

# Using Context-Sensitive Text Mining to Identify miRNAs in Different Stages of Atherosclerosis

Markus Joppich<sup>1</sup> Christian Weber<sup>2</sup> Ralf Zimmer<sup>1</sup>

<sup>1</sup> Department of Informatics, LFE Bioinformatics, Ludwig-Maximilians-Universität München, Munich, Germany

<sup>2</sup> Institute for Cardiovascular Prevention, Ludwig-Maximilians-Universität München, Munich, Germany

Address for correspondence Markus Joppich, MSc, LFE Bioinformatics, Department of Informatics, Ludwig-Maximilians-Universität München, Amalienstr. 17, Munich, Bavaria 80333, Germany (e-mail: joppich@bio.ifi.lmu.de).

Thromb Haemost 2019;119:1247–1264.

## Abstract

790 human and mouse micro-RNAs (miRNAs) are involved in diseases. More than 26,428 miRNA–gene interactions are annotated in humans and mice. Most of these interactions are posttranscriptional regulations: miRNAs bind to the messenger RNAs (mRNAs) of genes and induce their degradation, thereby reducing the gene expression of target genes. For atherosclerosis, 667 miRNA–gene interactions for 124 miRNAs and 343 genes have been identified and described in numerous publications. Some interactions were observed through high-throughput experiments, others were predicted using bioinformatic methods, and some were determined by targeted experiments. Several reviews collect knowledge on miRNA–gene interactions in (specific aspects of) atherosclerosis.

Here, we use our bioinformatics resource (atheMir) to give an overview of miRNA–gene interactions in the context of atherosclerosis. The interactions are based on public databases and context-based text mining of 28 million PubMed abstracts. The miRNA–gene interactions are obtained from more than 10,000 publications, of which more than 1,000 are in a cardiovascular disease context (266 in atherosclerosis). We discuss interesting miRNA–gene interactions in atherosclerosis, grouped by specific processes in different cell types and six phases of atherosclerotic progression. All evidence is referenced and easily accessible: Relevant interactions are provided by atheMir as supplementary tables for further evaluation and, for example, for the subsequent data analysis of high-throughput measurements as well as for the generation and validation of hypotheses. The atheMir approach has several advantages: (1) the evidence is easily accessible, (2) regulatory interactions are uniformly available for subsequent high-throughput data analysis, and (3) the resource can incrementally be updated with new findings.

## Keywords

- ▶ chemokines
- ▶ endothelial cells
- ▶ macrophage
- ▶ atherosclerosis

## Introduction

In almost any area of biomedical research the published knowledge is overwhelming. Scientific reviews try to give an overview over a more or less specific field of interest.

*Note: The review process for this paper was fully handled by Gregory Y. H. Lip, Editor-in-Chief.*

Even for domain experts delivering a good review requires a lot of effort and almost always a restrictive selection, assessment, and compilation of the established scientific facts. On the other hand, a review is maybe not as helpful for everyone because the selection is too distant from the needs of the individual researcher who wants to employ a review to set own research into context of the state-of-the-art.

received

March 14, 2019

accepted after revision

May 14, 2019

© 2019 Georg Thieme Verlag KG  
Stuttgart · New York

DOI <https://doi.org/10.1055/s-0039-1693165>.  
ISSN 0340-6245.

We investigate whether an automated approach can address both issues: it should not only help to write a compelling and up-to-date review but also to allow the user to obtain all the established facts related to the reviewed field of interest. For the case of miRNAs, it is clear that many miRNAs have been reported to play a role in almost any process under some condition. Moreover, in almost all cases an interaction between a miRNA and some target(s) are reported to be important. Thus, inherently, in this field long lists of findings have to be reviewed and any of these findings can yield a relevant hypothesis for the question investigated by the researcher or measured via some high-throughput technique. Often researchers want to check whether their brand new finding is really new and surprising and whether and how it adds to the established knowledge of the field. Thus, it is important to provide convenient access to facts, hypotheses, and the associated evidence in an as complete as possible way.

Another problem in any field of study, of course, is the dynamics of scientific progress. Continuously, findings are added to the published literature, such that incremental updates of reviews and the associated evidence should be mandatory to achieve the above-mentioned goals. Otherwise, review articles quickly become outdated.

Hundreds of miRNAs have been reported to be relevant for atherosclerosis via thousands of specific interactions with target genes. We have collected and derived context-based miRNA–target interactions from public databases and by sensitive context- and ontology-based text mining of PubMed abstracts.

miRNAs have often been identified as important posttranscriptional regulators<sup>1,2</sup> associated to various stages of complex human diseases, like diabetes<sup>3</sup> or cardiovascular diseases.<sup>4,5</sup> For atherosclerosis research, in particular, several relevant miRNAs have been identified and brought into context.<sup>6,7</sup> It has been found that miRNAs modulate the function of endothelial cells, smooth muscle cells (SMCs), and macrophages by controlling the expression levels of chemokines.<sup>8</sup>

Several reviews examine the interaction of miRNAs and genes in different stages and processes of atherosclerosis. Andreou et al<sup>9</sup> reviewed atherogenic and protective miRNAs in several processes and disease stages. Chemokines play a crucial role during initiation and progression of atherosclerosis. Besides homeostatic functions, chemokines have essential functions in leukocyte recruitment and govern the infiltration with mononuclear cells and macrophage accumulation. Hartmann et al<sup>8</sup> summarize miRNA–chemokine interactions in the context of atherosclerosis.

We use the interactions identified by Andreou et al and Hartmann et al in their reviews as benchmark to test the performance of our context-sensitive text mining approach, *atheMir*. Then, we increment the networks and stage-specific disease contexts by facts from *atheMir*. Thereby, we provide a data-driven collection of current knowledge of miRNA–gene interactions and miRNA–gene interactions in atherosclerosis. This is done via highly sensitive text mining on all PubMed (<https://www.ncbi.nlm.nih.gov/pubmed>) abstracts. Moreover, we add more context information to reflect the

fact that miRNA regulation is context-dependent<sup>10</sup>: (1) we distinguish between general interactions found by some high-throughput methods, which might be interesting to generate hypotheses for specific atherosclerosis experiments, and results from PubMed abstract text mining; (2) we add information on known regulations in specific cell types, which might be interesting to transfer to other cell types, especially as it is hypothesized that various cell types can differentiate into each other in specific conditions and disease stages<sup>11</sup> (3) we add condition-specific information as this could hint at particular disease states in which these regulations can be active; and (4) we retrieve information about miRNAs active in several cell types as this can hypothesize common regulatory mechanisms.

Additionally, we use the collected information from databases and text mining together with gene-regulatory networks proposed for a sequence of disease stages of atherosclerosis and the underlying Causal Biological Networks (CBNs).<sup>12</sup> These six networks, similar to the stages of atherosclerosis, summarize important genes per stage, which we use to highlight possible context-dependent changes of miRNA regulations of chemokines and other proteins. To achieve this, we enhance these networks of regulatory hypotheses with miRNA–gene interactions characterizing the respective phase and showcasing the relevant key players in each successive phase and in each relevant cell type. Every edge in the network is associated with the supporting evidence and context such that it can be utilized in the respective research question.

## Materials and Methods

Our approach, *atheMir*, collects information from public databases, recent reviews, and performs context-sensitive text mining on public PubMed abstracts. We use curated sets of synonyms derived from domain-specific ontologies for genes, diseases, species, cell types, experimental contexts, functional classes, and pathways. After detecting cooccurrences of synonyms for objects from these ontologies, we apply natural language processing techniques to analyze the sentence structure and to extract miRNA–gene interactions, as well as the associated context. This way we produce a list of miRNA–gene interactions which can afterwards easily be queried and restricted to specific diseases, for example, atherosclerosis, and specific contexts, for example, endothelial cells and processes in specific late atherosclerosis stages such as plaque destabilization. The list of miRNA:target relations has been extracted from public miRNA databases and from PubMed abstract via text mining. Text mining has been performed on the complete PubMed corpus (September 2018), which consists of 28,787,497 abstracts.

Our automated *atheMir* approach is evaluated via a detailed assessment using two reviews of the field. After evaluating the general performance of our data-driven review process, we focus on miRNA–gene interactions in specific cell types and specific stages of atherosclerosis thereby aiming at a more (context) specific view on active miRNA regulations.

### Gene and miRNA Vocabulary

For gene and miRNA vocabulary, controlled synonym lists are obtained from HUGO Gene Nomenclature Committee (HGNC)<sup>13</sup> (human) and Mouse Genome Informatics (MGI)<sup>14</sup> (mouse). Applying a Named Entity Recognition (NER) approach for finding interactions (using *syngrep*<sup>15</sup>), we rely on a good vocabulary. The vocabulary should be chosen such that it is exact, commonly used, easily maintainable, and mostly not ambiguous. Both the human HGNC and mouse MGI gene lists fulfill these criteria. Both lists provide approved and previous gene symbols as well as gene names, common synonyms, and name synonyms. For cross-species transfer, mouse gene identifiers are mapped on the human gene identifiers by gene symbol using mapping lists created from homologous gene lists provided by BioMart.<sup>16</sup>

The HGNC gene list contains 45,467 gene entries. For 43,518 gene entries synonyms are given. On average, about six different synonyms are provided per gene, totaling to 264,850 synonyms. For mouse, the corresponding gene file from MGI contains 66,933 entries. Synonyms are provided for 16,470 entries. On average, each of these gene identifiers has about 7 synonyms, 113,187 synonyms in total.

Of all PubMed abstracts, 6,373,603 abstracts mention at least one human gene and 7,578,721 a mouse synonym, respectively.

The miRNA vocabulary is derived from miRBase.<sup>17</sup> miRNAs are mentioned in 53,896 abstracts and miRNAs together with human genes are found in 40,312 abstracts of these, or 40,957 in mouse, respectively.

### Context-Based Text Mining

Each abstract is categorized with respect to five dimensions: species (human, mouse), disease (from disease ontology<sup>18</sup>), cell line (from Cellosaurus<sup>19</sup>), protein function (from National Cancer Institute Thesaurus<sup>20</sup>), and gene ontology.<sup>21</sup>

The context of a miRNA–gene interactions is defined by the document it is found in, thus for the list of recognized features per dimension, in the document. A document has a specific feature, if a synonym of this feature (e.g., ontology terms) was found. For miRNA identifiers, also the organism prefix is evaluated, that is, miRNAs hsa-miR-98 and mmu-miR-124 establish a human and mouse context, respectively. While this would also be possible for human/mouse genes, we have refrained from doing so, because frequently the only difference between human and mouse genes is the capitalization.

The second dimension of the context of a miRNA interaction is the disease. We extract disease synonyms as controlled vocabulary from the disease ontology. The Gene Ontology (GO) is used to categorize abstracts by molecular function, cellular compartment, and biological process. Finally, the NCIT ontology is subset for proteins grouped by function, to search for protein classes (e.g., cytokines or chemokines).

For each vocabulary of a dimension, derived from ontologies, common (English) words and manually curated words are excluded from the synonym lists. In general, taxonomic identifiers, cell names, and disease names are excluded, if

these are not relevant for the specific dimension. For example, creating the gene synonyms, taxonomic identifiers, cell names and disease names are excluded to avoid ambiguities. Common words leading to ambiguities within the same dimensions are also excluded.

### Aggregation of Text Mining Results

Using the controlled vocabularies, the PubMed abstracts are scanned for occurrences of synonyms. Building the *attheMir* database, the text mining result has to be summarized and aggregated.

With the custom gene and miRNA lists, we perform NER on these abstracts using an in-house tool, *syngrep*.<sup>15</sup> Using a sentence-based cooccurrence of a miRNA and a gene object with connecting relation (interaction verb), we identify a total of 41,930 miRNA–gene interactions within 36,505 PubMed abstracts (note that only 0.1% of all and 0.5% of human PubMed abstracts contain a valid miRNA interaction).

For a valid interaction, an appropriate verb phrase has to be identified as target verb connecting miRNA and gene. If both gene and miRNA are mentioned in an enumeration (without a connecting verb), the interaction is discarded. To detect the relation, we use *spaCy*<sup>22</sup> to analyze the grammatical structure of the sentence. *spaCy* has been shown to be quite accurate and it is one of the fastest tools available.<sup>23</sup> As a complete analysis of the PubMed including text mining and detection of valid interactions takes less than 12 h (on a laptop computer), this allows updating the text mining regularly. *spaCy* has several pretrained models available for analyzing sentences. Here, we make use of the *en\_core\_web\_1* g neural-network model. Using such models, *spaCy* can build dependency trees for full sentences. This allows to use the semantic structure of a sentence to accept or reject a miRNA–gene interaction (► **Supplementary Fig. S1**, available in the online version). For this dependency tree, we reconstruct the path from the target word (gene or miRNA) up to the root element, which in this case is also the connecting verb, confirmed. This path is called the stack and is further analyzed. For *KLF12*, the stack is *KLF12, confirmed*. For *miR-34a* the stack is: *miR-34a, 5p, of, targets, as, confirmed*. This analysis also shows the problems of natural language processing. The 5p part of *miR-486-5p* is detected as a separate noun and is thus returned as a separate element within the stack. This stack ends at the root element of the sentence, too. Following the stack construction, we perform three analyses.

First, we compare whether the stack contains any connecting verbs by intersecting the stacks of the gene and the miRNA. If a verb is found in the intersection, it is ensured that the sentence structure either resembles subject (S)–verb (V)–object (O) (SVO) or OVS, where gene and miRNA must be subject and object, or vice versa. A slightly weaker criterion for detecting an interaction, but required due to the aforementioned problems in natural language processing regarding the detection of the sentence structure, particularly with miRNAs, is the following. For every verb in the whole sentence it is checked whether the gene and miRNA are subject or object of the given verb. In this case, the only verb

detected which fulfills this condition is again *confirmed*. Finally, to exclude cases in which miRNA and gene are mentioned in an enumeration, for example, in the sentence “It has been shown that macrophages can communicate with endothelial cells via ICAM1 and miR-98” (► **Supplementary Fig. S2**, available in the online version), enumerations are also analyzed. We detect that ICAM1 and miR-98 are contained both in the same enumeration and particularly note that there is no connecting verb. Thus, a possible miR-98:ICAM1 interaction is thus rejected, here.

In general, the remaining dimensions are aggregated within *atheMir* independently from found interactions, because the abstract classification by features is independent from interactions. Thus, dimensional information exists for a PubMed abstract even if not all context features could be identified (e.g., missing disease). After combining all dimensions, *atheMir* provides information in at least one dimension for 22,286,667 PubMed articles, of which 8,813,329 have annotated diseases and 36,955 have miRNA–gene interactions.

### Additional Databases

In addition to text mining results, further databases with experimental data have been integrated into *atheMir*. These include interactions from miRTarBase,<sup>24</sup> miRecords,<sup>25</sup> and DIANA-TarBase.<sup>26</sup> Expression values for miRNAs in specific cells are provided by the FANTOM5 project.<sup>27</sup>

The CBN database provides gene-regulatory networks for several (biological) processes.<sup>12</sup> The CBN database contains network models for different processes and diseases in human, mouse, or rat. Among the 169 networks (version 1 + 2), 6 networks are specific to cardiovascular diseases. These 6 CBN model different stages of atherosclerosis development: (I) Endothelial cell activation, (II) Endothelial cell-monocyte interaction, (III) Foam cell formation, (IV) Plaque destabilization, (V) Platelet activation and (VI) Smooth muscle cell activation.

## Results and Discussion

### Benchmarking and Assessment Using Atherosclerosis Reviews

To assess *atheMir* with respect to precision, sensitivity, and false discovery rate (FDR), we use interactions from published reviews by Andreou et al<sup>9</sup> and Hartmann et al.<sup>8</sup> We restrict *atheMir* to context-specific miRNAs for several processes in atherosclerosis. In particular, we compare the *atheMir* miRNA–gene interactions with those described in endothelial cells (Andreou et al,<sup>9</sup> Fig. 1), those involved in plaque destabilization (Andreou et al,<sup>9</sup> Fig. 2), and miRNAs involved in the regulation of the initiation, progression, and thrombotic complications (Andreou et al,<sup>9</sup> Table 1).

Based on this assessment of *atheMir* for miRNAs in atherosclerosis, we more specifically analyze the miRNA interactions with chemokines in atherosclerosis. For this, we compare interactions identified herein with the review by Hartmann et al.<sup>8</sup>

We assess the sensitivity and precision of *atheMir* via a systematic comparison with (standard-of-truth as estab-

lished by) these reviews. We evaluate further whether the text mining problems mentioned above hamper our approach and goals, and show that our approach can recapitulate current knowledge and provide added value. The assessment also indicates that the approach can be used in other contexts and may also be successfully applied to other fields/diseases of interest.

For each stage from the CBNs, we identify the interacting miRNAs and the regulating processes. Thus, for each stage we identify the important cell types and the relevant miRNA–gene interactions in these cell types.

Ideally, we would see a sequential progression of the identified processes through these six stages. But, for this, the low resolution of the stages poses a problem. Also, only few relevant key players have been identified in each stage so far, and many miRNAs are involved in many stages. For many interactions, the necessary context has not yet been established. And, lastly, the text mining methods are not perfect and may produce both false positive and false negative hits even if relevant interactions have been described in the literature. Particularly the low resolution is problematic, since the six networks model both spatial (cell–cell migration and differentiation) and temporal development of various cells and tissues over years. Thus, we cannot expect a real time series of miRNA regulation in atherosclerosis from this resource.

### atheMir Database

The *atheMir* database contains text mining interactions for 6,244 genes and 1,375 miRNAs. A total of 26,428 interactions between these genes and miRNAs are recorded. Of these, 19,679 interactions are in a disease context. A total of 2,242 of these interactions are associated with cardiovascular system diseases (DOID:1287) or atherosclerosis (DOID:1936). In the atherosclerosis context, *atheMir* contains 643 miRNA–gene interactions from text mining. The number of PubMed abstracts per dimension, including additional databases, is listed in ► **Supplementary Table S1** (available in the online version).

### miRNAs Relevant in Atherosclerosis

For a first assessment of *atheMir*, all miRNAs involved in atherosclerosis-specific processes and respective cell types (derived from the processes defined by Andreou et al<sup>9</sup>) are analyzed (► **Supplementary Table S8**, available in the online version). Any miRNA–gene interaction must have been detected in the atherosclerosis (DOID:1936) context (► **Supplementary Table S8**, search parameters in ► **Supplementary Table S6**, available in the online version).

For SMCs (proliferation/migration), many missed miRNAs have been observed. Also for the cell types in angiogenesis, monocyte differentiation/macrophage activation, and cholesterol efflux, several miRNAs are missed. A miRNA is missed, if the gold standard (here Andreou et al review<sup>9</sup>) lists this miRNA, but it is not found by *atheMir*. We investigate why these miRNAs are missed, considering the wide search in our database. We used the keywords given in the review<sup>9</sup> to perform a manual search in PubMed, for example, for miR-378 in the angiogenesis context atherosclerosis



angiogenesis miR-378. Similarly, we evaluated the missed miRNAs interactions of miR-125 in the T cell differentiation and activation context. For most of the missed miRNAs, such manual searches returned no results.

The found miRNAs show that on the one hand an automated approach can retrieve many relevant miRNAs for specific processes in a disease context, but may miss some. This might be because information is not easily accessible from PubMed abstracts: some reported miRNAs might be involved in more general processes, which are not specific to atherosclerosis, and therefore the atherosclerosis keyword is not included in the specific abstracts.

Another reason for not finding relevant miRNAs in *atheMir* is its focus on miRNA–gene interactions—if an article mentions a miRNA in a specific context without a gene, it is not included in *atheMir*. In some cases, the vocabulary used may not be sufficient to detect specific diseases or processes. Furthermore, the interactions could only be present in the full text, which we currently do not analyze with text mining.

For the eight atherosclerotic processes defined by Andreou et al,<sup>9</sup> we provide an overview of the accepted, missed, and additional miRNAs (► **Supplementary Table S8**, available in the online version).

### Endothelial miRNAs Implicated in Atherosclerosis<sup>9</sup>

Analyzing Fig. 1 from Andreou et al,<sup>9</sup> we augment the identified interactions with *atheMir* (► **Table 1**). The search parameters for the *atheMir* query are outlined in ► **Supplementary Table S5** (available in the online version). For most interactions, corresponding literature was found.

For all missed interactions (e.g., miR-146:IRAK2), we find an interaction without the atherosclerosis context. For instance, the miR-31:SELE (E-selectin) interaction is only reported in cancer,<sup>34</sup> yet mentioned in the review. Also, the missed interactions miR-126:ETS1 and miR-17:ICAM1 can be explained: the cited literature only refers to endothelial cells in general, and does not mention any disease. Finally, the miR-146a/b:IRAK1/2 interactions are not mentioned in the abstract of the cited literature. Among the accepted interactions for KLF4 is also its interaction with miR-103.<sup>35</sup>

### miRNAs Implicated in Atherosclerotic Plaque Destabilization<sup>9</sup>

Similar to the previous benchmark, we also checked Fig. 2 from the Andreou et al review.<sup>9</sup> The results are very similar (► **Supplementary Table S2**, available in the online version). Some interactions are missed, because they are not (yet) reported in an atherosclerotic context. More specifically, the miR-29 and MMP2/3/9/13/14 interactions are only found in rotator cuff tears,<sup>36</sup> but no atherosclerosis-specific interactions are found in/by PubMed.

### Regulation of the Initiation, Progression, and Thrombotic Complications of Atherosclerosis by miRNAs in Mice<sup>9</sup>

In the same fashion as before table 1 by Andreou et al<sup>9</sup> has been analyzed (► **Table 1**). All interactions are found in the atherosclerosis context.

### Systematic Evaluation

For the endothelial miRNAs implicated in atherosclerosis (► **Table 1**) and regulation of the initiation, progression, and thrombotic complications of atherosclerosis by miRNAs in mice (► **Table 1**), we manually checked all interactions to systematically evaluate our text mining method for sensitivity  $\frac{TP}{TP+FN}$ , precision  $\frac{TP}{TP+FP}$ , and FDR  $\frac{FP}{TP+FP}$ . Furthermore, we calculate the  $F_1$  score as  $\frac{2 \cdot \text{sensitivity} \cdot \text{precision}}{\text{sensitivity} + \text{precision}}$ . The evaluations are presented in ► **Table 2**. Sensitivity and precision are best for the comparison of *atheMir* and PubMed. On the one hand, this is little surprising. Our approach has the same input as the PubMed search. Thus, the number of true predicted elements should be equal to the true positives. This shows that our context-based search does neither miss nor add too many interactions.

The comparison between Andreou et al and *atheMir* mostly lacks precision. This is due to *atheMir* finding many additional interactions. However, since these have been manually checked, it allows us to combine the ground truth from the Andreou et al review regarding sensitivity and the manual curation from PubMed regarding precision. The resulting  $F_1$  scores of 0.75 and 0.97 for the combined analyses show that *atheMir* can reliably be used for miRNA–gene interaction mining. For the endothelial miRNAs evaluation, the low  $F_1$  scores originate from many miRNA–gene interactions which we could not find in atherosclerosis using manual PubMed search.

### Chemokine-Specific Networks

In the previous section, *atheMir* could replicate and increment the presented networks. Here, we want to focus on the specific miRNA–chemokine interactions for endothelial cells and macrophages. Using *atheMir*, we extract a network of all miRNA–chemokine interactions (► **Fig. 1**, no other context) and increment the existing network from Hartmann et al (► **Fig. 2**). There are 742 interactions, 234 of these are derived from DIANA-TarBase only. Interestingly, only 20 interactions are recorded by both DIANA-TarBase and miRTarBase. The intersections of results from any experimental database, DIANA-TarBase, miRTarBase, and miRecords, with PubMed is relatively small (37, 25 and 5 interactions, respectively). Note that 261 miRNAs interact only with a single gene. In comparison to the original article by Hartmann et al,<sup>6</sup> 6 (missed) interactions could not be found by our approach. For all of these, but one, a PubMed search does not return any results for interactions or even for the miRNAs (miR-1843/1935) themselves. The miR-21:CXCR4 interaction evidences are not included in *atheMir*, because these occur in an enumeration, and are rejected according to our text mining rules.

While this gives a general overview over the miRNA–chemokine landscape, there are more specific processes relevant in atherosclerosis. Restricting the miRNA–gene interaction search to only such interactions which are found in cardiovascular disease (DOID:1287 [cardiovascular system disease] or DOID:1936 [atherosclerosis]), it becomes apparent that there are many miRNA–gene interaction not yet studied in an atherosclerotic context. In contrast to the

**Table 1** Comparison of miRNA-gene interactions

Analysis of endothelial miRNA-gene interactions in atherosclerosis (Andreou et al, <sup>9</sup> fig. 1). Interactions mentioned in the Andreou et al review and found by atheMir are counted in atheMir + Andreou. Those mentioned by Andreou et al, but not found by atheMir (missed) are counted in the Andreou column and listed as Missed miRNAs. Values in brackets represent manually curated true interactions found only by atheMir.				
Gene	atheMir + Andreou	atheMir	Andreou	Missed miRNAs
CXCL12	1	0	0	
DLK1	2	0	0	
ETS1	2	3 (3)	1	miR-126 <sup>a</sup>
F11R	1	0	1	miR-143 <sup>a</sup>
ICAM1	1	3 (3)	1	miR-17 <sup>a</sup>
IRAK1	0	0	2	miR-146a <sup>a</sup> , miR-146b <sup>a</sup>
IRAK2	0	0	2	miR-146a <sup>a</sup> , miR-146b <sup>a</sup>
KLF2	2	0	2	miR-145 <sup>a</sup> , miR-126 <sup>a</sup>
KLF4	2	2 (2)	0	miR-663b
KPNA4	1	2 (1)	0	
NANOS3	1	10 (10)	0	
PPARA	2	2 (0)	0	
SELE	0	2 (2)	1	miR-31 <sup>a</sup>
SIRT1	1	4 (4)	0	
SOCS5	1	0	0	
TAB1	1	0	0	miR-10 <sup>b</sup>
TIMP3	1	0	0	
TRAF6	0	0	2	miR-146a <sup>a</sup> , miR-146b <sup>a</sup>
VCAM1	1	6 (6)	0	
Analysis of miRNA-gene interactions involved in initiation, progression, and thrombotic complications of atherosclerosis (Andreou et al <sup>9</sup> , table 1)				
Gene	atheMir + Andreou	atheMir	Andreou	Missed miRNAs
ABCA1	2	29 (28)	0	
ABCG1	1	7 (6)	0	
AKT1	1	8 (8)	0 (1)	
BCL6	2	1 (1)	0	
CPT1A	1	1 (1)	0	
DLK1	2	0	0	
KLF2	1	1	0	
KLF4	2	3 (3)	0	
KPNA4	1	2 (1)	0	
MAP3K10	2	0	0	
MT-TP	1	0	0	
RGS16	1	0	0	
SOCS1	1	0	0	
SOCS5	1	0	0	
TIMP3	1	2 (1)	0	

Abbreviation: miRNA, micro-ribonucleic acid.

<sup>a</sup>Missed also in PubMed search.<sup>b</sup>miR-Xa found as additional.

**Table 2** Systematic evaluation of atheMir Text mining results against facts mentioned in the Andreou et al review<sup>9</sup>

Systematic evaluation for endothelial miRNAs in atherosclerosis (►Table 1)						
		PubMed Corr.		Andreou		
		Cond. Pos.	Cond. Neg.	Cond. Pos.	Cond. Neg.	
DB	Pred. (True)	20 + 31	3	20	34	
	Not Pred. (False)	0	—	12	—	
Systematic evaluation for miRNAs in regulation of the initiation, progression, and thrombotic complication (►Table 1)						
		PubMed Corr.		Andreou		
		Cond. Pos.	Cond. Neg.	Cond. Pos.	Cond. Neg.	
DB	Pred. (True)	20 + 49	5	20	54	
	Not Pred. (False)	0	—	0	—	
Evaluation of statistical measures for the above results						
	Endothelial miRNA–gene interactions (►Table 1)			Init/Progr/Thrombotic (►Table 1)		
Measure	DB / PubMed	DB / Andr.	Comb.	DB / PubMed	DB / Andr.	Comb.
Sensitivity	1.0	0.625	0.625	1.0	1.0	1.0
False Discovery	0.0556	0.6297		0.0676	0.7297	
Rate						
Precision	0.9444	0.3703	0.9444	0.9324	0.2703	0.9324
F <sub>1</sub>	0.9714	0.4651	0.7522	0.9650	0.4255	0.965

Abbreviation: miRNA, micro-ribonucleic acid.

general interactions, more missed interactions are observed and only a few chemokines have multiple interactions studied, like CCL2, CXCL10, CXCR4, and CXCL12.

### CCL2 Expression in Macrophages

Inspired by the review by Hartmann et al,<sup>8</sup> we investigate the miRNA-mediated CCL2 expression in macrophages. Searching only the macrophages context (►Fig. 3, search parameters in ►Supplementary Table S5, available in the online version), all but one interaction from the Hartmann et al<sup>8</sup> review are found. The missed interaction between miR-150:CCL2 can be explained such that Hartmann et al<sup>8</sup> show this regulation in their network, but state, that this regulation is indirect via KLF2 and miR-124a, which is found by atheMir.

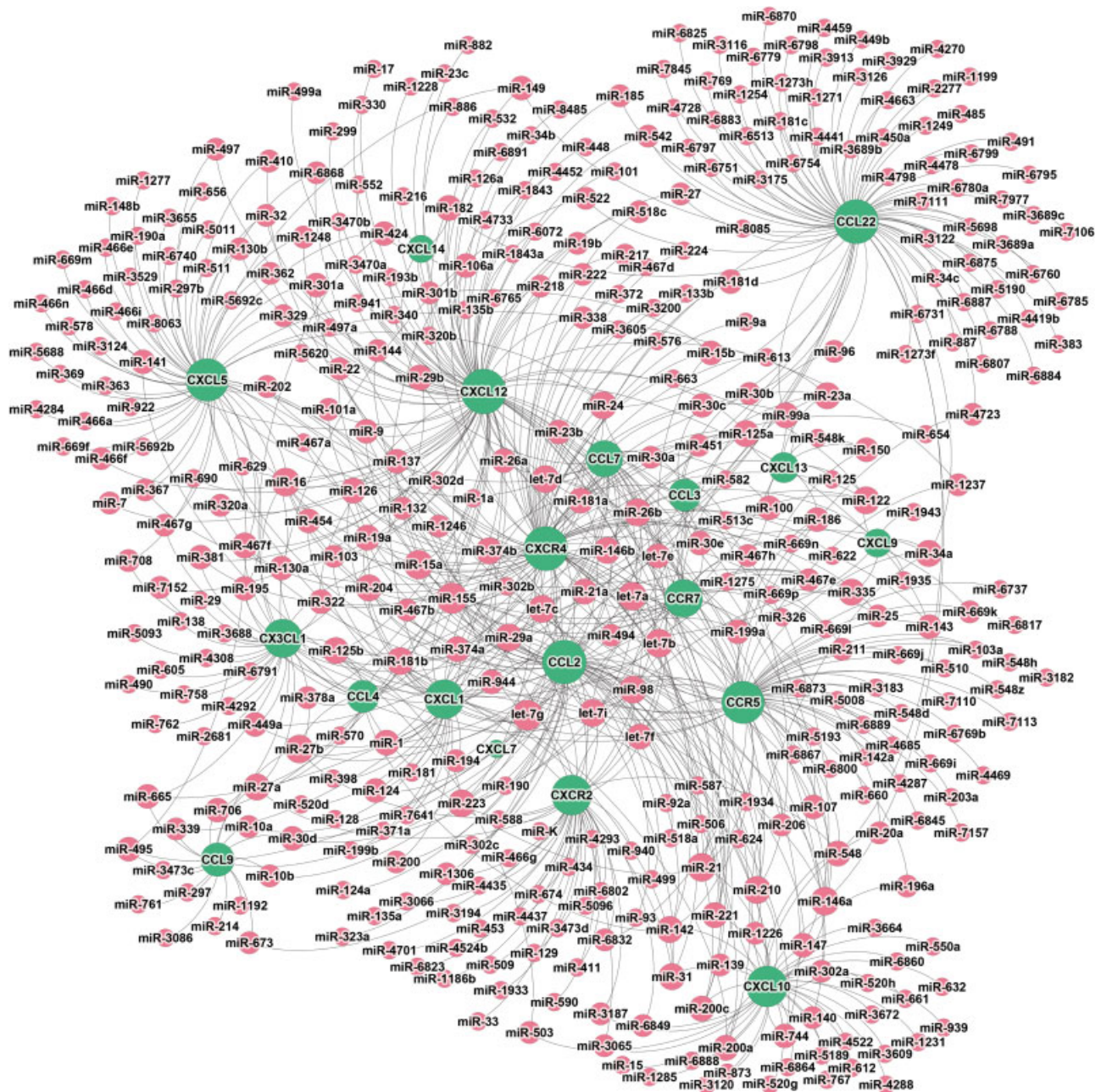
Applying the atherosclerosis context (DOID:1936) to macrophages, fewer additional interactions (►Supplementary Figure S8, available in the online version) are found in atheMir. Some interactions are not found, such as miR-24:CHI3L1, which is only reported in vascular diseases, but not explicitly in atherosclerosis. This also applies to the missed interaction of miR-146:IRAK1 which is, according to our database, only described in cardiac dysfunction.<sup>37</sup>

We will focus on the additional interactions in a cardiovascular disease or atherosclerosis context. Some additional interactions are shown in ►Supplementary Table S3 (available in the online version).<sup>38–40</sup> First, it is described that lipoprotein lipase (LPL) and CCL2 are both directly targeted by miR-590.<sup>41,42</sup> Repressing CCL2 prevents lipid accumulation, diminishing atherosclerosis. In-

creased LPL expression accelerates atherosclerosis by promoting lipid accumulation and inflammatory response.<sup>41</sup> miR-125b is known to regulate tumor necrosis factor receptor-associated factor 6 (TRAF6)<sup>43</sup> and CCL2 (via LACTB<sup>44</sup>) in atherosclerosis. This miRNA is particularly interesting, since it is normally depleted in leukocytes and monocytes.<sup>27</sup> Thus, an increase of miR-125b in these cell types could prevent atherosclerosis.

Toll-like receptor 4 (TLR4) can regulate CCL2 expression via nuclear factor kappa B (NF-κB) as stated in the original review by Hartmann et al.<sup>8</sup> Another interesting path is formed by TLR4→miR-124:CCL2, because miR-124 is naturally depleted in leukocytes.<sup>27</sup> It is known that TLR4 is repressed via miR-146a (enriched in leukocytes), while TLR4 is also a regulator of miR-124<sup>45</sup> in cocaine-mediated inflammation. It is also known from the literature that CCL2 is regulated via a miR-124-dependent pathway.<sup>46</sup> Thus, the TLR4→miR-124:CCL2 path could be directly controlled via miR-146a such that a higher miR-146a expression would lead to less repression of TLR4/miR-124/CCL2, while the pathway via NF-κB would be repressed (►Fig. 4A). In summary, a knockout of miR-146a would lead to more CCL2 via the NF-κB pathway, while the miR-124 pathway is repressed. On the other hand, an overexpression of miR-146a would lead to less CCL2 repression via TLR4→miR-124:CCL2. Under the assumption that both pathways are similar effective, miR-146a will likely not influence CCL2 expression in atherosclerosis. This matches the observations made by del Monte et al,<sup>28</sup> that there is no change in CCL2 levels when disturbing miR-146a.





**Fig. 1** The chemokine–miRNA interactome identified by atheMir: for each chemokine (green), all interacting miRNAs (red) are shown. The size of a node corresponds to the number of found interactions (representing its degree). Interactions are taken from text mining (PubMed abstracts), miRTarBase, and DIANA-TarBase.

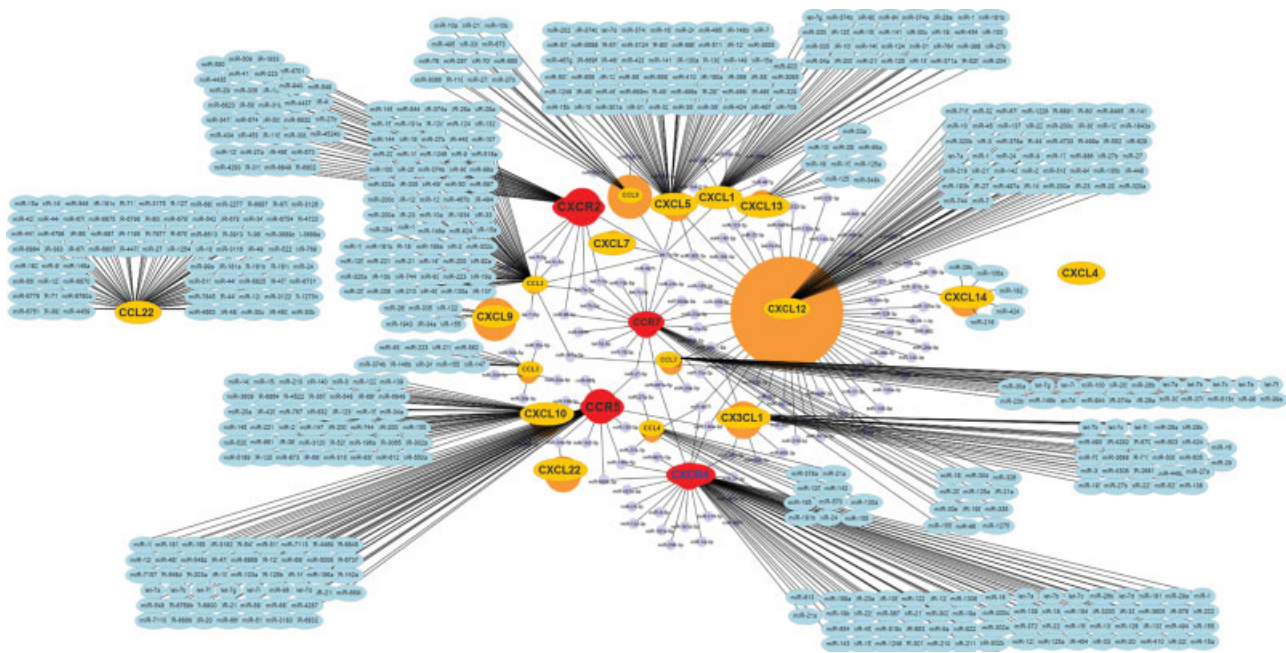
### miRNA-Mediated Inflammatory Response in Endothelial Cells

Additionally, we assessed our results regarding the miRNA-mediated inflammatory response in endothelial cells (► **Supplementary Fig. S7**, search parameters in ► **Supplementary Table S5**, available in the online version). Again, we first analyze the network of all known miRNA–gene interactions and find that there are no interactions missed by atheMir. Restricting the database search to the atherosclerosis context (DOID:1936, ► **Supplementary Fig. S9**, available in the online version), 10 interactions are not reported in atheMir: let-7g interacting with SIRT1, SMAD2, THBS1, and TGFBR1,

miR-146 interacting with HuR, TRAF6, and IRAK, and miR-10a interacting with TRC and TAK1 and miR-181b:KPNB3.

The miR-146:HuR interaction is reported for ELAVL1, the gene symbol for HuR. The miR-181b:KPNB3 interaction is found, but not within an explicit endothelial cell context. An interaction let-7g:TGFBR1 is found, but not within the atherosclerosis context. Regarding the other missed let-7g interactions, we checked the original reference.<sup>47</sup> First, SIRT1 is not recognized, because an uncommon symbol SIRT-1 is used, and to avoid confusions, we enforce that gene symbols are matched without error. SMAD2 and THBS1 are both recognized, however, the abstract has not triggered





**Fig. 2** The chemokine-miRNA interactome identified by atheMir. We show the increment to the original fig. 3 from Hartmann et al.<sup>8</sup> For each chemokine in the figure, a set of new interacting miRNAs is shown. The respective blocks of miRNAs exhibit the massive growth of knowledge on miRNA interactions in atherosclerosis since the Hartmann et al review in 2015 (original figure underlayed).

a cardiovascular disease context according to our vocabulary. For miR-146, it must be noted that the original article<sup>48</sup> cited by Hartmann et al<sup>8</sup> does not mention TRAF6 and IRAK1 in the abstract, but only in the full text. There exist other abstracts mentioning this interaction, but these do not focus on atherosclerosis.<sup>49</sup> Thus, with the current setup of text mining, the missed interactions are not found without curation of the underlying ontologies.

We further analyze atheMir's interactions for disturbing CCL2 expression in endothelial cells and have prepared selected additional interactions (► **Supplementary Table S4**, available in the online version).<sup>50–54</sup> First, we wanted to check results for miR-126 (► **Fig. 4B**) as this miRNA targets several of the genes in the pathway explained by Hartmann et al.<sup>8</sup> ETS1 is a transcription factor for miR-126,<sup>30</sup> and is controlled by miR-155, miR-221, and miR-222,<sup>29</sup> where the latter two are known to be enriched in endothelial cells.<sup>27</sup> If these miRNAs are downregulated, ETS1 expression is promoted. ETS1 itself influences CCL2 expression in several ways. First, ETS1 can directly coregulate CCL2<sup>29</sup> in an atherosclerotic context. Second, miR-126 is reported to reduce CCL2 expression in hCMC/D3 (brain) cells<sup>31</sup> directly. Finally, miR-126 also targets SIRT1 in artery disease<sup>32</sup> which represses the NF- $\kappa$ B pathway and, thus, reduces CCL2 levels.<sup>33</sup> In addition, the THBS1/TGFBR1/SMAD2 path is affected due to miR-126 interactions with THBS1 in ischemic hind limb.<sup>55</sup> While there is no reported evidence in atherosclerosis via this path, it is known that miR-126 is important in atherosclerosis, influencing endothelial cell proliferation.<sup>56</sup>

Another interesting factor in endothelial cell activation is miR-98. In contrast to miR-126, miR-98 has only CCL2 as target here, which was shown in the context of blood-brain barrier disease.<sup>57</sup> From expression data<sup>27</sup> it is known that

miR-98 typically is enriched in endothelial cells of the vascular tree, reducing oxidized low-density lipoprotein (LDL) uptake and, thus, apoptosis.<sup>58</sup> But since it also represses CCL2, fewer macrophages are attracted to the endothelial cells. An inhibition of miR-98 could thus increase atherosclerosis.

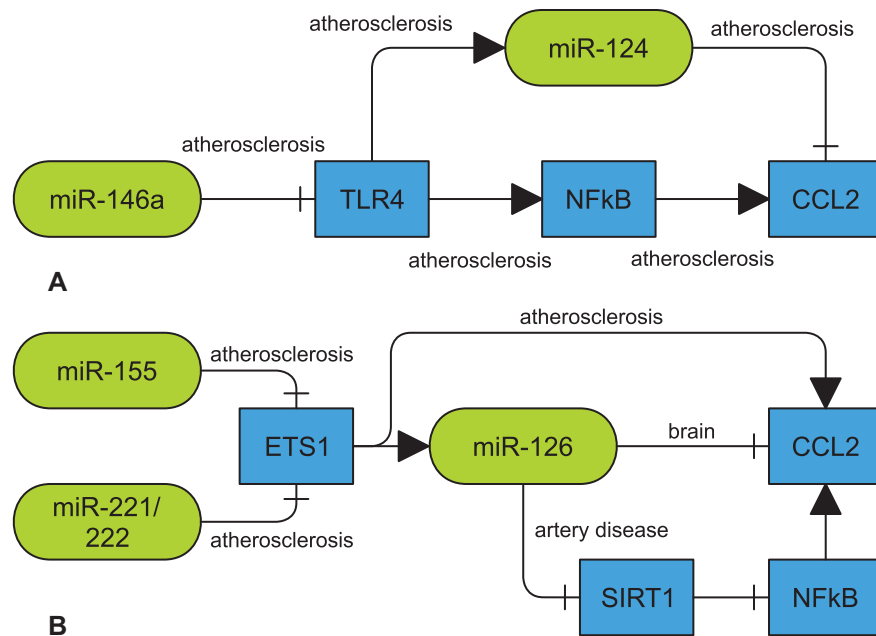
In conclusion, regarding the original pathways presented by Hartmann et al,<sup>8</sup> we could show that using available literature, combined with miRNA expression data, most known interactions described by domain experts could be reproduced and linked to respective evidence. The found miRNAs for these pathways are summarized in a UpSet plot<sup>59</sup>-like matrix (► **Fig. 5**). It can be seen that within the atherosclerosis context many miRNAs only have a single PubMed evidence. Many miRNAs have been detected in literature corresponding to multiple cell types. Regarding the endothelial cell pathway from Andreou et al,<sup>9</sup> and the endothelial inflammatory response pathway from Hartmann et al,<sup>8</sup> a huge overlap of miRNAs can be found. Furthermore, additional interactions have been proposed, which have either been already looked into, or represent mechanisms known in other disease contexts, and, therefore, could yield new hypotheses for atherosclerosis.

### Causal Networks

Similar to the specific chemokine networks, we analyze the six cardiovascular disease networks from the CBN database.<sup>12</sup> The networks contain nodes of different types of entities. Here, we filter the CBNs such that only nodes representing genes or proteins are contained.

First, for each network, we match it with the American Heart Association (AHA) atherosclerosis stages and determine which cells are most active in this stage. This allows to





**Fig. 4** (A) A model for the regulation of CCL2 in macrophages via miR-146a and miR-125. Depending on an over- or underexpression of miR-146a, Toll-like receptor 4 (TLR4) is either repressed, or regularly expressed. In the first case, the nuclear factor kappa B (NF- $\kappa$ B)-mediated pathway is reduced, but also less miR-124 represses CCL2. If TLR4 is expressed regularly, CCL2 is expressed via the NF- $\kappa$ B pathway, but also repressed via miR-124. Assuming that both paths are equally strong, the expression of TLR4 does not affect CCL2 expression. This matches the observations made by del Monte et al.<sup>28</sup> (B) A model for the regulation of CCL2 in endothelial cells via miR-126, which is coregulated via miR-155/221/222 and ETS1 (context information on edges). Literature reports three paths of regulation for CCL2. ETS1 directly regulates CCL2,<sup>29</sup> but is also a transcription factor upregulating miR-126,<sup>30</sup> which can directly downregulate CCL2.<sup>31</sup> Additionally, miR-126 can also downregulate SIRT1.<sup>32</sup> SIRT1 is an inhibitor of NF- $\kappa$ B,<sup>33</sup> which helps to upregulate CCL2.

126/146a/155/221 are among the top 10 regulating miRNAs in the first and second stage. Some of the miRNAs with a large number of interactions we found for this stage are not mentioned by Andreou et al in any associated process. Of these, for instance miR-499 is of interest, because it also regulates CCL2 and further genes directly associated with the attraction of monocytes, such as VCAM1, ICAM1, CXCL8, and CCL2.<sup>60</sup>

The interactions of miR-34a in atherosclerosis are also interesting. Some interactions are already reported in atherosclerosis, such as the interaction with SIRT1. From the literature it is known that miR-34 represses SIRT1 and thereby regulates apoptosis,<sup>61</sup> which is one of the main mechanisms during the second and third AHA stage. However, it also interacts with PDGFRA/PDGFRB, MEK1, and CDK4/6, regulating rat mesangial cell proliferation in glomerulonephritis.<sup>62</sup> Since these genes are also contained in the CBNs for atherosclerosis, these interactions could be a promising target in atherosclerosis.

Combining the first two stages/networks, miR-155 (31), miR-126 (25), miR-21 (16), miR-146a (13), and miR-124 (11) have the largest number of targets (in brackets) listed in *atheMir*.

### (III) Foam Cell Formation

During foam cell formation, other miRNAs play an important role (**Supplementary Fig. S11**, available in the online version): miR-21 (24), miR-155 (6), and miR-34a (4) are the most connected miRNAs.

With 24 found interactions, miR-21 appears to have a central role. Among its target genes are AKT1, MAPK8, FASLG, PPARA, CXCL2, and pTEN. While Andreou et al do not mention miR-199a in this stage (or a related process), *atheMir* predicts miR-199a interacting with both EGR1 and CD14 in this stage. It has been shown that EGR1 is a strong positive regulator of miR-199a.<sup>63</sup> In addition, CD14 is coexpressed with miR-199a and is known to regulate CXCL2, IL6, TNFA, and NO production.<sup>64</sup> miR-370 targets KDR and FOXO1 according to our database. It is known that both genes can block the AKT/FOXO1 signaling pathway in the context of cerebral aneurysm.<sup>65</sup> The inhibition of FOXO1 is atherogenic, as it leads to increased vascular calcification in atherosclerosis.<sup>66</sup>

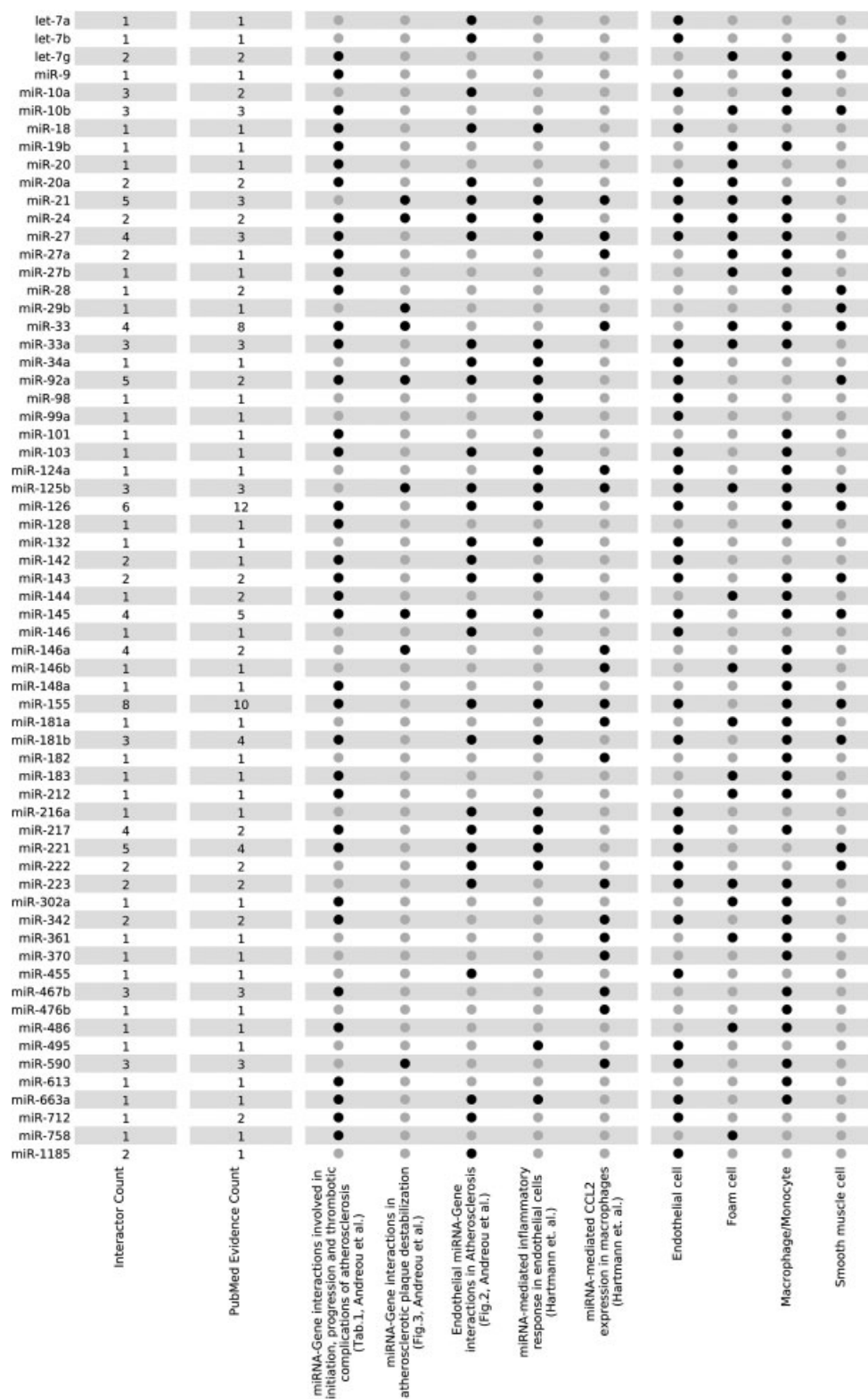
### (IV) Smooth Muscle Cell Activation

In the fourth stage of atherosclerosis, SMCs are activated, forming the lipid core and initiating fibrous cap formation. In addition to SMCs, endothelial cells are involved in the following.

During SMC activation, the miRNAs with more than 20 targets are miR-126 (31), miR-146a (30), miR-21 (27), and miR-155 (21). Indeed, for most genes in the enriched causal network, regulation by many miRNAs has been reported.

The miR-152 interactions are of special interest, because miR-152 is known to be relevant in atherosclerosis. In addition, we find further interactions in other contexts (diseases). In our network, miR-152 targets, among others, ESR1, ADAM17, KDR, and VEGFA. It has been shown that ESR1



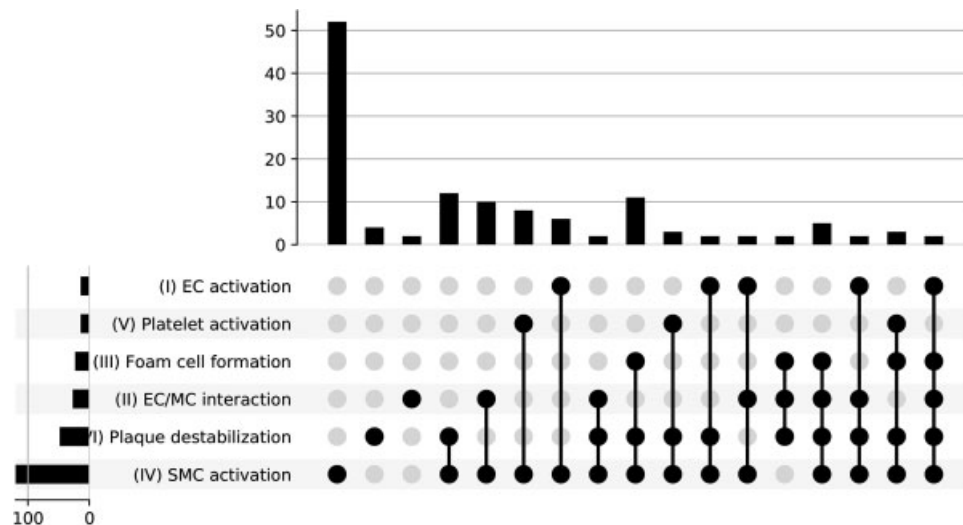


**Fig. 5** For each miRNA, its number of gene interactions in the atherosclerosis context and the number of corresponding PubMed evidences is shown. The middle and right block show a black dot if a miRNA is within the associated pathway or found within the cell type context.

expression is reduced via miR-152 repressing deoxyribonucleic acid methyltransferase. A high level of ESR1 is reported to protect against atherosclerosis.<sup>67</sup> Another target of miR-152 is ADAM17. A high ADAM17 expression is known to decrease lesion formation<sup>68</sup> and, thereby, functions in an

atheroprotective manner. Thus, miR-152 is an interesting target in atherosclerosis because it can reduce lesion formation via two different pathways. Furthermore, it regulates apoptosis in brain microvascular endothelial cells via PTEN and Bax.<sup>69</sup>





**Fig. 6** miRNA overlap between stages of atherosclerosis as proposed by the Causal Biological Networks (CBNs) when restricting the miRNA–gene interaction search to atherosclerosis and cardiovascular diseases.

### (V) Platelet Activation

The fifth stage of atherosclerosis is platelet activation with thrombus formation from the necrotic core. Compared with the previous stage, the enriched miRNA network is considerably smaller, however, also interesting (–**Supplementary Fig. S14**, available in the online version).

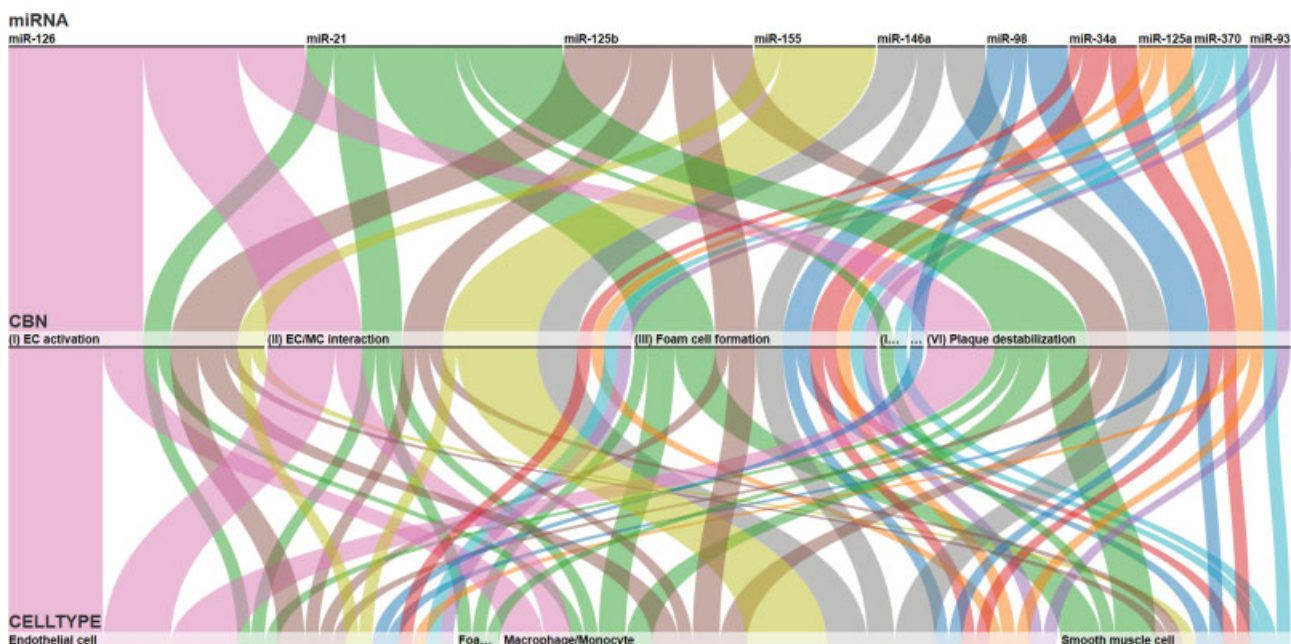
miR-20a regulates the AKT1, PTEN, EDN1, VEGFA, and NANOS3. NANOS3 and OLR1 are also targeted by let-7a/b. OLR1 is also targeted by miR-590, which also represses TP53 and BAX. Interestingly, there is a large gene overlap between the miR-20a and miR-590 targets and the miR-152 targets from the previous stage. In this stage of atherosclerosis, lipoproteins ABCA1 and LDL receptor are regulated by miR-143. This implies that repressing miR-143, lipoprotein uptake and necrotic core formation can be reduced.<sup>70</sup>

### (VI) Plaque Destabilization

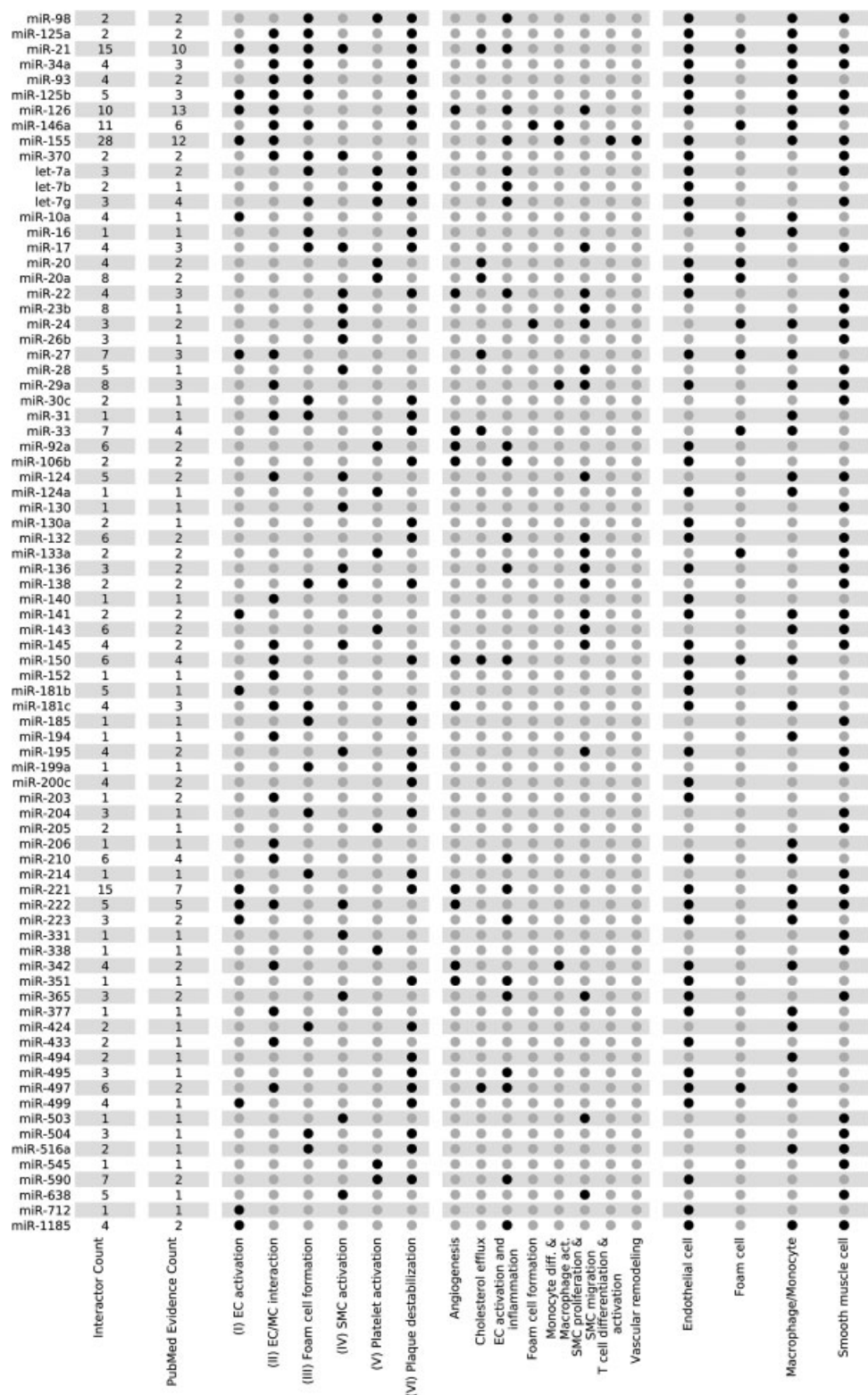
The final stage of atherosclerosis is plaque destabilization (–**Supplementary Fig. S15**, available in the online version). In this stage, the necrotic core breaks the artery wall and can form a thrombus.

In this causal network, miR-21 (26) and miR-155 (16) have most interactions. We want to further investigate the lipid uptake and find that miR-33 interacts with both ABCA1 and ABCG1 in this stage. While the miR-33 interaction with ABCA1 is already known in atherosclerosis, the interaction with ABCG1 has not yet been reported in atherosclerosis, specifically. miR-33 contributes to the regulation of cholesterol homeostasis by targeting both ABCA1 and ABCG1 directly.<sup>71</sup>

Having focused on chemokines earlier, we look at the CCL2 interactions in this stage. In contrast to earlier stages, CCL2



**Fig. 7** Number of interactions seen in Causal Biological Networks (CBNs)<sup>12</sup> and cell type context. Selected miRNAs appear in most CBNs (in a cardiovascular disease/atherosclerosis context).



**Fig. 8** For each miRNA in the atherosclerosis context, its number of gene interactions, the number of PubMed evidences, associated Causal Biological Networks (CBN)<sup>12</sup> stages, processes, as defined in ► **Supplementary Table S13**, and cell types, are shown. Overall, there are 114 miRNAs in the CBN stages and processes, of which 80 are shown here (min. 2 PubMed abstracts and must be in at least one CBN stage). The top 10 miRNAs are the most interacting ones in most CBN stages (► **Fig. 7**).

can also be regulated by different miRNAs, namely miR-494/495 and miR-10b. It could be shown that miR-494 induces inflammatory mediators, including CCL2.<sup>72</sup> Additionally, miR-495 directly targets CCL2, affecting proliferation and

apoptosis of human umbilical vein endothelial cells.<sup>73</sup> An effect on CCL2 via miR-499 and the NF-κB signaling pathway has also been reported.<sup>60</sup> Moreover, miR-10b seems to affect CCL2 expression in the context of renal allograft loss.<sup>74</sup> This

could indicate a role of miR-10b in detecting foreign objects near endothelial cells. Hypothetically, plaque formation could induce similar reactions.

### Abundant miRNA Regulators per Stage

After looking at the context-specific roles of individual miRNAs in the six CBN stages, we also investigate the most abundant miRNAs across all stages. The number of miRNA-gene interactions for each miRNA is listed in **►Supplementary Table S12** (available in the online version). We only consider interactions within the cardiovascular disease and atherosclerosis context.

Summarizing all stages, we find that the following miRNAs have the largest number of interactions (interaction counts and PubMed evidence counts in brackets) reported by *atheMir*: miR-126 (15, 17), miR-21 (15, 15), miR-155 (14, 8), miR-146a (14, 8), miR-125b (13, 7), miR-34a (9, 6), miR-499 (8, 2), miR-221 (7, 5), miR-370 (7, 4), and miR-504 (6, 2).

These miRNAs have a high overlap with the miRNAs appearing in most CBN stages: miR-21 (5), miR-125b (4), miR-370 (4), miR-93 (3), miR-98 (3), miR-125a (3), miR-126 (3), miR-146a (3), miR-155 (3), and miR-34a (3).

For each of these miRNAs, we evaluate in which CBN stages and cell types it can be found by how many PubMed evidences. This has been visualized in a parallel set plot (**►Fig. 7**). For the above miRNAs, the width of the connections shows the number of evidences found for the miRNA interacting in the given stages and cell types. It can be seen that miR-126 is well studied, particularly in endothelial cells and within stages (I) and (IV). In the following stages, more and more miRNAs become relevant. The foam cell formation stage has few reported miRNA interactions, yet it consists of miRNA interactions in foam cells, macrophages, and SMCs. Likewise, the plaque destabilization stage combines all cell types. In the platelet activation stage, none of the otherwise frequently occurring miRNAs is important, only miR-98 is active. Particularly miR-21 is mostly active in stages 2 to 4, while miR-125b seems to be involved in all stages. On the other hand, miR-155 and miR-98 seem to be specifically relevant in SMC activation.

### Prevalent Cell Types per Stage

Besides the miRNA interactions per CBN<sup>12</sup> stage, we are interested in the different cell types prevalent per stage. Exemplarily, we counted the occurrences of cell types for all miRNA-gene interactions in the two stages: endothelial cell/monocyte interaction (**►Supplementary Fig. S5**, available in the online version) and SMC activation (**►Supplementary Fig. S6**, available in the online version).

For the endothelial cell/monocyte interaction (**►Supplementary Fig. S5**, available in the online version), it can be seen that the most frequent cell types are monocytes and vascular endothelial cells. However, also other cell types relevant to atherosclerosis, such as SMCs, inflammatory macrophages, foam cells, natural killer cells, neutrophils, and platelets, are mentioned. However, also some references to cell types in the brain (brain microvascular endothelial cells, neoplastic cells) exist.

Similarly, most miRNAs in the SMC activation stage are occurring in SMCs. However, there are quite some occurrences in other cells which are involved in atherosclerosis, such as platelets, endothelial cells, monocytes, and foam cells.

### Cell Type-Based miRNA Cooccurrences

miRNAs occurring in more than one cell type within the same stage, could be interesting targets, because they could hypothetically resemble a similar mechanism in the affected cell types. We want to explore such cooccurrences again in the stages endothelial cell/monocyte interaction and SMC activation.

In the cell type cooccurrence figure, two cell types are connected (by a miRNA), if this miRNA is involved in both cell types. By definition, a miRNA defines a clique (fully connected network) of cell types it is active in. For the endothelial cell/monocyte interaction stage (**►Supplementary Fig. S17**, available in the online version) such cliques are formed. For instance, miR-145 forms a large clique of mainly SMCs. However, also cliques are formed for foam cells, microvascular endothelial cells, and monocytes. miR-126 is mainly active within endothelial cells, but is also detected in monocytes. miR-155 is reported in macrophages and monocytes. Finally, miR-222 is both described in SMCs and endothelial cells.

Switching to the SMC activation stage (**►Supplementary Table S9**, available in the online version), more and larger cliques can be found, indicating that miRNAs are active in multiple cell types. Here, particularly the small cliques with only a few cell types could be of interest, because these miRNAs are more cell-type specific. For instance, miR-10a is mentioned in monocytes, endothelial cells, and inflammatory cells. Monocytes and vascular SMCs have in common that, in both cell types, miR-33/181 and miR-516a interactions are reported. Also, miR-98 is of interest because it functions in monocytes, endothelial cells, and SMCs, and thus is involved in cells relevant for early atherogenesis. In combination with the previous finding in inflamed endothelial cells, it could be an interesting target for chemokine-mediated processes in atherosclerosis.

### Combining Stages and Processes of Atheroprogession

We focus again on rather broad contexts, namely the eight processes of atheroprogession described by Andreou et al<sup>9</sup>: endothelial cell activation and inflammation, monocyte differentiation and macrophage activation, foam cell formation, angiogenesis, vascular remodeling, T cell differentiation and activation, cholesterol efflux, and SMC proliferation and migration.

We have refined queries for *atheMir* using cell types and GO terms to find both miRNAs involved in these processes and their target genes (**►Supplementary Table S13**, available in the online version). For each process, we determined a range of 2 to 35 relevant miRNAs and 2 to 80 gene targets (**►Supplementary Table S10**, available in the online version). For T cell differentiation and activation, we set the disease context to cardiovascular system disease. Additionally we require the GO classes for SMC migration and proliferation in



the (IV) SMC activation stage, to distinguish this stage from the endothelial cell stages (I) and (II) (► **Supplementary Fig. S16**, available in the online version).

The miRNAs occurring in at least one CBN stage<sup>12</sup> and their detected presence in the processes as well as cell types is summarized in ► **Fig. 8**. In addition, the number of found target genes (interactors) and PubMed evidences from the combined CBN stage and process analysis are shown. This figure allows to make several interesting observations: Even though some miRNAs are associated to specific stages, they are not associated to any defined processes of atheroprogession. Many miRNA–gene interactions are only supported by one PubMed article. Most miRNAs are associated to multiple cell types. Only a few miRNAs occur in a majority of the CBN stages, and many miRNAs are relevant to only 2 or 3 CBN stages. The difference between the recognized miRNAs in the stages and processes of atherosclerosis shows the limits of the used NER approach: it relies on the quality and completeness of the synonym lists, and that authors make use of that vocabulary.

Thus, it is important to not only rely on one dimension (e.g., GO), because evidences may be missed during classification. Using further dimensions, such as disease, cell type, or protein class, can be used to characterize the context of literature, as has been shown in our evaluation. This underlines the importance of accessing the underlying evidences. With *atheMir*, these options can be explored, and evidence can be accessed, to make informed decisions on the found interactions.

Finally, particularly those miRNAs occurring in few stages and processes could be interesting targets for further research to determine their role in atherosclerosis, and prove the specificity of these miRNAs to certain phases of the disease.

## Conclusion

Here, we presented an approach to derive and collect results obtained from *atheMir*, which identifies miRNA–target interactions relevant to atherosclerosis and its broadly defined stages in a context-sensitive manner, mainly using text mining.

We provide an overview of miRNA–gene interactions in complex regulatory networks with a focus on specific pathways for inflammation in macrophages and endothelial cells as well as on the six stages of atherosclerosis represented by CBNs.<sup>12</sup>

With this focus, we evaluate *atheMir* based on two existing reviews about atherosclerosis and specifically chemokines. These expert curated benchmarks are used to assess the performance of *atheMir*. In summary,  $F_1$  scores of 0.75 and 0.97 demonstrate that *atheMir* performs well. Moreover, *atheMir* can add many facts and hypotheses on miRNA:gene interactions, indicating better explanation of the published literature but also the remarkable progress in the field in the last years.

We discussed a miR-146-mediated pathway in macrophages, where neither a knockout nor overexpression is

showing an effect on the outcome of atherosclerosis. We could formulate a hypothesis regarding the possible mechanism for the observed CCL2 expression. When the NF- $\kappa$ B pathway via TLR4 is repressed, the alternative pathway via miR-124 is enhanced, thus, CCL2 expression is not affected.

We combined several paths for CCL2 regulation from different contexts in endothelial cells using the ETS1/miR-126 path. While the combined pathway has not yet been validated in the literature, this path could realize similar compensating mechanisms as discussed above for the miR-124 example.

Furthermore, we emphasize the role of miR-98 in atherosclerosis, since it might be responsible for macrophage attraction to endothelial cells, by repressing CCL2 in normal state. This observation is supported by the finding that miR-98 plays a role in all three cell types, monocytes, SMCs, and endothelial cells.

Via a large-scale network analysis, we identified several interesting miRNAs which could be involved in atherosclerosis. During foam cell formation, we highlight miR-199a and miR-370 as interesting regulators, as miR-199a is regulated by EGR1 and CD14, which regulates TNFA production. The interaction of miR-370 with the AKT/FOXO1 signaling pathway could play an interesting role in atherogenesis.

In SMCs we identified miR-152 as interesting candidate, due to its known effects on ADAM17, but also on ESR1. In platelet activation we can observe that many targets of miR-152 are also targeted by miR-20a. Thus, both miRNAs may be good targets for further investigation.

During plaque destabilization, we emphasize the role of miR-494/-495 and miR-10b due to their influence on CCL2 expression. These miRNAs either target CCL2 directly or via NF- $\kappa$ B, thus modulating CCL2 expression in various ways and, thereby, possibly regulating macrophage attraction.

Finally, we compared the different stages of atherosclerosis and found miRNAs which are active in several cell types. During the two initial stages of atherosclerosis, miR-125b/126/155 and miR-222 are important in endothelial cells and monocytes. In the later stage of SMC activation, other miRNAs, like miR-98 and miR-504, are prevalent in endothelial cells, SMCs, and monocytes.

Using the context-based *atheMir* database, existing miRNA–gene interaction networks are largely reproduced. In all cases, the augmented models are significantly enhanced by additional interactions, indicating much more complex regulatory networks and hypotheses even as currently known from facts reported in experimental databases or the published literature. The final picture will likely be even more complex.

Context-based text mining methods can massively influence and support reviews from domain experts. While the information in *atheMir* is certainly not complete, it might be a good starting point for experts writing reviews but also for researchers investigating highly specific hypotheses in certain contexts. For both use cases, *atheMir* facilitates an easy access to individual, context-sensitive miRNA–gene interactions. Most importantly, it provides supporting evidence for each reported interaction. Using a context-based approach,



such as *atheMir*, is a helpful method to explore miRNA–gene interaction hypotheses in atherosclerosis and beyond.

#### Funding

This work has been supported by the DFG (Deutsche Forschungsgemeinschaft) via SFB1123/2 (projects A1 and Z2).

#### Conflict of Interest

R.Z. reports grants from DFG, during the conduct of the study. M.J. reports grants from DFG, during the conduct of the study.

#### References

- McManus MT, Sharp PA. Gene silencing in mammals by small interfering RNAs. *Nat Rev Genet* 2002;3(10):737–747
- Moss EG. MicroRNAs: hidden in the genome. *Curr Biol* 2002;12(04):R138–R140
- Poy MN, Eliasson L, Krutzfeldt J, et al. A pancreatic islet-specific microRNA regulates insulin secretion. *Nature* 2004;432(7014):226–230
- Urbich C, Kuehnbacher A, Dimmeler S. Role of microRNAs in vascular diseases, inflammation, and angiogenesis. *Cardiovasc Res* 2008;79(04):581–588
- Johnson Jason L. Elucidating the contributory role of microRNA to cardiovascular diseases (a review). *Vascul Pharmacol* 2019;114:31–48
- Jackson AO, Regine MA, Subrata C, Long S. Molecular mechanisms and genetic regulation in atherosclerosis. *Int J Cardiol Heart Vasc* 2018;21:36–44
- Donaldson CJ, Lao KH, Zeng L. The salient role of microRNAs in atherogenesis. *J Mol Cell Cardiol* 2018;122:98–113
- Hartmann P, Schober A, Weber C. Chemokines and microRNAs in atherosclerosis. *Cell Mol Life Sci* 2015;72(17):3253–3266
- Andreou I, Sun X, Stone PH, Edelman ER, Feinberg MW. miRNAs in atherosclerotic plaque initiation, progression, and rupture. *Trends Mol Med* 2015;21(05):307–318
- Erhard F, Haas J, Lieber D, et al. Widespread context dependency of microRNA-mediated regulation. *Genome Res* 2014;24(06):906–919
- Durham AL, Speer MY, Scatena M, Giachelli CM, Shanahan CM. Role of smooth muscle cells in vascular calcification: implications in atherosclerosis and arterial stiffness. *Cardiovasc Res* 2018;114(04):590–600
- Boue Stephanie, Talikka Marja, Westra Jurjen Willem, et al. Causal biological network database: a comprehensive platform of causal biological network models focused on the pulmonary and vascular systems. *Database (Oxford)*; 2015
- Yates B, Braschi B, Gray KA, Seal RL, Tweedie S, Bruford EA. Genenames.org: the HGNC and VGNC resources in 2017. *Nucleic Acids Res* 2017;45(D1):D619–D625
- Smith CL, Blake JA, Kadin JA, Richardson JE, Bult CJ; Mouse Genome Database Group. Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. *Nucleic Acids Res* 2018;46(D1):D836–D842
- Csaba Gergely. Personal communication on syngrep 2019
- Smedley D, Haider S, Durinck S, et al. The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res* 2015;43(W1):W589–W598
- Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 2014;42(Database issue):D68–D73
- Schriml LM, Mittraka E, Munro J, et al. Human Disease Ontology 2018 update: classification, content and workflow expansion. *Nucleic Acids Res* 2019;47(D1):D955–D962
- Bairoch A. The Cellosaurus, a cell-line knowledge resource. *J Biomol Tech* 2018;29(02):25–38
- Hartel FW, de Coronado S, Dionne R, Fragoso G, Golbeck J. Modeling a description logic vocabulary for cancer research. *J Biomed Inform* 2005;38(02):114–129
- Ashburner M, Ball CA, Blake JA, et al; The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nat Genet* 2000;25(01):25–29
- Honnibal Matthew, Montani Ines spaCy 2019
- Choi JD, Joel T, Amanda S It Depends: Dependency Parser Comparison Using a Web-based Evaluation Tool in Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers); 1 Stroudsburg, PA, USA:387–396; Association for Computational Linguistics; 2015
- Chou CH, Shrestha S, Yang CD, et al. miRTarBase update 2018: a resource for experimentally validated microRNA–target interactions. *Nucleic Acids Res* 2018;46(D1):D296–D302
- Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA–target interactions. *Nucleic Acids Res* 2009;37(Database issue):D105–D110
- Karagkouni D, Paraskevopoulou MD, Chatzopoulos S, et al. DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA–gene interactions. *Nucleic Acids Res* 2018;46(D1):D239–D245
- de Rie D, Abugessaisa I, Alam T, et al; FANTOM Consortium. An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat Biotechnol* 2017;35(09):872–878
- Del Monte A, Arroyo AB, Andrés-Manzano MJ, et al. miR-146a deficiency in hematopoietic cells is not involved in the development of atherosclerosis. *PLoS One* 2018;13(06):e0198932
- Zhu N, Zhang D, Chen S, et al. Endothelial enriched microRNAs regulate angiotensin II-induced endothelial inflammation and migration. *Atherosclerosis* 2011;215(02):286–293
- Harris TA, Yamakuchi M, Kondo M, Oettgen P, Lowenstein CJ. Ets-1 and Ets-2 regulate the expression of microRNA-126 in endothelial cells. *Arterioscler Thromb Vasc Biol* 2010;30(10):1990–1997
- Cerutti C, Edwards LJ, de Vries HE, Sharrack B, Male DK, Romero IA. MiR-126 and miR-126\* regulate shear-resistant firm leukocyte adhesion to human brain endothelium. *Sci Rep* 2017;7:45284
- Togliatto G, Trombetta A, Dentelli P, et al. Unacylated ghrelin induces oxidative stress resistance in a glucose intolerance and peripheral artery disease mouse model by restoring endothelial cell miR-126 expression. *Diabetes* 2015;64(04):1370–1382
- Fan W, Fang R, Wu X, et al. Shear-sensitive microRNA-34a modulates flow-dependent regulation of endothelial inflammation. *J Cell Sci* 2015;128(01):70–80
- Zhong L, Huot J, Simard MJ. p38 activation induces production of miR-146a and miR-31 to repress E-selectin expression and inhibit transendothelial migration of colon cancer cells. *Sci Rep* 2018;8(01):2334
- Hartmann P, Zhou Z, Ntarelli L, et al. Endothelial Dicer promotes atherosclerosis and vascular inflammation by miRNA-103-mediated suppression of KLF4. *Nat Commun* 2016;7:10521
- Leal MF, Caires Dos Santos L, Martins de Oliveira A, et al. Epigenetic regulation of metalloproteinases and their inhibitors in rotator cuff tears. *PLoS One* 2017;12(09):e0184141
- Echavarría R, Mayaki D, Neel JC, Harel S, Sanchez V, Hussain SNA. Angiopoietin-1 inhibits toll-like receptor 4 signalling in cultured endothelial cells: role of miR-146b-5p. *Cardiovasc Res* 2015;106(03):465–477
- Rajaram MVS, Ni B, Morris JD, et al. Mycobacterium tuberculosis lipomannan blocks TNF biosynthesis by regulating macrophage MAPK-activated protein kinase 2 (MK2) and microRNA miR-125b. *Proc Natl Acad Sci U S A* 2011;108(42):17408–17413
- Jiang W, Liu G, Tang W. MicroRNA-182-5p ameliorates liver ischemia-reperfusion injury by suppressing Toll-like receptor 4. *Transplant Proc* 2016;48(08):2809–2814

- 40 Cheng H-P, Gong D, Zhao Z-W, et al. MicroRNA-182 promotes lipoprotein lipase expression and atherogenesis by targeting histone deacetylase 9 in apolipoprotein E-knockout mice. *Circ J* 2017;82(01):28–38
- 41 He PP, Ouyang XP, Tang YY, et al. MicroRNA-590 attenuates lipid accumulation and pro-inflammatory cytokine secretion by targeting lipoprotein lipase gene in human THP-1 macrophages. *Biochimie* 2014;106:81–90
- 42 He P-P, Ouyang XP, Li Y, et al. MicroRNA-590 inhibits lipoprotein lipase expression and prevents atherosclerosis in apoE knockout mice. *PLoS One* 2015;10(09):e0138788
- 43 Ma H, Wang X, Ha T, et al. MicroRNA-125b prevents cardiac dysfunction in polymicrobial sepsis by targeting TRAF6-mediated nuclear factor  $\kappa$ B activation and p53-mediated apoptotic signaling. *J Infect Dis* 2016;214(11):1773–1783
- 44 Lu JB, Yao XX, Xiu JC, Hu YW. MicroRNA-125b-5p attenuates lipopolysaccharide-induced monocyte chemoattractant protein-1 production by targeting inhibiting LACTB in THP-1 macrophages. *Arch Biochem Biophys* 2016;590:64–71
- 45 Periyasamy P, Liao K, Kook YH, et al. Cocaine-mediated down-regulation of miR-124 activates microglia by targeting KLF4 and TLR4 signaling. *Mol Neurobiol* 2018;55(04):3196–3210
- 46 Gu W, Yao L, Li L, et al. ICAM-1 regulates macrophage polarization by suppressing MCP-1 expression via miR-124 upregulation. *Oncotarget* 2017;8(67):111882–111901
- 47 Liao YC, Wang YS, Guo YC, Lin WL, Chang MH, Juo SHH. Let-7g improves multiple endothelial functions through targeting transforming growth factor- $\beta$  and SIRT-1 signaling. *J Am Coll Cardiol* 2014;63(16):1685–1694
- 48 Cheng HS, Sivachandran N, Lau A, et al. MicroRNA-146 represses endothelial activation by inhibiting pro-inflammatory pathways. *EMBO Mol Med* 2013;5(07):1017–1034
- 49 Zilahi E, Tarr T, Papp G, Griger Z, Sipka S, Zeher M. Increased microRNA-146a/b, TRAF6 gene and decreased IRAK1 gene expressions in the peripheral mononuclear cells of patients with Sjögren's syndrome. *Immunol Lett* 2012;141(02):165–168
- 50 Manoharan P, Basford JE, Pilcher-Roberts R, Neumann J, Hui DY, Lingrel JB. Reduced levels of microRNAs miR-124a and miR-150 are associated with increased proinflammatory mediator expression in Krüppel-like factor 2 (KLF2)-deficient macrophages. *J Biol Chem* 2014;289(45):31638–31646
- 51 Prattichizzo F, Giuliani A, Recchioni R, et al. Anti-TNF- $\alpha$  treatment modulates SASP and SASP-related microRNAs in endothelial cells and in circulating angiogenic cells. *Oncotarget* 2016;7(11):11945–11958
- 52 Xu P, Zhao Y, Liu M, et al. Variations of microRNAs in human placenta and plasma from preeclamptic pregnancy. *Hypertension* 2014;63(06):1276–1284
- 53 Ferreira R, Santos T, Amar A, et al. Argonaute-2 promotes miR-18a entry in human brain endothelial cells. *J Am Heart Assoc* 2014;3(03):e000968
- 54 Xu Y-L, Zhang M-H, Guo W, et al. MicroRNA-19 restores vascular endothelial cell function in lower limb ischemia-reperfusion injury through the KLF10-dependent TGF- $\beta$ 1/Smad signaling pathway in rats. *J Cell Biochem* 2018;119(11):9303–9315
- 55 Mathiyalagan P, Liang Y, Kim D, et al. Angiogenic mechanisms of human CD34<sup>+</sup> stem cell exosomes in the repair of ischemic hind limb. *Circ Res* 2017;120(09):1466–1476
- 56 Schober A, Nazari-Jahantigh M, Wei Y, et al. MicroRNA-126-5p promotes endothelial proliferation and limits atherosclerosis by suppressing Dlk1. *Nat Med* 2014;20(04):368–376
- 57 Rom S, Dykstra H, Zuluaga-Ramirez V, Reichenbach NL, Persidsky Y. miR-98 and let-7g\* protect the blood-brain barrier under neuroinflammatory conditions. *J Cereb Blood Flow Metab* 2015;35(12):1957–1965
- 58 Chen Z, Wang M, He Q, et al. MicroRNA-98 rescues proliferation and alleviates ox-LDL-induced apoptosis in HUVECs by targeting LOX-1. *Exp Ther Med* 2017;13(05):1702–1710
- 59 Lex A, Gehlenborg N, Strobel H, Vuilleumot R, Pfister H. UpSet: visualization of intersecting sets. *IEEE Trans Vis Comput Graph* 2014;20(12):1983–1992
- 60 Zhang YH, He K, Shi G. Effects of microRNA-499 on the inflammatory damage of endothelial cells during coronary artery disease via the targeting of PDCD4 through the NF- $\kappa$ B/TNF- $\alpha$  signaling pathway. *Cell Physiol Biochem* 2017;44(01):110–124
- 61 Yamakuchi M, Ferlito M, Lowenstein CJ. miR-34a repression of SIRT1 regulates apoptosis. *Proc Natl Acad Sci U S A* 2008;105(36):13421–13426
- 62 Chen D, Li Y, Mei Y, et al. miR-34a regulates mesangial cell proliferation via the PDGFR- $\beta$ /Ras-MAPK signaling pathway. *Cell Mol Life Sci* 2014;71(20):4027–4042
- 63 Sakurai K, Furukawa C, Haraguchi T, et al. MicroRNAs miR-199a-5p and -3p target the Brm subunit of SWI/SNF to generate a double-negative feedback loop in a variety of human cancers. *Cancer Res* 2011;71(05):1680–1689
- 64 Du L, Rong H, Cheng Y, et al. Identification of microRNAs dysregulated in CD14 gene silencing RAW264.7 macrophage cells. *Inflammation* 2014;37(01):287–294
- 65 Hou W-Z, Chen X-L, Wu W, Hang C-H. MicroRNA-370-3p inhibits human vascular smooth muscle cell proliferation via targeting KDR/AKT signaling pathway in cerebral aneurysm. *Eur Rev Med Pharmacol Sci* 2017;21(05):1080–1087
- 66 Deng L, Huang L, Sun Y, Heath JM, Wu H, Chen Y. Inhibition of FOXO1/3 promotes vascular calcification. *Arterioscler Thromb Vasc Biol* 2015;35(01):175–183
- 67 Wang YS, Chou WW, Chen KC, Cheng HY, Lin RT, Juo SHH. MicroRNA-152 mediates DNMT1-regulated DNA methylation in the estrogen receptor  $\alpha$  gene. *PLoS One* 2012;7(01):e30635
- 68 Holdt LM, Thiery J, Breslow JL, Teupser D. Increased ADAM17 mRNA expression and activity is associated with atherosclerosis resistance in LDL-receptor deficient mice. *Arterioscler Thromb Vasc Biol* 2008;28(06):1097–1103
- 69 Cao YH, Li DG, Xu B, et al. A microRNA-152 that targets the phosphatase and tensin homolog to inhibit low oxygen induced-apoptosis in human brain microvascular endothelial cells. *Genet Mol Res* 2016;15(02):15
- 70 Sala F, Aranda JF, Rotllan N, et al. MiR-143/145 deficiency attenuates the progression of atherosclerosis in Ldlr<sup>-/-</sup> mice. *Thromb Haemost* 2014;112(04):796–802
- 71 Rayner KJ, Suárez Y, Dávalos A, et al. MiR-33 contributes to the regulation of cholesterol homeostasis. *Science* 2010;328(5985):1570–1573
- 72 Lan Y-F, Chen H-H, Lai P-F, et al. MicroRNA-494 reduces ATF3 expression and promotes AKI. *J Am Soc Nephrol* 2012;23(12):2012–2023
- 73 Liu D, Zhang XL, Yan CH, et al. MicroRNA-495 regulates the proliferation and apoptosis of human umbilical vein endothelial cells by targeting chemokine CCL2. *Thromb Res* 2015;135(01):146–154
- 74 Liu X, Dong C, Jiang Z, et al. MicroRNA-10b downregulation mediates acute rejection of renal allografts by derepressing BCL2L1. *Exp Cell Res* 2015;333(01):155–163