517.

Blick, T., Widodo, E., Hugo, H., Waltham, M., Lenburg, M.E., Neve, R.M., and Thompson, E.W. (2008). Epithelial mesenchymal transition traits in human breast cancer cell lines. Clin. Exp. Metastasis *25*, 629–642.

Bracken, C.P., Li, X., Wright, J.A., Lawrence, D.M., Pillman, K.A., Salmanidis, M., Anderson, M.A., Dredge, B.K., Gregory, P.A., Tsykin, A., et al. (2014). Genome-wide identification of miR-200 targets reveals a regulatory network controlling cell invasion. EMBO J. *33*, 2040–2056.

Bracken, C.P., Khew-Goodall, Y., and Goodall, G.J. (2015). Network-based approaches to understand the roles of miR-200 and other microRNAs in cancer. Cancer Res. *75*, 2594–2599.

Bracken, C.P., Scott, H.S., and Goodall, G.J. (2016). A network-biology perspective of microRNA function and dysfunction in cancer. Nat. Rev. Genet. *17*, 719–732.

Breuer, K., Foroushani, A.K., Laird, M.R., Chen, C., Sribnaia, A., Lo, R., Winsor, G.L., Hancock, R.E., Brinkman, F.S., and Lynn, D.J. (2013). InnateDB: systems biology of innate immunity and beyond – recent updates and continuing curation. Nucleic Acids Res. *41*, D1228–D1233.

Chen, Y., Lun, A.T., and Smyth, G.K. (2016). From reads to genes to pathways: differential expression analysis of RNA-seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. F1000Res *5*, 1438.

Chimge, N.O., Baniwal, S.K., Little, G.H., Chen, Y.B., Kahn, M., Tripathy, D., Borok, Z., and Frenkel, B. (2011). Regulation of breast cancer metastasis by Runx2 and estrogen signaling: the role of SNAI2. Breast Cancer Res. *13*, R127.

Cieply, B., Farris, J., Denvir, J., Ford, H.L., and Frisch, S.M. (2013). Epithelial-mesenchymal transition and tumor suppression are controlled by a reciprocal feedback loop between ZEB1 and Grainyhead-like-2. Cancer Res. *73*, 6299–6309.

Comijn, J., Berx, G., Vermassen, P., Verschueren, K., Van Grunsven, L., Bruyneel, E., Mareel, M., Huylebroeck, D., and Van Roy, F. (2001). The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. Mol. Cell *7*, 1267–1278.

Cowley, M.J., Pinese, M., Kassahn, K.S., Waddell, N., Pearson, J.V., Grimmond, S.M., Biankin, A.V., Hautaniemi, S., and Wu, J. (2012). PINA v2.0: mining interactome modules. Nucleic Acids Res. *40*, D862–D865.

Daemen, A., Griffith, O.L., Heiser, L.M., Wang, N.J., Enache, O.M., Sanborn, Z., Pepin, F., Durinck, S., Korkola, J.E., Griffith, M., et al. (2013). Modeling precision treatment of breast cancer. Genome Biol. *14*, R110.

Dave, N., Guaita-Esteruelas, S., Gutarra, S., Frias, A., Beltran, M., Peiro, S., and De Herreros, A.G. (2011). Functional cooperation between Snail1 and twist in the regulation of ZEB1 expression during epithelial to mesenchymal transition. J. Biol. Chem. *286*, 12024–12032.

Foroutan, M., Cursons, J., Hediyeh-Zadeh, S., Thompson, E.W., and Davis, M.J. (2017). A transcriptional program for detecting TGFbeta-induced EMT in cancer. Mol. Cancer Res. *15*, 619–631.

Frampton, A.E., Castellano, L., Colombo, T., Giovannetti, E., Krell, J., Jacob, J., Pellegrino, L., Roca-Alonso, L., Funel, N., Gall, T.M., et al. (2014). MicroRNAs cooperatively inhibit a network of tumor suppressor genes to promote pancreatic tumor growth and progression. Gastroenterology *146*, 268–277.e18.

Friard, O., Re, A., Taverna, D., De Bortoli, M., and Cora, D. (2010). CircuitsDB: a database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. BMC Bioinformatics *11*, 435.

Friedman, R.C., Farh, K.K., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. *19*, 92–105.

Gaidatzis, D., Burger, L., Florescu, M., and Stadler, M.B. (2015). Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. Nat. Biotechnol. *33*, 722–729.

Gene Ontology Consortium (2015). Gene Ontology Consortium: going forward. Nucleic Acids Res. *43*, D1049–D1056.

Goeman, J.J., and Buhlmann, P. (2007). Analyzing gene expression data in terms of gene sets: methodological issues. Bioinformatics *23*, 980–987.

Gosline, S.J., Gurtan, A.M., Jnbaptiste, C.K., Bosson, A., Milani, P., Dalin, S., Matthews, B.J., Yap, Y.S., Sharp, P.A., and Fraenkel, E. (2016). Elucidating MicroRNA regulatory networks using transcriptional, post-transcriptional, and histone modification measurements. Cell Rep. *14*, 310–319.

Gray, K.A., Yates, B., Seal, R.L., Wright, M.W., and Bruford, E.A. (2015). Genenames.org: the HGNC resources in 2015. Nucleic Acids Res. *43*, D1079–D1085.

Gregory, P.A., Bert, A.G., Paterson, E.L., Barry, S.C., Tsykin, A., Farshid, G., Vadas, M.A., Khew-Goodall, Y., and Goodall, G.J. (2008). The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat. Cell Biol. *10*, 593–601.

Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol. Cell *27*, 91–105.

Hanzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics *14*, 7.

Heiser, L.M., Sadanandam, A., Kuo, W.L., Benz, S.C., Goldstein, T.C., Ng, S., Gibb, W.J., Wang, N.J., Ziyad, S., Tong, F., et al. (2012). Subtype and pathway specific responses to anticancer compounds in breast cancer. Proc. Natl. Acad. Sci. USA *109*, 2724–2729.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with bioconductor. Nat. Methods *12*, 115–121.

Hunter, J.D. (2007). Matplotlib: a 2D graphics environment. Comput. Sci. Eng. *9*, 90–95.

Huttlin, E.L., Ting, L., Bruckner, R.J., Gebreab, F., Gygi, M.P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., et al. (2015). The BioPlex network: a systematic exploration of the human interactome. Cell *162*, 425–440.

Jolly, M.K., Tripathi, S.C., Jia, D., Mooney, S.M., Celiktas, M., Hanash, S.M., Mani, S.A., Pienta, K.J., Ben-Jacob, E., and Levine, H. (2016). Stability of the hybrid epithelial/mesenchymal phenotype. Oncotarget *7*, 27067–27084.

Jones, E., Oliphant, T., and Peterson, P. (2014). {SciPy}: open source scientific tools for {Python}. http://www.scipy.org/.

Khella, H.W., Bakhet, M., Allo, G., Jewett, M.A., Girgis, A.H., Latif, A., Girgis, H., Von Both, I., Bjarnason, G.A., and Yousef, G.M. (2013). miR-192, miR-194 and miR-215: a convergent microRNA network suppressing tumor progression in renal cell carcinoma. Carcinogenesis *34*, 2231–2239.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kundu, S.T., Byers, L.A., Peng, D.H., Roybal, J.D., Diao, L., Wang, J., Tong, P., Creighton, C.J., and Gibbons, D.L. (2016). The miR-200 family and the miR-183∼96∼182 cluster target Foxf2 to inhibit invasion and metastasis in lung cancers. Oncogene *35*, 173–186.

Lamouille, S., Xu, J., and Derynck, R. (2014). Molecular mechanisms of epithelial-mesenchymal transition. Nat. Rev. Mol. Cell Biol. *15*, 178–196.

Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15–20.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Liao, Y., Smyth, G.K., and Shi, W. (2013). The subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. *41*, e108.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Lim, Y.Y., Wright, J.A., Attema, J.L., Gregory, P.A., Bert, A.G., Smith, E., Thomas, D., Lopez, A.F., Drew, P.A., Khew-Goodall, Y., et al. (2013). Epigenetic modulation of the miR-200 family is associated with transition to a breast cancer stem-cell-like state. J. Cell Sci. *126*, 2256–2266.

Lin, K.Y., Ye, H., Han, B.W., Wang, W.T., Wei, P.P., He, B., Li, X.J., and Chen, Y.Q. (2016). Genome-wide screen identified let-7c/miR-99a/miR-125b regulating tumor progression and stem-like properties in cholangiocarcinoma. Oncogene *35*, 3376–3386.

Lu, M., Jolly, M.K., Levine, H., Onuchic, J.N., and Ben-Jacob, E. (2013). MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. Proc. Natl. Acad. Sci. USA *110*, 18144–18149.

Maetschke, S.R., Simonsen, M., Davis, M.J., and Ragan, M.A. (2012). Gene Ontology-driven inference of protein-protein interactions using inducers. Bioinformatics *28*, 69–75.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. *17*, 10–12.

Mayya, V.K., and Duchaine, T.F. (2015). On the availability of microRNA-induced silencing complexes, saturation of microRNA-binding sites and stoichiometry. Nucleic Acids Res. *43*, 7556–7565.

McCall, M.N., Kim, M.S., Adil, M., Patil, A.H., Lu, Y., Mitchell, C.J., Leal-Rojas, P., Xu, J., Kumar, M., Dawson, V.L., et al. (2017). Toward the human cellular microRNAome. Genome Res. *27*, 1769–1781.

McKinney, W. (2010). Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference. van der Voort S, Millman J, pp. 51–56.

Mooney, S.M., Talebian, V., Jolly, M.K., Jia, D., Gromala, M., Levine, H., and Mcconkey, B.J. (2017). The GRHL2/ZEB feedback loop - a key axis in the regulation of EMT in breast cancer. J. Cell. Biochem. *118*, 2559–2570.

Pal, B., Chen, Y., Bert, A., Hu, Y., Sheridan, J.M., Beck, T., Shi, W., Satterley, K., Jamieson, P., Goodall, G.J., et al. (2015). Integration of microRNA signatures of distinct mammary epithelial cell types with their gene expression and epigenetic portraits. Breast Cancer Res. *17*, 85.

Paraskevopoulou, M.D., Georgakilas, G., Kostoulas, N., Vlachos, I.S., Vergoulis, T., Reczko, M., Filippidis, C., Dalamagas, T., and Hatzigeorgiou, A.G. (2013). DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. Nucleic Acids Res. *41*, W169–W173.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). Scikit-learn: machine learning in python. J. Mach. Learn. Res. *12*, 2825–2830.

Perdigao-Henriques, R., Petrocca, F., Altschuler, G., Thomas, M.P., Le, M.T., Tan, S.M., Hide, W., and Lieberman, J. (2016). miR-200 promotes the mesenchymal to epithelial transition by suppressing multiple members of the Zeb2 and Snail1 transcriptional repressor complexes. Oncogene *35*, 158–172.

Phipson, B., and Smyth, G.K. (2010). Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. Stat. Appl. Genet. Mol. Biol. *9*. Article39. https://doi.org/10.2202/1544-6115.

Pinzón, N., Li, B., Martinez, L., Sergeeva, A., Presumey, J., Apparailly, F., and Seitz, H. (2017). The number of biologically relevant microRNA targets has been largely overestimated. Genome Res. *27*, 234–245.

Re, A., Cora, D., Taverna, D., and Caselle, M. (2009). Genome-wide survey of microRNA-transcription factor feed-forward regulatory circuits in human. Mol. Biosyst. *5*, 854–867.

Reczko, M., Maragkakis, M., Alexiou, P., Grosse, I., and Hatzigeorgiou, A.G. (2012). Functional microRNA targets in protein coding sequences. Bioinformatics *28*, 771–776.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Robinson, M.D., Mccarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Schult, D.A. and Swart, P. (2008). Exploring network structure, dynamics, and function using NetworkX. Proceedings of the 7th Python in Science Conferences (SciPy 2008), pp. 11–16.

Tam, W.L., and Weinberg, R.A. (2013). The epigenetics of epithelial-mesenchymal plasticity in cancer. Nat. Med. *19*, 1438–1449.

Tan, C.C., Li, G.X., Tan, L.D., Du, X., Li, X.Q., He, R., Wang, Q.S., and Feng, Y.M. (2016). Breast cancer cells obtain an osteomimetic feature via epithelial-mesenchymal transition that have undergone BMP2/RUNX2 signaling pathway induction. Oncotarget *7*, 79688–79705.

Tan, T.Z., Miow, Q.H., Miki, Y., Noda, T., Mori, S., Huang, R.Y., and Thiery, J.P. (2014). Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. EMBO Mol. Med. *6*, 1279–1293.

Telonis, A.G., Loher, P., Jing, Y., Londin, E., and Rigoutsos, I. (2015). Beyond the one-locus-one-miRNA paradigm: microRNA isoforms enable deeper insights into breast cancer heterogeneity. Nucleic Acids Res. *43*, 9158–9175.

The Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192.

Tian, X.J., Zhang, H., and Xing, J. (2013). Coupled reversible and irreversible bistable switches underlying TGFbeta-induced epithelial to mesenchymal transition. Biophys. J. *105*, 1079–1089.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. *31*, 46–53.

Verhaak, R.G., Tamayo, P., Yang, J.Y., Hubbard, D., Zhang, H., Creighton, C.J., Fereday, S., Lawrence, M., Carter, S.L., Mermel, C.H., et al. (2013). Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. J. Clin. Invest. *123*, 517–525.

Walt, S.V.D., Colbert, S.C., and Varoquaux, G. (2011). The NumPy array: a structure for efficient numerical computation. Comput. Sci. Eng. *13*, 22–30.

Witwer, K.W., and Halushka, M.K. (2016). Toward the promise of microRNAs - enhancing reproducibility and rigor in microRNA research. RNA Biol. *13*, 1103–1116.

Wu, D., Lim, E., Vaillant, F., Asselin-Labat, M.L., Visvader, J.E., and Smyth, G.K. (2010). ROAST: rotation gene set tests for complex microarray experiments. Bioinformatics *26*, 2176–2182.

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., et al. (2016). Ensembl 2016. Nucleic Acids Res. *44*, D710–D716.

Ye, X., and Weinberg, R.A. (2015). Epithelial-mesenchymal plasticity: a central regulator of cancer progression. Trends Cell Biol. *25*, 675–686.

Yuan, X., Liu, C., Yang, P., He, S., Liao, Q., Kang, S., and Zhao, Y. (2009). Clustered microRNAs' coordination in regulating protein-protein interaction network. BMC Syst. Biol. *3*, 65.

Zhang, J., Tian, X.J., Zhang, H., Teng, Y., Li, R., Bai, F., Elankumaran, S., and Xing, J. (2014). TGF-beta-induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. Sci. Signal. *7*, ra91.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. *9*, R137.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
E-Cadherin	BD Biosciences	Cat#610182; RRID:AB_397581
Zeb1	Cell Signaling Technologies	Cat#D80D3; RRID:AB_1904164
Vimentin	Cell Signaling Technologies	Cat#5741; RRID:AB_10695459
α-tubulin	Abcam	Cat#Ab7291; RRID:AB_2241126
Goat anti-rabbit HRP	Thermo Scientific	Cat#31460; RRID:AB_228341
Goat anti-mouse HRP	Thermo Scientific	Cat#31446; RRID:AB_228318
Goat anti-mouse AlexaFluor-488	Thermo Scientific	Cat#A-11029; RRID:AB_2534088
Goat anti-rabbit AlexaFluor-594	Thermo Scientific	Cat#A-11037; RRID:AB_2534095
Histone H3	Abcam	Cat#ab1791; RRID:AB_302613
H3K27Ac	Millipore	Cat#17-683; RRID:AB_1977529
H3K4me3	Abcam	Cat#ab8580; RRID:AB_306649
H3K9-14Ac	Millipore	Cat#06-599; RRID:AB_2115283
H3K27me3	Abcam	Cat#ab6002; RRID:AB_305237
Chemicals, Peptides, and Recombinant Proteins		
RIPA lysis buffer (for western analysis)	Abcam	Cat#Ab156034,
Protease Inhibitor Cocktail	Roche	Cat#1183617001
Clarity Western ECL substrate	Biorad	Cat#170-5060
Human TGFβ	R&D Systems	Cat#7754BH
HuMEC media (for HMLE cells)	Gibco	Cat#12752010
DMEM:F12 (1:1) media (for MesHMLE cells)	Gibco	Cat#11320033
rh EGF	R&D Systems	Cat#236-EG
Hydrocortisone	Sigma	Cat#H0888; Lot#119K1444
Human Insulin (rys)	Novo Nordisk	Actrapid Penfill 3mL 100IU
Lipofectamine RNAi Max	Invitrogen	Cat#13778030
Fibronectin	Roche	Cat#10638039001
Trizol	Thermo Scientific	Cat#15596026
ProLong gold antifade mounting reagent with DAPI	Molecular Probes	Cat#P36935
Fluorescent mounting medium	DAKO	Cat#S3023
Bis-Tris Bolt gels	Life Technologies	Cat#NW00100BOX
miRvana negative control miRNA inhibitor (antagomiR)	Ambion	Cat# 4464076; Lot# AS028H0P
miRvana miR-141 inhibitor (antagomiR)	Ambion	Cat# 4464084; ID# MH10860; Lot# AS028TFR
miRvana miR-200c inhibitor (antagomiR)	Ambion	Cat# 4464084; ID# MH117174; Lot# AS028TFS
miRvana miR-182 inhibitor (antagomiR)	Ambion	Cat# 4464084; ID# MH12369 Lot# AS028TFT
miRvana miR-183 inhibitor (antagomiR)	Ambion	Cat# 4464084; ID# MH12830; Lot# AS028TFU
Critical Commercial Assays		
CyQuant Cell proliferation assay kit	Invitrogen	Cat#C7026
PolyA+ RNA isolation kit	New England Biolabs	Cat#E7490S

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Clarity Western ECL substrate	BioRad	Cat#170-5060
QuantiTect Reverse Transcription Kit	Qiagen	Cat#205313
Deposited Data		
RNA-seq data	This paper. Please refer to Table S4 for further details.	ENA: PRJEB8225 ENA: PRJEB25061 ENA: PRJEB25042
Small RNA-seq	This paper. Please refer to Table S4 for further details.	ENA: PRJEB25116
MicroRNAome primary cell miRNA abundance data	(McCall et al., 2017)	https://genome.cshlp.org/content/suppl/2017/09/06/gr.222067.117.DC1/Supplemental_Table_S5.xlsx *Accessed 13th Dec 2017*
TargetScan (v7.1)	(Friedman et al., 2009; Grimson et al., 2007; Lewis et al., 2005)	http://www.targetscan.org/ *Accessed 25th May 2016*
DIANA-microT CDS ***NB**: account creation is required to download the DIANA-microT CDS database*	(Paraskevopoulou et al., 2013; Reczko et al., 2012)	http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=microT_CDS/index *Accessed 11th Jan 2016*
HUGO Gene Nomenclature Committee database	(Gray et al., 2015)	http://www.genenames.org/ *Accessed 15th Dec 2017*
Gene Ontology database	(Ashburner et al., 2000)	http://www.geneontology.org/ *Accessed 6th July 2016*
ENSEMBL Biomart data tables	(Yates et al., 2016)	http://ensembl.org/biomart *Accessed 31st March 2016 – 20th Oct 2016*
Breast cancer cell line RNA-seq data (Daemen et al., 2013)	(Daemen et al., 2013; Heiser et al., 2012)	GEO: GSE48213
TCGA Breast Cancer RNA-seq and miRNA-seq data	(The Cancer Genome Atlas Network, 2012)	https://portal.gdc.cancer.gov/projects/TCGA-BRCA *mRNA data accessed* Feb 2015 *miRNA data accessed* Dec 2016
hg19 reference genome	-	http://sapac.support.illumina.com/sequencing/sequencing_software/igenome.html *UCSC version used*
Experimental Models: Cell Lines		
HMLE cells	ATCC	CRL-4010
Oligonucleotides		
qPCR primers	GenePharma	See Table S5 for primer sequences
Software and Algorithms		
ImageJ	Wayne Rasband, NIH	https://imagej.nih.gov/ij/ v. 1.51p RRID:SCR_003070
Image Lab	Bio-Rad	RRID:SCR_014210
Zen	Zeiss	RRID:SCR_013672
WinPython (64-bit)	The Winpython development team	http://winpython.github.io/ v. 3.5.2.1
NumPy + Intel® Math Kernel Library	(Walt et al., 2011)	*via.* WinPython v. 1.11.1+mkl
Matplotlib	(Hunter, 2007)	*via.* WinPython v. 1.5.2
Pandas	(McKinney, 2010)	*via.* WinPython v. 0.18.1
SciPy	(Jones et al., 2014)	*via.* WinPython v. 0.18.0rc2

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
scikit-learn	(Pedregosa et al., 2011)	*via* WinPython v. 0.17.1
networkx	(Schult and Swart, 2008)	*via* WinPython v. 1.11
R	The R Foundation	https://www.r-project.org v. 3.4
BioConductor	(Huber et al., 2015)	https://bioconductor.org/ v. 3.5
limma (with ROAST/FRY)	(Goeman and Buhlmann, 2007; Phipson and Smyth, 2010; Ritchie et al., 2015; Wu et al., 2010)	*via* R/Bioconductor v. 3.30.13
Rsubread (with featureCounts)	(Liao et al., 2013)	*via* R/Bioconductor v. 1.24.2
edgeR	(Robinson et al., 2010)	*via* R/Bioconductor v. 3.16.5
GSVA (with ssGSEA)	(Hanzelmann et al., 2013; Verhaak et al., 2013)	*via* R/Bioconductor v. 1.26
TopHat	(Kim et al., 2013)	v. 2.0.9
Integrative Genomics Viewer	(Thorvaldsdottir et al., 2013)	http://software.broadinstitute.org/software/igv/ v. 2.3.80
Cuffdiff	(Trapnell et al., 2013)	v. 2.1.1
BWA	(Li and Durbin, 2009)	http://bio-bwa.sourceforge.net/ v. 0.6.2
MACS	(Zhang et al., 2008)	http://liulab.dfci.harvard.edu/MACS/ v. 1.4.2
FastQC	Babraham Bioinformatics	http://www.bioinformatics.babraham.ac.uk/projects/fastqc v. 0.10.1 – v. 0.11.5
Other		
Epithelial and mesenchymal gene lists for cell lines and tissues	(Tan et al., 2014)	http://embomolmed.embopress.org/content/embomm/6/10/1279/DC2/embed/inline-supplementary-material-2.pdf *Accessed Jun 2016*
Chemidoc Touch imaging system	BioRad	1708370
Confocal microscope	Zeiss	LSM 700

CONTACT FOR REAGENT AND RESOURCE SHARING

Requests for further information should be directed to the Lead Contact Dr Melissa Davis (davis.m@wehi.edu.au).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Tissue Culture Cell Lines

The Human Mammary Epithelial Cells (HMLE) cell line is derived from a female patient and was obtained from the ATCC. The HMLE cell line was grown in HuMEC Ready Media. A HMLE subline with a mesenchymal phenotype (MesHMLE) was achieved by culture in DMEM:F12 media (1:1) supplemented with 10 mg/ml insulin, 20 ng/ml EGF, 0.5 mg/ml hydrocortisone, and 5% fetal calf serum (FCS) and treatment with 2.5 ng/ml of TGFB1 for 24 days. MesHMLEs were then maintained in this supplemented media without TGFB1. Cell line identity was authenticated by short tandem repeat profiling and mycoplasma testing was performed. Cells were grown at 37°C with 5% CO2 and saturated air humidity.

Selection of Target miRNAs

To consider miRNAs for perturbation in the context of EMT a literature search was performed, identifying over 130 miRNAs directly implicated in regulating EMT (Table S1). As detailed below, experimental validation focussed on miRNAs which also undergo extensive changes in the HMLE to MesHMLE transition.

METHOD DETAILS

Tissue Culture

MesHMLEs were seeded into T25 flasks at a low density to allow for expansion and transfected with miRNA mimics using Lipofectamine RNAi-Max for 96h. Cells were then transfected with fresh reagents for an additional 24h before being used in proliferation and migration assays. HMLEs were plated at $8x10^4$ cells/mL in 12 well plates and transfected with antagomiRs using Lipofectamine RNAi-Max for 72h. Where indicated, HMLE cells were transfected with miRNA inhibitor negative control, or miRNA inhibitors against miR-200c, miR-141, miR-182, and/or miR-183 at a final concentration of 100nM. Individual antagomiRs were transfected at 25nM with the negative control used to bring the final concentration to 100nM. For immunofluorescence, glass coverslips were inserted into each well and incubated in fibronectin solution (25 μg per well at 50 μg/mL) at 37°C for 30 minutes. Excess fibronectin solution was removed using a pipette and cells were plated onto the coverslips.

Isolation of RNA and Real-Time PCR

Total RNA was extracted using Trizol according to manufacturer instructions. Reverse transcription was performed using the QuantiTect reverse transcription kit. Primers are listed in Table S5.

Western Blotting

Transfected MesHMLE cells were lysed in RIPA buffer with protease inhibitor cocktail and purified to produce whole cell extracts. 30 μg total protein was fractionated on 10% Bis-Tris Bolt gels. A 100 mM Tris-Cl, 150mM NaCl, and 0.05% Tween (TNT) buffer at pH 7.5 was used. Proteins were transferred onto nitrocellulose membranes, blocked (5% skim milk powder in TNT) and probed in blocking buffer with the following primary antibodies: α-Tubulin (1:5000), ZEB1 (1:500), E-Cadherin (1:1000) and Vimentin (1:1000) overnight at 4°C. Blots were washed in TNT then probed with secondary antibodies (goat anti-rabbit HRP or goat anti-mouse HRP) and visualized on the Chemidoc Touch imaging system following application of Clarity Western ECL substrate.

Immunofluorescence

Following transfection, cells were washed with warm PBS and fixed in 4% paraformaldehyde (PFA) for 10 minutes. The cells were then permeabilized using wash buffer (0.1% Triton X-100/TBS) for 10 minutes, and blocked with 2% BSA in wash buffer for 1 hour at RT. Primary antibodies for E-Cadherin (1:500,) and Vimentin (1:100) were diluted in block buffer and applied to the coverslips which were incubated overnight at 4°C. Following three 5 minute washes, the coverslips were incubated with goat anti-mouse AlexaFluor 488 and anti-rabbit AlexaFluor 594 for 1 hour at RT. After washing, coverslips were mounted using ProLong Gold antifade reagent with DAPI. Six images were taken per coverslip using a Zeiss LSM 700 confocal microscope.

Cellular Proliferation Assay

Transfected MesHMLE cells were dissociated, counted, and replated into four 96 well plates in quadruplicate at a density of $2x10^3$ cells per well. Cells were left to settle at 37°C for 4 hours, after which plate 1 was removed (t=0h). The media was aspirated and the plate was wrapped in parafilm and stored at -20°C. This was repeated for plates 2 (t=24h), 3 (t=48h) and 4 (t=72h). The CyQUANT Cell Proliferation Assay Kit was then used to quantify DNA content and thus, measure proliferation as per manufacturer instruction. CyQUANT GR dye was applied to each well and sample fluorescence immediately measured using the Fluostar, with filters set for excitation at 480nm and emission at 520nm.

Transwell Migration Assay

Transwell inserts were humidified with serum free media at 37°C for an hour. Meanwhile, transfected MesHMLE cells were dissociated, counted and resuspended in serum free media (DMEM + EGF + insulin + hydrocortisone). Media was aspirated from transwells, which were then placed above 600μL full media (5% added FCS). $2x10^5$ transfected cells were added to individual transwells in triplicate, and incubated at 37°C for 4 hours. Each transwell was rinsed in PBS, then fixed in buffered formalin overnight at 4°C. Following a PBS rinse, cells were removed from the inside membrane of the transwells using a cotton tip, and washed off with PBS. Transwells were incubated with 0.1% Triton X for 5 minutes to permeabilise the cells, which were subsequently rinsed with water and incubated with DAPI for 30 minutes at room temperature. The transwells were washed once more, the membranes dried and mounted onto microscope slides with fluorescent mounting medium. DAPI was imaged in 6 distinct locations per membrane.

RNA-sequencing

Poly(A) enriched mRNA was extracted from HMLE and MesHMLE cells (biological triplicates), and miR-200c transfected MesHMLE cells (biological duplicates). Poly(A) depleted mRNA was extracted from HMLE cells (biological duplicate) and MesHMLE cells (biological triplicate). All mRNA RNA-seq was collected using the Illumina HiSeq 2500 platform with a stranded paired end protocol, and read length of 100. On average, 95 million raw paired reads were obtained for each sample (Table S4A). Small RNA-seq was performed as previously described (Bracken et al., 2014). For data accession numbers please refer to the Key Resources Table.

Histone Modification ChIP-seq

Histone modification ChIP-Seq in HMLE and MesHMLE cells was performed as previously described (Lim et al., 2013; Attema et al., 2013). Antibodies are listed in the Key Resources Table.

QUANTIFICATION AND STATISTICAL ANALYSIS

Real-Time Quantitative PCR

Real-time PCR values were quantified relative to the average values for GAPDH and HPRT. Where indicated, significance was determined using a two-sided t-test for independent samples without assuming equal variance (*i.e.* Welch's t-test), with the null hypothesis that the sample means are equal. Statistical testing was performed using the SciPy (Jones et al., 2014) stats module.

Immunofluorescence

Across the six images taken, FIJI software was used to calculate the fluorescence intensity (integrated density) for each channel. These values were normalized to the DAPI stain and averaged to obtain the mean fluorescence intensity per condition.

Cellular Proliferation Assay

CyQUANT fluorescence intensity data were blank corrected, averaged between technical replicates and normalised against the t=0h time point for each transfection. Data are represented as mean fold change and standard deviation.

Transwell Migration Assay

Using image data from the transwell membranes, the number of cells per image was counted and the average per filter was then averaged and normalised to the cell density of the original suspensions (as determined by the CyQUANT assay described above).

Histone Modification ChIP-seq

ChIP-seq analysis was performed as previously described (Attema et al., 2013; Lim et al., 2013). Briefly, read data were aligned to the hg19 genome using BWA (Li and Durbin, 2009) and MACS (Zhang et al., 2008) was used to call peaks and determine differential enrichment above local anti-Histone H3 background. Scores (Figure S7C) reflect the confidence of enrichment for each modification of interest in a region -1 kb to +1kb from the annotated transcription start site of each gene.

RNA-sequencing

Raw reads were adapter trimmed and filtered using cutadapt v1.8 (Martin, 2011), setting the minimum-length to 18, the error-rate to 0.2, and allowing an overlap of 5 bases. The resulting FASTQ files were analysed and quality checked using FastQC. Filtered reads were mapped against the human reference genome (hg19) using the TopHat spliced alignment algorithm (Kim et al., 2013) producing an average alignment rate of 91% (Table S4). Alignments were visualised and interrogated using the Integrative Genomics Viewer (Thorvaldsdottir et al., 2013). Cuffdiff (Trapnell et al., 2013) was used to quantify gene expression and estimate significance of differential transcript abundance.

RNA-seq data were also processed using a parallel alignment/quantification pipeline. The trimmed read data (*described above*) were aligned to the hg19 reference genome using the R/Bioconductor package (Huber et al., 2015) Rsubread (Chen et al., 2016). For gene set scoring (Figures 1D and S1B) and batch analysis (Figure S8), logCPM values were obtained using the R/Bioconductor package edgeR without TMM normalisation (Robinson et al., 2010). These data were also used for a parallel EISA analysis (*details below*).

Exon-Intron Split Analysis

The exon-intron split analysis (EISA) was performed essentially as described in Gaidatzis et al. (Gaidatzis et al., 2015). Custom python scripts were used for all steps. HMLE and MesHMLE data from PRJEB8225 and PRJEB25061 were used (as these data contain more intronic reads), while MesHMLE+miR-200c data were taken from PRJEB25042. Briefly, using only non-overlapping genes and uniquely mapped reads, we quantified the number of reads in intronic and exonic regions in a strand-specific manner for all UCSC RefSeq mRNA transcripts from each gene. Read pairs were 'exonic' if the 5' end of the first read was mapped within an exon of any UCSC transcript or 'intronic' if the first read mapped entirely within an intron and not within 10 base pairs of an exon. Genes with insufficient read coverage (less than 24 reads in either intron or exonic read count) were discarded and the data was log2 transformed after adding a pseudo-count of 8. For each sample, normalisation was performed separately for intronic and exonic counts, dividing through by the total number of reads. To allow comparison between experiments, counts were then multiplied by a standardised ratio of coverage of 40 for exons and 2 for introns (the coverage, in millions of reads, for a representative polyA-selected sample). The analysis was performed for all possible combinations of pairs of samples where one sample was from each experimental group (e.g. Group A replicate 1 vs Group B replicate 1, A1 vs B2, A2 vs B1, A2 vs B2). For each pair of individual replicates, Δexon and Δintron were calculated as the difference in log2 read counts between experimental conditions. To summarise across replicates, the means of the Δintron or Δexon values were calculated; in text these are referred to as 'Δintron' and 'Δexon'. The data were further filtered to ensure all of the examined genes were expressed (above the minimum coverage above) in all replicates in at least one experimental group.

For the parallel exon-intron split analysis, intronic and exonic counts from Rsubread were summarised using featureCounts (Liao et al., 2014). Comparing Δintron and Δexon values from this alternative mapping to the values obtained by the standard EISA analysis some quantitative differences were observed, although the results were largely consistent, particularly for the Δexon data (Figure S2).

Computational Analysis

Subsequent computational analyses were performed using python (v 3.6) with pandas (McKinney, 2010), NumPy (Walt et al., 2011), SciPy (Jones et al., 2014), scikit-learn (Pedregosa et al., 2011), Matplotlib (Hunter, 2007), and networkx (Schult and Swart, 2008).

Gene Identifiers

Gene identifiers were mapped against the HUGO Gene Nomenclature Committee (HGNC) database (Gray et al., 2015) and ENSEMBL BioMart data tables (Aken et al., 2017; Yates et al., 2016) using python dictionaries.

Gene Set Scoring

A full description of the gene set scoring approach is given in (Foroutan et al., 2017). Briefly, a single-sample gene set enrichment analysis (Verhaak et al., 2013) (ssGSEA) was applied through the R/Bioconductor package GSVA (Hanzelmann et al., 2013) to score each sample against the epithelial and mesenchymal gene sets for tumour samples or for cell lines from Tan et al (Tan et al., 2014). The pandas corr function was used to calculate the Pearson's correlation between miRNA abundance and epithelial score or mesenchymal score, and between the abundance of miRNAs.

Partitioning of Gene Sets by mRNA Transcript Differential Expression

To partition gene sets containing mRNAs with evidence of differential regulation through alternative transcriptional and post-transcriptional processes (Figure 2A), the top/bottom 15th percentile of transcripts by Δintron and Δexon-Δintron were taken for regulated gene sets (Table 1). The genes with 'no change' were selected through Δintron, Δexon-Δintron and differential transcript abundances centred around zero, with the percentile values indicated (Table 1, 'No change'). The resulting number of gene transcripts within each set is shown on Figure 2B.

Gene Ontology Analysis

Gene ontology (GO) (Ashburner et al., 2000; Gene Ontology Consortium, 2015) annotations were downloaded and using the underlying network structure (*i.e.* parent/child terms) GO membership was propagated up the annotation inheritance tree using networkx (Schult and Swart, 2008) to improve coverage (Maetschke et al., 2012). To test the enrichment of membership for specific GO processes a χ^2-test was applied with one degree of freedom (gene is a member/not a member) using expected values derived from genome-wide frequencies/marginal probabilities. GO annotations with less than 20 or more than 500 members were discarded. For multiple hypothesis testing a Bonferroni correction was applied using the number of remaining GO annotations (n=3691).

GO Enrichment with Limma and FRY

A limitation of the χ^2 test for gene set enrichment is a loss of quantitative information beyond defining transcripts as differentially expressed (or not). An alternative GO enrichment analysis was performed with Δexon and Δintron values from the Rsubread alignment using the FRY method for ROAST (Goeman and Buhlmann, 2007; Phipson and Smyth, 2010; Wu et al., 2010) in the R/Bioconductor package limma (Ritchie et al., 2015), with GO categories derived through propagation as described above (but here, performed using R). The analysis of Δexon-Δintron data was achieved by specifying an appropriate design matrix for FRY.

The methods used and hypotheses tested by FRY are slightly different to those employed for the Gene Ontology enrichment performed within the main text, and thus small differences in the selected are to be expected, although in general the results were similar (Tables S3B–S3D).

HMLE System miRNA Ranking

To identify relatively-high abundance miRNAs which underwent large differential changes during the HMLE to MesHMLE transition (Figure 3A) we ranked miRNAs by the product of their summed abundance (x; across both HMLE and MesHMLE states) and the absolute value of the $\log_2$ fold change:

$$\text{miRNA score} = \log_{10}(x_{\text{HMLE}} + x_{\text{MesHMLE}}) * \log_2|(x_{\text{HMLE}}/x_{\text{MesHMLE}})|$$

Predicted miRNA Targets

Predicted miRNA targets were taken from TargetScan (Friedman et al., 2009; Grimson et al., 2007; Lewis et al., 2005) for miRNA families, and DIANA-microT CDS (Paraskevopoulou et al., 2013; Reczko et al., 2012) for individual miRNAs. To select for relatively high-confidence putative targets we took the high-confidence (top 5 percentile of) predictions from TargetScan (ranked by *context+*), and DIANA-microT (ranked by *miTG-score*). Distributions of these metrics are shown in Figure S6.

There were some minor discrepancies where miRNAs mapped within the HMLE system data could not be directly mapped to the TargetScan database. In particular: hsa-miR-381 (data) was mapped to hsa-miR-381-3p; hsa-miR-411-5p was mapped to

both hsa-miR-411-5p.1 and hsa-miR-411-5p.2; and hsa-miR-183-5p was mapped to both hsa-miR-183-5p.1 and hsa-miR-183-5p.2. It should be noted, however, that at least in the case of miR-183-5p, different isomiRs have shown enrichment in triple negative breast cancer which is dependent upon patient race (Telonis et al., 2015), and these effects are not considered here.

MicroRNA Targeting Analysis

To examine the enrichment of high confidence miRNA targets within the EISA-filtered gene sets (Figure 3C), the relative fraction of transcripts within each EISA-partitioned group (Figure 2A) was determined, and for each miRNA this was used to calculate the number of targets expected within each EISA-filtered set. The relative enrichment was calculated as: $(n_{observed} - n_{expected})/n_{expected}$.

A schematic illustration of target overlap quantification (Figure 3E) is shown in Doc S1, together with an expanded description of the method. Briefly, to examine the significance of direct miRNA target set overlap (Figure 3E, *at top right*), the frequency of targets for a miRNA was used to calculate an expected frequency of shared targets when compared to another miRNA, under the assumption of independence between miRNA target sets. A χ^2-statistic was calculated using the observed and numbers of shared targets, and a χ^2-test was applied with 1 degree of freedom to estimate a *p*-value for evidence against the null hypothesis that miRNA target sets are independent. A Bonferroni correction was applied, and adjusted *p*-values were calculated by multiplying the estimated *p*-value by the total number of miRNA pairs: $(n_{miRNAs}{}^{*}(n_{miRNAs} - 1))/2$.

Next, a protein-protein interaction (PPI) network was constructed by combining experimentally observed interactions from BioPlex (Huttlin et al., 2015), PINA v2.0 (Cowley et al., 2012) and InnateDB (Breuer et al., 2013). For each pair of miRNAs (Figure 3E, bottom left) we used the target lists to construct a bipartite graph, and calculated the resulting graph density ($\rho_{bipartite}$) - a function of the number of edges between bipartite sets ($n_{edgesBetween}$), the number of nodes targeted only by miR-*x* ($n_{Targ(miR\text{-}x\ \&\ \sim miR\text{-}y)}$) and the number of nodes targeted only by miR-y ($n_{Targ(miR\text{-}y\ \&\ \sim miR\text{-}x)}$):

$$\rho_{bipartite} = n_{edgesBetween} \Big/ \left(n_{Targ(miR-x\ \&\sim \mathrm{miR-y})}\right)\left(n_{Targ(miR-y\ \&\sim \mathrm{miR-x})}\right)$$

The distribution of density values was Gaussian and approximately Normal, thus density values (ρ) were converted to a *Z*-score using the sample mean $(\hat{\mu})$ and standard deviation $(\hat{\sigma})$: $Z(\rho_i) = \rho_i - \hat{\mu}(\rho)/\hat{\sigma}(\rho)$. As illustrated, this approach ignores nodes (mRNA transcripts) targeted by both miRNAs and edges (protein interactions) between targets of an individual miRNA; however both of these features will to an extent be captured by the direct miRNA target overlap (Figure 3E *at top right*).

DATA AND SOFTWARE AVAILABILITY

HMLE System RNA-seq Data

HMLE cell line model data used in this study are available from the European Nucleotide Archive. For further details please refer to the Key Resources Table and Table S4.

Machine readable versions of supplementary tables and data are available through the project GitHub repository (*see below*).

Computational Scripts

Computational scripts associated with this report are available under an MIT license, from: http://github.com/DavisLaboratory/Combinatorial_miRNAs.