Sequencing of Captive Target Transcripts Identifies the Network of Regulated Genes and Functions of Primate-Specific miR-522

Shen Mynn Tan, 1,2 Rory Kirchner, 3 Jingmin Jin, 4 Oliver Hofmann, 3 Larry McReynolds, 4 Winston Hide, 3,5 and Judy Lieberman 1,2,*

1Cellular and Molecular Medicine Program, Boston Children’s Hospital, Boston, MA 02115, USA
2Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA
3Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA
4New England Biolabs, Ipswich, MA 01938, USA
5Harvard Stem Cell Institute, Cambridge, MA 02138, USA
*Correspondence: judy.lieberman@childrens.harvard.edu

http://dx.doi.org/10.1016/j.celrep.2014.07.023

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

SUMMARY

Identifying microRNA (miRNA)-regulated genes is key to understanding miRNA function. However, many miRNA recognition elements (MREs) do not follow canonical “seed” base-pairing rules, making identification of bona fide targets challenging. Here, we apply an unbiased sequencing-based systems approach to characterize miR-522, a member of the oncogenic primate-specific chromosome 19 miRNA cluster, highly expressed in poorly differentiated cancers. To identify miRNA targets, we sequenced full-length transcripts captured by a biotinylated miRNA mimic. Within these targets, mostly noncanonical MREs were identified by sequencing RNase-resistant fragments. miR-522 overexpression reduced mRNA, protein levels, and luciferase activity of >70% of a random list of candidate target genes and MREs. Bioinformatic analysis suggested that miR-522 regulates cell proliferation, detachment, migration, and epithelial-mesenchymal transition. miR-522 induces G1 cell-cycle arrest and causes cells to detach without anoikis, become invasive, and express mesenchymal genes. Thus, our method provides a simple but effective technique for identifying miRNA-regulated genes and biological function.

INTRODUCTION

Identification of microRNA (miRNA) target genes and the specific sequences they bind (miRNA recognition elements [MREs]) is important for understanding miRNA function (Bartel, 2009). Target prediction algorithms and experimental methods can be combined to identify candidate miRNA-regulated genes (Bartel, 2009; Thomas et al., 2010). However, even the best methods leave room for improvement. Target prediction algorithms are mostly based on a set of canonical rules for miRNA-mRNA interactions, in the 3′ UTR of the miRNA, mediated by complementarity to the seed region (nucleotides 2–8 of the miRNA). However, increasing evidence suggests that noncanonical MREs, involving bulges and G:U wobbles, pairing outside the seed and even “seedless” interactions, in the 3′ UTR as well as CDS, are not uncommon (Lal et al., 2009; Tay et al., 2008; Xia et al., 2012). The canonical rules may apply only to some miRNAs and to only a quarter of all miRNA-mRNA interactions overall (Helwak et al., 2013).

Biochemical isolation of target mRNAs bound to AGO and the RNA-induced silencing complex has improved MRE identification. AGO pull-down, combined with crosslinking and RNase digestion (high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation [HITS-CLIP], crosslinking and immunoprecipitation followed by high-throughput sequencing [CLIP-Seq], and photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation [PAR-CLIP]), generates a global snapshot of miRNA-mRNA interactions (Chi et al., 2009; Hafner et al., 2010; Loeb et al., 2012; Zisoulis et al., 2010). However, these methods, even when combined with manipulation of miRNA expression, may not be adequate for identifying the mRNAs regulated by a particular miRNA. In fact, they usually rely on seed-based analysis or predictive algorithms to match the isolated MRE sequences to miRNAs. Recently, a novel unbiased method of mapping RNA-RNA interactions, termed crosslinking ligation and sequencing of hybrids (CLASH), was used to identify miRNA-MRE pairs by ligating bound MREs to their cognate miRNAs (Helwak et al., 2013). These global miRNA-binding methods are technically challenging, require large numbers of cells, and are most effective for identifying highly expressed miRNAs and their targets. UV crosslinking, which is central to these methods, is inefficient and can introduce background. For example, it increases the false discovery rate in genome-wide studies of RNA-protein interactions (Keene et al., 2006). Although use of photoactivatable nucleosides, as in PAR-CLIP, increases the specificity of target identification, it restricts crosslinking to the photoreactive nucleotide, potentially causing sequence bias (Kishore et al., 2011).
Figure 1. Identification of miR-522 Targets by Pulldown-Seq and Microarray Analysis

(A) Schematic of the strategy used to identify miR-522 target transcripts, MREs, and function.

(B) Volcano plot of gene enrichment in miR-522 versus control Pulldown-seq data set. Red box denotes genes enriched >2-fold with p < 0.05.

(C) Bubble plot of genes in the red box in (B). Genes significantly downregulated after miR-522 overexpression by microarray analysis (FDR < 0.1) are in red, those not downregulated are in blue, and those not represented on the microarray are in gray. The size of each bubble is proportionate to the extent of downregulation. Note the higher proportion of red bubbles and larger bubbles with increasing enrichment.
Bona fide targets for specific miRNAs can be identified without bias by isolating mRNAs bound to a transfected miRNA mimic biotinylated at the 3’ end of the active strand (Lal et al., 2011; Ørom et al., 2008). These transfected miRNA mimics are incorporated into the miRNA-induced silencing complex (miRISC) and bind to and downregulate known miRNA targets. Their association with target mRNAs is not a postlysis artifact, because they do not pull down miRNA targets when added to lysed cells (Lal et al., 2011). These pull-downs do not require crosslinking and can be performed with just a million cells (~20-fold fewer than crosslinking protocols). Typically hundreds to thousands of transcripts are identified as candidate genes regulated by an individual miRNA. Despite these large numbers, the method has high specificity (90% in one study; Lal et al., 2011). However, this method has some shortcomings. Because it uses mRNA microarrays to identify enriched candidate targets, it only identifies mRNAs probed on the array and does not directly identify MREs. Identifying the key miRNA targets and function by altering the protocol and subsequent functional analysis to generate a simple, systems-level analysis of the transcription factors that regulate the target genes regulated by an individual miRNA. Despite these large numbers, the method has high specificity (90% in one study; Lal et al., 2011). However, this method has some shortcomings. Because it uses mRNA microarrays to identify enriched candidate targets, it only identifies mRNAs probed on the array and does not directly identify MREs. Identifying the key miRNA targets and function by altering the protocol and subsequent functional analysis to generate a simple, systems-level analysis of the transcription factors that regulate the target genes regulated by an individual miRNA. Despite these large numbers, the method has high specificity (90% in one study; Lal et al., 2011). However, this method has some shortcomings.

RESULTS

miR-522 Is Overexpressed in TNBCs and Other Aggressive Cancers, and Its Amplification Is Associated with Poor Prognosis

Four C19MC miRNAs (miR-517c, miR-519a, miR-521, and miR-522) are overexpressed in TNBCs (Enerly et al., 2011), but only miR-522 is consistently significantly overexpressed in poorly differentiated estrogen receptor (ER)-negative or TNBC, relative to more differentiated ER+ and luminal breast cancers or normal breast tissue, in four additional independent data sets (Biagioni et al., 2012; Buffa et al., 2011; Lowery et al., 2009; Cancer Genome Atlas Network, 2012; Figure S1A). The difference in miR-522 expression between luminal breast cancer and TNBCs was confirmed using ten representative cell lines (Figure S1B).

Identification of miR-522 Target Transcripts by Pulldown-Seq

To identify genes regulated by miR-522, we used the human TNBC cell line MDA-MB-468, which expresses miR-522, as a model system. The RNAs pulled down with biotinylated miR-522 and cel-miR-67 (control miRNA) were compared by RNA sequencing (Pulldown-seq). We obtained 550 and 403 million uniquely aligned bases from two biological replicate experiments, each sequenced in duplicate, from miR-522 and control miRNA samples, respectively, of which about 75% aligned to annotated transcripts. miR-522 target fold enrichment (FE) and p values (p) compared to control miRNA, calculated based on the fragments per kilobase of exon per million fragments mapped (FPKM) of each data set, were depicted in a volcano plot (Figure S1C). A total of 547 RNAs (mostly miRNAs [85%] but also novel transcripts, pseudogenes, large intergenic noncoding RNAs [lincRNAs], and miRNAs) satisfied the cutoff conditions of FE > 2 and p < 0.05 (Figure S1D; Table S1). Depletion of cell lysates with pan-AGO antibody and qRT-PCR analysis of three captured miR-522 target mRNAs showed that Pulldown-seq mostly identified AGO-associated target mRNAs, because the AGO antibody depleted 70%–80% of these targets (Figure S1E).

The mRNA of most bona fide miRNA targets is diminished by miRNA overexpression (Bartel, 2009; Lal et al., 2011). Therefore, to assess the quality of mRNA target identification, Illumina beadchip expression microarrays were used to identify genes downregulated by miR-522 overexpression. The accuracy of the microarray data was confirmed by qRT-PCR analysis of 12 randomly selected genes (R² = 0.76; Figure S1F). Of the 547 miR-522 targets, 428 were probed on the microarray and 255 (60%) had significantly downregulated mRNA (false discovery rate [FDR] < 0.1; Figure 1C; Table S1). Cumulative frequency

(D) Cumulative gene expression after miR-522 overexpression, comparing miR-522 targets identified by Pulldown-seq using three different cutoffs. FE, fold enrichment (‘‘K-S test p < 0.05; ‘‘‘p < 0.001).

(E) Cumulative gene expression plot comparing miR-522 targets identified by Pulldown-seq (FE > 2; p < 0.05) with those predicted by target prediction algorithms (‘‘‘p < 0.001).

See also Figure S1.
plots compared mRNA changes after miR-522 overexpression of genes enriched in the pull-down, defined by different cutoffs, with changes in all genes or algorithm-predicted target genes (Figures 1D and 1E). The decrease in mRNA of the 47 genes selected from the pull-down data using a FDR of <0.1 (40 on the microarray) was similar to the decrease in gene expression of the 197 genes (130 on the microarray) chosen using a cutoff of FE > 4 and p < 0.01. The less stringent cutoff (FE > 2; p < 0.05) used to define the 547 miR-522 target genes also highly enriched for downregulated miRNAs, outperforming all five algorithm-predicted gene lists (TargetScan, miRanda, PITA, DIANA, and RNA22). Of the algorithms, TargetScan-predicted targets were the most significantly downregulated after miR-522 overexpression (Figure 1E). Pulldown-seq significantly outperformed TargetScan, even though it identified more targets (428 versus 331; p < 0.001). Of the 547 candidate target genes, only 53 (47 on the array) were also predicted by TargetScan.

**Most miR-522-Associated Genes Had Reduced mRNA and Protein after miR-522 Overexpression**

To evaluate the specificity of the pull-down, a random list of 30 genes representing the full range of FE and p values of identified miR-522 targets was chosen. The effect of miR-522 overexpression on mRNA and protein levels was assessed by microarray and immunoblot densitometry, respectively (Figure 2; Table S1). Eighty-three percent (25 of 30) showed significantly reduced mRNA, and 73% (22 of 30) had significantly reduced protein. Thus, Pulldown-seq is very specific for identifying miRNA targets. To test whether endogenous miR-522 suppresses expression of putative Pulldown-seq target genes, we used qRT-PCR to quantify selected target gene miRNAs in cells transfected with anti-miR-522 or a control hairpin. Antagonizing endogenous miR-522 significantly increased three of four miR-522 target genes also encoded in lincRNAs, miRNAs, pseudogenes, and antisense RNAs. Of the 2,467 3′ UTR MREs, only 111 were predicted by PITA, miRanda, and/or TargetScan (Tables S1 and S2).

IMPACT-seq miR-522 targets were also significantly downregulated by miR-522 overexpression but to a lesser extent than Pulldown-seq targets (Figure 3D). The extent of downregulation was similar to the miranda gene set, but not as good as the TargetScan-predicted genes. Of the 4,848 IMPACT-seq MREs, 743 were in 56% (306 of 547) of Pulldown-seq transcripts. The 275 genes identified by both Pulldown-seq and IMPACT-seq performed comparably to the Pulldown-seq target set (Figure 3D). The lack of MREs for 44% of Pulldown-seq miRNAs and the less-striking downregulation of IMPACT-seq genes suggest that IMPACT-seq could be improved.

To analyze the regions of miR-522 complementarity in these sequences, the GLAM2 tool in the MEME suite of motif-based sequence analysis tools (which allows for gaps or bulges) was used to discover motif(s) (Frith et al., 2008). Reasoning that MREs between 25 and 35 bases in length would contain only single-miR-522-binding sites, we limited analysis to these 1,887 sequences. We found that 1,639 sequences contained an enriched motif partially complementary to residues 13–20 of miR-522 (score: 9,426), and 1,753 contained an enriched motif partially complementary to residues 2–9 in the seed region of miR-522 (score: 9,113). One hundred scrambled versions of the RNAse-protected sequences had a significantly lower GLAM2 score of 2,576 ± 1,245 (Figure 3E). Most sequences contained motifs partially complementary to both the 5′ and 3′ ends of miR-522. Thus, miR-522 appears to belong to a recently described class of miRNAs, whose MREs are imperfectly complementary to both mRNA ends (Helwak et al., 2013). Fifty-nine percent of all transcripts containing these GLAM2 motifs were significantly downregulated after miR-522 overexpression. Their cumulative frequency plots were also statistically indistinguishable, supporting our conclusion that miR-522 MREs contain imperfect complementary matches at both ends (Figure 3D; data not shown).

To test the accuracy of IMPACT-seq MRE identification, we cloned a random set of 30 putative MREs (representing the range of read number and FE scores, most of which had a GLAM2 motif) into luciferase reporter plasmids and performed reporter assays (Table S3). Only four were predicted targets and five (17%) had a perfect seed match. Four sequences of similar length that were not identified as MREs, but were encoded in MRE-containing genes, were used as controls. Twenty-five of
the candidate MREs (83%) reduced luciferase activity (Figure 3F), suggesting that IMPACT-seq is very specific. To confirm that these genes were indeed direct targets of miR-522, we cloned the full-length 3’UTRs of eight genes with 3’UTR MREs validated in Figure 3F and performed luciferase reporter assays (Figure 3G). Only one of these did not reduce luciferase activity upon miR-522 overexpression, indicating that most of these MREs function within the endogenous 3’UTR.

**miR-522 Target Functional Analysis**

To uncover the function of miR-522, we combined several bioinformatics tools to analyze the Pulldown-seq data set. First, we...
### Figure A

<table>
<thead>
<tr>
<th></th>
<th>BANF1</th>
<th>ADIPOR1</th>
<th>MAL2</th>
<th>SOX4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pulldown-seq reads</td>
<td>Control</td>
<td>miR-522</td>
<td>Control</td>
<td>miR-522</td>
</tr>
<tr>
<td>Control</td>
<td>5</td>
<td>34</td>
<td>64</td>
<td>166</td>
</tr>
<tr>
<td>miR-522</td>
<td>95</td>
<td>163</td>
<td>514</td>
<td>777</td>
</tr>
</tbody>
</table>

### Figure B

![Graph showing number of reads vs sequence length](image)

### Figure C

- mRNA (3'UTR)
- mRNA (5'UTR)
- lincRNA
- miRNA
- Pseudogenes
- antisense RNA
- Other RNA

### Figure E

miR-522: 3' TGTTGAGATTTCCCTTGTTAAAAA 5'

Scrambled

### Figure F

Graph showing relative luciferase activity (% of control) for different genes

#### MRE Legend:
- 5'UTR
- CDS
- 3'UTR
- Predicted 3'UTR

### Figure G

Graph showing relative luciferase activity (% of control) for different genes

- Control
- miR-522
used TRANSFAC, a well-curated knowledge base of eukaryotic transcription factors (Matys et al., 2006), to search for overrepresented transcription factors predicted to bind to the promoter regions upstream of miR-522 target genes. Binding sites for four transcription factors, ELK1, E2F, TEAD2, and PAX3, which are all expressed in MDA-MB-468, were significantly enriched in the promoters of the 547 Pulldown-seq genes (Figure 4A). ELK1 and E2F3 were also identified as miR-522 target genes that were enriched 17- and 2.8-fold in the pull-down, respectively. The mRNA and protein levels of both were significantly downregulated by miR-522 overexpression (Table S1; Figure 2). ELK1 and PAX3 both enhance the epithelial state. ELK1-regulated genes are downregulated during transforming growth factor β (TGF-β)-induced epithelial mesenchymal transition (EMT) (Venkov et al., 2011), whereas PAX3 expression enhances mesenchymal epithelial transition (MET) (Wiggan et al., 2002). Of note, TEAD2 transcriptionally activates PAX3 (Milewski et al., 2004). This analysis suggests that miR-522 might foster EMT. E2F transcription factors promote G1/S progression in the cell cycle (Fan and Bertino, 1997), suggesting that miR-522 might inhibit progression at this phase of the cell cycle.

miRNAs can target multiple genes that participate in common pathways (Lal et al., 2009, 2011). We therefore next used ingenuity pathway analysis (IPA), which curates biological interactions and function (Ingenuity Systems; http://www.ingenuity.com), to identify all directly related miR-522 target genes (Figure S2). This interactome of 221 genes was then analyzed using IPA to identify the top network-associated functions (Figure 4B). These included cellular movement and morphology (functional components of EMT), embryonic development (of which EMT is an essential component), as well as cell cycle, corroborating the TRANSFAC analysis. To include possible indirect functional effects of miR-522, we next identified common molecular functions overrepresented in both the Pulldown-seq interactome and set of 205 directly connected genes whose mRNAs were significantly downregulated after miR-522 overexpression (p < 0.0001 and fold change > 1.5; Table S4). The p values for each common overrepresented molecular function were displayed in a bubble plot, where the number of miR-522 target genes identified by Pulldown-seq for each function is proportional to the bubble size (Figure 4C). The most significantly enriched functions, for which more than ten target genes were annotated, fell into four large functional categories: proliferation, apoptosis, cell cycle, and cell morphology/migration/transformation. The latter two categories also stood out in both the TRANSFAC and IPA analyses. Based on these bioinformatics, we hypothesized that miR-522 might regulate G1/S cell cycle progression, cell transformation and survival, cell motility, and EMT, which we next investigated in MDA-MB-468 TNBC cells.

**miR-522 Overexpression Induces Mesenchymal Properties**

Loss of cellular adhesion is an easily detectable phenotype that involves changes in cell morphology, movement, and cytoskeletal organization and is an important step in EMT (Yang and Weinberg, 2008). In tissue culture, most MDA-MB-468 cells are adherent, but some detach and remain viable. We first assessed by TaqMan PCR whether miR-522 levels might be different in adherent versus nonadherent cells. miR-522 was expressed at significantly higher levels in nonadherent than adherent cells, but another C19MC miRNA (miR-519a), miR-16, and let-7a were expressed similarly (Figure 5A). Overexpression of miR-522 decreased the number of live adherent cells and increased the number of live cells in suspension (Figure 5B). To assess whether miR-522 expression was responsible for loss of adhesivity, we collected suspension cells that were transfected 48 hr previously with either control miRNA or miR-522 and re-transfected an equal number of each with control miRNA, miR-522, or anti-miR-522. Only cells transfected with anti-miR-522 formed adherent colonies (Figure 5C). Thus, miR-522 promotes detachment and viability of detached cells.

We next investigated whether miR-522 promotes EMT. Changes in important mesenchymal gene (ZEB2, TWIST1, FOXO2, SNAI1, SNAI2, and VIM) transcripts were assessed by qRT-PCR in adherent and nonadherent cells 3 and 5 days after miR-522 or control miRNA transfection (Figures 5D and 5E). Expression of all six genes strongly and progressively increased in both cell populations but most dramatically in nonadherent cells. However, epithelial genes encoding E-cadherin (CDH1) or the cytokeratins (KRT14, KRT18, and KRT19) did not significantly change (Figure S3A). These results are consistent with the notion that poorly differentiated cancer cells exhibit plasticity in epithelial and mesenchymal properties and can simultaneously show features of both states (Granit et al., 2013). Overexpression of miR-522, and to a lesser extent miR-517c, but not the other two C19MC miRNAs abundant in TNBCs (Enerly et al., 2014)
Figure 4. Bioinformatics Analysis of miR-522 Target Genes

(A) Promoter regions of miR-522 target genes that were enriched for binding sites of four transcription factors by TRANSFAC analysis.

<table>
<thead>
<tr>
<th>Transcription factor</th>
<th>p-value</th>
<th>Motif</th>
<th>Number of genes</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELK1</td>
<td>1.77x10^{-5}</td>
<td>CCGGAART</td>
<td>113</td>
<td>Regulates cytoskeleton and migration</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Identified here as a target of miR-522</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Targets downregulated upon EMT</td>
</tr>
<tr>
<td>E2F</td>
<td>9.83x10^{-5}</td>
<td>GGCXSG</td>
<td>281</td>
<td>Regulator of G1/S cell cycle progression</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E2F identified here as a target of miR-522</td>
</tr>
<tr>
<td>TEAD2</td>
<td>8.40x10^{-4}</td>
<td>GVGGMGG</td>
<td>184</td>
<td>Hippo signaling transcriptional regulator</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Positive regulator of PAX3</td>
</tr>
<tr>
<td>PAX3</td>
<td>3.76x10^{-1}</td>
<td>CGTCACGSTY</td>
<td>101</td>
<td>Developmental regulator</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Positive regulator of MET</td>
</tr>
</tbody>
</table>

(B) Top enriched functions of directly associated miRNA-522 targets by IPA. Score: -log10(p value). EMT-related functions are indicated in red.

<table>
<thead>
<tr>
<th>Top associated network functions</th>
<th>IPA score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cellular movement, embryonic development, cell morphology</td>
<td>71</td>
</tr>
<tr>
<td>Gene expression, cell cycle, connective tissue development and function</td>
<td>44</td>
</tr>
<tr>
<td>Cell signaling, connective tissue disorders, developmental disorder</td>
<td>35</td>
</tr>
<tr>
<td>Cellular movement, hematological system development and function, immune cell trafficking</td>
<td>32</td>
</tr>
<tr>
<td>Cardiovascular system development and function, embryonic development, organ development</td>
<td>17</td>
</tr>
</tbody>
</table>

(C) Bubble plot of the IPA molecular function p values of directly associated miR-522 targets plotted against interactome genes that were downregulated by miR-522 overexpression (IPA p < 0.0001; FC > 1.5). The size of each bubble is proportional to the number of miR-522 target genes assigned to that molecular function. Bubbles are selectively annotated based on low p values and large size and annotated by function: cell cycle (black), proliferation (blue), cell death (green), cancer (purple), EMT-related (cell morphology, movement, transformation, and differentiation: red), and others (gray).

See also Figure S2.
TGIF1 of these genes was knocked down at least 3-fold, except for miR-522’s ability to promote detachment without anoikis. Each whether knockdown of these genes individually recapitulated BMI1 group genes TPD52 suspension (Figure 6B). Knockdown of four other genes TIMP3, FOXP1, suppressors [Lv et al., 2012]—or which repress TGF- and DNAJC25, cycling membrane protein ERGIC1, polycomb group genes BM1 and RN2, Ccr4-NOT deadenylase subunit CNOT7, and antifolate drug target DHFR. We first asked whether knockdown of these genes individually recapitulated miR-522’s ability to promote detachment without anoikis. Each of these genes was knocked down at least 3-fold, except for TGF1, whose expression was reduced 2-fold (Figure S4). Knockdown of 10 of the 21 genes, including all five adhesion-related genes and half of the EMT-related genes, significantly decreased the number of attached viable cells 3 days later (Figure 6A). However, knockdown of only three of these (ZFYPE21, TMP3, and SPSB1) also increased the number of live cells in suspension (Figure 6B). Knockdown of four other genes (FOX1, DED2, TP52, and BM1) only increased the number of viable nonadherent cells.

Next, we examined by qRT-PCR whether individual knockdown of any of the 21 genes enhanced expression of the mesenchymal transcription factors ZEB2, FOXC2, and SNAI2 upregulated by miR-522 (Figure 6C). Knocking down TMP3, PNF1, FOX1, TGIF2, SPSB1, or BM1 upregulated all three transcription factor mRNAs 5 days later, although less potently than miR-522 overexpression. These data suggest that miR-522 regulates invasivty and mesenchymal gene expression by suppressing the expression of multiple genes identified by Pull-down-seq.

DISCUSSION

Identifying miRNA targets and function is challenging, requiring a lot of trial and error (Thomas et al., 2010). This is especially true of nonconserved miRNAs and miRNAs that recognize their targets by noncanonical binding. Here, we describe a relatively straightforward systems-based strategy for miRNA target gene identification and functional characterization. We used it to identify miR-522 target genes, MREs, and function, directly and readily, with a high degree of specificity. Although identified MREs did not follow canonical rules, most target genes had reduced mRNA and protein levels when miR-522 was overexpressed, and most identified MREs tested were active by luciferase assay (Figure 7A). The method is simple and can be performed with only a million cells. Other advantages include freedom from assumptions about target recognition rules, reduced sequence biases from crosslinking, direct genome-wide identification of regulated mRNAs and MREs, and the ability to identify noncoding transcripts as miRNA targets. Functional miRNAs are associated with all four AGO homologs (Hafner et al., 2010; Su et al., 2009). Our method identifies miRNA targets and MREs, regardless of which AGO protein(s) are involved. Combining biotinylated miRNA pull-down with AGO immunoprecipitation might improve target identification and could be used to investigate whether individual AGOs obey distinct binding rules.

We added TRANSFAC promoter analysis to more commonly used bioinformatics tools (interactome and gene ontology functional analysis) to analyze both pulled-down and downregulated genes to facilitate the tricky task of identifying the biological function of a previously unstudied miRNA. This bioinformatics work flow led to robust predictions of miR-522 function that directed experimental efforts into fruitful investigations and avoided unproductive searching. This streamlined approach should be useful for uncovering the hidden meaning of large genome-wide data sets.

Recent studies have suggested that individual miRNAs differ in how they recognize their targets. miR-522 recognizes its mRNA targets largely by binding to both an imperfect seed match and a 3’ supplementary motif. In fact, globally noncanonical binding may be more prevalent than canonical seed pairing (Loeb et al., 2012; Xia et al., 2012). Seed-only interactions were found in fewer than one in five endogenous miRNA-MRE pairs identified by CLASH (Helwak et al., 2013). The experimentally verified miR-522 MREs were located in the CDS and 3’ and 5’ UTRs of target mRNAs, corroborating previous studies that MREs occur outside the 3’ UTR (Lal et al., 2011; Tay et al., 2008). Although the 3’ UTR was overrepresented and was where most (56%) MREs were located, this is only a modest overrepresentation, because 3’ UTRs make up 43% of RefSeq sequences. The PAR-CLIP data set also reported moderate genome-wide overrepresentation of MREs in the 3’ UTR (Hafner et al., 2010).
The cutoffs chosen to define targets are somewhat arbitrary. Increasing the stringency reduces false positives but also reduces the sensitivity for identifying bona fide targets. For example, changing the fold enrichment cutoff in the pull-down from 2 to 4 significantly improved the cumulative gene expression plot but at the cost of dramatically reducing the number of candidate targets by about a third. However, the most important targets, those most downregulated by miR-522, may be more enriched in the pull-down (Figure 1D). With our current cutoffs, we identified more potential miR-522 targets with IMPACT-seq than with Pulldown-seq. This apparent discrepancy can be explained by the fact that IMPACT-seq looks at a much-smaller sequence window. If a transcript contains MREs for both control miRNA and miR-522 in different locations, Pulldown-seq will not identify this transcript as a target unless it is pulled down with different efficiencies (1,704 transcripts have at least one MRE in both control and miR-522). However, we identified MREs for more than half of the miR-522 target transcripts, an impressive statistic considering that the data sets were generated from distinct experiments using two different sequencing platforms. To improve the overlap of regulated mRNAs and MRE-containing transcripts, the stringency of Pulldown-seq cutoffs could be relaxed and/or the sequencing coverage of IMPACT-seq could be improved. With our current method and cutoffs, we view these two methods of identifying miRNA targets as complementary.

miR-522 caused G1 cell-cycle arrest and loss of adhesion without anoikis and induced mesenchymal genes and properties in a TNBC cell line (Figure 7B). Acquisition of mesenchymal properties is thought to promote cancer progression and metastasis, especially in breast cancer (Yang and Weinberg, 2008). Although miR-522 induced mesenchymal genes and mesenchymal properties, it did not downregulate epithelial gene expression in this cell line. Breast cancer circulating tumor cells (CTCs), which are thought to be the intermediary between primary tumor and secondary metastatic cells, highly express mesenchymal markers even though most primary solid tumors and their metastases are epithelial (Aktas et al., 2009; Kallergi et al., 2011). However, breast cancer CTCs, like the miR-522-overexpressing cells in this study, often coexpress epithelial and mesenchymal genes. In fact, CTCs in mice implanted with metastasizing MDA-MB-468 xenographs express both mesenchymal genes and epithelial E-cadherin (Bonnomet et al., 2012). Rather than undergoing an EMT, it might be more accurate to characterize metastasizing cancer cells as occupying a metastable state between these two fixed lineages. Expression of miR-522 may turn on this metastatic program to increase motility, survival, and numbers of CTCs. Supporting our hypothesis that miR-522 instigates metastasis, miR-522 was more highly expressed in CTCs from patients with TNBC tumors that are prone to metastasize than in better prognosis ER+ tumors (Sieuwerts et al., 2011). There is currently no targeted therapy for TNBC, the breast cancer with the worst prognosis. In the future, miR-522 antagonists could be evaluated for treating TNBC patients, especially for tumors with amplified C19MC.

Several miRNAs, including miR-200, miR-205, and miR-34, induce epithelial transition (Lamouille et al., 2013). Some of these induce this dramatic change in cellular state by suppressing key transcriptional activators of mesenchymal genes, such as Snai1 and Snai2, or transcriptional repressors of epithelial genes, such as Zeb1/Zeb2. More recently, miRNAs, such as miR-9, miR-22, and miR-661, have been identified that accomplish the reverse—reduce expression of epithelial genes and induce mesenchymal programs (Ma et al., 2010; Song et al., 2013; Vetter et al., 2010). miR-22 targets the TET family of methylcytosine dioxygenases that may globally suppress epithelial gene expression indirectly by inhibiting demethylation of miR-200 family promoters, thereby suppressing Zeb expression. Some of the mesenchymal-promoting miRNAs suppress the expression of one or two epithelial genes, such as genes encoding E-cadherin (miR-9) or nectin-1 (miR-661), which help maintain tight junctions. However, modulation of a single target rarely explains most of a miRNA’s effect. Many miRNAs target multiple transcripts in the same networks or pathways to regulate biological processes. An apt analogy might be transcription factors, which regulate many genes; one regulated gene rarely reproduces the effect of the transcription factor. Thus, miRNAs regulate the epithelial-mesenchymal transitions most likely by targeting other genes in the key pathways that contribute to this transition, as we found for miR-522. In fact, a miR-200 pull-down (R. Perdígão-Henriques, F. Petrocca, G. Altschuler, M.P. Thomas, M.T.N. Le, S.M.T., W.H., and J.L., unpublished data) identified miR-200 target genes that interact with Zeb transcription factors to

Figure 5. Overexpression of miR-522 Induces Hallmarks of Epithelial-Mesenchymal Transition

(A) qRT-PCR analysis of miRNA levels in MDA-MB-468 cells showed a higher level of miR-522 in nonadherent cells. Mean ± SD of three independent experiments (*p < 0.01 compared to control).

(B) Trypan blue exclusion cell counts of adherent and nonadherent cells 3 days after transfection with miR-522 or control miRNA showed a higher number of miR-522-transfected live cells in suspension. Mean ± SD of three independent experiments (*p < 0.05 compared to control).

(C) Nonadherent cells transfected with control or miR-522 as in (B) were retransfected with anti-miR-522, control antisense miRNA, or miR-522. Only antagonizing miR-522 led to adherent colonies (representative photos).

(D and E) qRT-PCR analysis of EMT genes, normalized to GAPDH, in adherent and nonadherent cells 3 (D) and 5 (E) days after transfection with miR-522 or control miRNA. Note the difference in the scales. Mean ± SD of three independent experiments (*p < 0.05; **p < 0.01 compared to control).

(F–H) Transwell invasion assay demonstrates that transient miR-522 overexpression increases invasion. (F) Representative photos of crystal violet-stained invading cells. (G) Quantification by calcein AM staining. (H) Quantification of invaded live cells in suspension in the bottom chamber by CellTiter-Glo. Mean ± SD of three independent experiments (*p < 0.01 compared to control).

(i) Transwell invasion assay as in (F) but with MDA-MB-468 cells stably infected with pBABE-Hygro or pBABE-Hygro-miR-522.

(J and K) Cell-cycle analysis of MDA-MB-468 cells transfected with miR-522 or control miRNA, showing an increase in cells in the G1 phase. (J) shows a representative profile. (K) shows mean ± SD of three independent experiments (*p < 0.05 compared to control). PI, propidium iodide.

See also Figure S3.
Figure 6. Knockdown of Individual Target Genes Partly Recapitulates Effects of mir-522 Overexpression

(A and B) CellTiter-Glo was used to quantify live adherent (A) and nonadherent cells (B) 3 days after small interfering RNA (siRNA) or miRNA transfection, normalized to nontargeting siRNA control (siNT). Bars in purple indicate genes whose knockdown both decreased live adherent cells and increased live nonadherent cells.

(C) qRT-PCR analysis of relative EMT gene mRNA in adherent cells 5 days after siRNA or miRNA transfection. Relative mRNA level normalized to GAPDH was compared to cells transfected with siNT. Knockdown of genes in the red boxes significantly enhanced the expression of all three mesenchymal transcription factors. Mean ± SD of three independent experiments (*p < 0.05; compared to control).

See also Figure S4.
The advantage of Pulldown-seq is its ability to identify without bias bona fide miRNA targets. Our results suggest that miR-522 targets a large number of transcripts, of which probably only a subset are biologically important in a given cell at a particular time. It will be worthwhile to define at a systems level the genes regulated by the EMT- and MET-regulating miRNAs. The method described here should be useful for accomplishing this goal.

**EXPERIMENTAL PROCEDURES**

**Cell Culture and Transfection**

Breast cancer cell lines were obtained from American Type Culture Collection and cultured as recommended. Unless otherwise stated, DharmaFECT 1 reagent (Dharmacon) was used for transfection. Control miRNA (cel-miR-67) or miR-522 mimic (50 nM, Thermo Scientific) was used to transfect 100,000 cells/well in suspension in 12-well dishes as per manufacturer’s protocol, and medium was changed after 24 hr.

**Microarray**

Total RNA, extracted using Trizol (Invitrogen) 48 hr after transfection from biological triplicate wells, was purified with the RNeasy kit (QiAGEN). cDNA and cRNA were generated and hybridized to HT-12 v4 beadchips (Illumina) according to the manufacturer’s protocols. Cubic spline-normalized (without background normalization) data were analyzed by the NIA Array Analysis tool (http://igsun.gcc.nia.nih.gov/ANOVA) to identify transcripts downregulated after miR-522 overexpression.

**Pulldown-Seq**

MDA-MB-468 pull-down experiments with control miRNA (cel-miR-67) and miR-522 mimic (Thermo Scientific) were conducted as described previously but modified to include molecular crowders in the lysis buffer (Lai et al., 2011; Tan et al., 2014). Briefly, cell pellets were collected 24 hr posttransfection, washed twice with cold PBS, and incubated with lysis buffer (20 mM Tris [pH 7], 25mM EDTA [pH 8], 100 mM KCl, 5 mM MgCl₂ [all Ambion], 2.5 mg/ml Ficoll PM400, 7.5 mg/ml Ficoll PM70 [GE Healthcare], 0.25 mg/ml dextran sulfate 670k, 0.3% NP-40 [Fluka], 50 U each of RNase OUT and SUPERase In [Invitrogen], and complete protease inhibitor cocktail [Roche Applied Science]) on ice for 20 min. The cytoplasmic lysate, isolated by centrifugation at 5,000 g for 5 min, was added to 1 mg/ml yeast tRNA and 1 mg/ml BSA (Ambion)-blocked streptavidin (SA)-coated magnetic beads (Invitrogen) and rotated for 4 hr at 4°C. The beads were washed five times with 1 ml lysis buffer and bead-bound RNA extracted using Trizol LS (Invitrogen). Ribosomal RNA was depleted using the Ribo-Zero rRNA removal kit (Epipenter). Libraries were generated using the NEBNext Multiplex Small RNA Library Prep Set with modified adaptors and primers compatible for Ion Torrent Sequencing Platform (New England Biolabs) and sequenced on the Ion Torrent platform using the 314 Chip (Invitrogen), according to the manufacturer’s protocols.

**IMPACT-Seq**

The IMPACT-seq protocol used the Pulldown-seq protocol to isolate miRNA-bound RNAs with SA beads as above. The cytoplasmic lysate-SA bead mixture was incubated with rotation overnight at 4°C before adding RNase T1 (25 U/µl) for 10 min on-bead digestion at 37°C. RNA was then extracted from washed beads using Trizol LS as above. The RNA was treated with T4 Polynucleotide Kinase (NEB) to obtain 5’-phosphate ends for subsequent ligations and then passed through NuCAway columns (Ambion) to remove RNAs <20 nt in length. Libraries were generated using the NEBNext Small RNA Sample Prep Set for Illumina chemistry (NEB). cDNA fragments with an insert size of 20–60 nt were gel extracted and sequenced on the HiSeq platform (Illumina) according to the manufacturer’s protocol.

**RNA-Seq Quality Control and Alignment**

Pralignment and postalignment libraries were screened for quality, specificity, and contaminating sequences using FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), RSeQC (Wang et al., 2012), and RNA-Seq QC (DeLuca et al., 2012). Prior to alignment, low-quality bases, homopolymer sequences, sequences matching the first 13 bp, and the reverse complement of the adaptor sequences for Ion Torrent or Illumina were trimmed using cutadapt version 1.2.1 (Martin, 2011). Trimmed reads smaller than 30 nt for Pulldown-seq or 20 nt for IMPACT-seq were discarded. Pulldown-seq used reads that uniquely mapped to the GRCh37 assembly of the human genome augmented with the Ensembl 68 genome annotation using Novoalign (http://www.novocraft.com) with the following parameters: -H -k -n 250 -F STDFQ -r all 10 -e 10 -g 15 -x 4. IMPACT-seq used reads that uniquely mapped (with greater than or equal to two mismatches) to the GRCh37 assembly of the human genome augmented with the Ensembl 72 genome annotation (Flicek et al., 2013), using Tophat version 2.0.8b (Trapnell et al., 2009). Quality control, trimming, and alignment were performed using the bipy (https://github.com/roary/bipy) and bcBio-nextgen (https://github.com/chapmanb/bcBio-nextgen) automated sequencing analysis pipelines.

**Pulldown-Seq Target Identification**

Postalignment gene counts were generated using htseq-count 0.5.4p3 with the counts aggregated by gene_id of the Ensembl 68 annotation. Differentially expressed transcripts were called with DESeq version 1.9.2 using the default parameters: -p 0.05 -q 0.01 -O FDR. Peaks were called on both control miRNA and miR-522 IMPACT-seq samples using CLIPper (https://github.com/YeoLab/clipper/) with the following options: -a 0.05 -b 0.01 -c 0.01 -d 0.05 -e 0.05 -f 0.05

**Figure 7. Model of miR-522 Functions**

(A) Summary of experimental results to validate the specificity of Pulldown-seq and IMPACT-seq for identifying miR-522 target mRNAs and MREs. (B) Schematic showing important miR-522 targets and their biological functions.
mean p < 0.05 was considered significant. *p < 0.05; **p < 0.01; ***p < 0.001. The distribution plots were analyzed using the Kolmogorov-Smirnov test (K-S test). In vitro data were analyzed using unpaired Student's t test; cumulative distribution plots were analyzed using the Kolmogorov-Smirnov test (K-S test), p < 0.05 was considered significant. *p < 0.05; **p < 0.01; ***p < 0.001. The mean ± SD of three or more independent experiments is reported.

**Statistical Analysis**

In vitro data were analyzed using unpaired Student’s t-test; cumulative distribution plots were analyzed using the Kolmogorov-Smirnov test (K-S test), p < 0.05 was considered significant. *p < 0.05; **p < 0.01; ***p < 0.001. The mean ± SD of three or more independent experiments is reported.

**ACCESSION NUMBERS**

Pull-down, IMPACT RNA-seq, and overexpression microarray datasets are available in the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress) under accession numbers E-MTAB-2112, E-MTAB-2119, and E-MTAB-2110, respectively.

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes Supplemental Experimental Procedures, four figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.07.023.

**AUTHOR CONTRIBUTIONS**

S.M.T. and J.L. designed the experiments, analyzed the data, and wrote the manuscript. S.M.T. performed the experiments. J.J. and L.M. obtained the RNA-seq data. R.K., O.H., and W.H. did the bioinformatics analysis.

**ACKNOWLEDGMENTS**

We thank J.L. laboratory members for critical discussions. S.M.T. is supported by the Department of Defense Breast Cancer Research Program. Microarray experiments were performed by the Molecular Genetics Core Facility at Boston Children’s Hospital supported by NIH-P50-NS40828 and NIH-P30-HD18656. The bioinformatics analyses, supported by the HSCI Center of Stem Cell Bioinformatics, were run on the Odyssey cluster supported by the Harvard University FAS Research Computing Group. J.J. and L.M. are employees of New England Biolabs, a company that sells deep sequencing kits for RNA and DNA research.

Received: November 21, 2013
Revised: May 7, 2014
Accepted: July 16, 2014
Published: August 14, 2014

**REFERENCES**


