RNA BIOLOGY
2017, VOL. 14, NO. 7, 831–834
https://doi.org/10.1080/15476286.2017.1320469

Taylor & Francis
Taylor & Francis Group

POINT OF VIEW

Issues in current microRNA target identification methods

Hervé Seitz

Institut de Génétique Humaine UMR 9002 CNRS-Université de Montpellier, 141, rue de la Cardonille, 34396 Montpellier CEDEX 5, France

ABSTRACT

microRNAs are currently believed to control a large diversity of physiologic processes, through the collective repression of thousands of target genes. Both experimental and computational analyses indeed suggest that each microRNA regulates tens or hundreds of genes. But some observations suggest that the phenotypic consequences of many published miRNA/mRNA interactions are dubious. For example, the reported amplitude of miRNA-guided repression is very small, while biologic processes tend to be robust to small changes in gene expression. We recently showed, on one particular miRNA, that for most predicted targets, miRNA-guided repression is even smaller than inter-individual variability among wild-type specimens. We also put forward several sources of computational false positives. These issues are generally neglected by the scientific community, probably resulting in the frequent publication of irreproducible or misinterpreted results regarding microRNA function. We propose novel types of analyses, easily accessible to the community, that could help improve microRNA target identification.

ARTICLE HISTORY
Received 27 February 2017
Revised 11 April 2017
Accepted 13 April 2017

KEYWORDS
False positives; inter-individual variability; microRNA; over-conservation; robustness; target identification

Identification of microRNA targets

microRNAs (miRNAs) have attracted a lot of attention from the community over the last 16 years, essentially because of their great versatility: according to both high-throughput experiments and computational predictions, each miRNA is believed to regulate tens or hundreds of genes. They have therefore been proposed to control an extraordinarily large array of physiologic processes, both in healthy conditions and in disease.

Current high-throughput experimental methods for miRNA target identification are based on an immuno-precipitation of the miRNA effector complex, followed by deep-sequencing-based identification of its interacting RNAs.[1] Computational prediction programs are diverse, but most of them rely on the identification of phylogenetically conserved matches to the miRNA "seed" (nucleotides 2–7 of the miRNA): a perfect seed match appears to be the best available predictor of miRNA-guided repression and phylogenetic conservation (above conservation levels that would be expected by chance) is seen as a proof of the biologic importance of the regulation.[2]

Known inconsistencies in the genome-wide targeting theory — and how they are addressed

At least 2 main issues suggest that the genome-wide role of miRNAs may not be so well established:

1. Results of high-throughput experiment-based identification of miRNA targets usually overlap very partially with computationally-predicted targets,[3-6] casting doubt on the validity of the experimental technique, of computational predictions, or both.
2. When quantified, the effect of miRNAs on their targets is very small (less than 2-fold in general), while macroscopic biological phenotypes tend to be robust to such small changes in gene expression. When miRNA mutants control a macroscopic phenotype and when the mechanistic causes of the phenotype can be found, they tend to be due to the misregulation of a few dose-sensitive genes, rather than to a global misregulation of every target.[7-10]

The first issue can be resolved quite simply, at least in theory: it is possible that many experimentally-detected interactions are phenotypically inconsequential (for the reason given in issue #2), hence they would not be conserved in evolution, thus not predicted computationally. Reciprocally, some biologically important miRNA/mRNA interactions may be predicted yet unobserved experimentally, just because the experiment was performed in a tissue or a cell line where that interaction does not occur. Additional sophistication can be considered when reconciling experimental and computational analyses; for example, interactions that do not involve a perfect seed match (the so-called "non-canonical" interactions) may be stable and detected, yet unable to trigger target mRNA repression.[11]

The second issue is classically addressed using phylogenetic conservation: indeed, target repression is surprisingly small, but if an interaction has been conserved, then it must be important at some point. While we cannot predict the phenotypic consequences of a moderate upregulation for most genes, the mere fact that they seem to be under a selective pressure to keep their miRNA binding sites is perceived as a strong hint that organismal fitness is exquisitely sensitive to such small changes in gene expression.[2]

CONTACT Hervé Seitz herve.seitz@igh.cnrs.fr Institut de Génétique Humaine UMR 9002 (CNRS-Université de Montpellier), 141, rue de la Cardonille, 34396 Montpellier, France.

Our recent work sheds a new light on these questions, suggesting that miRNAs likely have just a few biologically important targets, and most computationally or experimentally identified targets would not be functionally repressed.[12]

Phenotypic robustness vs. miRNA-guided repression

By comparing the effect of miRNA-guided repression to inter-individual variability in gene expression, we observed that most predicted targets for miR-223 in mouse neutrophils appear to be functionally insensitive to the miRNA: the difference in target expression among wild-type mice is frequently larger than miR-223-guided repression. That kind of experiment is quite simple; anyone aiming at measuring the effect of a miRNA (*e.g.*, by a transcriptomics experiment in wild-type vs. miRNAΔ mutants) should be able to also perform a comparison among wild-type individuals. If our results are generalizable, then these experiments would also show that miRNA-guided repression does not exceed inter-individual variability in gene expression for most predicted targets, even though they bear phylogenetically conserved miRNA binding sites.

Because we observed large inter-individual variability in target gene expression among phenotypically wild-type mice, we proposed that these genes are not functionally sensitive to miR-223. But many mRNAs bear seed matches for multiple miRNAs (and sometimes, multiple seed matches for the same miRNA). It could be argued that the collective effect of multiple sites on a given target mRNA could result in a large, biologically relevant repression, even though individual sites cannot achieve functional repression.

Such a scenario would not explain why each individual site has been conserved in evolution: mutations occur randomly and independently on these sites. If indeed the contribution of each individual site is too weak to control a selectable phenotype *per se*, then it will accumulate mutations at the same rate than non-functional genomic sequences. Only the coordinated loss of several sites would be subjected to natural selection, but natural mutations are uncoordinated: they occur randomly and independently (simultaneous co-mutation is only observed within ≈50 bp, excluding the possibility that multiple miRNA target sites are coordinately mutated[19-21]). Therefore, the observed conservation of miRNA seed matches implies that each of them exerts a selectable function, independently of the activity of other sites. Note that other interpretations of our experimental results are possible, but they also suggest that only few genes are functionally sensitive to fine-tuning.[12]

Our result thus opens the possibility that a large fraction of known or predicted miRNA binding sites are phenotypically inconsequential. It would then remain to be explained why these sites have been conserved in evolution, if they cannot mediate a biologically relevant repression.

Could mRNA-mediated miRNA titration explain seed match conservation?

By an absolute measurement of the intracellular levels of miRNAs and their targets in C2C12 cultured cells, we found a few mRNAs whose abundance seems to be sufficient to titrate miRNAs by more than 10%. We verified the miRNA titrating activity for one of them (the *Tmsb4x* mRNA) by mutating its miR-1a/miR-206 binding site by genome editing: miR-1a/miR-206 was found to be more active on a reporter gene after that mutation, suggesting that it was indeed titrated by the *Tmsb4x* mRNA. That result thus demonstrates the possibility that a miRNA binding site could be conserved for its miRNA-sequestering activity, rather than for its mRNA-silencing activity. In other words, computational target predictions may be contaminated by such "pseudo-targets," whose binding sites for miRNAs have been conserved despite their functional insensitivity to the regulation.

Yet we only identified a small number of potential miRNA titrators in C2C12 cells (2 for miR-1a/miR-206 and 5 for miR-133). We cannot exclude that some other predicted targets for these 2 miRNA families may behave as miRNA titrators in other cell types: perhaps many cell types express detrimental levels of miR-1a/miR-206 or miR-133, and some of their highly abundant mRNAs may act as titrators, thus providing a beneficial function for the organism's overall fitness. Even though it is formally possible, we do not favor that hypothesis. It is now clear that an efficient titration of miRNAs by individual mRNAs must be rare *in vivo*.[13-16] Among the rigorous studies investigating the necessary conditions for mRNA-mediated miRNA titration, only one suggested the existence of endogenous titrators — under specific conditions regarding miRNA and target abundance and their mutual affinity.[17] It thus seems unlikely that a large fraction of predicted miRNA targets would have been under a selective pressure to titrate miRNAs: the few titrators we observed for miR-1a/206 and miR-133 may very well be the only ones.

Because several articles concluded that individual mRNAs are generally unable to mediate substantial titration,[13-16] it is also important to confront our results on mRNA-mediated miRNA titration to the existing literature. Some of these conclusions are derived from textbook-based estimations or imprecise statements on mRNA abundance, but others were deduced from *in vivo* measurements of competition between targets for miRNA binding.[15,16] These 2 studies concluded for example that miR-122 (which is strongly hepatocyte-specific) is not titrated to a noticeable extent by any mRNA in hepatocytes: there are so many miR-122-interacting mRNAs that each of them contributes little to the overall titration. In other words, it does not really affect miR-122 availability, and it would need to be expressed at unrealistically high levels to impart some functional modulation of miR-122.

While we do not question that conclusion, we reasoned that miRNA titration by mRNAs, whenever it occurs, plays miRNA regulatory functions. Therefore, titrators probably repress specific miRNAs in cells where they would have detrimental effects. miRNAs strongly expressed in a given cell type (*e.g.*, miR-122 in hepatocytes, miR-155 in haematopoietic cells,) probably play a biological role in that cell type, and titration of these miRNAs is expected to be beneficial only in other cell types. This is how we interpret the observation that miR-122 is not efficiently titrated by any individual mRNA in mature hepatocytes: we propose that miR-122 is titrated by some mRNAs in other cell types or other differentiation stages, reinforcing the specificity of its expression pattern.

Phylogenetic analyses also challenge the current view

Because there may not be many miRNA-titrating mRNAs, we had to consider another possible explanation to the frequent

conservation of miRNA seed matches. We thus explored the possibility that there could be a fundamental bias in the comparative genomics approach. We observed that seed matches for some miRNA families can be more deeply conserved in evolution than the miRNA family itself. In other words, some miRNA binding sites are conserved… even in species that do not express any miRNA with that seed.

It is currently not clear why these binding sites are conserved: we can imagine that these sites bind an RNA-binding protein or a DNA-binding protein, whose binding is under selective pressure — and they would fortuitously be complementary to a miRNA seed. Perhaps, too, some parts of the genome appear to be conserved even in the absence of any selective pressure (because of a locally low mutation rate,[18] or for reasons which are still not known).

Such "over-conserved" sites (*i.e.*: miRNA binding sites that are more conserved than the cognate miRNA family) are more frequent for poorly conserved miRNA families than for deeply conserved families. This is mainly due to a technical artifact: because of the arborescent structure of the phylogenetic tree, it is easier to observe a site being conserved outside a shallow clade than outside a deep clade. There are more outer species outside shallow clades (hence increasing the chances of observing the site in at least one of them), and the various lineages diverged for a shorter time (hence increasing the chances that the site did not have enough time to diverge if inconsequential). When re-analyzing predicted miRNA binding sites for the most popular target prediction programs, we thus observed a higher rate (≈50 to 70%) of over-conserved sites for Hominidae-specific miRNA families, than for more deeply conserved families. It is likely that the causes of over-conservation (*e.g.*, conserved protein-binding sites that exhibit complementarity to an miRNA seed) apply similarly to miRNAs specific to deeper clades, and the observed over-conservation rate would be lower only because of such technical artifact. These high ($\geqslant$50%) rates might therefore extrapolate to deeply conserved miRNA families.

Importantly, these "over-conserved" sites certainly do not constitute the totality of false positive predictions: some seed matches may be conserved for miRNA-independent reasons, without being more conserved than the miRNA family. Some may even be more conserved than the miRNA family, but their over-conservation would not be visible in our analysis because they are conserved in outer species whose genome is not sequenced yet (*e.g.*, only one non-Euteleostomi vertebrate genome is currently available: the *Petromyzon marinus* lamprey genome; some sites for Euteleostomi-specific miRNAs may be conserved in non-Euteleostomi genomes, but not that of *Petromyzon marinus*). Reciprocally, it is conceivable that some sites identified by our approach are in fact true positives: one could imagine a scenario where a protein-binding site has been conserved for a long time, then a miRNA family with a complementary seed would appear — and the site would gain a novel beneficial function by this newly acquired repression. Such a "regulatory relay" has never been documented so far. It is more conservative to consider the functionality of over-conserved seed matches with caution, and not to assume the existence of such a scenario. Overall, the over-conservation analysis shows that phylogenetic conservation of a seed match is no guarantee for their functionality — over-conserved seed matches are not more convincing than non-conserved seed matches.

Conclusion

The notion that miRNAs control a large diversity of physiological phenotypes is well established. Our results challenge that view, suggesting that most miRNA "targets" identified by current high-throughput experimental or computational methods may not be repressed enough to trigger macroscopic phenotypes. The generality of our findings will now have to be assessed independently, and on other experimental systems. Fortunately, the required analyses are rather simple, and they appear to be accessible to our colleagues interested in miRNA-mediated regulation. An investigator comparing target expression levels between wild-type individuals and miRNA mutants, should also be able to measure inter-individual variability among wild-type specimens. An analyst evaluating the phylogenetic conservation of a given miRNA binding site should also be able to probe its conservation in species that do not express that miRNA family.

Questioning the validity of current miRNA target identification schemes could have enormous consequences on a large part of the published scientific literature which relied on these experimental or computational methods. Approaches aimed at identifying false positives, like the ones described above, could help isolating biologically relevant targets from the long lists of computationally or experimentally identified candidates. But it is important to realize that there is a sort of inertia in scientific practice, sometimes even against published evidence. Perhaps a convincing example is the case of the miR-34 family, which is regulated by p53, and which was initially proposed to participate in p53-mediated control of cell proliferation in mammals.[22] Rigorous assessment of the p53 response in miR-34 mutant mice failed to observe any defect,[23] yet several years later, many articles still refer to miR-34 as a component of the p53 pathway, without any reference to the *in vivo* mouse study.

But fortunately, the recent development of genome editing tools now allows the *in vivo* exploration of functional consequences of the loss of individual miRNA/mRNA interactions.[9] It can be anticipated that such precise *in vivo* assessment could disprove many of the published functions for miRNAs: biological roles which were inferred from reporter assays or from automatic annotation of predicted miRNA targets, may not stand the test of rigorous biological investigation.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

Acknowledgments

I thank I. Busseau, G. Dupin, A. Gato and B. Li for critical reading of the manuscript.

Funding

This work was supported by the Human Frontier Science Program under Grant CDA-00017/2010-C; and CNRS and Sanofi under Grant ATIP-Avenir.

ORCID

Hervé Seitz http://orcid.org/0000-0001-8172-5393

References

1. Mittal N, Zavolan M. Seq and CLIP through the miRNA world. Genome Biol 2014; 15(1):202; PMID:24460822; https://doi.org/10.1186/gb4151
2. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell 2009; 136(2):215-33; PMID:19167326; https://doi.org/10.1016/j.cell.2009.01.002
3. Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature 2009; 460(7254):479-86; PMID:19536157; https://doi.org/10.1038/nature08170
4. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jr, Jungkamp AC, Munschauer M, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 2010; 141(1):129-41; PMID:20371350; https://doi.org/10.1016/j.cell.2010.03.009
5. Zisoulis DG, Lovci MT, Wilbert ML, Hutt KR, Liang TY, Pasquinelli AE, Yeo GW. Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. Nat Struct Mol Biol 2010; 17(2):173-9; PMID:20062054; https://doi.org/10.1038/nsmb.1745
6. Leung AK, Young AG, Bhutkar A, Zheng GX, Bosson AD, Nielsen CB, Sharp PA. Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. Nat Struct Mol Biol 2011; 18(2):237-44; PMID:21258322; https://doi.org/10.1038/nsmb.1991
7. Johnston RJ, Hobert O. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. Nature 2003; 426(6968):845-9; PMID:14685240; https://doi.org/10.1038/nature02255
8. Seitz H. Redefining microRNA targets. Curr Biol 2009; 19(10):870-3; PMID:19375315; https://doi.org/10.1016/j.cub.2009.03.059
9. Ecsedi M, Rausch M, Großhans H. The *let-7* microRNA directs vulval development through a single target. Dev Cell 2015; 32(3):335-44; PMID:25669883; https://doi.org/10.1016/j.devcel.2014.12.018
10. Drexel T, Mahofsky K, Latham R, Zimmer M, Cochella L. Neuron type-specific miRNA represses two broadly expressed genes to modulate an avoidance behavior in *C. elegans*. Genes Dev 2016; 30(18):2042-7; PMID:27688400; https://doi.org/10.1101/gad.287904.116
11. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife 2015; 4; PMID:26267216; https://doi.org/10.7554/eLife.05005; https://www.ncbi.nlm.nih.gov/pubmed/26267216
12. Pinźon N, Li B, Martinez L, Sergeeva A, Presumey J, Apparailly F, Seitz H. microRNA target prediction programs predict many false positives. Genome Res 2017; 27(2):234-45; PMID:28148562; https://doi.org/10.1101/gr.205146.116
13. Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP. Weak seed-pairing stability and high target-site abundance decrease the proficiency of *lsy-6* and other microRNAs. Nat Struct Mol Biol 2011; 18(10):1139-46; PMID:21909094; https://doi.org/10.1038/nsmb.2115
14. Wee LM, Flores-Jasso CF, Salomon WE, Zamore PD. Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. Cell 2012; 151(5):1055-67; PMID:23178124; https://doi.org/10.1016/j.cell.2012.10.036
15. Denzler R, Agarwal V, Stefano J, Bartel DP, Stoffel M. Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. Mol Cell 2014; 54(5):766-76; PMID:24793693; https://doi.org/10.1016/j.molcel.2014.03.045
16. Denzler R, McGeary SE, Title AC, Agarwal V, Bartel DP, Stoffel M. Impact of microRNA levels, target-site complementarity, and cooperativity on competing endogenous RNA-regulated gene expression. Mol Cell 2016; 64(3):565-79; PMID:27871486; https://doi.org/10.1016/j.molcel.2016.09.027
17. Bosson AD, Zamudio JR, Sharp PA. Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. Mol Cell 2014; 56(3):347-59; PMID:25449132; https://doi.org/10.1016/j.molcel.2014.09.018
18. Baer CF, Miyamoto MM, Denver DR. Mutation rate variation in multicellular eukaryotes: causes and consequences. Nat Rev Genet 2007; 8(8):619-31; PMID:17637734; https://doi.org/10.1038/nrg2158
19. Tian D, Wang Q, Zhang P, Araki H, Yang S, Kreitman M, Nagylaki T, Hudson R, Bergelson J, Chen JQ. Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. Nature 2008; 455(7209):105-8; PMID:18641631; https://doi.org/10.1038/nature07175
20. Schrider DR, Hourmozdi JN, Hahn MW. Pervasive multinucleotide mutational events in eukaryotes. Curr Biol 2011; 21(12):1051-4; PMID:21636278; https://doi.org/10.1016/j.cub.2011.05.013
21. Harris K, Nielsen R. Error-prone polymerase activity causes multinucleotide mutations in humans. Genome Res 2014; 24(9):1445-54; PMID:25079859; https://doi.org/10.1101/gr.170696.113
22. He L, He X, Lim LP, de Stanchina E, Xuan Z, Liang Y, Xue W, Zender L, Magnus J, Ridzon D, et al. A microRNA component of the p53 tumour suppressor network. Nature 2007; 447(7148):1130-4; PMID:17554337; https://doi.org/10.1038/nature05939
23. Concepcion CP, Han YC, Mu P, Bonetti C, Yao E, D'Andrea A, Vidigal JA, Maughan WP, Ogrodowski P, Ventura A. Intact p53-dependent responses in miR-34-deficient mice. PLoS Genet 2012; 8(7):e1002797; PMID:22844244; https://doi.org/10.1371/journal.pgen.1002797