



Computational prediction of miRNA/mRNA duplexomes at the whole human genome scale reveals functional subnetworks of interacting genes with embedded miRNA annealing motifs

Claude Pasquier, Alain Robichon

► To cite this version:

Claude Pasquier, Alain Robichon. Computational prediction of miRNA/mRNA duplexomes at the whole human genome scale reveals functional subnetworks of interacting genes with embedded miRNA annealing motifs. Computational Biology and Chemistry, 2020, 88, pp.107366. <10.1016/j.combiolchem.2020.107366>. <hal-02999731>

HAL Id: hal-02999731

<https://hal.science/hal-02999731v1>

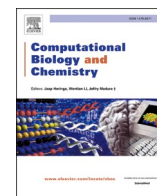
Submitted on 11 Nov 2020

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.



HAL Authorization



Review Article

Computational prediction of miRNA/mRNA duplexomes at the whole human genome scale reveals functional subnetworks of interacting genes with embedded miRNA annealing motifs

Claude Pasquier^{a,*}, Alain Robichon^b

^a Université Côte d'Azur, CNRS, I3S, France

^b Université Côte d'Azur, INRAE, CNRS, ISA, France

ARTICLE INFO

Keywords:

miRNA
mRNA collection
RNA annealing
subclasses of mRNA

ABSTRACT

Perfect annealing between microRNAs (miRNAs) and messenger RNAs (mRNAs) was computationally searched at a broad scale in the human genome to determine whether theoretical pairing is restrictively represented in functional subnetworks or is randomly distributed. Massive RNA interference (RNAi) pairing motifs in genes constitute a remarkable subnetwork that displays highly genetically and biochemically interconnected genes. These analyses show unexpected repertoires of genes defined by their congruence in comatching with miRNAs at numerous sites and by their interconnection based on protein/protein interactions or proteins regulating the activity of others. This offers insights into the putatively coregulated homeostasis of large networks of genes by RNAi, whereas other networks seem to be independent of this regulatory mode. Genes accordingly defined by theoretical RNAi pairing cluster mainly in subnetworks related to cellular, metabolic and developmental processes and their regulation. Indeed, genes harboring numerous potential sites of hybridization with miRNAs are highly enriched with GO terms depicting the abovementioned processes and are grouped in a subnetwork of genes that are significantly more highly connected than they would be according to a random distribution. The significant number of interacting genes that present numerous potential comatches with miRNAs suggests that they may be under the control of the integrative and concerted action of multiple miRNAs.

1. Introduction

miRNAs usually act in the regulation of mRNAs by pairing with complementary sequences; most of their targeting takes place in the 3'-untranslated region (3' UTR) of mRNAs, which initiates degradation via the Dicer/Ago machinery and/or prevents translation via interference with the mRNA reading process [Bartel, 2004 and 2009; He and Han-non, 2004]. Authors have proposed that 60 % of human protein-coding genes may theoretically have maintained pairing with miRNAs, and many of the genes within this catalog have been found to be experimentally controlled by miRNAs [Friedman et al., 2009; Lewis et al., 2005]. This suggests that miRNAs intervene in most signaling pathways and molecular processes in a cell. On the side of biogenesis, the pre-miRNA precursor molecules are transferred to the cytoplasm by the Exportin 5/RAN-GTP transport system and are then processed by the endonuclease Dicer, which releases mature miRNA duplexes. One strand of the miRNA duplex, loaded into the Argonaute protein (AGO) within

the silencing complex (miRISC), is used to select the matching mRNAs for degradation [Snead and Rossi, 2012; Snead et al., 2013; Sun et al., 2008]. Approximately 2000 miRNA genes have been identified in the human genome [Kozomara and Griffiths-Jones, 2014]. The observations that a particular miRNA can hypothetically target hundreds or thousands of different mRNAs and that an individual mRNA can inversely be targeted by hundreds of miRNAs [Agarwal et al., 2015] shed light on the role of miRNAs in shaping very large gene regulatory networks.

Approximately 40 % of human miRNAs are located in clusters within the genome, and the others seem to be randomly distributed [Kozomara and Griffiths-Jones, 2014; Agarwal et al., 2015; Tan et al., 2019]. On the other hand, a total of 591 miRNAs have been calculated to be located within genes, referred to as "host" genes; as consequence, some intragenic miRNAs are positively correlated with their host genes in terms of their levels of expression [Tan et al., 2019]. Although the effective regulation of mRNAs can be demonstrated and assessed by the existence of pairing with miRNAs, their overall regulation of large networks likely

* Corresponding author.

E-mail address: claude.pasquier@univ-cotedazur.fr (C. Pasquier).

<https://doi.org/10.1016/j.compbiolchem.2020.107366>

Received 27 July 2020; Received in revised form 18 August 2020; Accepted 19 August 2020

Available online 22 August 2020

1476-9271/© 2020 The Authors.

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

involves many secondary and downstream events, such as gene-gene interactions, which complicates the analysis of the real contribution of miRNAs in the global landscape of gene network equilibrium [Tan et al., 2019; Gennarino et al., 2009; He et al., 2012; Franca et al., 2016]. When miRNAs are inserted in a host gene, their coexpression is usually guided by a unique promoter. The landscape appears even more complicated than previously thought because of the discovery that intronic miRNAs may have their own independent intronic promoter, which suggests a double mode of expression regulation [Monteys et al., 2010; Ozsolak et al., 2008]. Moreover, in most cases, miRNAs have been reported to stimulate the expression of genes by binding directly to the promoter region and interfering with transcriptional regulatory regions [Place et al., 2008; Majid et al., 2010; Huang et al., 2012]. Although down-regulation by RNAi is the major effect that has been observed and quantified, positive correlations exist between the level of miRNAs and target genes in some cases as an indirect consequence of the down-regulation of an upstream suppressor [Tan et al., 2019]. Some individual miRNAs play a regulatory role for a large number of genes; for instance, miR-150 takes part in the regulation of approximately 1000 genes, and miR-150-dependent genes are nearly evenly distributed across the whole genome [Tan et al., 2019].

In this study, we focused on one major function of miRNAs: pairing with mRNAs as a trigger for Dicer/Ago degradation. We explored the broad-scale genome distribution of miRNA targets, addressing the questions of how many genes are targeted by a unique miRNA and how many genes are targeted by multiple miRNAs. In this report, a series of computational analyses demonstrate the existence of a discriminative distribution of genes presenting motifs for miRNA annealing and hot spots of miRNA/mRNA matches within subcategories of genes. Our analysis also revealed that the genes with multiple miRNA-annealing motifs present large discrepancies among GO categories related to biological processes pertaining to development, neurogenesis and differentiation. Furthermore, we identified a cohort of hundreds of genes that cluster within a very large network of interacting genes that share motifs for potential co-miRNA targeting. In parallel, we show that a very large gene network is fully independent of any miRNA action, which supports the paradigm of “layers” of genes that are selected and built through evolution over hundreds of millions of years, with their own network regulatory modes.

2. Methods

2.1. Computational programs used in this study

Reliable identification of miRNA targets is still a major challenge. Reporter assays and Western blotting are the most reliable methods for demonstrating direct interactions between an miRNA and its target but can still generate false positives [Kuhn et al., 2008]. In addition, the interactions that are supported by strong experimental evidence are limited. MiRTarBase database release 7.0 from September 15, 2017 [Chou et al., 2018], which includes all experimentally verified interactions, comprises the interactions between 735 miRNAs and 2766 targets obtained by reporter assays or Western blotting for *Homo sapiens*. This is far too small a scope for our purposes since, according to these data, the majority of genes are not considered to be targeted by miRNAs. To overcome this limitation, we use computationally predicted interactions, which is a strategy that has been employed by many other researchers. We base our study on the data provided by miRDB v6 [Chen and Wang, 2020], in which the results of a target prediction algorithm (MirTarget [Liu and Wang, 2019]) are improved with RNA-seq data. Each predicted interaction is associated with a probability score in the range of 0–100, which reflects the statistical confidence of the prediction. MiRDB includes all results with a score ≥ 50 , but the authors advise caution if the score is less than 60. They indicate that a prediction score greater than 80 is most likely to be real, and this is the threshold that we use.

2.2. Collection of miRNA-target interactions

The content of miRDB database v6 was downloaded using the link provided on the miRDB website (http://mirdb.org/download/miRDB_v6.0_prediction_result.txt.gz). The data are presented in a single file that lists for each miRNA its predicted targets and the associated score. For *Homo sapiens*, miRDB records more than 15 million interactions between 2656 miRNAs and 17,464 targets. After removing the predictions associated with a score below 80, 825,626 associations between 2638 miRNAs and 16,732 targets remained for *Homo sapiens*. As targets are identified by their NCBI gene IDs, we used the following file downloaded from the NCBI website to map the NCBI identifiers to gene symbols: <ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene2refseq.gz>. This file also allowed us to obtain a list of all the genes indexed by NCBI (which number 20,227); thus, by differentiating between this list and the list of genes referenced in miRDB, we identified 3,495 genes predicted to not be targeted by miRNAs.

2.3. Analysis of the distribution of miRNA target associations

With a Python script that we developed, we counted the predicted number of miRNAs that target each gene. This allowed us to draw the plot presented in Fig. 1 and to note that the shape of the distribution is typical of a power law function. We performed the same processing steps to determine the number of predicted targets for each miRNA (data are presented in Supplemental Table S2) and to draw the plot shown in Fig. 2. To study possible correlations between the susceptibility of genes to be controlled by miRNAs and their functions, we categorized the genes according to their degree of targeting by miRNAs. The 3495 genes that were not targeted by any miRNA constituted the first category. To avoid a population size bias, we defined the other categories such that each category contained a similar number of genes, i.e. a number of genes as close as possible to 3495. Thus, the second category was represented by genes that were very weakly targeted by miRNAs (i.e., targeted by between 1 and 5 miRNAs), which represents 3722 genes (Fig. 1a). The third category, illustrated in Fig. 1b, groups together 3303 weakly targeted genes (i.e., targeted by between 6 and 11 miRNAs). The fourth category consisted of 3409 genes targeted by between 12 and 21 miRNAs, which we refer to as moderately targeted genes (Fig. 1c). Continuing on this principle, we then identified 3211 highly targeted genes (i.e., targeted between 22 and 41 miRNAs) and the 3087 very highly targeted genes (i.e., targeted by more than 41 miRNAs), as illustrated in Fig. 1d and Fig. 1e. The details of the different categories of genes are given in Supplemental Table S1.

2.4. Gene ontology enrichment

GO enrichment analysis was performed for each category of genes using the enrichment analysis tool available at <http://geneontology.org/>. The enrichments obtained for each category were combined using a Python script to generate a table summarizing the results obtained for each GO term (Supplemental Table S3). The data for 15 terms considered to be the most representative are graphically depicted in Fig. 3.

2.5. PPI enrichment

To determine whether each of the categories includes genes that exhibit more interactions with each other than could be due solely to chance, using the data from the String database [Franceschini et al., 2012], we determined the number of corresponding proteins for each category and the number of interactions between the proteins in the same category associated with a high confidence score (minimum required interaction score ≥ 0.700). We performed the same procedure for 1000 randomly picked sets of genes of similar size and averaged the number of interactions found (Supplemental Table S4). The average

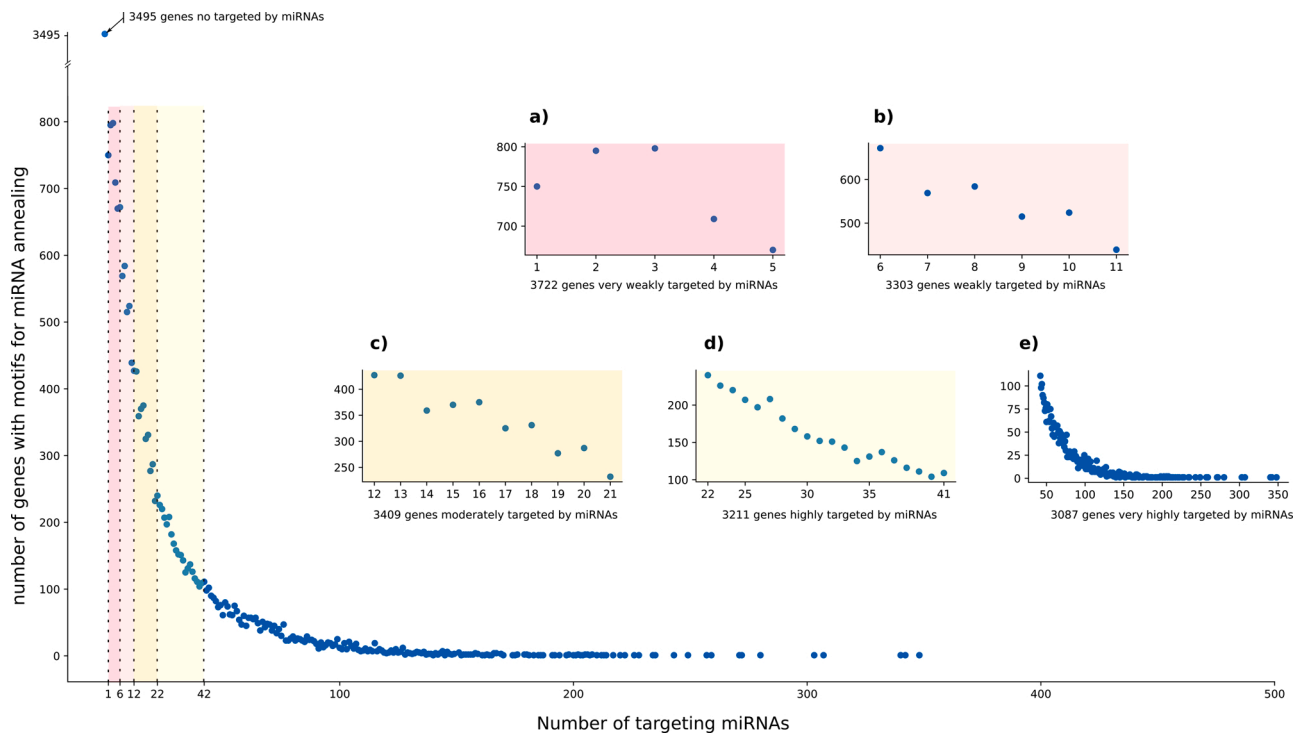


Fig. 1. Distribution of genes according to the number of miRNAs that target them. a) Magnified view of the genes very weakly targeted by miRNAs (i.e., targeted by between 1 and 5 miRNAs). b) Genes weakly targeted by miRNAs (i.e., targeted by between 6 and 11 miRNAs). c) Genes moderately targeted by miRNAs (i.e., targeted by between 12 and 21 miRNAs). d) Highly targeted genes (i.e., targeted by between 22 and 41 miRNAs). e) Very highly targeted genes (i.e., targeted by more than 41 miRNAs).

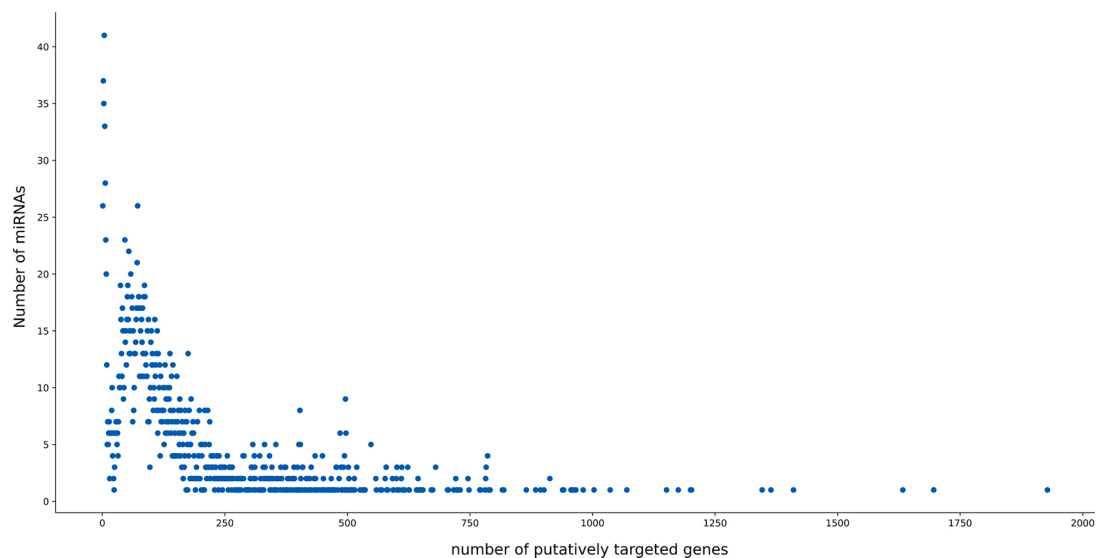


Fig. 2. Distribution of the number of genes targeted by miRNAs.

degree of the networks constructed from the genes in each category was significantly higher than that of random networks of the same size. To determine if this difference is sufficient to be considered representative, we computed the PPI enrichment p-value for the six categories as described by Franceschini et al. [Franceschini et al., 2012]. The obtained p-values were very low, well below $10E-16$, meaning that the genes in each category show more interactions among themselves than would be expected for a random set of genes of similar size. It can also be seen that the average degree correlates with the number of miRNAs that target genes in each category. This number ranges from 5.82 for highly targeted genes to 3.65 for nontargeted genes. However, although the

“nontargeted genes” network is less dense, it still presents more connections than a random network, and it is apparent from the visualization of this network using interaction data from the String database with a high confidence (minimum required interaction score ≥ 0.700) that there are modules of genes that seem to present a definite arrangement (Supplemental Figure S1).

2.6. Analysis of cotargeted genes

Among the genes that were most frequently targeted by miRNAs, some were targeted by the same set of microRNAs. To visualize these

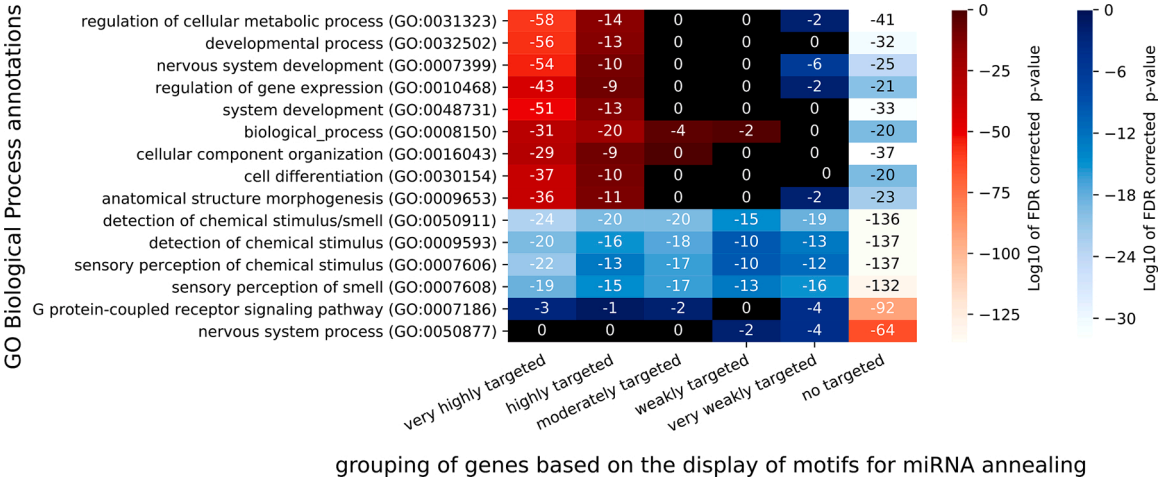


Fig. 3. Illustration of the differences in GO enrichment with biological process annotations obtained for the different categories of genes. Each table row represents a different GO term, and each column represents a different category of genes. The numbers in the cells are the log10 p-values with FDR correction of the enrichment of a category of genes with a specific GO term. Cells with a red background represent overrepresented annotations, while cells with a blue background represent underrepresented annotations. Black cells represent categories with no significant enrichment.

cotargeted genes, we built a network in which each node was a gene and a link existed between two genes if they were targeted by more than 50 shared miRNAs (Fig. 4a). From the genes in this network, we reproduced the known protein-protein interaction network using the interactions described in the String database [Franceschini et al., 2012] associated with a high confidence score (minimum required interaction score ≥ 0.700) (Fig. 4b). The interaction enrichment p-value was lower than $10E-16$. This means that the proteins encoded by the genes cotargeted by

more than 50 miRNAs exhibit more interactions among themselves (687 edges in the network) than what would be expected for a random set of proteins drawn from the genome (368 edges).

3. Results

The analysis of the computationally predicted miRNA-target interactions throughout the human genome was conducted to determine

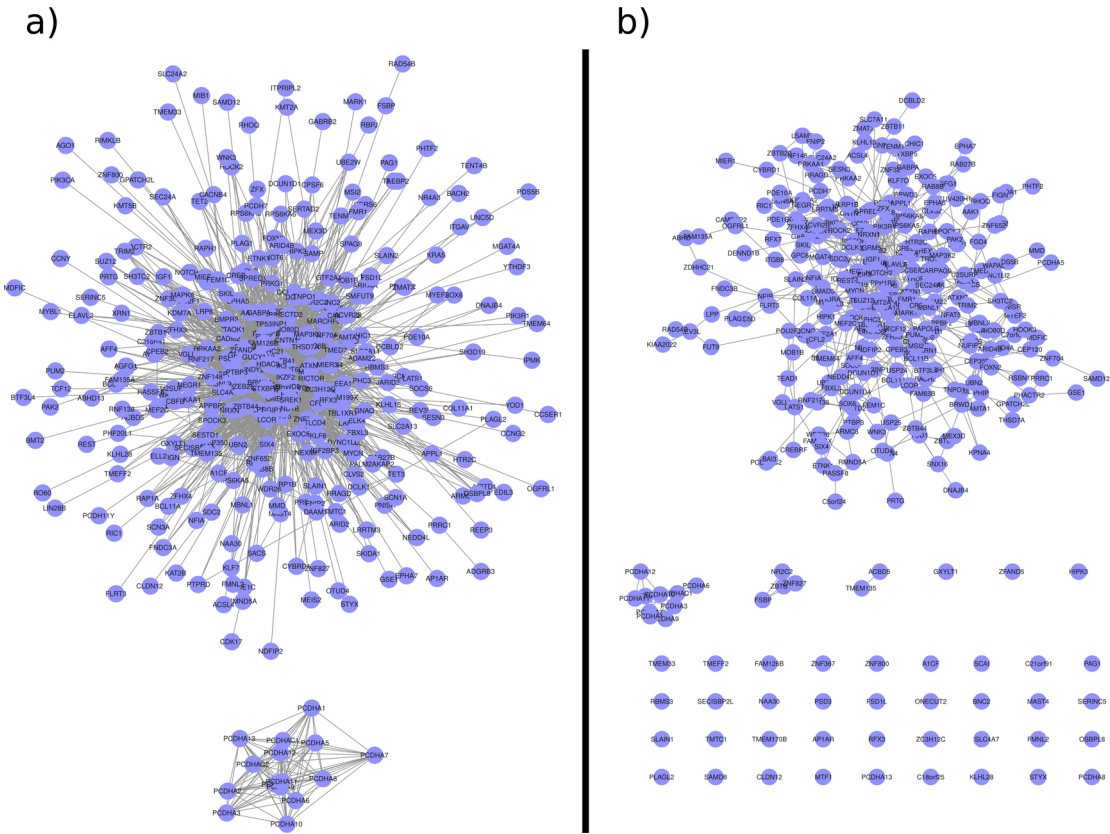


Fig. 4. a) Network of cotargeted genes. In this network, each node is a gene, and a link exists between two genes if they are targeted by more than 50 shared miRNAs. b) PPI network for the same set of genes as in panel a, built using the interactions described in the String database associated with high confidence (minimum required interaction score ≥ 0.700).

the distribution of these pairings. The results are summarized in Fig. 1 and Supplemental Table S1. While roughly one-sixth of the genes were not targeted by miRNAs, a number of genes were targeted by hundreds of miRNAs (approximately 450 genes showed potential pairing with more than one hundred miRNAs). In parallel, we determined the reverse distribution, i.e., the number of miRNAs per gene (Fig. 2 and Supplemental Table S2). As shown in Fig. 1, a power law distribution was observed, indicating that few genes were hypothetically targeted by a large number of miRNAs with different combinatorial assortments for each gene in this category. In this study, we evaluated the correlation between the number of matching sequences with miRNAs within genes and the functions of these genes. To this end, we partitioned the genes into six categories of roughly equal size that included potentially very highly targeted, highly targeted, moderately targeted, weakly targeted, very weakly targeted and nontargeted genes (see the methods section for details about the composition of each group). We performed a comprehensive GO enrichment analysis of these different categories of genes by distinguishing the biological functions that appeared more frequently than expected by chance and those that were underrepresented (Supplemental Table S3). Considering the functions that were related to these 6 subdivisions of miRNA target gene populations, some striking differences were observed. The two extreme subgroups of genes (the genes with the most motifs for miRNA annealing and nontargeted genes) presented an inverse distribution of GO terms, showing overrepresentation in one case and underrepresentation in the other (Fig. 3, Supplemental Table S3). The genes that were potentially highly targeted by miRNAs were positively enriched in functions related to development, neurogenesis and differentiation. Genes related to nervous system development (GO:0007399), developmental processes (GO:0032502), system development (GO:0048731), cell differentiation (GO:0030154), and the regulation of gene expression (GO:0010468) were very overrepresented among the genes that were very highly theoretically targeted by miRNAs. For instance, the probability (with false discovery rate - FDR correction) of obtaining the same number of genes related to nervous system development (GO:0007399) by chance is below $10E-54$. Genes annotated with the same GO terms were overrepresented in the second category (highly targeted genes), with an FDR lower than $1E-10$. The 3 central categories (moderately targeted, weakly targeted and very weakly targeted genes) contained a number of genes related to development, neurogenesis and differentiation, in line with their average occurrence throughout the genome. Conversely, the most overrepresented functions among the genes that did not match any miRNAs were related to sensory perception, particularly the sensory perception of smell, G protein-coupled receptors and synaptic processes (Fig. 3). Following the same pattern, these functions were expectedly underrepresented in the genes that were most targeted by miRNAs. Supplemental Table S3 lists all of the enriched terms identified per subgroup of miRNA annealing motifs with attributed FDR values. Overall, the most significant enrichment was observed for genes that were either potentially targeted by a large number of miRNAs or were not targeted at all. Genes that were moderately to very weakly targeted by miRNA showed a distribution that was relatively in line with a normal distribution of genes in the genome. However, this does not mean that these genes are randomly distributed. The evaluation of the interactions between the genes of each category showed that the obtained networks were significantly more connected than would be the case for randomly selected groups of genes of the same size (Supplemental Table S4). The genes that were potentially the most targeted by miRNAs exhibited the most interactions with other genes in the same category (genes belonging to the category of 'highly targeted genes', showing 5.82 interactors on average). This number decreased as the potential targeting of the genes by miRNAs decreased. However, the density of connections, even for nontargeted genes, was much greater than for a randomly picked set of genes (Supplemental Table S4). In addition, even in the network of the interactions of the nontargeted gene category, modules of genes that seemed to present a definite

arrangement could be observed (Supplemental Figure S1). This suggests that functional classes of genes depend on a large number of miRNAs and recognition motifs within the genes that potentially control them. The density of miRNA/transcript annealing motifs was correlated with development, neurogenesis, and differentiation genes and was inversely correlated with chemical detection genes (Fig. 3). However, gene expression and biochemical interactions vary among developmental times and stimuli in tissues and cell types, suggesting the rewiring of regulatory networks. These limits of the whole-genome computational search approach not distinguishing between the expressed and nonexpressed candidate genes were overcome by the observation of strong discrepancies in the distribution of miRNA recognition motifs within the genes assembled in association with the GO terms. This suggests that the overall approach adequately assessed the role of miRNAs in the global organization of integrative networks. Future investigations that consider the variable integrative role of miRNAs across tissues and developmental stages may offer a new perspective regarding how the corresponding networks are coregulated at the broad-scale genome level. An issue that is more complicated to address is that we might still assume that each cell type has a history imprinted by differences in epigenetic wiring, in which miRNA/mRNA pairing plays a prominent role as a new class of driver that has been unexplored in the framework of our current knowledge.

Furthermore, the actual paradigm of a large panel of genes that are coorganized and orchestrated within a network of interacting genes was expanded to include the miRNA component acting as an extra layer of coregulation. The pairing motifs within the network and with the full miRNA catalog allowed us to assess the miRNA contribution in some subnetworks and not in others. As an example, a gene network was constructed based on a minimum of 50 comatches with the same miRNAs between two genes. The network architecture was composed of a large number of nodes according to the selection term; i.e., two nodes linked by a vertex refer to two genes sharing at least 50 identical motifs theoretically annealing with different miRNAs (Fig. 4a). All of these comatching genes (two by two) constitute the bricks for subsequently building a gene network using interactions stored in the String database with high confidence. The graphical results are presented in Fig. 4b. The PPI enrichment p-value of this network is below $10E-16$, meaning that these genes exhibit more interactions among themselves than would be expected for a random set of genes of similar size drawn from the genome. The two overlapping networks (the network of genes with miRNA motifs and the network of interacting genes taken from the former network) indicate that in this large group of genes, biochemical/genetic interactions are likely orchestrated by a layer of miRNAs acting at the network level. The statistical enrichment of the density of theoretical miRNA pairing sites allowed us to assess the relative contributions of the miRNA machinery in the interactive networks. This finding argues in favor of a networking role as a whole that would remain ignored by restricting molecular analysis to individual genes.

4. Discussion

Most known miRNAs exhibit conserved sequences in closely related mammalian species such as humans and mice. Intriguingly, roughly one-third of *C. elegans* miRNAs have close homologs among human miRNAs. For instance, the let-7 family has four members in *C. elegans* and at least 15 in humans but only one in *Drosophila* [Bartel, 2018]. miRNAs and associated proteins appear to be the most common type of complexes within cells. miRNAs are approximately 22 nucleotides (nt) long and inhibit protein synthesis by annealing with their seed region of 2–8 nt or with the 3'UTR of target mRNAs and may lead to the breakdown of mRNAs via the endoribonuclease activity of Ago after full perfect miRNA/mRNA annealing [Quevillon Huberdeau and Simard, 2019; Liu et al., 2004; Fabian and Sonenberg, 2012; Guo et al., 2010; Eichhorn et al., 2014]. Dicer is a key endoribonuclease enzyme that matures metazoan miRNA by cleaving double-stranded RNA from a precursor

molecule that presents a hairpin structure. The mature miRNA has a 5' phosphate and a 2 nt 3' overhang, characteristic of RNase III activity. The miRNA pathways of plants, fungi and animals appear to be biochemically indistinguishable in terms of mRNA silencing pathways and inhibitory translation [Quevillon Huberdeau and Simard, 2019; Liu et al., 2004; Fabian and Sonenberg, 2012; Guo et al., 2010; Eichhorn et al., 2014]. miRNAs are subsequently incorporated into a ribonucleoprotein complex known as the RNA-induced silencing complex (RISC), from which the “passenger” strand will be removed, leaving the “guide” mono-strand to mediate the downregulation of mRNA targets in situ. Argonaute proteins, which form the core of the RISC, are highly conserved 100-kDa proteins that contain a signature consisting of a PAZ domain (anchoring the single-stranded 3' end of small RNAs) and PIWI domains (conferring mRNA-cleaving “Slicer” activity through RNA-RNA duplexes) [Quenvillon and Simard, 2019; Liu et al., 2004; Fabian and Sonenberg, 2012; Guo et al., 2010; Eichhorn et al., 2014; Weinmann et al., 2009]. All human Ago isoforms bind both miRNAs and small interfering RNAs (siRNAs), but only Ago2 has been shown to carry out mRNA cleavage [Bartel, 2018; Quevillon Huberdeau and Simard, 2019]. Both siRNAs and miRNAs, when perfectly base-paired to target mRNAs, direct the cleavage of a single phosphodiester bond in the target mRNA via the “Slicer” activity of Ago in the RISC complex [Liu et al., 2004; Fabian and Sonenberg, 2012; Guo et al., 2010; Eichhorn et al., 2014]. The miRNA-guided gene repression system in bilaterian animals is associated with the TNRC6 protein, which interacts with PABPC. Then, the *de novo*-formed heterocomplex recruits either the PAN2–PAN3 deadenylase complex or the CCR4–NOT deadenylase complex to shorten the mRNA poly(A) tail and stop translation in both cases [Bartel, 2018; Quevillon Huberdeau and Simard, 2019]. Other proteins have been reported to associate with RISC and modify Ago cleavage properties [Weinmann, 2009]. In parallel, upon miRNA binding to the 3'-UTR of an mRNA target, the asymmetric miRNA/mRNA duplex induces the inhibition of translation initiation without cleavage activity [Bartel, 2018; Quevillon Huberdeau and Simard, 2019]. The miR/miR* duplexes observed in plants and animals are similar: they are ~22 nt long and show imperfect complementarity between the two strands and a 2 nt 3'-overhang after Dicer cleavage [Moran et al., 2017]. The degree of complementarity between an miRNA and its mRNA target is a determinant that predicts the mode of target repression. High complementarity promotes target cleavage by Ago, and restrictive seed matching (7 to 8-base sequences at one extremity of an miRNA) leads to translational inhibition by pairing with the 3' UTR of mRNAs, which is referred to as noncanonical pairing [Khorshid et al., 2013; Seok et al., 2016; Helwak et al., 2013; Djuranovic and Nahvi, 2012]. Conversely, many other mRNAs acting as “neutral targets” would escape downregulation by default in the absence of matching miRNAs. This strongly suggests that “anti-targets” avoid fortuitous pairing with the large number of miRNAs present in the cells where they are coexpressed via an evolutionary process. Under selective pressure over millions of years, complementarity that would dampen their expression has not arisen. Evolutionary selection appears to have successfully achieved the fine tuning of miRNA catalogs to maximize targeting specificity and avoid disastrous interference with other bystander genes [Bartel, 2009 and 2018; Quevillon Huberdeau and Simard, 2019]. Here, we expand a previous analysis to quantitatively evaluate the enrichment of miRNA/target gene networks on a genome-wide scale. Our approach integrates the sequence profiles of miRNAs (~2500 miRNAs) and their matching counterparts among RNAs to determine the degree of enrichment among different networks. Although recently developed high-throughput sequencing technologies have revealed that the catalog of miRNA and transcriptome expression profiles in tissues differ greatly [De Rie et al., 2017; Liao et al., 2010], we performed a systematic global computational search to assess the categorization of the identified networks. The purpose was to further decipher the theoretical contribution of miRNAs to the coorganization and coregulation of networks, whereas some other networks seem to function without miRNA annealing, likely through evolutionarily selective

processes. The analysis of GO term enrichment in miRNA motifs highlighted that neurogenesis, cell differentiation and morphogenesis, along with related metabolic processes, involve a strong component of miRNA pairing. Conversely, chemical detection organs, G protein-coupled receptors and synaptic process seem to not show an miRNA contribution. This overall discrepancy between GO annotation categories was reinforced by the construction of a network in which two genes linked by a vertex presented 50 comatching miRNAs. This network overlapped with the network of the same genes constructed on the basis of proven genetic or biochemical interactions. Overall, the data highlight an ancestral role of miRNAs in the organization large gene networks, whereas other large panels of genes appear to be insensitive and unrelated to miRNA interference.

Our computational search led to highly discriminatory GO categories selected by miRNA/mRNA matching terms. A challenge in miRNA research revealed by the computational construction of networks will be to evaluate the specific component governed by miRNA interactions within interconnected gene networks within the temporal cascade; all of these integrated individual events orchestrate gene homeostasis within a network. Different algorithms have been developed for bioinformatic searches to predict miRNA target genes, limited by the fact that bioinformatic methods may provide false-positive pairing candidates because these molecules might or might not be coexpressed in the same cell [Singh, 2017; Stanhope et al., 2009; Tarang and Weston, 2014]. One of the most sensitive computational scoring tools, developed by Weijun Liu and Xiaowei Wang in 2019, has allowed the systematic retrieval of a large catalog of candidates in several vertebrates (human, mouse, rat, dog and chicken) [Lim et al., 2003; Lee et al., 2003]. We relied on the results of this tool, stored in the mirDB database, to perform our analyses [Lee et al., 2003]. Noncanonical pairing might undermine powerful prediction tools, and additional experiments will be needed to validate the prediction results.

CRedit authorship contribution statement

Claude Pasquier: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **Alain Robichon:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by an ANR grant, “Methylclonome” ANR-12-BSV6-006-01, to Alain Robichon and Claude Pasquier. This work was also supported by the French National Research Agency (ANR) through the LABEX SIGNALIFE program (reference # ANR-11-LABX-0028-01).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.compbiolchem.2020.107366>.

References

- Agarwal, V., Bell, G.W., Nam, J.W., Bartel, D.P., 2015. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 4, e05005.
- Bartel, D.P., 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297.

- Bartel, D.P., 2009. MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233.
- Bartel, D.P., 2018. Metazoan MicroRNAs. *Cell* 173, 20–51.
- Chen, Y., Wang, X., 2020. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.* 48 (D1), D127–D131.
- Chou, C.H., Shrestha, S., Yang, C.D., Chang, N.W., Lin, Y.L., Liao, K.W., Chiew, M.Y., 2018. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* 46 (D1), D296–D302.
- De Rie, D., Abugessaisa, I., Alam, T., Arner, E., Arner, P., Ashoor, H., Astrom, G., Babina, M., Bertin, N., Burroughs, A.M., et al., 2017. An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat. Biotechnol.* (2017) 35, 872–878.
- Djuranovic, S., Nahvi, A., 2012. Green R. miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* 336, 237–240.
- Eichhorn, S.W., Guo, H., McGeary, S.E., Rodriguez-Mias, R.A., Shin, C., Baek, D., Hsu, S. H., Ghoshal, K., Villén, J., Bartel, D.P., 2014. mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Mol. Cell* 56, 104–115.
- Fabian, M.R., Sonenberg, N., 2012. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat. Struct. Mol. Biol.* 19, 586–593.
- Franca, G.S., Vrbancovski, M.D., Galante, P.A., 2016. Host gene constraints and genomic context impact the expression and evolution of human microRNAs. *Nat. Commun.* 7 (11438).
- Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Jensen, L.J., 2012. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41 (D1), D808–D815.
- Friedman, R.C., Farh, K.K., Burge, C.B., Bartel, D.P., 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19, 92–105.
- Gennarino, V.A., Sardiello, M., Avellino, R., Meola, N., Maselli, V., Anand, S., et al., 2009. MicroRNA target prediction by expression analysis of host genes. *Genome Res.* 19, 481–490.
- Guo, H., Ingolia, N.T., Weissman, J.S., Bartel, D.P., 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466, 835–840.
- He, L., Hannon, G.J., 2004. MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* 5, 522–531.
- He, C., Li, Z., Chen, P., Huang, H., Hurst, L.D., Chen, J., 2012. Young intragenic miRNAs are less coexpressed with host genes than old ones: implications of miRNA-host gene coevolution. *Nucleic Acids Res.* 40, 4002–4012.
- Helwak, A., Kudla, G., Dudnakova, T., Tollervey, D., 2013. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 153, 654–665.
- Huang, V., Place, R.F., Portnoy, V., Wang, J., Qi, Z., Jia, Z., et al., 2012. Upregulation of Cyclin B1 by miRNA and its implications in cancer. *Nucleic Acids Res.* 40, 1695–1707.
- Khorshid, M., Hausser, J., Zavolan, M., Van Nimwegen, E., 2013. A biophysical miRNA/mRNA interaction model infers canonical and noncanonical targets. *Nat. Methods* 10, 253–255.
- Kozomara, A., Griffiths-Jones, S., 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–73.
- Kuhn, D.E., Martin, M.M., Feldman, D.S., Terry, A.V., Nuovo, G.J., Elton, T.S., 2008. Experimental validation of miRNA targets. *Methods* 44, 47–54.
- Lee, P., Lim, Lau N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., David, P., Bartel, D.P., 2003. The microRNAs of *Caenorhabditis elegans*. *Genes Dev.* 17, 991–1008.
- Lewis, B.P., Burge, C.B., Bartel, D.P., 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120, 15–20.
- Liao, J.Y., Ma, L.M., Guo, Y.H., Zhang, Y.C., Zhou, H., Shao, P., Chen, Y.Q., Qu, L.H., 2010. Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 30 trailers. *PLoS One* 5, e10563.
- Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B., Bartel, D.P., 2003. Vertebrate microRNA genes. *Science* 299, 1540.
- Liu, W., Wang, X., 2019. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. *Genome Biol.* 20, 1–10.
- Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., Hannon, G.J., 2004. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305, 1437–1441.
- Majid, S., Dar, A.A., Saini, S., Yamamura, S., Hirata, H., Tanaka, Y., et al., 2010. MicroRNA-205-directed transcriptional activation of tumor suppressor genes in prostate cancer. *Cancer* 116, 5637–5649.
- Monteys, A.M., Spengler, R.M., Wan, J., Tecedor, L., Lennox, K.A., Xing, Y., et al., 2010. Structure and activity of putative intronic miRNA promoters. *RNA* 16, 495–505.
- Moran, Y., Agron, M., Praher, D., Technau, U., 2017. The evolutionary origin of plant and animal microRNAs. *Nat. Ecol. Evol.* 1 (27).
- Ozsolak, F., Poling, L.L., Wang, Z., Liu, H., Liu, X.S., Roeder, R.G., et al., 2008. Chromatin structure analyses identify miRNA promoters. *Genes Dev.* 22, 3172–3183.
- Place, R.F., Li, L.C., Pookot, D., Noonan, E.J., Dahiya, R., 2008. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc Natl Acad Sci U S A* 105, 1608–1613.
- Quevillon Huberdeau, M., Simard, M.J., 2019. A guide to microRNA-mediated gene silencing. *FEBS J.* 286, 642–652.
- Seok, H., Ham, J., Jang, E.S., Chi, S.W., 2016. MicroRNA target recognition: insights from transcriptome-wide noncanonical interactions. *Mol. Cell* 39, 375–381.
- Singh, N.K., 2017. miRNAs target databases: developmental methods and target identification techniques with functional annotations. *Cell. Mol. Life Sci.* 74, 2239–2261.
- Snead, N.M., Rossi, J.J., 2012. RNA interference trigger variants: getting the most out of RNA for RNA interference-based therapeutics. *Nucleic Acid Ther.* 22, 139–146.
- Snead, N.M., Wu, X., Li, A., Cui, Q., Sakurai, K., Burnett, J.C., Rossi, J.J., 2013. Molecular basis for improved gene silencing by Dicer substrate interfering RNA compared with other siRNA variants. *Nucleic Acids Res.* 41, 6209–6221.
- Stanhope, S.A., Sengupta, S., den Boon, J., Ahlquist, P., Newton Michael, A., 2009. Statistical use of argonaute expression and RISC assembly in microRNA target identification. *PLoS Comput. Biol.* 5 (9), e1000516.
- Sun, X., Rogoff, H.A., Li, C.J., 2008. Asymmetric RNA duplexes mediate RNA interference in mammalian cells. *Nat. Biotechnol.* 26, 1379–1382.
- Tan, H., Huang, S., Zhang, Z., Qian, X., Sun, P., Zhou, X., 2019. Pancancer analysis on microRNA-associated gene activation. *EBioMedicine* 43, 82–97.
- Tarang, S., Weston, M.D., 2014. Macros in microRNA target identification: a comparative analysis of in silico, in vitro and in vivo approaches to microRNA target identification. *RNA Biol.* 11, 324–333.
- Weinmann, L., Hock, J., Ivacevic, T., Ohrt, T., Mutze, J., Schwille, P., Kremmer, E., Benes, V., Urlaub, H., Meister, G., 2009. Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. *Cell* 136, 496–507.