Making use of transcription factor enrichment to identify functional microRNA-regulons

Pacôme B. Prompsy a,b,c,d, John Toubia e,f,g, Linden J. Gearing a,b, Randle L. Knight a,b, Samuel C. Forster a,b, Cameron P. Bracken e,g,h, Michael P. Gantier a,b,*

a Centre for Innate Immunity and Infectious Diseases, Hudson Institute of Medical Research, Clayton, Victoria 3168, Australia
b Department of Molecular and Translational Science, Monash University, Clayton, Victoria 3168, Australia
c CNRS UMR3244, Institut Curie, PSL Research University, Paris 75005, France
d Translational Research Department, Institut Curie, PSL Research University, Paris 75005, France
e Centre for Cancer Biology, SA Pathology and University of South Australia, Adelaide, South Australia 5000, Australia
f ACRF Cancer Genomics Facility, Centre for Cancer Biology, SA Pathology and University of South Australia, Frome Road, Adelaide, South Australia 5000, Australia
g School of Molecular and Biomedical Science, University of Adelaide, Adelaide, South Australia 5005, Australia
h Department of Medicine, University of Adelaide, Adelaide, South Australia 5005, Australia

A R T I C L E   I N F O

Article history:
Received 28 February 2021
Received in revised form 20 August 2021
Accepted 20 August 2021
Available online 25 August 2021

Keywords:
- microRNAs
- Transcription factors
- microRNA regulons
- Target prediction
- microRNA pathways

A B S T R A C T

microRNAs (miRNAs) are important modulators of messenger RNA stability and translation, controlling wide gene networks. Albeit generally modest on individual targets, the regulatory effect of miRNAs translates into meaningful pathway modulation through concurrent targeting of regulons with functional convergence. Identification of miRNA-regulons is therefore essential to understand the function of miRNAs and to help realise their therapeutic potential, but it remains challenging due to the large number of false positive target sites predicted per miRNA. In the current work, we investigated whether genes regulated by a given miRNA were under the transcriptional control of a predominant transcription factor (TF). Strikingly we found that for ~50% of the miRNAs analysed, their targets were significantly enriched in at least one common TF. We leveraged such miRNA-TF co-regulatory networks to identify pathways under miRNA control, and demonstrated that filtering predicted miRNA-target interactions (MTIs) relying on such pathways significantly enriched the proportion of predicted true MTIs. To our knowledge, this is the first description of an in-silico pipeline facilitating the identification of miRNA-regulons, to help understand miRNA function.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

microRNAs (miRNAs) are 18–26 nt short single stranded RNAs controlling messenger RNA stability and translation. Directly pertaining to their reliance on short complementary target sites to control messenger RNAs, miRNA-target interactions (MTIs) are very frequent with >10,000 genes putatively regulated per miRNA [1]. Early studies in the field established the importance of the 5’-end of miRNAs (known as the seed region) in target recognition, prompting the development of MTI prediction tools filtering target sites based on the quality of interactions with the seed region and the interspecies conservation of these regions [2]. Nonetheless, recent studies based on high-throughput sequencing of miRNA target sites have now shown that seed dependence is not essential for at least 20% of the target sites, underlining the potential bias of seed-based MTI predictions [3,4].

Although more than 70 strategies of functional MTI identification have been proposed in the past 15 years [5], the identification of true miRNA targets which modulate cell function remains difficult, questioning whether their role has often been overestimated [6]. Indeed, miRNA binding to its target does not necessarily relate to translational repression [7], and when it does, the impact on individual targets is generally less than two fold [8,9]. Nonetheless, miRNAs can have important regulatory activities on pathways due to their concurrent regulation of genes with converging function [10–12]. While approaches leveraging pathway analyses to identify the function of biologically relevant MTIs have been reported, they are also limited by the high number of pathways predicted for each microRNA [13,14].

https://doi.org/10.1016/j.csbj.2021.08.032
2001-0370© 2021 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
With recent FDA approval of the first three small interfering RNAs, the perspective of successful development of therapeutic miRNAs based on similar delivery strategies and already pursued in clinical trials is now more realistic than ever [15,16]. Yet, as illustrated with the case of miRNAs in the field of inflammation, the functional impact of miRNAs is most often context dependent [12]. As such, characterisation of miRNA function is usually inferred from that of a few selected targets that do not represent the broad network of genes modulated by a miRNA. With in vivo delivery of therapeutic miRNAs unlikely to be entirely specific to target cells, it is therefore essential to better define pathway-centric effects of miRNAs contemplated for clinical use.

In this work, we set out to define whether co-transcriptional regulation of miRNA targets could be used to stratify miR-regulons (i.e. a set of genes regulated by a miRNA), and prioritise the pathways they modulate to better define miRNA function. Relying on predicted miRNA and validated transcription factor (TF) targets, we demonstrate that TF co-regulations can be used to identify prevalent pathways. Critically, the use of tripartite miRNA-TF-pathway associations to filter predicted MTIs significantly enriched the proportion of true MTIs, while decreasing the predicted interactions by more than 85%. Finally, we demonstrate that our targeting predictions are associated with a decreased expression of target miRNAs, confirming the biological significance of our approach.

2. Material and methods

2.1. Regulatory network and pathway network

2.1.1. miRNA datasets

Experimentally validated MTIs were downloaded from miRTarBase v7.0 human database, which is a manually curated database of experimentally validated MTIs (based on luciferase assays, qPCRs, micro-arrays and high-throughput RNA sequencing approaches such as CLASH) [17]. To ensure that sufficient numbers of MTIs were available, we restricted our analyses to the use of miRNAs with a minimum of 75 MTIs, encompassing 660 miRNAs and 6,578 genes (representing a matrix of 99,375 interactions – Supplementary Table S1).

Predicted MTIs were sourced from miRDIP 4.1 [5] and Targetscan 7.2 [7]. miRDIP integrates MTI predictions across 30 different sources, including Targetscan, and ranks the confidence of those interactions into classes ranging from “very high” to “low” based on an integrative score. Both databases were first filtered to keep only the 660 miRNA present in the validation dataset, miRTarBase (Supplementary Table S1). miRDIP high MTIs resulted from the merging of rows marked “very high” and “high” confidence MTIs from miRDIP (82,199 MTIs for 639 miRNAs), while “low” confidence MTIs were taken from rows marked “low” from miRDIP. For step 3 of the miRSTATION analyses presented below we filtered total miRDIP MTIs supported by ≥ 5 sources and with an integrative score >0.1, to generate a dataset of 2.3 million MTI predictions, referred to as miR-DIP* predictions herein. For Targetscan, the MTIs used were derived from both the conserved and non-conserved human MTIs, which represented respectively 184 and 453 miRNAs. In Figs. 1 and 3 where miRDIP high/low or Targetscan MTIs were compared with miRTarBase MTIs, both predictive sets were filtered to have the same number of targets as in miRTarBase, keeping the top predictions (e.g. if mir-X had 100 MTIs in miRTarBase, only the best 100 MTIs for this miRNA were used from miRDIP high/low or Targetscan).

2.1.2. Transcription factor (TF) datasets

For the initial studies shown in Fig. 1A and 1B, experimentally validated TF interactions from the ENCODE project [18] were retrieved for the human genome as “.bigwig” files for hg19 and hg38. The files were transformed into “.bed” files relying on the bedGraph algorithm (https://genome.ucsc.edu/goldenpath/help/bedgraph.html). TF sites were restricted to those present in the 1500 base-pair region upstream of transcriptional start sites. For this purpose genome coordinates were normalised using Perl scripts and CrossMap (http://crossmap.sourceforge.net/) to convert genome coordinates from one assembly to another. The method resulted in the use of 124 TFs targeting a total of 15,653 genes for a total of 59,306 interactions (Supplementary Table S2).

For all other analyses, experimentally validated TF interactions were retrieved from the “ITfTargets” R package (https://github.com/slowkow/ITfTargets) that aggregates multiple validated TF interactions datasets. We selected four comprehensive TF-target interaction datasets: ENCODE [18], IFFP [19], TRED [20] and TRUST [21]. The aggregation of these TF target datasets resulted in the use of 1249 TFs targeting a total of 13,500 genes for a total of 86,070 interactions (Supplementary Table S3).

2.1.3. Pathway datasets

For the pathway dataset, we used MSigdb [22], a collection of annotated gene sets utilized for gene set analyses. Only the most general and functionally informative categories of pathways were selected: “C2_curated”, “C5_GO” and “Hallmark”. Additionally, we arbitrarily filtered out gene sets with low (below 30) or high (above 1000) gene numbers, reasoning that small gene sets would miss overlaps of TF and miRNA regulons, while the too large gene sets would be too unspecific for TF and miRNAs regulons. This gave us 21,319 pathways with 1.3 M interactions in total, with gene clustered into biological processes informing on their high-level functions. The database was retrieved through the MSigdb API accessible through R with the package “msigdbr”.

2.1.4. miRNA Selection of TArgeting by Transcriptional co-regulation (miRSTATION)

2.1.4.1. Step 1: TF enrichment analyses. TF target and miRNA target from miRTarBase, miRDIP high or low matrices were built with genes as row and regulators (i.e. TF or miRNA) in columns. Multiplication of the matrices was used to identify the number of common targets in each miRNA-TF combination. We then compared the proportion of TF targets in the miRNA-regulated gene sets to the proportion of the TF targets in the “genome background”, which is the number of unique genes in our dataset (i.e. 25,367). The Fisher’s exact test was used with the following alternative hypothesis, for each miRNA set of genes: the proportion of genes with both miRNA and TF sites is greater than the proportion of sites for this TF in the genome background. Associations having a p-value lower than 0.05 and with 5 or more common targets were considered as significantly enriched (Supplementary Table S4 - for miRDIP high).

2.1.4.2. Step 2: Pathway enrichment. Identification of pathways enriched in miRNA-TF associations was carried out next. Target matrices of pathways were built using the MSigdb pathway database as described for miRNA and TF matrices in Step 1. For each miRNA-TF association identified in Step 1, pathways significantly enriched in the miRNA and/or in the associated TF regulatory networks were searched separately. A pathway was considered as significantly enriched in the miRNA-TF association if it passed the Fisher’s exact test with an adjusted p-value lower than 0.1 in both regulatory programs. When miRNAs were associated with more than 1 TF, we selected the TF that had the lowest pathway enrichment p-value. This resulted in miRNA-pathway associations. This approach was chosen over calculating enrichment in common miRNA-TF targets, since this strategy strongly reduced gene num-
Fig. 1. Functional miRNA targets are enriched in transcription factor binding sites. (A) Cumulative histogram of miRNA counts, function of their number of targets, based on the functional MTIs from miRTarBase (Supplementary Table S1). 660 miRNAs with more than 75 targets were included in all our analyses – shown in blue. (B) The cumulative counts of miRNA networks co-regulated are shown for the top 15 TFs (out of 72). Above each bar, the rank of each TF based on its amount of its total targets is indicated. (C) Violin plot of the number of enriched TF per miRNA for each of the 660 miRNAs and their targets, predicted by each approach: miRDIP high (Predicted MTIs high confidence), miRDIP low (Predicted MTIs low confidence) and miRTarBase (Validated MTIs). (The lines in the violin plots represent the median – data based on Supplementary Table S4 for miRDIP high). Wilcoxon ranking tests between datasets are shown. **** p < 0.0001, ns: non-significant. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Results

3.1. Functional miRNA targets are enriched for co-regulations by transcription factor

Early in silico analyses of gene network co-regulation by miRNA-TF pairs have previously suggested a functional overlap between select miRNA targets and TF targets [23]. This indicates that certain miRNAs more specifically modulate TF activities, consistent with their effect in buffering protein expression from transcriptional bursts [24,25]. Since the previously identified miRNA-TF interactions exclusively relied on predicted miRNA targets and TF targets, we initially decided to revisit the frequency of miRNA-TF interactions relying on datasets of experimentally validated miRNA targets (miRTarBase V7 [17]) and TF targets (ENCODE [18]). miRTarBase compiles a vast array of experimentally derived miRNA-target interactions (MTIs), originating from both low-throughput assays (e.g. Western blots, 3’UTR luciferase reporter assays, RT-qPCR) and high throughput approaches (e.g. RNA-seq approaches following cross-linking immunoprecipitation, such as CLIP-seq, HITS-CLIP, etc.). Within the MTIs of miRTarBase, we selected 660 miRNAs displaying greater than 75 targets for detailed analysis, as lower thresholds did not increase the number of miRNAs significantly associated with TF binding (Fig. 1A, Supplementary Table S1 and Material and Methods; this represented 99,375 MTIs). TF binding site enrichment was calculated for the promoters of the targets of these 660 miRNAs, based on validated
3.2. Enrichment for co-regulations by transcription factor is more frequent for high-ranking predicted and validated miRNA targets

To define the biological relevance of TF enrichment in miRNA targets, we next tested whether it was influenced by the functional nature of the MTIs, i.e. whether the MTIs were true targets or not. For this purpose, we investigated how relying on in silico predicted MTIs (which have a higher proportion of non-functional MTIs than miRTarBase, as they strictly rely on predictions) would impact TF enrichment. To obtain more comprehensive analyses, we relied on validated binding sites of 1249 TFs from ENCODE (Supplementary Table S2), affording a larger dataset of TF-target sites than the ENCODE dataset itself (86,070 target sites with T Targets, versus 59,306 with ENCODE).

First, high-ranking MTI predictions from miRDIP (referred to as miRDIP high herein), which collates MTIs from across 30 different predictive resources [5], were analysed for TF enrichment, limited to the targets of the 660 miRNAs defined in Supplementary Table S1 (to allow for subsequent comparisons with miRTarBase below). The miRDIP high targets of 369 miRNAs were significantly associated with at least one TF, when compared to our “genome background” set (i.e. 55.9% of the 660 miRNAs analysed here) (the 369 miRNAs and associated enriched TFs are provided in Supplementary Table S4). 477 out of the 1249 TFs networks were enriched in the targets of the 369 miRNAs from miRDIP high, and enrichment of the TFs was not strongly correlated with the number of their targets in T Targets (Supplementary Fig. 1). Finally, 186 (50.4%) of the 369 miRNAs with enriched TFs displayed no more than two enriched TFs in their gene networks (Supplementary Table S4 and Supplementary Fig. 1–).

Second, we repeated the analyses described above using miRDIP low. Our rationale was that if the miRNA-TF co-regulations obtained previously with miRDIP high were only observed by chance, low-ranking predicted MTIs would be expected to yield similar numbers of enriched miRNA-TF associations, for the predicted targets of the 660 miRNAs studied above. Critically, using similar target population sizes for each one of the 660 miRNAs, significant TF enrichment was much lower in the targets predicted with a low miRDIP confidence score, compared to those with a high score – 70 of the miRNAs from miRDIP low had at least one associated TF, versus 186 for miRDIP high targets. We also re-ran our analyses with miRTarBase relying on T Targets. Direct comparison of the number of enriched TFs per miRNA between miRTarBase, miRDIP high and miRDIP low revealed that there were more enriched TF sites in validated miRTarBase targets and high-ranking predicted targets of miRDIP high, compared to lower ranking predicted targets of miRDIP low (Fig. 1C). However, there was no significant difference between the number of enriched TF associations from miRTarBase and miRDIP high. Collectively, these analyses supported the notion of functional overlap between select miRNA targets and TF targets previously suggested [23].

3.3. miRNA-TF associations overlap with specific pathways

Functional association of the targets of a given miRNA has previously been suggested by several groups as a strategy to help understand the biological activities of miRNAs at a network level. DIANA-miRPath [13] and miRPathDB [14,26] for instance allow the identification of enriched pathways in predicted and experimentally validated miRNA targets. Nonetheless, this approach is limited by the large number of enriched pathways it can identify for a miRNA network, making it hard to prioritise the most important pathways. Since we were able to identify a significant overlap between select miRNA and TF-regulated genes, we hypothesised that TF co-regulations could be informative to prioritise the most important pathways regulated by a miRNA.

This was tested by analysing MSigdb pathway enrichment for the targets of the 477 TFs selected previously, in parallel to pathway enrichment for miRDIP high MTIs of the 369 miRNAs (noting that some of the TFs targets considered in these enrichments were not miRNA targets, and vice versa). Based on the previously defined
miRNA-TF co-regulations identified above for miRDIP high, we next looked for pathways concurrently enriched in co-regulated miRNAs and TFs. This strategy naturally decreased the number of enriched pathways obtained for the targets of the miRNAs as shown in Fig. 2A with the example of the targets of miR-5582-3p (here from 93 enriched pathways, decreased down to 23 when restricting the pathways to those also enriched in AR [Androgen Receptor] regulated genes).

Relying on this approach, we further selected up to 3 highest ranking enriched pathways associated with each one of our 369 miRNAs (significantly decreasing the number of enriched pathways for each miRNA network from 83,359 pathways down to 1,105 pathways for 365 miRNAs at \( p < 0.1 \)) (see Material and Methods). Importantly, this analysis identified 552 unique pathways enriched over the 369 miRNAs, underlying a good diversity of pathway enrichment across the TFs, independent of the number of their target genes. Accordingly, filtering the pathways associated with miRNA regulons using the TF-pathway enrichment strategy lowered the relation between the total number of targets in a pathway and the number of miRNA targets enriched for this pathway, going from a Pearson’s correlation of 0.931 to 0.467 (Fig. 2B, 2C).

### 3.4. miR Selection of Targeting by Transcriptional co-regulation (miRSTATION)

To gain further insights into the biological relevance of the miRNA-regulated pathways identified above through TF prioritization, we next investigated whether these pathways could help enrich true positive MTIs from predicted MTIs, which have a high rate of false positives. Our motive was to try to broaden miRNA predictions to encompass a wide range of possible interactions that are omitted by phylogenetic-conservation and seed weighting, yet have biological relevance, while decreasing the number of false positives that low stringency predictions generate. As such, we reasoned that biologically relevant pathways might be used to filter in silico predicted MTIs, to enrich the number of true MTIs. For this purpose, we generated a pipeline, which we refer to as miRNA Selection of Targeting by Transcriptional co-regulation (miRSTATION), starting from 2.3 million miRDIP* MTI predictions with low stringency for the 369 miRNAs with enriched pathways associations under TF co-regulation (Fig. 3A). Restricting the predicted MTIs to those overlapping with enriched MSigdb pathways reduced the number of miRDIP* MTIs by 87.2%, from an average...
of 4436.8 down to 566.54 MTIs per miRNA (Fig. 3A and Supplementary Table S5). Critically, pathway filtering significantly enriched the proportion of true MTIs validated in miRTarBase for 291 out of 369 miRNAs assessed here (with 10 predicted targets – at confidence level of 0.95). This demonstrates that pathway filtering based on co-transcriptional regulations helps increase the proportion of true MTIs from miRDIP*. Accordingly, our pipeline led to the selection of 187,503 MTIs for 368 miRNAs and 721 pathways (with 10 targets), and 6,117 (3.26%) of these MTIs were present in miRTarBase (Fig. 3A and Supplementary Table S5).

In addition, the proportion of predicted MTIs present in miRTarBase was compared between miRSTATION, miRDIP high and Targetscan (best scores) for the 368 miRNAs – relying on the same number of MTIs per miRNA (see Material and Methods) (Fig. 3B, 3C). miRSTATION filtering of miRDIP* significantly increased the selection of true MTIs compared to miRDIP high both for precision (number of true positive on total number of predicted targets, Fig. 3B) and recall (number of true positive on total number of true positive, Fig. 3C). In addition, miRSTATION also significantly outperformed the precision of Targetscan, though it showed decreased recall (Fig. 3B, 3C). The different results seen between miRDIP high and Targetscan were unexpected since miRDIP high comprises predictions from Targetscan, but probably relate to the ranking system used in miRDIP that gives priority to the number of targets concurrently predicting an MTI. Collectively, these results supported the biological function of the tripartite miRNA-TF-pathway associations identified, since they could be used to enrich the proportion of predicted true MTIs.

3.5. miRSTATION predicted targets are negatively correlated with miRNA levels in several cancers

To validate further the capacity of our approach to identify biologically relevant MTIs in an unbiased manner, we next assessed the correlation between the expression levels of miRNAs and their MTIs, in datasets from The Cancer Genome Atlas (TCGA). Since miRNAs predominantly impact mRNA translation by decreasing mRNA levels [27], we were interested to test whether correlations between miRNA levels and those of the MTIs predicted by miRSTATION were more often negative correlations than positive ones.

First, we created a correlation analysis of miRNA and mRNA expression levels for each patient with Breast Cancer (BRCA), Prostate Adenocarcinoma (PRAD), Lung Adenocarcinoma (LUAD), Kidney renal clear cell carcinoma (KIRC), and Colon adenocarcinoma (COAD), in the selected datasets (having concurrent small and mRNA sequencing data available), restricted to the 368 miRNAs and the 187,503 MTIs identified by miRSTATION. We restricted our analyses to these 5 cancer types out of 32 from the TCGA on the basis that they had enough samples to obtain meaningful results. In parallel, we also created control analyses based on the MTIs predicted by miRDIP high, miRDIP low and Targetscan for these 368 miRNAs.

We subsequently compared the obtained correlations between the different prediction tools for the 368 miRNAs studied, for each of the 5 cancer groups [Supplementary Table S6 [miRSTATION], Table S7 [miRDIP high], Table S8 [miRDIP low] and Table S9 [Targetscan]]. Overall, Targetscan predictions performed the best and were more often negatively correlated with miRNA levels (as seen with the negative mean values) in each cancer type (Fig. 4A, 4B). miRSTATION predictions also performed relatively well with the exception of the analysis of LUAD, where its predicted MTIs were rather positively regulated with miRNA levels (as seen with the positive mean values – Fig. 4B). Critically, the correlations obtained between predicted MTIs and miRNA expression with miRSTATION significantly outperformed the ones obtained with miRDIP high and miRDIP low in all 5 cancer types (Fig. 4A, 4B), as revealed by the lower means obtained for miRSTATION in all cases. In addition, and further validating the relevance of this approach, the correlations obtained with miRDIP low were the worst performing and resulted in mean positive correlations between MTIs and miRNAs for 3 out of 5 cancers, while being greater than those obtained with the three other tools (Fig. 4A, 4B). This indicates that correlations between MTIs and miRNA expression levels analysed here were directly associated with the strength/quality of the MTIs considered, giving further support to the biological relevance of the MTIs predicted by miRSTATION.

Finally, we were interested to see whether the miRSTATION predicted regulons could be used to inform on miRNA function. As a proof of principle of this concept, we looked at mir-122-5p, which is a well-studied miRNA predominantly expressed in liver hepatocytes. We noted that in our miRSTATION analyses mir-122-5p was associated with the pathway “GOBP_RESPONSE_TO_CYTOKINE” from MSigdb (Supplementary Table S5). Importantly, this pathway/gene signature was not identified in the top three pathways enriched from the best MTIs of miRDIP high and Targetscan. Similarly, mir-122-5p was not predicted as an enriched miRNA for the pathway “Response To Cytokine” in mirPathBD [26].

Type-I interferons (IFN) are a family of cytokines which control expression of thousands genes, and are essential to the control of viral infections [28]. In patients, mir-122-5p intra-hepatic levels are negatively correlated with the expression of several type-I interferon (IFN) regulated genes (IRGs – STAT1, IFI27, CXCL10 and USP18), suggesting that mir-122-5p could regulate IFN responses [29]. Critically, we noted that miRSTATION predicted targeting of the IFN receptor 1 and 2 (IFNAR1 and IFNAR2) and JAK1, which are directly involved in the molecular complex sensing type-I IFN [30], along with STAT1 and STAT2 which transduce the signal to activate expression of genes exhibiting an IFN Stimulated Response Element (ISRE) in their promoter (Fig. 4C). Similarly, miRSTATION predicted targeting of many other antiviral IRGs including STAT3, OAS1, OAS2, OAS3, MX1, IFIT1, IFIT2, IFIT3, IFI16, CCL5, and RNASEL (Fig. 4C). The convergence of these predictions suggested that mir-122-5p could functionally regulate sensing of type-I IFN (by down-regulating the expression of the receptor complex), but also through direct targeting of key antiviral effectors of the pathway. Aligning with this, transfection of synthetic mir-122-5p significantly reduced IFN signalling, as measured by ISRE-luciferase reporter [31]. Conversely, inhibition of mir-122-5p in hepatic Huh7 cells (which have high levels of mir-122-5p) significantly enhanced IFN response, collectively suggesting that mir-122-5p does impact IFN responses [31].

4. Discussion

Until recently, the characterisation of miRNA function has predominantly been inferred from the function of a few of their target genes, with canonical seed targeting. While useful in some instances, this widespread target-centric approach has generally limited our understanding of miRNA function, which is overall very mild at the level of individual targets [8,9]. This has also probably led to the over-claiming of significance of specific target genes, whilst under-claiming the true effects of miRNAs as modulators of regulons and genetic networks. However, the discovery that miRNA-mRNA interaction sites could be captured in chimera reads by small RNA sequencing has revolutionised the landscape of miRNA-target interactions, leading to the identification of more than a hundred thousand miRNA target sites, often independent of seed targeting [3,32]. Although some of these miRNA-mRNA target sites may not functionally modulate expression of their target [7], we reasoned that large sets of miRNA-mRNA interactions derived from such high-throughput screens and other more dis-
crete approaches collectively compiled in miRTarBase, could offer a novel opportunity to perform unbiased characterisation of miRNA function.

Our analyses of the miRTarBase targets of 660 miRNAs suggest that for greater than 50% of these (i.e. 348 miRNAs) an overlapping layer of transcriptional regulation can be identified. This overlapping effect of miRNAs on mRNAs that are co-regulated by the same TFs supports the concept that miRNAs help buffer transcriptional programs [23–25]. Critically, there were more enriched miRNA-TF associations when using validated miRTarBase and high quality predicted MTIs from miRDiP* high, compared to poorer predicted MTIs from miRDiP low, underlining that these associations are not a mere result of chance.

Relying on the miRNA-TF associations identified for high quality predicted MTIs from miRDiP* high, we next proposed to stratify MSigdb pathways enriched amongst the targets of each one of the 369 miRNAs. As such, we hypothesised that concurrent regulation at transcriptional and posttranscriptional levels converged towards a specific cellular function, and that these co-regulations could help prioritise the prevalent functional networks regulated by a miRNA. This approach allowed us to select up to three MSigdb pathways significantly enriched for each mRNA network of targets. Most importantly, we show that the MSigdb pathways selected through this approach can be used to filter poorly stringent miRDiP* MTI predictions, and significantly enriched these in true positive MTIs (based on miRTarBase validated targets). It is also noteworthy that this approach significantly increased the selection of true MTIs compared to miRDiP high MTIs for both precision and recall. Since miRDiP high relies on high quality MTIs concurrently predicted by several different algorithms, and miRDiP* MTIs encompass those of miRDiP high, these findings suggest that co-transcriptional prioritisation of functionally relevant MTIs performs better than algorithms existing to date – noting that compared to TargetsCan, miRSTATION showed a decreased recall level. Importantly, pathway ranking based on TF co-regulations performed better than relying solely on the p-values of MSigdb pathways enrichment, for its capacity to enrich in true positive MTIs in miRDiP* (preliminary analyses not shown).

In addition to decreasing the amount of predicted MTIs by greater than 85% for the miRNAs studied, one important particularity of our approach is that it did not make any assumption on the type of miRNA-mRNA target sites considered – although our analyses were restricted to sites predicted by ≥ 5 different tools in miRDiP. Since several MTI prediction tools in miRDiP are predominantly relying on energy of miRNA-mRNA interactions, they can encompass non-canonical binding sites. This contrasts with current approaches that favour binding to the 5′-end region of the miRNA (the seed), the mRNA sequence context and also interspecies site conservation to filter the most relevant sites from the greater than 10,000 possible predictions otherwise available per miRNA, such as TargetsCan or DIANA-microT [7,33]. For this reason, our approach encompasses a very large set of possible miRNA target sites independent of their adherence to the canonical definition. In doing so we can capture miRSTATION validated MTIs that would not otherwise be considered. We also demonstrate that the expression of miRSTATION predicted targets is significantly more negatively correlated with miRNA expression levels, than that seen with predicted targets obtained with miRDiP high, relying on 5 cancers. This unbiased global analysis gives further support to the biological significance of our approach relying on enriched pathways to filter large numbers of MTIs, independent of the type of miRNA-mRNA interactions. Nonetheless, the TargetsCan MTIs performed best in this correlative analysis, which we attribute to the fact that its prioritisation is based on conservation of sites between orthologous species, and that such sites are associated with stronger repression of the target [8].

To our knowledge, this is the first demonstration that leveraging experimentally validated TF networks of genes can help prioritise pathways enriched in miRNA predicted targets, which in turn can be used to identify novel MTIs. We propose that this can help define a bird’s-eye view of the functional regulation of a miRNA based on a set of key regulated genes, the individual contribution of which may only be very limited to the overall miRNA function. As such, in addition to target genes that are under strong miRNA control and may individually contribute to another specific function of the miRNA, our approach has the capacity to define less obvious miRNA-regulons, with converging activities. We illustrate this concept with the identification of IFNAR1, IFNAR2, JAK1, STAT1 and STAT2 as potential targets of miR-122-5p with converging activity on the sensing of type-I IFNs. Although further experiments will be necessary to confirm direct targeting of these genes by miR-122-5p, the previous reports that modulation of miR-122-5p levels controlled the response to type-I IFNs [31], and that miR-122-5p levels were negatively correlated with IRG expression in human hepatocytes [29], do support the concept of a regulon with converging activity controlling type-I IFN responses, aligned with the “RESPONSE TO CYTOKINE” pathway. Importantly, miR-122-5p was not associated with this pathway relying on TargetsCan and miRDiP high predictions, or in miRPathDB or DIANA-miPath, illustrating the unique potential of miRSTATION.

In conclusion, we demonstrate here the feasibility to identify functional miRNA-regulon based on overlapping transcriptional and translational co-regulations. By avoiding standard MTI prediction biases, our approach represents a paradigm shift in the definition of miRNA targets and miRNA function. Although currently limited by the number of miRNAs targets present in miRTarBase (i.e. 660 miRNAs used to benchmark our analyses), and the use of miRDiP as a starting point in the MTIs selected, we demonstrate proof-of-principle that this approach can identify miRNA-regulons directly informing on miRNA function for 368 miRNAs. It will be interesting to test, in further analyses, whether some of the TF-miRNA co-regulatory networks identified through our approach can be further modulated through targeting of the TF itself by the miRNA it associates with, since predicted miRNA targets are enriched for TFs [34]. Collectively our findings should help the miRNA community move away from target-centric definitions of miRNA function, which have prevailed to date in the field [12].

Funding

This work was funded in part by the Australian National Health and Medical Research Council (1062683 to M.P.G. and 1159239 to S.C.F.); the Australian Research Council (FT140100594 Future Fellowship to M.P.G., FT190100544 Future Fellowship and DP190103333 to C.P.B.); and the Victorian Government’s Operational Infrastructure Support Program. Funding for open access charge: Hudson Institute of Medical Research.

CRediT authorship contribution statement

Pacôme B. Prompsy: Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing. John Toubia: Methodology, Formal analysis, Writing - review & editing. Linden J. Gearing: Formal analysis, Writing - review & editing. Randle L. Knight: Data curation. Samuel C. Forster: Formal analysis, Writing - review & editing. Cameron P. Bracken: Funding acquisition, Project administration, Supervision, Writing - review & editing. Michael P. Ganter: Conceptualization, Methodology, Funding acquisition, Project administration, Supervision, Writing - original draft, Writing - review & editing.
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.

Acknowledgement

We thank J. Revote and Monash eResearch for assistance with the servers used for our analyses; and F. Cribbin for help with the editing of this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.08.032.

References

[1] Friedmann RC, Burge CB. MicroRNA target finding by comparative genomics. Methods Mol Biol 2014;1097:457–76.
[2] Lewis BP, Shih I-H, Jones-Rhoades MW, Bartel DB, Burge CB. Prediction of mammalian microRNA targets. Cell 2003;115(7):787–98.
[3] Helwak A, Kudla C, Dudnalova T, Tollervey D. Mapping the human mRNA interactome by CLASH reveals frequent noncanonical binding. Cell 2013;153(3):654–65.
[4] Grosswendt S, Filipchyk A, Manzano M, Kliounnos M, Schilling M, Herzog M, et al. Unambiguous identification of mRNA-target site interactions by different types of ligation reactions. Mol Cell 2014;54(6):1042–54.
[5] Tokar T, Pastrello C, Rossos AEM, Abovsky M, Hauschild AC, Tsay M, et al. Ambiguous target prediction programs predict many false positives. Genome Res 2017;27(2):234–45.
[6] Agarwal, V., Bell, G.W., Nam, J.W. and Bartel, D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. Elife, 4.
[7] Selbach M, Schwanhausser B, Thierfelder N, Fang Z, K汉语 R, Rajewsky N. Widespread changes in protein synthesis induced by microRNAs. Nature 2008;455(7209):58–63.
[8] Baek D, Vilién J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. Nature 2008;455(7209):64–71.
[9] Peng KJ, Halberg C, Yoshida M, Tavazoie SF. A microRNA regulon that mediates endothelial recruitment and metastasis by cancer cells. Nature 2012;481(7380):190–4.
[10] Gantier MP, Stundon HJ, McCoy CE, Belheke MA, Wang D, Kaparakis-Liakos M, et al. A miR-19 regulon that controls NF-kappaB signaling. Nucleic Acids Res 2012;40:4084–58.
[11] Nejad C, Stundon HJ, Gantier MP. A guide to miRNAs in inflammation and innate immune responses. FEBS J 2018;285(20):3695–716.
[12] Vlachos IS, Zagganas K, Paraskevopoulos MD, Georgakilas G, Karagkouni D, Vergoulis T, et al. DIANA-miRPath v3.0: deciphering microRNA function with experimental support. Nucleic Acids Res 2015;43:W460–466.
[13] Backes C, Kehl T, Stöckel D, Fehlmann T, Schneider L, Meese E, et al. miRPathDB: a new dictionary on microRNAs and target pathways. Nucleic Acids Res 2017;45(D1):D90–6.
[14] Mullard, A. (2018) FDA approves landmark RNAi drug. Nat Rev Drug Discov, 17, 613.
[15] Chang MM, Bahali R, Rasmussen TP, Manautou JE, Zhong X-B. The growth of miRNA-based therapeutics: Updated clinical studies. Biochem Pharmacol 2021;114432.
[16] Chou, C.H., Shrestha, S., Yang, C.D., Chang, N.W., Lin, Y.L., Liao, K.W., Huang, W. C., Sun, T.H., Yu, S.J., Lee, W.H. et al. (2018) miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic Acids Res, 46, D296-D302.
[17] Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57–74.
[18] Zheng G, Yu K, Yang Q, Xiong Y, Wei C, Xie L, et al. ITTFP: an integrated platform of mammalian transcription factors. Bioinformatics 2008;24(20):2416–7.
[19] Jiang C, Xuan Z, Yao F, Zhang MQ. TRED: a transcriptional regulatory element database, new entries and other development. Nucleic Acids Res 2007;35(Database):D137–40.
[20] Han H, Shin H, Shin D, Shim JE, Ko Y, Shin J, et al. TRRUST: a reference database of human transcriptional regulatory interactions. Sci Rep 2015;5(1):13.
[21] Liberezon A, Beger C, Theovalsdottir H, Ghandi M, Mieusot JP, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. Cell Systems 2015;1(6):417–25.
[22] Zhou Y, Ferguson J, Chang JT, Kluger Y. Inter- and intra-combinatorial regulation by transcription factors and microRNAs. BMC Genomics 2007;8(1):396. https://doi.org/10.1186/1471-2164-8-396.
[23] Schmediel Jörn M, Klemm SL, Zheng Y, Sahay A, Blüthgen N, Marks DS, et al. Gene expression. MicroRNA control of protein expression noise. Science 2015;348(6230):126–32.
[24] Ebert MS, Sharp PA. Roles for microRNAs in conferring robustness to biological processes. Cell 2012;149(3):515–24.
[25] Kehl T, Kern F, Backes C, Fehlmann T, Stöckel D, Meese E, et al. miRPathDB 2.0: a novel release of the miRNA Pathway Dictionary Database. Nucleic Acids Res 2020;48:D142–7.
[26] Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 2010;466(7308):835–40.
[27] Rusinova I, Forster S, Hu S, Kannan A, Masse M, Cummimg H, et al. INTERFEROME v2.0: an updated database of annotated interferon-regulated genes. Nucleic Acids Res 2012;41:D140–6.
[28] Sarasin-Filipowicz M, Krol J, Markiewicz I, Heim MH, Filipowicz W. Decreased levels of microRNA miR-122 in individuals with hepatitis C responding poorly to interferon therapy. Nat Med 2009;15(1):31–3.
[29] Zanin N, Viaris de Lesegno C, Lamae C, Blouin CM. Interferon Receptor Trafficking and Signalling: Journey to the Cross Roads. Front Immunol 2021;11.
[30] Yoshikawa T, Takata A, Otsuka M, Kishikawa T, Kojima K, Yoshida H, et al. Silencing of microRNA-122 enhances interferon-α signaling in the liver through regulating SOCS3 promoter methylation. Sci Rep 2012;2(1):3.
[31] Pacôme B. Prompsy, J. Toubia, L.J. Gearing et al. Computational and Structural Biotechnology Journal 19 (2021) 4896–4903