



A new perspective on microRNA-guided gene regulation specificity, and its potential generalization to transcription factors and RNA-binding proteins

Hervé Seitz

► To cite this version:

Hervé Seitz. A new perspective on microRNA-guided gene regulation specificity, and its potential generalization to transcription factors and RNA-binding proteins. *Nucleic Acids Research*, 2024, 52 (16), pp.9360-9368. <10.1093/nar/gkae694>. <hal-04669737>

HAL Id: hal-04669737

<https://hal.science/hal-04669737v1>

Submitted on 9 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

PAPER

A new perspective on microRNA-guided gene regulation specificity, and its potential generalization to transcription factors and RNA-binding proteins

Hervé Seitz^{1,*}¹Institut de Génétique Humaine (UMR 9002), CNRS, 141, rue de la Cardonille, 34396, France

*Corresponding author. herve.seitz@igh.cnrs.fr

FOR PUBLISHER ONLY Received on Date Month Year; revised on Date Month Year; accepted on Date Month Year

Abstract

Our conception of gene regulation specificity has undergone profound changes over the last 20 years. Previously, regulators were considered to control few genes, recognized with exquisite specificity by a “key and lock” mechanism. But recently, genome-wide exploration of regulator binding site occupancy (whether on DNA or RNA targets) revealed extensive lists of molecular targets for every studied regulator. Such poor biochemical specificity suggested that each regulator controls many genes, collectively contributing to biological phenotypes. Here I propose a third model, whereby regulators’ biological specificity is only partially due to “key and lock” biochemistry. Rather, regulators affect many genes at the microscopic scale, but biological consequences for most interactions are attenuated at the mesoscopic scale: only a few regulatory events propagate from microscopic to macroscopic scale, others are made inconsequential by homeostatic mechanisms. That model is well supported by the microRNA literature, and data suggests it extends to other regulators. It reconciles contradicting observations from biochemistry and comparative genomics on one hand, and *in vivo* genetics on the other hand, but this conceptual unification is obscured by common misconceptions and counter-intuitive modes of graphical display. Profound understanding of gene regulation requires conceptual clarification, and better suited statistical analyses and graphical representation.

Key words: specificity, gene regulation, microRNA, transcription factor, RNA-binding protein, homeostasis

Introduction: a shift in the definition of regulatory targets

The development of deep-sequencing-based methods allowed high-throughput identification of physical interactors for transcription factors (“TFs”) [1, 2], microRNAs (“miRNAs”) [3] and RNA-binding proteins (“RBPs”) [4, 5]. Each of these analyses revealed that these gene regulators have very poor intrinsic specificity, with each TF, miRNA or RBP interacting with hundreds to thousands of distinct targets in mammals.

That discovery profoundly modified our representation of gene regulation. Before the advent of high-throughput methods, gene regulation was commonly described by a very restrictive “key and lock” mechanism, where various biochemical discriminants would guarantee that the regulator physically binds to a very limited number of targets (see [6] for the example of transcriptional regulation). In stark contrast with that view, gene regulation is now essentially presented as a complex, highly-branched network, with each TF, miRNA or RBP controlling simultaneously a large number of genes whose coordinated (typically mild) regulation ultimately results in macroscopic phenotypes [7, 8, 9, 10].

Yet the biological functionality of many of these regulator-target interactions appears questionable: (i) A large fraction of

experimentally-identified binding sites for TFs [11, 12], miRNAs [13] or RBPs [14] do not appear to be under selective pressure and are therefore not phylogenetically conserved. (ii) When a regulator controls a biological phenotype, and when the targets responsible for that phenotype are identified, they generally constitute a small minority of experimentally-detected or phylogenetically-conserved binding sites (see Table 1; such *in vivo* assessment of the functional contribution of individual binding sites has only been performed for miRNAs so far).

In these experiments, the phenotypes of miRNA mutants was compared to the phenotypes caused by precise mutation of miRNA binding sites on individual targets. These studies found that abolishing the repression of 1 or 2 target(s) essentially phenocopies the miRNA mutant, suggesting that all the other targets contribute little to the organism-scale phenotype. The tested targets were usually selected by an educated guess, with the exception of *egl-1*, which was identified by an unsupervised screen among every predicted binding site for *mir-35* family miRNAs in *C. elegans* [15].

For four analyzed miRNAs in Table 1, mutating these binding sites appeared to phenocopy every reported defect of the miRNA mutant. For the remaining four miRNAs, mutants in miRNA binding sites were reported to recapitulate most,

Table 1. Identified regulator-target interactions involved in biological phenotypes.

Regulator	Functional target(s)	Reference
<i>Mus musculus</i> miR-155	<i>AID</i> and <i>SOCS1</i>	[16, 20, 17]
<i>Cænorhabditis elegans</i> <i>let-7</i>	<i>lin-41</i>	[21, 22]
<i>C. elegans</i> miR-791	<i>akap-1</i> and <i>cah-3</i>	[23]
<i>M. musculus</i> miR-200	<i>Zeb1</i>	[24]
<i>C. elegans</i> miR-35	<i>egl-1</i>	[15]
<i>Drosophila melanogaster</i> miR-iab-4/8	<i>hth</i>	[25]
<i>M. musculus</i> miR-140-5p	<i>Wnt11</i>	[18]
<i>Oryzias latipes</i> miR-202	<i>Tead3b</i>	[19]

but not all phenotypes observed in the miRNA mutant (class-switch recombination phenotype in *miR-155*^{-/-} B cells is not recapitulated in AID site mutant cells: [16]; Treg cell deficiency in *miR-155*^{-/-} mice is not recapitulated in SOCS1 site mutant mice: [17]; reduction in brood size and embryonic viability in *egl-1* site mutant *C. elegans* only partially phenocopies *mir-35* family mutants: [15]; some skeletal defects in *miR-140*^{-/-} mice are not recapitulated in *Wnt11* site mutant mice: [18]; decrease in medaka egg developmental success rate, as observed in *miR-202*^{-/-} females, is not recapitulated in *Tead3b* site mutants: [19]). These observations keep open the possibility that additional targets may also contribute to the observed phenotypes; alternatively, differences in genetic background between miRNA and target mutants, and miRNA-independent alteration of 3' UTR regulation, may explain differences between these mutants.

These observations deepen the paradox highlighted by [26] in the pre-deep-sequencing era, and which was coined the “futility theorem” (namely, that almost every genomic occurrence of a sequence motif known to bind a TF is in fact devoid of a biological function). At the time, it was already clear that sequence motifs for any TF are highly frequent in metazoan genomes, while only $\approx 1/1000$ of these sites were believed to be functionally important. It was proposed that chromatin structure and other factors could restrict site accessibility *in vivo*, therefore partially explaining that discrepancy (even though the theorem was referring to TFs only, that concept could be trivially generalized to miRNAs and RBPs, whose RNA target sites can also be made inaccessible by RNA structures and by competing protein or RNA partners). Phylogenetic conservation was presented as an efficient filter in order to exclude unimportant sites.

Nowadays, after 20 more years of work (including high-throughput exploration of regulator binding *in vivo*), these notions can be revisited. Not only can the futility theorem be extended to other classes of regulators (at least miRNAs and RBPs), whose binding sites are also ≈ 6 bp long (hence: similarly frequent in genomes); but it even applies to (i) binding sites which do bind efficiently the regulator *in vivo*; (ii) phylogenetically conserved binding sites.

Of note, the primary determinant for miRNA binding appears to be a perfect sequence match to the miRNA “seed” (nt 2–7 of the miRNA), potentially supplemented by additional context features [27]. In first approximation, it is therefore straightforward to predict whether a gene is a credible miRNA target, simply by verifying whether its 3' UTR contains, or not, a perfect seed match. Such a “digital” description of miRNA targets contrasts with the fuzzier prediction of TF or RBP targets, whose binding sites are more aptly described by a position weight matrix (PWM), which predicts molecular targets in an analogical manner.

The simplicity and clarity of miRNA target recognition greatly facilitated the functional exploration of miRNA-target interactions,

which is more advanced than for TFs and RBPs. As a consequence, most of the currently available data on this topic deals with miRNA-mediated regulation, and most of the literature cited in this review centers on miRNAs. Another consequence is that the limitations in functional assignment for gene regulators are also better characterized for miRNAs (see for example [28, 29]). Yet, the concepts discussed here are formally the same for every class of regulator, and it can be anticipated that the TF and RBP literature will soon mature in the same direction as the miRNA literature does.

Significance is not relevance

It therefore appears that most physical interactors for these gene regulators are not functionally affected by the regulator. This does not mean that their expression is insensitive to the regulator: many of these genes are actually regulated at the molecular scale, and transcription tends to be measurably affected by TF binding; RNA stability and translatability tend to be measurably affected by miRNA or RBP binding (*e.g.*, see [30, 31, 32, 33, 34]).

But such consequences at the microscopic scale do not necessarily translate into macroscopic phenotypes. Alterations in abundances of molecules do not constitute biological phenotypes: only phenotypes visible to natural selection (*e.g.*, those affecting organismal fitness, or reproductive ability of the kin) should be considered biologically relevant. That definition excludes traits which may be clinically relevant (*e.g.*, phenotypes affecting lifespan after the end of reproductive age) or macroscopically visible (*e.g.*, morphological variation which is neutral in selective terms), but whose effects would be inexistent on the species' fitness.

Perhaps some confusion originated from a sloppy usage of the term “significant”. Significance of a molecular effect (*e.g.*, a change in mRNA abundance) means that the effect appears to be real, and (provided that the control conditions are adequate) it is unlikely to be due to confounding variables or to random fluctuations in the experimental measurement. It certainly does not mean that the observed molecular effect has a biologically important function.

Changes at the microscopic scale do not necessarily trigger macroscopic effects visible to natural selection. First, the propagation of perturbations from the microscopic to the macroscopic scale is frequently buffered by homeostatic mechanisms, as exemplified by the large proportion of reported haplo-sufficient genes in animals [35, 36, 37], and more generally, by the general robustness of biological phenotypes after small (≈ 2 -fold) changes in gene expression [38, 39]. Second, even perturbations reaching macroscopic outcomes may be neutral in terms of natural selection [40].

Measured yet functionless changes in gene expression levels are often called “noise”. But it is important to realize that functionless fluctuations in gene expression involve the same molecular actors (in particular, gene regulators) as biologically useful processes shaped

by natural selection. It therefore appears meaningless to distinguish useless regulatory events from biological noise [8] — they do constitute “biological noise”. Similarly, it is misleading to name them “inadvertent regulatory events” [8], as if there were “inadvertent” vs. “intentional” biological processes: this is a finalistic description, and molecules do not have any intention — hence from an objective point of view, functional and non-functional regulatory events are equally real, mechanistically indistinguishable, and they do not differ by any sort of intelligent desire.

Advocating for intuitive data presentation

Graphical representation of the effect of a regulator on its molecular targets can influence the reader’s perception of the biological phenomenon, probably contributing to a misrepresentation of gene regulation efficiency. Here I will discuss how the mode of graphical representation and the type of statistical tests being performed may alter the perception of the biological importance of gene regulation.

For example, endogenous miRNA-guided repression is very limited (typically a few percent decrease in mRNA abundance, and only exceptionally more than 50%; [30, 31, 41, 13]). But for some reason, gene expression fold-changes are frequently represented with poorly-intuitive cumulative plots, rather than with the more familiar barplots or boxplots. Cumulative plots display (for any value of x) the proportion of genes whose fold-change (or more frequently: its logarithm) is lower than x . They typically exhibit a sigmoid shape, with very few genes having $\log(\text{fold-changes})$ lower than low x values, most genes having $\log(\text{fold-changes})$ lower than large x values, and a steep increase around 0 because most genes have a $\log(\text{fold-change})$ close to 0.

Differences in fold-changes which are so small that they are hardly visible on a boxplot usually appear much more clear on a cumulative plot. This is exemplified by Figure 1a, showing the effect of miR-9 transfection on the abundance of its predicted target mRNAs in HeLa cells (dataset taken from [42]). Candidate miR-9 target genes are stratified by miRNA complementarity site type (with 6mer sites being the least efficient and 8mer sites being the most efficient at guiding target repression; [27]). Cumulative curves are neatly distinct from each other, suggesting a clear difference in mRNA repression among targets with various site types. On the other hand, a boxplot of the same dataset yields highly-overlapping boxes, clearly showing that the measured fold-change values do not differ much among various site types. It is certainly preferable to use the more intuitive representation of gene expression fold-changes with boxplots, to make sure that the readership gets a correct intuition of the amplitude of gene regulation and how much it differs between various gene sets.

Unlike with miRNAs (which all share the same effectors: Ago proteins, and their downstream effectors), the amplitude of regulation by TFs and RBPs can be quite diverse, reflecting the multiplicity of their mechanism of action. But reported fold-changes in the expression of their direct targets are also usually close to 1.5 [43, 44, 45, 46, 34]. The mere fact that regulation strength can be variable among TFs and RBPs also implies that the extent of gene regulation by these factors should be made explicit: knowing that a gene is regulated is hardly useful if one cannot tell by how much it is regulated.

Besides graphical display, analytical strategies can also obscure the reader’s perception of gene regulation. In the scientific literature, it is now increasingly common to measure the significance of differences by performing a two-sample Kolmogorov-Smirnov test, rather than a t-test or its derivatives (Welch’s t-test, Wilcoxon test, Dunnett’s test or ANOVA). Such a choice is hard to rationalize,

because a statistical test should be chosen based on the specific hypothesis that is put to the test, irrespectively of its apparent popularity in the field.

The Kolmogorov-Smirnov test assesses the likelihood that the two datasets are sampled from the same underlying distribution: a low p -value may be due to any sort of difference between the two distributions (mean, or standard deviation, or shape). On the other hand, a low t-test (or one of its derivatives) p -value allows concluding on the equality of the means of the two sampled populations, which is what the reader generally has in mind when assessing whether a regulator tends to have either a positive or negative effect on gene expression (see Figure 1b for a comparison: values for both conditions “x” and “y” tend to be close to 10, and a t-test on these datasets does not detect any significant difference; but distribution shape differs between x and y, and a Kolmogorov-Smirnov test finds a significant difference between those distributions). The Kolmogorov-Smirnov test is certainly useful in specific circumstances (*e.g.*, when theoretical reasons suggest a difference in shape for the two distributions), but it is hard to interpret when assessing the effect of a regulator on its targets.

Implications for functional assignment: with great sensitivity comes great responsibility

Because current high-throughput methods are so sensitive, it is now possible to detect trace amounts of nucleic acids, and minuscule changes in molecule abundances. The fact that a molecule is observed does not imply that it is abundant enough to be functional [47]. Before the development of such sensitive techniques, it was often implicitly assumed that an RNA abundant enough to be detected by Northern blotting, or a protein abundant enough to be detected by Western blotting, was functional in that biological sample — hence the classification into “expressed” and “non-expressed” genes (in concrete terms, “expressed genes” were those whose RNA or protein product yields a dark, neat band on a blot).

Now that deep-sequencing methods are available, it is likely that every RNA will appear to be expressed in any cell type, provided that its transcriptome has been sequenced deep enough. Categorizing genes into “expressed” and “non-expressed” thus became meaningless, and such a qualitative notion is being replaced by a quantitative assessment (“how much is this gene expressed in that sample ?”). Obviously then, functionality of a gene in a given biological sample can no longer be inferred from the fact that its expression has been detected — and there is no universal expression level threshold, above which a gene would play a biological function and below which it would not.

These notions are not specific to molecule abundance, they also generalize to biological processes (the fact that a protein is measurably phosphorylated does not imply functionality of that phosphorylation event; the fact that alternative splicing isoform distribution is modified does not mean that it plays a biological function; *etc.*).

In particular, physical interaction between molecules is frequently presented as a proof of functionality for that molecular complex. It should be stressed, however, how arbitrary the definition of a molecular complex can be: molecules keep interacting, more or less stably, and the detection of their interaction depends on experimental choices (salinity of buffers, temperature, ...). Defining the composition of a complex simply means that the reported interactions were stable enough for co-purified molecules to be detected above some arbitrary cutoff, after being purified with arbitrarily-stringent experimental conditions. With extreme stringency, complexes would be defined as single molecules, because

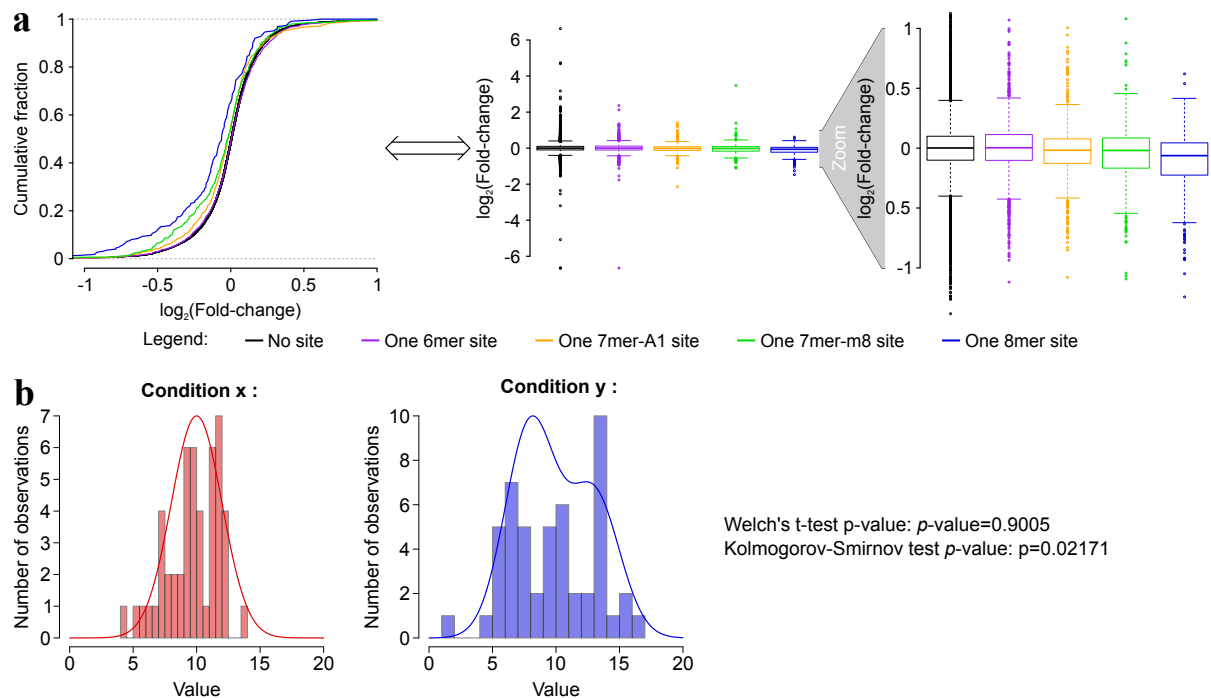


Fig. 1. Graphical display and methodological choices can alter the reader's perception of gene expression fold-changes. **a.** The same dataset (mRNA fold-change 24h after miR-9 transfection in HeLa cells) is shown as a cumulative distribution function plot (left) and as a boxplot (right). **b.** Two computationally-generated datasets, centered on the same mean but differing by their shape (“Condition x” was sampled from a monomodal distribution, “Condition y” from a bimodal distribution), were compared with a t-test (assessing the likelihood of equality of means in the sampled distributions) and with Kolmogorov-Smirnov test (assessing any type of difference between the two sampled distributions).

all their interactants would have been undetected; with extreme sensitivity, the whole organism would appear as a giant complex (all the molecules eventually interact indirectly with each other, giving physical consistency to the whole body). Within that range, there is no natural stability threshold defining the biological functionality of complexes (*e.g.*, codon-anticodon interactions are extremely weak, yet essential to every living organism). Consequently, the detection of a physical interaction between a regulator and a candidate target does not constitute a proof of biological functionality for that interaction.

In vivo veritas

In order to confidently probe whether a regulator-target interaction plays a biological role, it is thus necessary to assess it *in vivo* (with “*in vivo*” meaning: “in the whole organism”, in contrast to “*ex vivo*”, meaning: in cultured cells). Genome-editing techniques now allow precise mutation of regulators or their binding sites in individual targets, in a complete *in vivo* setting (*e.g.*, for TFs: [48]; for miRNAs, see examples listed in Table 1).

When a binding site for a regulator has been mutated, there is always a risk that reported phenotypes are due to additional, unwanted effects of the mutation (*e.g.*, mutating the site also perturbed a neighbouring binding site for another regulator, or local chromatin organization, *etc.*). At least for miRNAs (whose target recognition rules are based on nucleic acid pairing, hence precisely predictable), a convincing control can help ruling out that possibility: mutating both the miRNA and the target site in a compensatory manner should restore that individual interaction, and rescue those phenotypes that are due to that particular interaction [21, 22, 15].

When such *in vivo* data is not available, it is common practice to try to infer a regulator's biological function from the automatically-annotated functions of its molecular targets (*e.g.*, GO terms or KEGG terms). Comparing annotation terms for molecular targets and for the whole gene complement identifies terms which are over- or under-represented in molecular targets respectively to other genes. Significantly enriched, or depleted, terms among the numerous target genes for a TF or a miRNA are widely assumed to reflect that regulator's biological activity.

That method suffers from major weaknesses: first, inadequate controls frequently generate false positives by systematic biases in the automatic annotation of predicted targets [49].

Second, many GO or KEGG terms do not distinguish agonists from antagonists (*e.g.*, “regulation of endocytosis” applies equally well to activators and repressors of endocytosis): the effects of a regulator on agonists and antagonists would partly cancel each other, and eventually the phenotypic effect could be weaker than if a single gene in the pathway were regulated.

Third, even though such enrichment analysis is widely used, it typically fails at predicting a regulator's biological activity. Figure 2 shows a few examples where an actual knock-out model allowed faithful assessment of a regulator's macroscopic effects. Knockout of the murine miR-33 miRNA triggers various changes in cholesterol and lipid metabolism [50, 51], obesity and liver steatosis [52] and fibrosis [53]. Knockout of the murine and human miR-96 miRNA triggers hearing loss [54, 55], probably due to deficient differentiation of ciliated cells in the sensory epithelium [56]. Knockout of the nematode *lin-15* locus, encoding the Lin-15A and Lin-15B transcription factors, perturbs vulval development [57]. In these examples, enriched or depleted GO terms for candidate targets can rarely be linked to the phenotype: whether one selects the most enriched annotation terms [flagged with numbers 1–3],

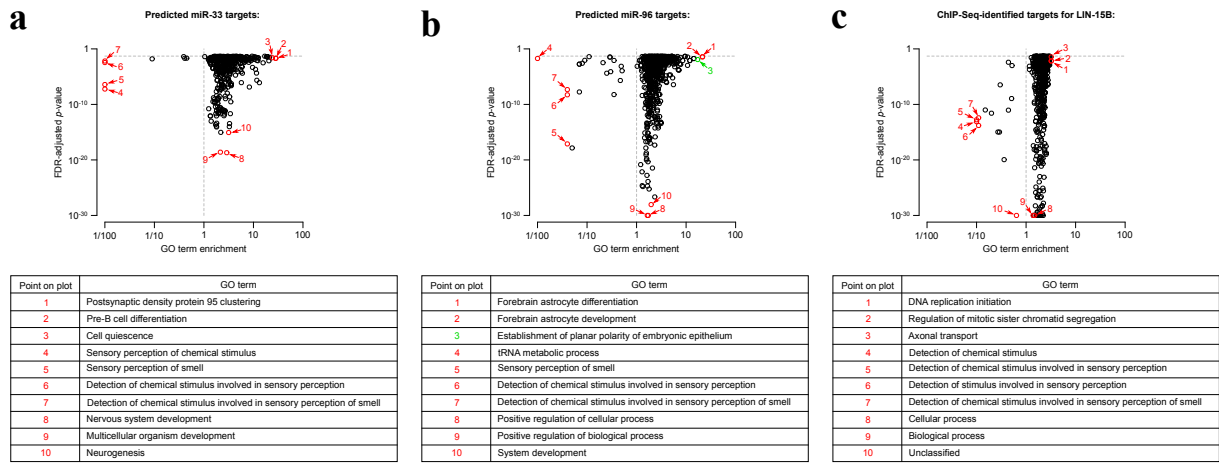


Fig. 2. GO term enrichment is a poor predictor of a regulator’s biological activity. TargetScan-predicted targets for the miR-33 (a) and miR-96 (b) miRNA families, and ChIP-Seq-identified targets for the LIN-15B TF (c) appear highly significantly enriched or depleted for several GO terms using PANTHER overrepresentation test (in red and green: top-scoring GO terms, either in terms of fold-change or p -value), yet most of them are unrelated to the observed *in vivo* phenotypes (in green: GO terms related to the phenotypes).

or the most depleted terms [flagged with numbers 4–7], or the lowest enrichment p -values [flagged with numbers 8–10], the selected terms are usually unrelated to the actual phenotype [these terms are represented by red circles] and only exceptionally can they be logically linked [represented by a green circle]).

Methodological improvements are therefore urgently needed. Biological functionality is safely assessed *in vivo* (e.g., after mutating regulator binding sites), while it cannot be extrapolated from molecular biology data and from poorly-controlled computational tools and databases. The current scientific literature relies heavily on TF, RBP and miRNA databases, whose quality is sometimes questionable (e.g., see [58]); on target prediction algorithms with high false positive rates or strong dependency to experimental protocol details [59, 60]; and on pathway analysis methods which perform very poorly at predicting phenotypic outcomes [61].

It has been proposed that phylogenetically conserved binding sites reveal functional interactions whose biological advantage may be invisible in lab conditions (e.g., they only matter in rare environmental conditions) [62]. In line with this view, comparative genomics has been heavily used in an attempt to identify biologically important interactions [63, 64]. Under the reasonable assumption that biologically-important sequences are better conserved in evolution, much effort has been devoted in the identification of phylogenetically-conserved binding sites for miRNAs, TFs or RBPs.

But that approach is limited by the possibility that a sequence motif is conserved for other reasons, while fortuitously matching a consensus sequence for a regulator’s binding site [59]. Because binding sites for miRNAs, TFs and RBPs all tend to be similarly-sized (≈ 6 bp long in general), because there are only $4^6 = 4,096$ distinct possible hexamers, and because a few thousands regulators are known in the best annotated vertebrate genomes, chances are high for a given regulator binding site motif to be also recognized by other regulators — not even counting other sources of phylogenetic conservation (binding sites for the spliceosome on RNA, for chromatin fiber organization on DNA, ...). It therefore appears that phylogenetic conservation of a candidate binding site is far from demonstrating biological functionality of the regulation, and it is certainly preferable that the burden of the proof lies with whoever wants to assign a biological function to a given regulator-target interaction.

One major drawback of *in vivo* assessment is its low throughput. But a large-scale screen for the biological importance of miRNA binding sites has already been performed for a miRNA family in *C. elegans* [15], and that method should be applicable to every genetically-tractable species, as long as the phenotype of interest is clearly defined and easy to score.

Improvements in the identification of functionally important targets could also rely on a better understanding of the determinants of phenotypic sensitivity to gene expression levels. There is currently no high-throughput experimental method in metazoans for such a screen. But proxies could be easier to measure: for example, assuming that dose-sensitive genes are tightly regulated (otherwise, their fluctuating expression could trigger adverse phenotypes), measuring inter-individual variability in gene expression could identify genes with large fluctuations among wild-type individuals (these genes would be unlikely to be functionally sensitive to a moderate regulation) and genes with little inter-individual variability (these genes would be the most promising candidates for phenotypically responding to a moderate regulation) [59, 19].

This idea may also generalize to the cellular level: assuming that dose-sensitive genes are also tightly regulated among cells of a given cell type, single-cell RNA-Seq may reveal genes whose expression is most stable across cells, hence, likely to be functionally sensitive to their expression level. Such an experiment faces several limitations (regarding the validity of its assumptions; regarding the relevant definition of a “cell type”, which may be artificial and arbitrary), but it is now readily doable for many species and tissues — future will tell how useful such results could be in the identification of functional regulatory targets.

For a new definition of regulatory targets

It is possible to reconcile microscopic and macroscopic observations: if indeed regulators bind many targets at the molecular scale, while most of these interactions are inconsequential at the organism scale, it means that the consequences of most microscopic interactions are attenuated at the mesoscopic (i.e., intermediary) scale. And indeed, homeostatic mechanisms, which can efficiently buffer fluctuations due to external causes, must similarly attenuate changes of internal origin, like gene regulation.

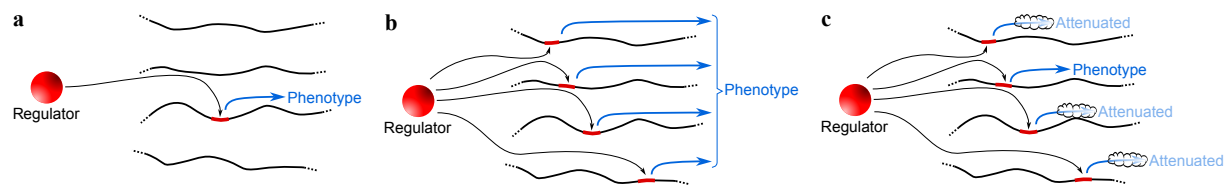


Fig. 3. Models for regulators' specificity of action. **a.** In the historical, “key and lock” model, gene regulators exhibit a great biochemical specificity. The few targets they regulate are responsible for the regulator's biological phenotype. **b.** In the current model, regulators coordinately control many targets, which collectively contribute to the phenotype. **c.** In the newly proposed model, regulators exhibit poor biochemical specificity and many genes respond at the microscopic scale. But homeostatic mechanisms attenuate the consequences of most of these events, and just a few regulator-target interactions actually contribute to the phenotype.

Such mechanisms can involve negative feedback loops (*e.g.*, if the pure effect of a regulator on a target gene is a 2-fold repression, a negative feedback could re-increase target expression and the final fold-change may just be 1.5; hence the regulation signal would be attenuated along its propagation from the microscopic to the macroscopic scale). They could also involve simple properties based on stoichiometry (*e.g.*, in conditions where an enzyme is in large excess compared to its substrate, a small change in the enzyme's expression would not have much of an effect on the biochemical flux because the enzyme would still be in large excess — meaning that the enzyme's gene would be functionally insensitive to a small change in its expression level).

In contrast to the “key and lock” model (where regulators interact with a limited set of targets; see Figure 3a), regulators would interact with large numbers of molecular targets. But in contrast with the current, “coordinated network” model (see Figure 3b), just a small number of interactions would have consequences propagating from the molecular to the organism scale, the other ones being attenuated by homeostatic mechanisms (see Figure 3c).

There is a natural antagonism between the mechanisms of developmental transitions (which tend to modify biological responses, *e.g.*, cell differentiation or migration) and homeostasis (which tends to keep them constant). Both are defining features of life, but they have not yet been analyzed under the same perspective. Quantitatively confronting these two phenomena therefore holds promises for the identification of biologically relevant regulator-target interactions: whenever a regulatory interaction is proposed, it is certainly useful to ask how efficiently its consequences can propagate from the microscopic to the organism scale, through various layers of homeostatic attenuation at the mesoscopic scale. From this point of view, it would be preferable to define “regulatory targets” as the genes whose regulation triggers organismal consequences, rather than defining them by molecular properties (“genes whose expression is affected”, “genes whose mRNA is physically bound by the regulator”, ...).

Conclusion

Technological advances in high-throughput molecular biology have revolutionized our perception of gene regulation, with evidence that most TFs, RBPs and miRNAs affect the expression of hundreds of genes each. That observation inspired the concept that regulators control biological phenotypes through the coordinated action of a large number of regulator-target interactions. The success of this theory may have been facilitated by disputable choices regarding graphical display and statistical testing of the effect of gene regulation, which fail to convey the information that gene regulation is usually modest in amplitude. At least when phenotypes are assessed at the organism scale, most of these interactions appear

inconsequential, with just a few target genes seemingly responsible for regulator-controlled phenotypes.

Proper genetic invalidation of regulator binding sites, followed by an assessment of its macroscopic consequences *in vivo*, therefore appears indispensable for a convincing functional assignment of regulator-target interactions. It is important to realize that even *in vivo* data can be misleading, if the interaction was not abolished genetically. Genetic invalidation of a regulator's binding site may not translate into any obvious macroscopic phenotype in lab conditions, even when abundant *in vivo* data suggested biological functionality (compare [65, 66]).

In the argumentation in favor of the third model that I propose here (with homeostasis acting as a filter among consequential and inconsequential interactions), several limitations have to be addressed. It is likely that some interactions play biological roles in the wild (explaining why they have been conserved), while no phenotype has been reported in lab conditions when they have been experimentally mutated [62]. It is also possible that some regulators control biological phenotypes through the coordinated regulation of many genes, yet such cases have not yet been observed because of the technical difficulty of experimentally manipulating many genes simultaneously [67].

At least that third model has the merit of reconciling molecular data and comparative genomics on one hand, and *in vivo* genetics on the other hand. It emphasizes the role of homeostasis (a universal feature in living organisms) in the convolution of regulatory signals. It also suggests that functionally important targets could be predicted more faithfully if the effects of regulatory network architectures are better understood, with perhaps specific network branches undergoing regulation attenuation through negative feedback loops while other branches may undergo regulation amplification through positive feedback loops, ultimately leading to macroscopic phenotypes.

Competing interests

No competing interest is declared.

Author contributions statement

H.S. wrote the manuscript.

Acknowledgments

The author is grateful to Isabelle Busseau, Denis Dupuy, Kazufumi Mochizuki, Alain Péliesson, Bénédicte Py and Yukihide Tomari for critical reading of the manuscript.

Data availability

Scripts and intermediary data files used to generate Figures 1 and 2 are available at https://github.com/HKeyHKey/Seitz_2024 and at <https://www.igh.cnrs.fr/en/research/departments/genetics-development/systemic-impact-of-small-regulatory-rnas#programmes-informatiques>.

References

1. G Robertson, M Hirst, M Bainbridge, M Bilenky, Y Zhao, T Zeng, G Euskirchen, B Bernier, R Varhol, A Delaney, N Thiessen, O L Griffith, A He, M Marra, M Snyder, and S Jones. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods*, 4(8):651–657, 2007.
2. D S Johnson, A Mortazavi, R M Myers, and B Wold. Genome-wide mapping of in vivo protein-DNA interactions. *Science*, 316(5830):1497–1502, 2007.
3. S W Chi, J B Zang, A Mele, and R B Darnell. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, 460(7254):479–486, 2009.
4. M Hafner, M Landthaler, L Burger, M Khorshid, J Hausser, P Berninger, A Rothballer, M Ascano, Jr., A C Jungkamp, M Munschauer, A Ulrich, G S Wardle, S Dewell, M Zavolan, and T Tuschl. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, 141(1):129–141, 2010.
5. E L Van Nostrand, P Freese, G A Pratt, X Wang, X Wei, R Xiao, S M Blue, J Y Chen, N A L Cody, D Dominguez, S Olson, B Sundararaman, L Zhan, C Bazile, L P B Bouvrette, J Bergalet, M O Duff, K E Garcia, C Gelboin-Burkhart, M Hochman, N J Lambert, H Li, M P McGurk, T B Nguyen, T Palden, I Rabano, S Sathe, R Stanton, A Su, R Wang, B A Yee, B Zhou, A L Louie, S Aigner, X D Fu, E Lécuyer, C B Burge, B R Graveley, and G W Yeo. A large-scale binding and functional map of human RNA-binding proteins. *Nature*, 583(7818):711–719, 2020.
6. B M Emerson. Specificity of gene regulation. *Cell*, 109(3):267–270, 2002.
7. D P Bartel and C Z Chen. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat Rev Genet*, 5(5):396–400, 2004.
8. M D Biggin. Animal transcription networks as highly connected, quantitative continua. *Dev Cell*, 21(4):611–626, 2011.
9. E L Sternburg and F V Karginov. Global approaches in studying RNA-binding protein interaction networks. *Trends Biochem Sci*, 45(7):593–603, 2020.
10. C P Bracken. "Crowd-control" by RNA: a pervasive theme in biology. *RNA*, 29(7):885–888, 2023.
11. D T Odom, R D Dowell, E S Jacobsen, W Gordon, T W Danford, K D MacIsaac, P A Rolfe, C M Conboy, D K Gifford, and E Fraenkel. Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nat Genet*, 39(6):730–732, 2007.
12. D Schmidt, M D Wilson, B Ballester, P C Schwalie, G D Brown, A Marshall, C Kutter, S Watt, C P Martinez-Jimenez, S Mackay, I Talianidis, P Flicek, and D T Odom. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science*, 328(5981):1036–1040, 2010.
13. V Agarwal, G W Bell, J W Nam, and D P Bartel. Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, 4:doi: 10.7554/eLife.05005, 2015.
14. A J Litterman, R Kageyama, O Le Tonqueze, W Zhao, J D Gagnon, H Goodarzi, D J Erle, and K M Ansel. A massively parallel 3' UTR reporter assay reveals relationships between nucleotide content, sequence conservation, and mRNA destabilization. *Genome Res*, 29(6):896–906, 2019.
15. B Yang, M Schwartz, and K McJunkin. In vivo CRISPR screening for phenotypic targets of the *mir-35-42* family in *C. elegans*. *Genes Dev*, 34(17-18):1227–1238, 2020.
16. Y Dorsett, K M McBride, M Jankovic, A Gazumyan, T H Thai, D F Robbani, M Di Virgilio, B Reina San-Martin, G Heidkamp, T A Schwickert, T Eisenreich, K Rajewsky, and M C Nussenzweig. MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. *Immunity*, 28(5):630–638, 2008.
17. L F Lu, G Gasteiger, I S Yu, A Chaudhry, J P Hsin, Y Lu, P D Bos, L L Lin, C L Zawislak, S Cho, J C Sun, C S Leslie, S W Lin, and A Y Rudensky. A single miRNA-mRNA interaction affects the immune response in a context- and cell-type-specific manner. *Immunity*, 43(1):52–64, 2015.
18. C Young, M Caffrey, C Janton, and T Kobayashi. Reversing the miRNA -5p/-3p stoichiometry reveals physiological roles and targets of miR-140 miRNAs. *RNA*, 28(6):854–864, 2022.
19. S Janati-Idrissi, M R de Abreu, C Guyomar, F de Mello, T Nguyen, N Mechkouri, S Gay, J Montfort, A A Gonzalez, M Abbasi, J Bugeon, V Thermes, H Seitz, and J Bobe. Looking for a needle in a haystack: de novo phenotypic target identification reveals Hippo pathway-mediated miR-202 regulation of egg production. *Nucleic Acids Res*, 52(2):738–754, 2024.
20. G Teng, P Hakimpour, P Landgraf, A Rice, T Tuschl, R Casellas, and F N Papavasiliou. MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. *Immunity*, 28(5):621–629, 2008.
21. M Ecsedi, M Rausch, and H Großhans. The *let-7* microRNA directs vulval development through a single target. *Dev Cell*, 32(3):335–344, 2015.
22. F Aeschmann, A Neagu, M Rausch, and H Großhans. *let-7* coordinates the transition to adulthood through a single primary and four secondary targets. *Life Sci Alliance*, 2(2):e201900335, 2019.
23. T Drexel, K Mahofsky, R Latham, M Zimmer, and L Cochella. Neuron type-specific miRNA represses two broadly expressed genes to modulate an avoidance behavior in *C. elegans*. *Genes Dev*, 30(18):2042–2047, 2016.
24. A C Title, S J Hong, N D Pires, L Hasenöhrl, S Godbersen, N Stokar-Regenscheit, D P Bartel, and M Stoffel. Genetic dissection of the miR-200-Zeb1 axis reveals its importance in tumor differentiation and invasion. *Nat Commun*, 9(1):4671, 2018.
25. D L Garaulet, B Zhang, L Wei, E Li, and E C Lai. miRNAs and Neural Alternative Polyadenylation Specify the Virgin Behavioral State. *Dev Cell*, 54(3):410–423.e4, 2020.
26. W W Wasserman and A Sandelin. Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet*, 5(4):276–287, 2004.
27. D P Bartel. MicroRNAs: target recognition and regulatory functions. *Cell*, 136(2):215–233, 2009.
28. K W Witwer and M K Halushka. Toward the promise of microRNAs — Enhancing reproducibility and rigor in microRNA research. *RNA Biol*, 13(11):1103–1116, 2016.
29. S Mockly and H Seitz. Inconsistencies and limitations of current microRNA target identification methods. *Methods Mol Biol*, 1970:2019:1970:291–2019:1970:314, 2019.

30. M Selbach, B Schwanhäusser, N Thierfelder, Z Fang, R Khanin, and N Rajewsky. Widespread changes in protein synthesis induced by microRNAs. *Nature*, 455(7209):58–63, 2008.
31. D Baek, J Villén, C Shin, F D Camargo, S P Gygi, and D P Bartel. The impact of microRNAs on protein output. *Nature*, 455(7209):64–71, 2008.
32. S Kadauke, M I Udugama, J M Pawlicki, J C Achtman, D P Jain, Y Cheng, R C Hardison, and G A Blobel. Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. *Cell*, 150(4):725–737, 2012.
33. L D Boxer, B Barajas, S Tao, J Zhang, and P A Khavari. ZNF750 interacts with KLF4 and RCOR1, KDM1A, and CTBP1/2 chromatin regulators to repress epidermal progenitor genes and induce differentiation genes. *Genes Dev*, 28(18):2013–2026, 2014.
34. D Sharma, L L Zagore, M M Brister, X Ye, C E Crespo-Hernández, D D Licatalosi, and E Jankowsky. The kinetic landscape of an RNA-binding protein in cells. *Nature*, 591(7848):152–156, 2021.
35. R A Fisher. *The genetical theory of natural selection*, chapter III, pages 51–52. Oxford University Press, 1st edition, 1930.
36. G Jimenez-Sanchez, B Childs, and D Valle. Human disease genes. *Nature*, 409(6822):853–855, 2001.
37. J Hodgkin. *WormBook*, chapter Karyotype, ploidy and gene dosage. The *C. elegans* Research Community, 2005. doi/10.1895/wormbook.1.3.1.
38. C C Laurie-Ahlberg, A N Wilton, J W Curtsinger, and T H Emigh. Naturally occurring enzyme activity variation in *Drosophila melanogaster*. I. Sources of variation for 23 enzymes. *Genetics*, 102(2):191–206, 1982.
39. V G Cheung, L K Conlin, T M Weber, M Arcaro, K Y Jen, M Morley, and R S Spielman. Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat Genet*, 33(3):422–425, 2003.
40. J Zhang. Neutral theory and phenotypic evolution. *Mol Biol Evol*, 35(6):1327–1331, 2018.
41. H Guo, N T Ingolia, J S Weissman, and D P Bartel. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, 466(7308):835–840, 2010.
42. A Grimson, K K Farh, W K Johnston, P Garrett-Engele, L P Lim, and D P Bartel. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell*, 27(1):91–105, 2007.
43. J H Paik, R Kollipara, G Chu, H Ji, Y Xiao, Z Ding, L Miao, Z Tothova, J W Horner, D R Carrasco, S Jiang, D G Gilliland, L Chin, W H Wong, D H Castrillon, and R A DePinho. FoxOs are lineage-restricted redundant tumor suppressors and regulate endothelial cell homeostasis. *Cell*, 128(2):309–323, 2007.
44. P Chi, Y Chen, L Zhang, X Guo, J Wongvipat, T Shamu, J A Fletcher, S Dewell, R G Maki, D Zheng, C R Antonescu, C D Allis, and C L Sawyers. ETV1 is a lineage survival factor that cooperates with KIT in gastrointestinal stromal tumours. *Nature*, 467(7317):849–853, 2010.
45. J Morf, G Rey, K Schneider, M Stratmann, J Fujita, F Naef, and U Schibler. Cold-inducible RNA-binding protein modulates circadian gene expression posttranscriptionally. *Science*, 338(6105):379–383, 2012.
46. M Yamaji, M Jishage, C Meyer, H Suryawanshi, E Der, M Yamaji, A Garzia, P Morozov, S Manickavel, H L McFarland, R G Roeder, M Hafner, and T Tuschl. DND1 maintains germline stem cells via recruitment of the CCR4-NOT complex to target mRNAs. *Nature*, 543(7646):568–572, 2017.
47. G Mullokandov, A Baccarini, A Ruzo, A D Jayaprakash, N Tung, B Israelow, M J Evans, R Sachidanandam, and B D Brown. High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. *Nat Methods*, 9(8):840–846, 2012.
48. P H Wu, Y Fu, K Cecchini, D M Özata, A Arif, T Yu, C Colpan, I Gainetdinov, Z Weng, and P D Zamore. The evolutionarily conserved piRNA-producing locus pi6 is required for male mouse fertility. *Nat Genet*, 52(7):728–739, 2020.
49. T Bleazard, J A Lamb, and S Griffiths-Jones. Bias in microRNA functional enrichment analysis. *Bioinformatics*, 31(10):1592–1598, 2015.
50. T Horie, K Ono, M Horiguchi, H Nishi, T Nakamura, K Nagao, M Kinoshita, Y Kuwabara, H Marusawa, Y Iwanaga, K Hasegawa, M Yokode, T Kimura, and T Kita. MicroRNA-33 encoded by an intron of sterol regulatory element-binding protein 2 (Srebp2) regulates HDL in vivo. *Proc Natl Acad Sci USA*, 107(40):17321–17326, 2010.
51. J Kim, H Yoon, T Horie, J M Burchett, J L Restivo, N Rotllan, C M Ramírez, P B Verghese, M Ihara, H S Hoe, C Esau, C Fernández-Hernando, D M Holtzman, J R Cirrito, K Ono, and J Kim. microRNA-33 regulates ApoE lipidation and amyloid- β metabolism in the brain. *J Neurosci*, 35(44):14717–14726, 2015.
52. T Horie, T Nishino, O Baba, Y Kuwabara, T Nakao, M Nishiga, S Usami, M Izuhara, N Sowa, N Yahagi, H Shimano, S Matsumura, K Inoue, H Marusawa, T Nakamura, K Hasegawa, N Kume, M Yokode, T Kita, T Kimura, and K Ono. MicroRNA-33 regulates sterol regulatory element-binding protein 1 expression in mice. *Nat Commun*, 4:2013:4:2883, 2013.
53. M Nishiga, T Horie, Y Kuwabara, K Nagao, O Baba, T Nakao, T Nishino, D Hakuno, Y Nakashima, H Nishi, F Nakazeki, Y Ide, S Koyama, M Kimura, R Hanada, T Nakamura, T Inada, K Hasegawa, S J Conway, T Kita, T Kimura, and K Ono. MicroRNA-33 controls adaptive fibrotic response in the remodeling heart by preserving lipid raft cholesterol. *Circ Res*, 120(5):835–847, 2017.
54. M A Lewis, E Quint, A M Glazier, H Fuchs, M H De Angelis, C Langford, S van Dongen, C Abreu-Goodger, M Piipari, N Redshaw, T Dalmay, M A Moreno-Pelayo, A J Enright, and K P Steel. An ENU-induced mutation of miR-96 associated with progressive hearing loss in mice. *Nat Genet*, 41(5):614–618, 2009.
55. A Mencía, S Modamio-Hoybjør, N Redshaw, M Morín, F Mayo-Merino, L Olavarrieta, L A Aguirre, I del Castillo, K P Steel, T Dalmay, F Moreno, and M A Moreno-Pelayo. Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nat Genet*, 41(5):609–613, 2009.
56. S Kuhn, S L Johnson, D N Furness, J Chen, N Ingham, J M Hilton, G Steffes, M A Lewis, V Zampini, C M Hackney, S Masetto, M C Holley, K P Steel, and W Marcotti. miR-96 regulates the progression of differentiation in mammalian cochlear inner and outer hair cells. *Proc Natl Acad Sci USA*, 108(6):2355–2360, 2011.
57. R K Herman and E M Hedgecock. Limitation of the size of the vulval primordium of *Caenorhabditis elegans* by *lin-15* expression in surrounding hypodermis. *Nature*, 348(6297):169–171, 1990.
58. B Fromm, T Billipp, L E Peck, M Johansen, J E Tarver, B L King, J M Newcomb, L F Sempere, K Flatmark, E Hovig, and K J Peterson. A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human microRNAome. *Annu Rev Genet*, 49:213–242, 2015.
59. N Pinzón, B Li, L Martinez, A Sergeeva, J Presumey, F Apparailly, , and H Seitz. microRNA target prediction programs predict many false positives. *Genome Res*, 27(2):234–245, 2017.

-
60. M Horlacher, G Cantini, J Hesse, P Schinke, N Goedert, S Londhe, L Moyon, and A Marsico. A systematic benchmark of machine learning methods for protein-RNA interaction prediction. *Brief Bioinform*, 24(5):bbad307, 2023.
 61. T M Nguyen, A Shafi, T Nguyen, and S Draghici. Identifying significantly impacted pathways: a comprehensive review and assessment. *Genome Biol*, 20(1):203, 2019.
 62. D P Bartel. Metazoan microRNAs. *Cell*, 173(1):20–51, 2018.
 63. A Ureta-Vidal, L Ettwiller, and E Birney. Comparative genomics: genome-wide analysis in metazoan eukaryotes. *Nat Rev Genet*, 4(4):251–262, 2003.
 64. R C Friedman and C B Burge. MicroRNA target finding by comparative genomics. *Methods Mol Biol*, 1097:457–476, 2014.
 65. I Becam, N Rafel, X Hong, S M Cohen, and M Milán. Notch-mediated repression of bantam miRNA contributes to boundary formation in the *Drosophila* wing. *Development*, 138(17):3781–3789, 2011.
 66. A R Bassett, G Azzam, L Wheatley, C Tibbit, T Rajakumar, S McGowan, N Stanger, P A Ewels, S Taylor, C P Ponting, J L Liu, T Sauka-Spengler, and T A Fulga. Understanding functional miRNA-target interactions in vivo by site-specific genome engineering. *Nat Commun*, 5:4640, 2014.
 67. J T Mendell and E N Olson. MicroRNAs in stress signaling and human disease. *Cell*, 148(6):1172–1187, 2012.