Epigenetics 6:3, 268-272; March 2011; © 2011 Landes Bioscience

The 3 prime paradigm of the miR-200 family and other microRNAs

Graham J. Brock,[1,*] Sterghios Moschos,[2] Simon D. Spivack[3] and Gregory J. Hurteau[4]

[1]Pfizer; Division of Molecular Medicine; Groton; CT USA; [2]Pfizer; Epigenetics Unit; Sandwich, UK; [3]Albert Einstein College of Medicine; Bronx; [4]Ordway Research Institute; Albany, NY USA

The number of predicted human microRNAs in Sanger miRBase currently stands at over 1,000, with each of these in turn predicted to target numerous mRNAs. However, those microRNAs for which mRNA targets have been evaluated, verified and reported in the literature are still in the minority and the bulk of microRNA/mRNA interactions are yet to be confirmed. Confirmation of microRNA interaction with predicted mRNA targets represents a considerable undertaking, made more complex by potential synergistic effects of multiple microRNAs and the three possible outcomes (translational repression, degradation or a mixture of both). In addition, contrasting results obtained when either stably expressing or transiently transfecting members of the miR-200 family illustrate limitations in the verification methods currently in use. In this article we suggest that instead of allowing computational predictions to drive investigation, it would be desirable, when possible, to systematically evaluate microRNA targets using inducible, stable, ectopic expression. The advantage of stable lines ectopically expressing microRNA(s) is that they allow an analysis of changes to both the proteome and the transcriptome. This would allow verification of targets, improve the design of prediction algorithms and greatly increase our understanding of the outcome of microRNA/mRNA interaction.

Key words: microRNA, target prediction, 3'UTR, mir-200 family, miR-200c

Abbreviations: 3'UTR, 3' prime untranslated region; MRE, microRNA recognition element; HITS-CLIP, high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation

Submitted: 11/23/10

Accepted: 12/22/10

DOI: 10.4161/epi.6.3.14635

*Correspondence to: Graham J. Brock and Gregory J. Hurteau;
Email: graham.brock@pfizer.com and GHurteau@ordwayresearch.org

Introduction

The most studied region of interaction between microRNAs and their mRNA target(s) is the 3' untranslated region (3'UTR). The process is thought to involve the 7–8 bp at the 5' end of the microRNA and its complementary sequence in the mRNA, called a "seed-match" or microRNA recognition element (MRE).[1,2] Due to the short recognition sequence and resultant frequency in the transcriptome, the identification of seed matches rapidly became a bioinformatics-led undertaking. Numerous predictive programs have been generated, several freely available (recently reviewed in ref. 3), and credit is due for the endeavors of those who developed them. While these programs have been invaluable for the guidance of microRNA research, they nevertheless have limitations. For example, as a consequence of limited flexibility in the choice of mRNA sequence used and subtle differences in the algorithms, there is rarely complete concordance between prediction programs. Examples of public domain gene entries with multiple putative 3'UTRs or different entries across databases abound; yet, limited efforts have been documented in rationalizing and appropriately contextualizing the choice of entry used in bioinformatic prediction.

In addition to the perceived shortcomings of these informatics programs, the methods used for experimental verification may also be flawed or inadequate. Although the widespread use of luciferase assays has led to the reporting of numerous targets, studies often select a single mRNA from the list of multiple predicted targets. Often, this highly subjective selection process is justified on theoretical target relevance to the biological context of interest rather than experimental, evidence-based selection of putative targets. Fragments or the whole 3'UTR of the target of choice

are then coupled to the reporter and co-transfected with the microRNA. The ability of such an approach to reproduce the conditions for effective interaction between approximately one third (7–8 bp seed match) of a microRNA in a protein complex with the 3'UTR of an mRNA is questionable, as this might require folding or secondary structure. Indeed, documented evidence of the impact of secondary structure on MRE accessibility is reported with respect to *miR-223* interactions with a number of its putative target genes.[4] If such secondary structure can be maintained when either a fragment or the entire 3'UTR is fused to luciferase is questionable, yet multiple papers using large amounts of transfected microRNA have reported such interactions as conclusive evidence of miRNA/mRNA targeting.

Another consideration is the potential importance of seed match location, both relative to other seed matches and in the context of the 3'UTR. This has been described in the context of both micro and short interfering RNA (siRNA) studies. For example, the hepatitis C + RNA genome contains two juxtaposed *miR-122* MRE's in the 5' end of the internal ribosomal entry site (IRES), whose occupancy is mutually exclusive due to the <10 nt distance between the two seed matches.[5] Evidence now suggests that the structure adopted by the IRES is entrenched with *miR-122* interaction and is crucial in translational activity of the virus.[6,7] In many cases the miRNA/mRNA interactions have been verified using alternative approaches, but doubts still remain as to the ability of these methods to reliably reproduce the interactions being studied. For example, there is often a substantial difference between synthetic miRNA analogue intracellular concentrations following transfection and endogenous, phenotypically relevant levels. Crucially, these approaches may have a propensity to describe miRNA/mRNA interactions that might not be manifest under physiologically relevant conditions.[8]

Finally, in addition to potential misinterpretation due to failure to reproduce biologically relevant levels of microRNA, the issues of secondary structure and MRE placement,[9] these approaches may not address the potentially confounding issues of other mRNA targets. In most cases they list several hundred predicted mRNAs and would be expected to dilute the effect of the transiently transfected microRNA. Indeed, natural and synthetic microRNA sponges have been described and already proposed as potential therapeutics.[10] We observed the global impact of a single differentially expressed mRNA targeted by an RNAi mechanism when exploring the utility of a NFkappaB-driven, luciferase reporter cell system as a platform for RNAi experiments. Under the well-described A549 lung epithelial IL-1beta-induced IL-8 release cell culture model, absence or presence of the reporter transgene had a profound, dose- and siRNA sequence-dependent impact on the inflammatory response profile.[8] Thus, a commercially available siRNA specific for luciferase (Dharmacon sequence 2; IC_{50} < 0.5 nM) resulted in IL-8 release inhibition when used in the parent cell line in the absence of the reporter gene, suggesting that the off-target activity observed was specific to the sponge effect of the luciferase mRNA.[8]

The commonly used verification methods of microRNA/mRNA interaction also fail to address the issue of synergistic action of other microRNAs predicted to target the mRNA being investigated. This is commonly referred to as the rheostat hypothesis, where a given phenotypic impact might result from multiple microRNA or mRNA changes which whilst individually apparently negligible, collectively serve to modulate a specific system/pathway. If we are to believe the prediction programs, then there is a complex regulatory network with multiple microRNAs regulating numerous mRNAs through translational repression or mRNA transcript degradation. However, as numerous examples of single microRNAs strongly regulating a single mRNA have been reported, perhaps the problem lies with current prediction programs being unable to determine the orchestrated actions of all co-expressed microRNAs, and investigators pursuing single molecules or linear pathways of interest as opposed to "omic," system-wide investigations. Although not ideal, the case of *miR-200c* provides additional evidence of current limitations and further evidence that an alternative approach is indeed required.

The Case of miR-200c

The paradigm of prediction versus experimental results is illustrated by discordant data obtained in studies of *miR-200c* and other members of the miR-200 family. Initially, through the use of a mixture of available bioinformatic data in the form of target predictions and relative quantitative measurement in the form of qRT-PCR, several mRNA targets of *miR-200c* were reported to be subject to degradation.[11] Subsequent laboratory experiments using stable ectopic expression of *miR-200c* in a breast cancer cell line indicated that the endogenous mRNA of *ZEB1* (or *TCF8*, as it is also known) was indeed degraded.[12,13] When the 3'UTR of *TCF8*/*ZEB1* was examined this revealed that it contained five "seed matches," or miRNA target sites, for *miR-200c.* Notably, and this has now been reported by others, most prediction programs use a shorter transcript variant of *ZEB1* which contains only three sites. As a consequence, *ZEB1* is not listed as one of the primary targets of *miR-200c.* However, the related transcription factor, *ZEB2,* which also has five predicted *miR-200c* seed matches is listed as a primary target of *miR-200c* (as most programs used the longer 3'UTR for *ZEB2*). Our studies focused on *miR-200c* and *ZEB1* so we did not investigate any other potentially underestimated mRNA targets attributable to the use of a shorter transcription variant. Nor did we investigate if such underestimates are widespread among microRNA predictions.

Following our initial publication, other members of the miR-200 family (as defined by seed match similarity) were then reported to also interact with *ZEB1* and *ZEB2*.[14,15] These papers and other reports that claimed additional mRNA targets of *miR-200c* all relied upon transient transfection of a double stranded nc-RNA. The transfected material is reportedly processed by endogenous components of the microRNA biogenesis machinery and in the case of *miR-200c,* a phenotype similar to that observed with the stable expression was reported

(albeit in 48 h as opposed to three weeks). The main discordance arose when the other members of the family (*miR-200a*, *miR-141* and *miR-200b*), which have identical or almost identical seed regions, also generated a similar phenotypic appearance upon transient transfection. When these family members were stably expressed using the same techniques used to generate stable expression of *miR-200c*, the resultant mature microRNAs produced no change in phenotype.[12] Although other possible explanations for the discordant results could be proposed, perhaps the most obvious is that transient transfection is akin to an RNAi experiment. Thus, in addition to being unable to reflect the endogenous mechanism, it is potentially capable of an off target effect, which may explain the phenotypic change seen with transient transfections of *miR-141* and *miR-200a*.

While it is usually deemed obvious when the point is raised that stable expression is more robust and biologically relevant than transient, the latter continue to be a mainstay of microRNA research, often without verification or, crucially, reviewer question. Another major issue is noted in the computationally predicted number of mRNA targets for *miR-200c* which exceeds 600. Although a small number of these mRNA targets have been reported again, the experimental verification relied on transient transfection of *miR-200c* without follow-up experiments involving a stable system.[16] Regardless, this still leaves over 590 "targets" and, as is the case with many microRNA studies (including our own), there is often a failure to even acknowledge the other predictions, focusing instead on a single mRNA/microRNA target pairing and its verification. The other notable omission in many of these studies is illustrated by the predicted sites for microRNAs other than *miR-200c* in the 3'UTR of the ZEB transcription factors. Depending on the algorithm used, there could be as many as fifty other microRNAs that target *ZEB1*. However, our data indicates that *miR-200c* can act alone and targets only the ZEB family of transcription factors: so, how do we reconcile the potential discrepancy between computational prediction and experimental results? While we cannot state it definitively, as an analysis of the proteomic and transcriptomic data has not been undertaken, based on our experimental results we could propose the following:

Firstly, *miR-200c* has a specific mode of action and targets and degrades only the ZEB family of transcription factors. The mechanism for this specificity lies in the multiple target sites for *miR-200c* present in the 3'UTRs of both transcription factors (*miR-200c* is not predicted to have multiple seed matches in any of its other mRNA targets listed in the prediction programs). However, assuming that this model is correct and the only two targets for *miR-200c* are *ZEB1* and *ZEB2*, there is a second point. Analysis using the miRanda prediction program suggests that ~50 other microRNAs are involved in the regulation of *ZEB1*. For clarity, **Figure 1A** shows only those microRNAs predicted to have two or more sites in the 3'UTR of *ZEB1*. This appears to indicate a complex regulatory system involving multiple microRNAs. However, if we examine the expression patterns of these other microRNA(s) in cell lines in which endogenous levels of *miR-200c* are either high (MCF7) or below detection (MDA-MB-231 and PC3) we see that only *miR-200c* and *miR-141* have an inverse expression pattern to that of *ZEB1* (**Fig. 1B**). In the cell line that ectopically expresses *miR-200c* and was previously shown to degrade *ZEB1* and produce the phenotypic appearance associated with expression of E-cadherin, the levels of these other microRNAs are not significantly altered compared to the MDA-MB-231 parent cell type. Again, the proposed prerequisite for a mechanism requiring multiple sites could explain why the predictions of regulation by other microRNAs may be inaccurate. The remaining targets could be transcriptionally repressed or expression could be tissue or development-stage specific. To effectively demonstrate the former, a thorough examination of the proteome of both MCF7 (endogenous *miR-200c*) and cells ectopically expressing *miR-200c* is required. In the later case, to demonstrate tissue or development specific expression of microRNAs, which individually or in groups are capable of regulating the ZEB family, will require additional experimentation.

What Should We Do?

Our proposed model for *miR-200c* has some limited in vitro experimental evidence; however, the reality is that none of the methods used to date to examine the function of this family of microRNA are adequate. For example, while *miR-200c* alone may be able to degrade ZEB1, there is also a strong argument for the synergistic involvement of *miR-141*. What is required is an alternative to current methods, one involving inducible stable expression of individual or, if possible, clusters of microRNAs. Expression should first be verified using qRT-PCR to ensure that the levels of expression are approximate to the endogenous levels in a reference cell line. The inducible expression system would allow sufficient quantities of cells to be generated before induction to enable both detailed proteomic and transcriptomic analysis and enable an examination of the concordance between prediction and experimental evidence. Although such an approach would be painstaking it does not need to be undertaken en masse. The best candidates are already known and reported in the literature and those microRNAs which have disease associated expression patterns should be examined first. In addition to an examination of the effects of microRNA at the protein level, techniques are now available that may enable identification of microRNA/mRNA interactions. One method utilizes the protein complex which facilitates microRNA/mRNA action. The technique, termed HITS-CLIP, simultaneously identifies Ago-bound microRNAs and the nearby mRNA sites.[17] Alternatively, methods for affinity purification of the 3'UTR have also been used to capture active microRNAs (personal communication with S. Spivack and ref. 18). Through the use of these or other robust methods, gradually putative rules will begin to reveal themselves and further experimentation using mutational analysis may then confirm or validate issues such as spatial arrangement or number of sites. Perhaps we will even find that a "well placed" single target site can lead to mRNA instability. The availability of vectors, many inducible, with which cell lines can be altered to ectopically express a microRNA of interest

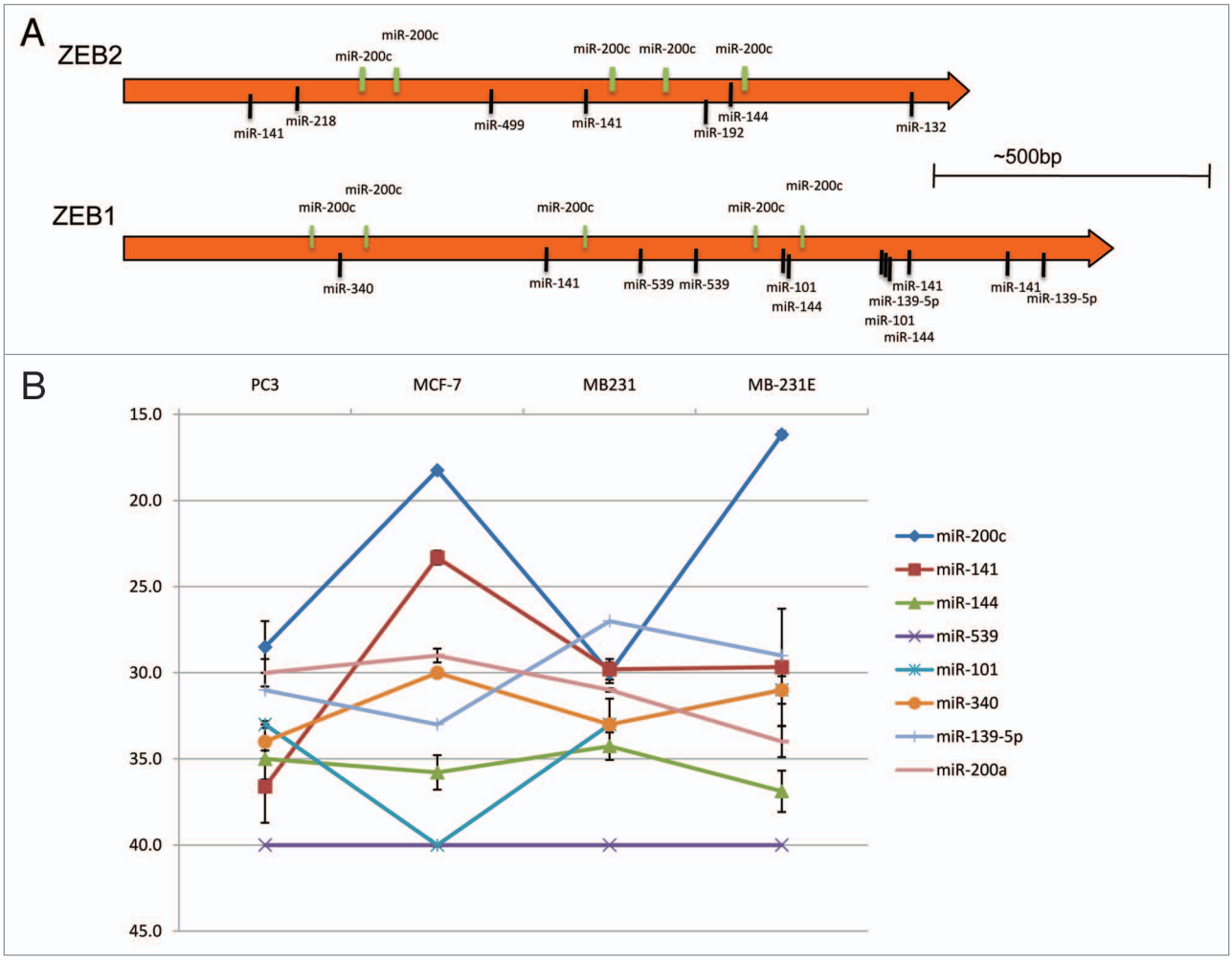

Figure 1. (A) The 3'UTRs of *ZEB1* and *ZEB2* showing some of the predicted sites for microRNAs. For clarity, only those microRNAs with two or more seeds matches are shown. Prediction maps were generated using miRanda at microRNA.org, as TargetScan and Pictar both use a shorter version of the *ZEB1* transcript. (B) Expression profiles of microRNAs predicted to target ZEB1 in cell lines. MicroRNA expression was measured using ABI-TLDA. Average Ct of three replicates with SD is shown on the y axis. All cell lines from ATCC: PC3, prostate; MCF-7, breast; MDA-MB-231, breast; MDA-MB-231E ectopically expressing *miR-200c*, breast.

should be considered as a potentially more valid approach than transient transfection. Ultimately, the pressing need is for the community to come to an agreement regarding cohorts of experiments that together confirm the microRNA/mRNA networks through rigorous pharmacologic and biologic assessments with for example animal models.[19] We should perhaps adopt a MIAME-styled[20] set of requirements in confirming and reporting microRNA target interactions. These should include experiments that address known and hypothesized off-target effects, as well as thorough investigation of biological interactions and consequences at an "omic" scale.

Concluding Remarks

Huge strides have been made and viable interactions identified between many microRNA and their mRNA targets using some current methods guided by prediction programs. This point of view has tried to both present alternative explanations rather than cite specific examples as good or bad experiments. Rather, we accept that all current methods are inadequate and suggest that the field now needs to raise the bar and accept that we need more robust techniques closer to the biological reality, with less reliance on computational prediction until we have verified that all seed matches are indeed created equally. We are entering the era of experimentation where hypotheses and endpoints should be performed at the network level as opposed to a single-point or across a linear pathway. Next generation sequencing techniques are rapidly identifying more microRNAs and the methods used to identify their modes of interaction should also be subject to more rigor. In this way we may identify the exact functional mechanism of action of a truly extraordinary group of non-coding RNAs. However, even if such experimental findings do indicate that each microRNA only targets a handful of mRNA targets, they are still a remarkable molecular species "punching well above their weight."

Acknowledgements

We thank Katrina Loomis and Tim Schaiff for helpful comments and critical reading of the manuscript.

References

1. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. Cell 2003; 115:787-98.
2. Lai EC. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. Nat Genet 2002; 30:363-4.
3. Alexiou P, Maragkakis M, Papadopoulos GL, Reczko M, Hatzigeorgiou AG. Lost in translation: an assessment and perspective for computational microRNA target identification. Bioinformatics 2009; 25:3049-55.
4. Sun G, Rossi JJ. Problems associated with reporter assays in RNAi studies. RNA Biol 2009; 6:406-11.
5. Jopling CL, Norman KL, Sarnow P. Positive and negative modulation of viral and cellular mRNAs by liver-specific microRNA miR-122. Cold Spring Harb Symp Quant Biol 2006; 71:369-76.
6. Diaz-Toledano R, Ariza-Mateos A, Birk A, Martinez-Garcia B, Gomez J. In vitro characterization of a miR-122-sensitive double-helical switch element in the 5' region of hepatitis C virus RNA. Nucleic Acids Res 2009; 37:5498-510.
7. Jangra RK, Yi M, Lemon SM. Regulation of hepatitis C virus translation and infectious virus production by the microRNA miR-122. J Virol 2010; 84:6615-25.
8. Perry MM, Moschos SA, Williams AE, Shepherd NJ, Larner-Svensson HM, Lindsay MA. Rapid changes in microRNA-146a expression negatively regulate the IL-1beta-induced inflammatory response in human lung alveolar epithelial cells. J Immunol 2008; 180:5689-98.
9. Saetrom P, Heale BS, Snove O Jr, Aagaard L, Alluin J, Rossi JJ. Distance constraints between microRNA target sites dictate efficacy and cooperativity. Nucleic Acids Res 2007; 35:2333-42.
10. Ebert MS, Neilson JR, Sharp PA. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. Nat Methods 2007; 4:721-6.
11. Hurteau GJ, Spivack SD, Brock GJ. Potential mRNA degradation targets of hsa-*miR-200c*, identified using informatics and qRT-PCR. Cell Cycle 2006; 5:1951-6.
12. Hurteau GJ, Carlson JA, Roos E, Brock GJ. Stable expression of miR-200c alone is sufficient to regulate TCF8 (ZEB1) and restore E-cadherin expression. Cell Cycle 2009; 8:2064-9.
13. Hurteau GJ, Carlson JA, Spivack SD, Brock GJ. Overexpression of the microRNA hsa-miR-200c leads to reduced expression of transcription factor 8 and increased expression of E-cadherin. Cancer Res 2007; 67:7972-6.
14. Burk U, Schubert J, Wellner U, Schmalhofer O, Vincan E, Spaderna S, et al. A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. EMBO Rep 2008; 9:582-9.
15. Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, et al. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat Cell Biol 2008; 10:593-601.
16. Schickel R, Park SM, Murmann AE, Peter ME. miR-200c regulates induction of apoptosis through CD95 by targeting FAP-1. Mol Cell 2010; 38:908-15.
17. Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature 2009; 460:479-86.
18. Vo NK, Dalton RP, Liu N, Olson EN, Goodman RH. Affinity purification of microRNA-133a with the cardiac transcription factor, Hand2. Proc Natl Acad Sci USA 2010; 107:19231-6.
19. van Rooij E, Sutherland LB, Liu N, Williams AH, McAnally J, Gerard RD, et al. A signature pattern of stress-responsive microRNAs that can evoke cardiac hypertrophy and heart failure. Proc Natl Acad Sci USA 2006; 103:18255-60.
20. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 2001; 29:365-71.