miRetrieve-an R package and web application for miRNA text mining

Friedrich, Julian; Hammes, Hans-Peter; Krenning, Guido

Published in:
NAR genomics and bioinformatics

DOI:
10.1093/nargab/lqab117

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

miRetrieve—an R package and web application for miRNA text mining

Julian Friedrich [1,2,*]**, Hans-Peter Hammes**[2,3] **and Guido Krenning**[1]

[1]Cardiovascular Regenerative Medicine (CAVAREM), Department of Pathology and Medical Biology, University Medical Center Groningen, University of Groningen, Hanzeplein 1 (EA11), 9713 GZ Groningen, The Netherlands, [2]5th Medical Department, Section of Endocrinology, Medical Faculty Mannheim, University of Heidelberg, 68167 Mannheim, Germany and [3]European Center of Angioscience, Medical Faculty Mannheim, University of Heidelberg, 68167 Mannheim, Germany

ABSTRACT

microRNAs (miRNAs) regulate gene expression and thereby influence biological processes in health and disease. As a consequence, miRNAs are intensely studied and literature on miRNAs has been constantly growing. While this growing body of literature reflects the interest in miRNAs, it generates a challenge to maintain an overview, and the comparison of miRNAs that may function across diverse disease fields is complex due to this large number of relevant publications. To address these challenges, we designed miRetrieve, an R package and web application that provides an overview on miRNAs. By text mining, miRetrieve can characterize and compare miRNAs within specific disease fields and across disease areas. This overview provides focus and facilitates the generation of new hypotheses. Here, we explain how miRetrieve works and how it is used. Furthermore, we demonstrate its applicability in an exemplary case study and discuss its advantages and disadvantages.

INTRODUCTION

microRNAs (miRNAs) are small, non-coding RNAs that regulate gene expression (1). By targeting complementary mRNAs, miRNAs repress the translation of genes into proteins and consequently control cellular functions (1,2). As a result, miRNAs add another layer of complexity to the regulation of cellular behavior, which is why an increased understanding of miRNAs may provide deeper insights into the pathophysiology of many diseases (3).

The field of miRNA research has expanded substantially over the past decade, as evidenced by the rapidly growing body of miRNA literature. Over the past decade, >110 000 publications containing the MeSH term 'miRNA' have appeared on PubMed (https://ncbi.nlm.nih.goc/pubmed/)! This large number of publications reflects the global interest in miRNAs across research fields, yet it also generates the challenge to maintain an overview of which miRNAs have been investigated and which functions they exert in a specific disease. Whereas this is already challenging in one field, it is even more challenging to compare miRNAs across disease fields function wise.

At the same time, comparing miRNAs across fields may extend the understanding of miRNA function per se—for example solving the key question on how miRNAs select their target genes, which is not exactly to date (4)—and facilitate the classification of miRNAs into functional groups based on the cellular behavior they influence.

This wealth of information has been summarized in manually curated databases such as miRTarBase (5), while tools such as fireflybio (https://www.fireflybio.com/portal/search), miRSel (6), miRTex (7) or emiRIT (8) help to automatically extract associations between miRNAs and genes and diseases from text. To further address the challenges above, we have designed miRetrieve. miRetrieve is an R package (9) and complements the previous tools by combining text mining algorithms to retrieve key information on miRNAs directly generated from text, such as their current frequency of investigation or likely function in a disease. Furthermore, miRetrieve also visualizes and compares miRNA:mRNA interactions by integrating information from external databases such as miRTarBase (5). Additionally, miRetrieve's most important functions are available as a web application (10).

Here we discuss how miRetrieve works and how it is used. Furthermore, we demonstrate its applicability in an exemplary case study and discuss its advantages and disadvantages.

---

*To whom correspondence should be addressed. Tel: +49 621 3832942; Fax: +49 621 3832999; Email: julian.friedrich@medma.uni-heidelberg.de
Present address: Julian Friedrich, Universitätsmedizin Mannheim, 5th Medical Department, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim, Germany.

MATERIALS AND METHODS

A detailed explanation of miRetrieve and step-by-step guide with code is given in the vignette (Supplementary File S1). A complete overview of all functions is in miRetrieve's documentation. This section explains the key tools in a miRetrieve workflow that are also provided in the web application.

Input

miRetrieve uses text as input and works with all databases that provide a MEDLINE or JATS format, but is optimized to work with PubMed abstracts (https://www.ncbi.nlm.nih.gov/pubmed/). We chose to design miRetrieve to text mine abstracts and not perform deep text mining of full papers to ensure access without restriction (compared to full text non-open access publications). Additionally, as abstracts summarize the major research findings, these may serve as a filter which facilitates the incorporation of miRNAs investigated in-depth and limits the incorporation of miRNAs only identified in array or sequencing experiments.

One key idea of miRetrieve is to text mine abstracts related to one specific disease field, e.g. *miRNAs in atherosclerosis*. That way, miRetrieve can identify and characterize miRNA functions distinct for this field, and later compare field-related miRNA functions in subsequent analysis steps. To use abstracts with miRetrieve, abstracts are downloaded from PubMed in a MEDLINE or JATS format and read into R.

Datamining

Extracting miRNAs. After the abstracts are loaded in miRetrieve, analysis is prepared by extracting miRNA names from text using several regular expressions (regex).

First, abstracts are prepared for miRNA extraction, e.g. by converting any non-ASCII character to its ASCII equivalent or inserting a '-' between any combination of *RNA* and a digit, combining the regex *(?< = [Rr][Nn][Aa])([0–9])* and *-\\1*. The abstracts are then screened for miRNAs, using the regex *[Mm][Ii][Cc]?[Rr][Oo]?[Rr]?[Nn]?[Aa]?[\\- \\s]?\\d+[a-zA-Z]?[\\-]?[12345]?[\\-]?[35]?[Pp]?*. This regex detects single miRNAs, such as *miR-146a-5p*, *MiRNA-146a*, *microRNA-146* or *MIR146*. Furthermore, it detects the first miRNA of enumerations, such as *miR-23b* from *miR-23b, -146a* and *-155* or *miR-23b/-146a/-155,* but misses subsequent miRNAs, like *-146a and -155* or */-146a/-155* in this case. Therefore, any miRNA detected by this initial regex is used to create an individual regex with a positive lookbehind to catch enumerated miRNAs, using the pattern *(?< = miRNA)([/,;and\\s?]+[-]?\\d+[a-z]?[-]?[1–5]?[-]?[53]?[pP]?\\s?)+* (e.g. *(?< = miR-23b)([/,;and\\s?]+[-]?\\d+[a-z]?[-]?[1– 5]?[-]?[53]?[pP]?\\s?)+* to catch *-146a, and -155* or */-146a/-155*).

Additionally, the regex *([Mm][Ii][Cc]?[Rr][Oo]?[Rr] ?[Nn]?[Aa]?[\\-\\s]?\\d+)([a-z][/]?)+* identifies miRNAs with enumerated lettered suffixes such as *miR-33a/b*, which are subsequently divided into their stem (*miR-33*) and lettered suffixes (*a/b*) by matching the regex *([Mm][Ii][Cc]?[Rr][Oo]?[Rr]?[Nn]?[Aa]?[\\- \\s]?\\d+)(.*)*. Afterwards, the stem is vectorized over the lettered suffixes to obtain the separated miRNAs, e.g. vectorizing *miR-33* over *a/b* to obtain *miR-33a* and *miR-33b*.

The identified miRNAs are cleaned, amongst others by removing trailing *-3p* and *-5p* or trailing digits such as *-1*, using the regex *[\\-\\s]?3[Pp]|[\\-\\s]?5[Pp]* and *(?< = \\d[a-zA-Z]?)(-[12345])*, respectively. The remaining numbers are miRNA numbers +/- lettered suffixes that are extracted with *\\d+[a-zA-Z]?* and concatenated with *miR-* to bring miRNA names into a cohesive form.

miRNA extraction is complemented by using the regex *[Ll][Ee][Tt][\\-\\s]?\\d+[a-zA-Z]?[\\-]?[35]?[Pp]?* to extract *let-7* variants, where *-3p* and *-5p* are again removed and *let-* is converted to lower case.

Finally, any lettered suffix is converted to lower case with *(?< = \\d)([A-Z])* and *\\L\\1*, and outdated miRNA names are replaced by their most recent names according to miRBase (11) version 22 (e.g. *miR-97* and *miR-102* are replaced by *miR-30a* and *miR-29b*, respectively).

The user can specify whether to extract miRNA names annotated with or without their lettered suffixes (e.g. *miR-146a* or *miR-146*). Extracting miRNA names with their lettered suffixes differentiates between miRNAs of similar sequence yet different genomic location and improves accuracy of the downstream analysis, but this accuracy critically depends on the correct and consequent usage of miRNA nomenclature (e.g. *hsa-miR-146a-3p*). As miRNA nomenclature may be inconsistently used throughout literature, extracting miRNA identifiers without these suffixes increases the consistency of extraction. For this reason, we designed miRetrieve to ignore extensions such a species prefix, -3p/-5p or genomic suffixes, but to focus on the main miRNA name instead.

By default, miRetrieve is programmed to extract all unique miRNA names in an abstract once to reflect the true count of abstracts mentioning one miRNA. This setting avoids the creation of bias that may be introduced when a miRNA is mentioned multiple times in the same abstract. However, as repeated mentioning of a miRNA name may reflect the depth a certain miRNA was investigated on, miRetrieve offers the possibility to extract only miRNAs mentioned with a minimal frequency in one abstract. Enabling this option ignores miRNAs mentioned seldomly or once and increases the visibility of well-investigated miRNAs.

The extracted miRNA names and their associated count files are the corner stone of subsequent analysis as detailed in the subsections below.

Frequency-distribution of miRNAs. Based on the extracted miRNA names, miRetrieve provides a frequency distribution of all identified miRNAs. Determining which miRNAs are frequently or seldomly mentioned across abstracts ascertains which miRNAs may be well investigated or under-investigated in a certain research field. Thus, the frequency distribution chart of miRNAs may pose as a surrogate measure of how large the influence

of a certain miRNA is in the determination of the cellular or disease phenotype under investigation.

Term-association of miRNAs. miRetrieve summarizes which terms are associated with a specific miRNA over several abstracts by tokenization. Tokenization refers to a process where text is split into smaller pieces, also called 'tokens'. By default, miRetrieve splits text on white space and strips punctuation. A token then may either represent a single word, such as *inflammation* or *tnf,* or adjacent words/numbers, such as *mir 146* or *low density lipoprotein* (12), also referred to as *n-grams* depending on the number *n* of words/numbers included in a token.

Per miRNA, all abstracts of the provided input containing this miRNA are split into tokens. Afterwards, frequent, but irrelevant tokens such as *a*, *is*, *western* or *expression level*, are removed, using a manually curated dictionary for common PubMed expressions and a dictionary of the *tidytext* package (12) for common English words. The remaining tokens are then displayed in decreasing frequencies, thus showing the frequency order of words used in the same abstracts as the respective miRNA.

Consequently, miRetrieve supports miRNA annotation to putative functional classes (e.g. inflammation and proliferation): A miRNA that is often mentioned in abstracts containing *inflammation*, *tnf* and *macrophages*, for example, may also play a role in inflammation.

Alternatively, miRetrieve can also do the reverse and identify which miRNAs are commonly associated with a user-defined term. Determining which miRNAs are frequently mentioned in combination with drugs or pathways, for example, may indicate possible miRNA-drug or miRNA-pathway interactions.

Class-association of miRNAs. Next to establishing which miRNAs are commonly associated with a specific term, miRetrieve can also dissect which miRNAs are commonly associated with multiple terms at once. Here, all abstracts of the text input are split into tokens, and only abstracts containing defined terms remain. If these terms summary a larger class (e.g. a combination of *circulating*, *biomarker*, *extracellular vesicles*, *urine*, *body fluid* for *biomarker*, or a combination of *patients*, *biopsies*, *children*, *women*, *men*, *humans* for *clinical studies*), miRetrieve assists in identifying miRNAs that may serve as biomarker, or that have already been investigated in patients. This feature can support the estimation of the potential clinical translatability of miRNAs and direct further research.

Visualizing validated miRNA–target interactions. Besides the functional association of a miRNA based on term associations, miRetrieve is able to visualize miRNA:mRNA interactions by relying on external databases such as miRTarBase (5). miRTarBase is a manually curated database that contains validated miRNA:mRNA interactions and the PubMed IDs of the publications these were reported in (5). Here, the strength of miRetrieve lies in matching the PubMed IDs of miRTarBase to the PubMed IDs of given abstracts. miRetrieve includes the latest miRTarBase version 8 (5), but any database

providing PubMed IDs is suitable. When these abstracts summarize miRNAs in a specific disease, e.g. *miRNAs in atherosclerosis*, miRetrieve consequently visualizes only validated miRNA:mRNA interactions in this field while ignoring unrelated interactions. This feature thus supports the identification of key miRNA:mRNA interactions in any specific disease field.

Comparing miRNA frequency, function and gene targets across research fields. As detailed above, miRetrieve provides an overview of miRNA frequency, function, and gene targets for any pre-specified research field. These functions can be easily extended to analyze and compare miRNAs across multiple fields. This cross-field comparison may be especially interesting in the identification of common miRNA signatures in multiple cancer subtypes, between multiple inflammatory disease and so on. By combining the same functions detailed above to compare miRNA frequency distributions, miRNA-term associations, and/or miRNA-target interactions across fields, miRetrieve may uncover common and distinct functions of miRNAs depending on disease context.

Usage and reliability of miRetrieve

Demonstration cases. The following case study demonstrates a typical miRetrieve workflow by illustrating how miRetrieve extracts information from thousands of abstracts. Here, miRetrieve identifies miRNAs frequently investigated in atherosclerosis and their putative functions. Furthermore, the interactions between miRNAs and their validated gene targets are visualized by the integration of miRTarBase (5) in miRetrieve. Throughout this demonstration, results generated by miRetrieve are manually validated to determine the accuracy and reliability of the program.

In a second part, a similar workflow is applied to compare miRNAs involved in atherosclerosis to miRNAs involved in lung cancer. Additionally, miRetrieve is used to identify biomarker miRNAs in both diseases. As atherosclerosis and lung cancer are largely independent from each other, these analyses provide an example of disease context-dependent and -independent features of the miRNAs under investigation.

The code for these demonstrations is available in Supplementary File S2. The graphs are created with ggplot2 (13) and patchwork (14).

Data preparation. To investigate the role of miRNAs in atherosclerosis, 1750 PubMed abstracts (https://www.ncbi.nlm.nih.gov/pubmed/) in total matching the keywords *atherosclerosis miRNA* as of 11 February 2021, were loaded as *Atherosclerosis* into R with read_pubmed().

Accuracy of miRNA extraction and frequency distribution. To determine the precision of the miRNA name extraction algorithm, miRNA names without and with lettered suffixes were extracted with extract_mir_df(). Afterwards, 200 abstracts were randomly selected and automatic miRNA name extraction was compared against manual extraction by computing the precision (*[Correctly extracted*

miRNAs] / ([Correctly extracted miRNAs] + [Falsely extracted miRNAs])) and recall (*[Correctly extracted miRNAs] / ([Correctly extracted miRNAs] + [Missed miRNAs])*) of automatic miRNA name extraction.

Additionally, extracted miRNA names were counted with count_mir(). Frequency distribution of miRNAs was validated by comparing the count of the top thirty miRNAs that carry no lettered suffix by definition against a PubMed search, following the pattern of '(*miR-(number) atherosclerosis) AND (('1990/01/01'[Date - Entry] : '2021/02/11'[Date - Entry]))*' as of 11 February 2021.

Per miRNA, miRetrieve's and PubMed's frequency count was compared by calculating a score with *[miRetrieve count] / [PubMed count]*. Afterward, all frequency scores were summarized by the mean, where a mean > 1 indicated a higher miRNA count with miRetrieve, on average, while a mean <1 indicated a higher miRNA count with PubMed.

Finally, the same steps were repeated for extracted miRNA names with lettered suffixes.

Text mining of miRNAs in atherosclerosis. From the original *Atherosclerosis* data frame, abstracts were filtered for research articles with subset_research() to focus on original research studies and not to bias the results with articles of other types such as *Reviews* or *Letters*. Next, miRNA names mentioned at least twice in an abstract were extracted with extract_mir_df(), resulting in 1003 abstracts and 1432 extracted miRNA names. The miRNA extraction threshold was set to >1 to increase the chance of focusing on relevant abstracts.

Frequency-distribution of miRNAs in atherosclerosis. To identify frequently investigated miRNAs in atherosclerosis, the ten most frequently mentioned miRNAs were plotted with plot_mir_count().

To estimate the translatability between miRNA frequency distribution and their degree of investigation, abstracts mentioning miR-155, miR-126, miR-146, miR-21 and miR-33—the top 5 miRNAs in atherosclerosis (see also Figure 2)—at least twice were compared to the number of abstracts indicating a relevant role of the respective miRNA in atherosclerosis, where 'relevant' was defined as the respective miRNAs either targeting key proteins or being differently expressed.

miRNA extraction thresholding. The threshold for extracting miRNAs may be altered by the user. Setting the miRNA extraction threshold to >1 may increase the scientific relevance of an extracted miRNA, but ignores abstracts that mention the respective miRNA only once, and thereby limits the diversity of the results. To examine if a higher miRNA extraction threshold improves the 'precision-to-recall ratio' for relevant abstracts, all abstracts mentioning miR-155, miR-126, miR-146, miR-21 and miR-33 were compared to the number of abstracts indicating a relevant role of the respective miRNA in atherosclerosis. For miRNA and extraction thresholds ('none' or >1), precision (*[Relevant abstracts] / ([Relevant abstracts] + [Irrelevant abstracts])*) and recall (*[Relevant abstracts] / ([Relevant abstracts] + [Relevant abstracts missed])*) were calculated.

Term-association of miRNAs in atherosclerosis. To determine the putative roles of miR-155 and miR-21, two frequently mentioned miRNAs in atherosclerosis, their most frequently associated single and two adjacent words were plotted with plot_mir_terms(). miR-155 was investigated as it is the most frequently mentioned miRNA in atherosclerosis, whereas miR-21 is frequently mentioned in atherosclerosis and the most frequently mentioned miRNA in lung cancer, the second main topic of this case study. Afterwards, these suggested miRNA annotations were validated by manually cross-referencing abstracts and corresponding articles.

Visualizing validated miRNA–target interactions in atherosclerosis. miRNA characterization was concluded by visualizing validated miRNA:mRNA interactions for miR-155 and miR-21 in atherosclerosis with join_targets() and plot_target_mir_scatter(), using the latest version of miRTarBase (5).

Comparing miRNA frequency, function and gene targets between atherosclerosis and lung cancer. To contrast the roles and functions of miRNAs in atherosclerosis and lung cancer, 6228 PubMed abstracts matching the keywords *lung cancer miRNA* as of 11 February 2021, were loaded as *Lung cancer* into R, filtered for original research articles, and miRNA names mentioned at least twice in an abstract were extracted, resulting in 4405 abstracts and 5865 miRNAs. To compare the most frequently investigated miRNAs in lung cancer to miRNAs in atherosclerosis, the ten most frequently mentioned miRNAs in lung cancer were plotted with plot_mir_count().

To exemplary compare the role of miR-21 in atherosclerosis (also see above) and lung cancer, the *Atherosclerosis* and *Lung cancer* data frames were combined with combine_df(). Afterward, the frequency of the shared terms of miR-21 were juxtaposed with compare_mir_terms().

The comparison of miRNAs in atherosclerosis and lung cancer was concluded by identifying potential biomarker miRNAs with calculate_score_biomarker(). calculate_score_biomarker() pinpointed abstracts containing at least five marker words for biomarker (*bio-marker, biological marker, biomarker, body fluid, bodyfluid, circulating, diagnostic, exosomal, exosomes, extracellular vesicles, plasma, serum, urinary, urine*). The choice of marker words for biomarker is empirical and reflects that *biomarker* miRNAs are usually *circulating* as *exosomes* or *extracellular vesicles* in easily accessible *bodyfluids* such as the *plasma, serum* or *urine* (15).

The seven most frequently reported miRNAs in these abstracts were visualized and compared with plot_mir_count(). The resulting frequency distribution was manually validated for selected miRNAs.

RESULTS

Accuracy of miRNA extraction

To validate miRetrieve's miRNA extraction, automatic miRNA name extraction without and with lettered suffixes

was compared to manual extraction from 200 randomly selected abstracts.

Without lettered suffixes, manual extraction yielded 351 miRNAs. Of these 351 miRNAs, miRetrieve correctly extracted 350 miRNAs, while missing 1 miRNA and falsely extracting 1 miRNA instead, which amounts to a recall and precision of 99.7%. With lettered suffixes, manual extraction totaled 370 miRNAs. Of these 370 miRNAs, miRetrieve correctly extracted 369 miRNAs, while missing 1 miRNA and falsely extracting 1 miRNA instead, corresponding to a recall and precision of 99.7% (Supplementary File S3). The discrepancy is due to a typing error that is easily compensated for by a human, as *miR-221/22* was extracted as *miR-221* and *miR-22*, when *miR-222* was meant and therefore missed ([16]).

Throughout, miRetrieve handled different miRNA spelling types and coerced them into one form, even within the same abstract (e.g. *MIRN125B*, *miR-125b* and *miR125b* to *miR-125b; miR-19a,-21,-126*[...] to *miR-19a, miR-21, miR-126* [...], Supplementary File S3). Furthermore, combined miRNA names such as miR-33a/b were split into miR-33a and miR-33b and counted accordingly (Supplementary File S3).

Thus, miRetrieve reliably detects and extracts miRNA names in abstracts, which provides a trustworthy foundation for subsequent analysis steps.

Accuracy of miRNA frequency distribution

Comparing miRetrieve's count of miRNA names to a distinct PubMed search per miRNA, miRetrieve arrives, on average, at a slightly higher miRNA count than PubMed.

When miRNA names carrying no lettered suffixes are extracted, miRetrieve identifies a higher count per miRNA than PubMed (Figure 1A), as indicated by a mean frequency score of 1.58.

Also, when miRNA names that carry lettered suffixes are extracted, miRetrieve again identifies a higher count per miRNA than PubMed (Figure 1B), as indicated by a mean frequency score of 1.82. This indicates that miRetrieve counts miRNAs accurately, while it reflects their occurrence in abstracts, on average, more precisely than PubMed.

Opposed to PubMed, miRetrieve recognizes miRNAs regardless of possible lettered suffixes during extraction. This offers the advantage that miRetrieve summarizes miRNAs with the same enumeration when lettered suffixes are ignored (e.g. *let-7a, let-7b, let-7c*, ... are summarized as *let-7*), which adds flexibility when the specific miRNA name is irrelevant/not known, or a general overview of a certain miRNA number is preferred. Additionally, miRetrieve extracts individual miRNAs from chained formats (such as *miR-221* and *miR-222* from *miR-221/222*, or *miR-33a* and *miR-33b* from *miR-33a/b*), which reflects miRNA occurrence more precisely.

Frequency-distribution of miRNAs in atherosclerosis

After validating miRetrieve's miRNA extraction and frequency distribution, miRNAs mentioned at least twice in an abstract were extracted to characterize the miRNA landscape in atherosclerosis.

Here, miR-155 is the most frequently mentioned miRNA, referenced at least twice in 80 abstracts (Figure 2). Next to miR-155, miR-126, miR-146, miR-21 and miR-33 are other frequently mentioned miRNAs in atherosclerosis, described in at least 50 abstracts each (Figure 2).

Of these abstracts, about 94% of abstracts, on average, report a relevant role of the respective miRNA in atherosclerosis, as evidenced by manual validation (Supplementary File S4). This demonstrates that the frequency distribution of miRNAs reflects their level of investigation in a research field, showing that frequently mentioned miRNAs are well investigated, whereas rarely mentioned miRNAs lack research and characterization vice versa.

miRNA extraction thresholding

In the approach above, the frequency distribution of miRNAs was based on a miRNA extraction threshold >1, meaning that per miRNA, only abstracts mentioning that miRNA at least twice were considered. Although this yielded relevant abstracts, it risks losing relevant abstracts that mention the respective miRNA only once.

Indeed, setting a miRNA extraction threshold >1 decreases the recall from 100% to, on average, 88% relevant abstracts, indicating a loss of relevant abstracts (Supplementary File S4). This decrease in recall, however, is contrasted by an increase in precision from 90% to, on average, >94% relevant abstracts for an extraction threshold >1 (Supplementary File S4). This implies that a higher miRNA extraction threshold increases precision for relevant abstracts, and therefore increases the chance to focus on relevant abstracts.

Term-associations of miR-155 in atherosclerosis

In atherosclerosis, miR-155 is the most frequently mentioned miRNA and thus the miRNA that has received most attention. Hence, its role in atherosclerosis is relevant (e.g. ([17–25])).

In atherosclerosis, miR-155 is often associated with terms related to inflammation (e.g. *inflammation*, *inflammatory response*, *tnf α*, *il 6*, *nf κb*) (Figure 3A,B). Furthermore, almost 30% of abstracts link miR-155 to macrophages (*macrophage*, *macrophages*) (Figure 3A), while about 20% and 10% of abstracts describe miR-155 alongside *miR-21* and *miR-126*, respectively (Figure 3B). Taken together, these terms indicate that miR-155 is a key player of *inflammation* in atherosclerosis, likely by affecting macrophage behavior. Additionally, miR-155 might be connected to miR-21 or miR-126.

Validating these propositions with a manual search reveals that miR-155 indeed drives *inflammation* by targeting SOCS1 ([26,27]) and by activating the NLRP3 inflammasome via the ERK1/2 pathway ([28]). Moreover, miR-155 polarizes *macrophages* to their pro-inflammatory M1 state ([29]), thus aggravating inflammation ([18,30]).

In addition, miR-155 correlates positively with *IL-6* and *TNF-α* ([31]), two other terms referenced in the associations. miR-155 increases both proteins ([32,33]) by targeting SOCS1 ([34]). Vice versa, *TNF-α* stimulates miR-155 expression ([25]) and releases miR-155 rich particles ([35]).
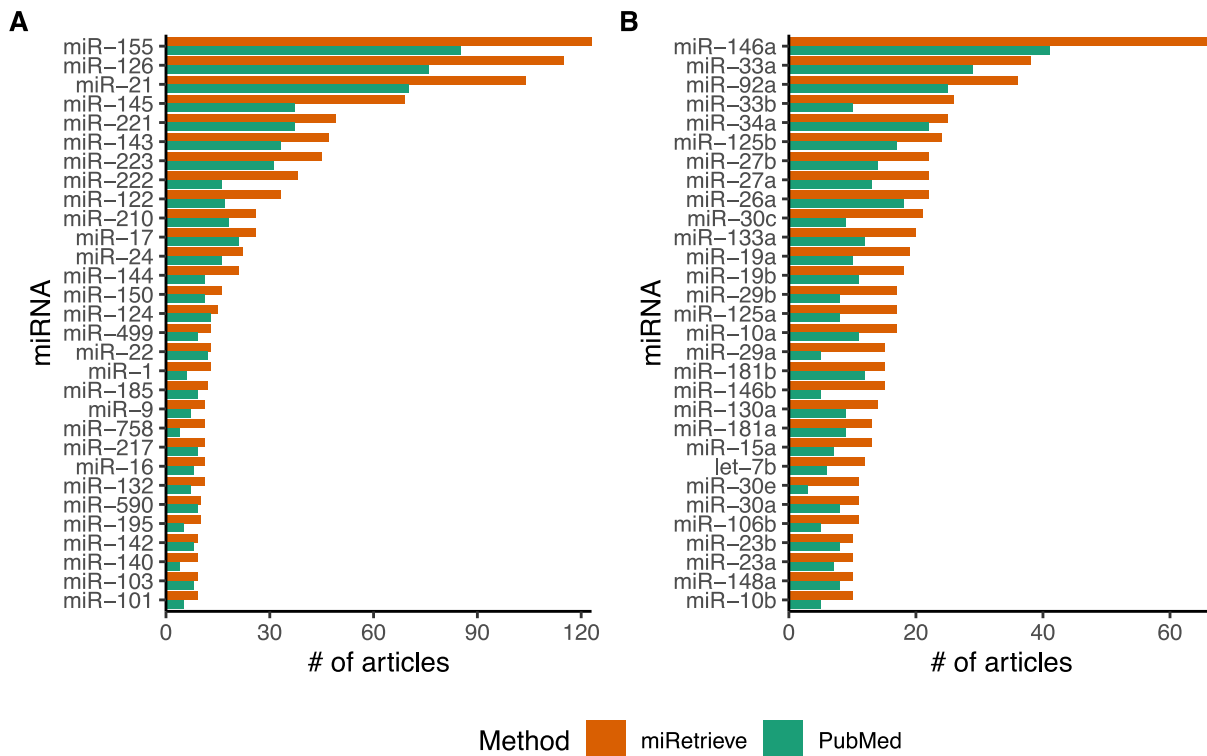
Figure 1. Comparison of frequency distribution of miRNAs as obtained by miRetrieve and PubMed. miRNA names without (**A**) and with lettered suffixes (**B**) were extracted from *atherosclerosis* abstracts. miRNA names were counted with miRetrieve and PubMed's search algorithm following the pattern '*miR-(number) atherosclerosis*', and top 30 miRNAs according to miRetrieve were visualized. Bargraphs indicate number of abstracts containing miRNAs as obtained by miRetrieve (orange) and PubMed's search algorithm (green).
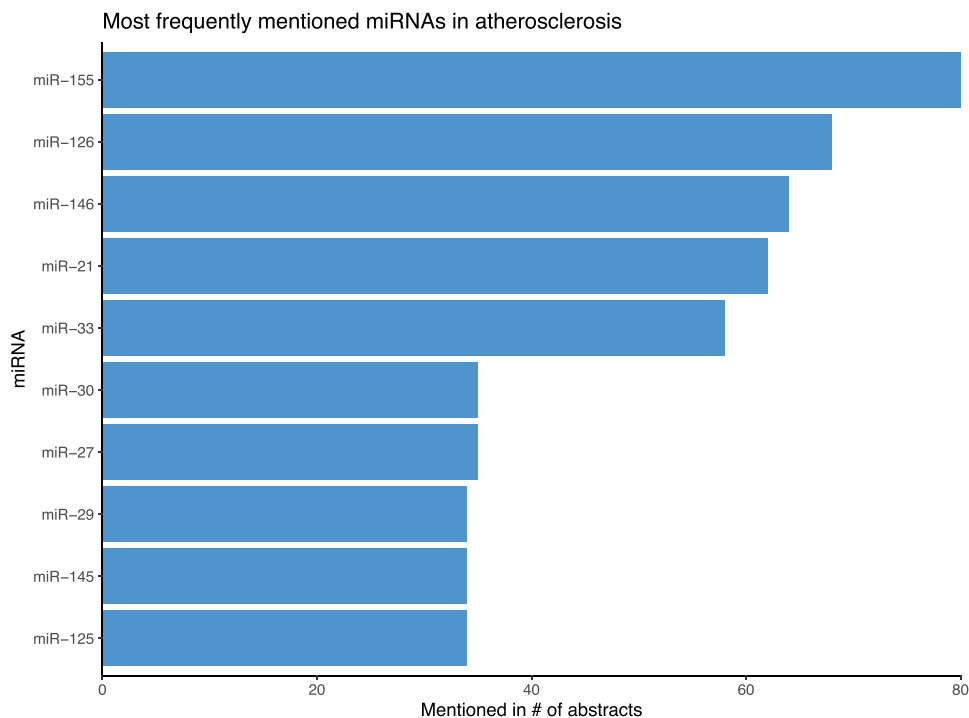

Figure 2. Most frequently mentioned miRNAs in atherosclerosis. *Atherosclerosis* abstracts were confined to research articles and miRNA names mentioned at least twice in an abstract were extracted. Ten most frequently mentioned miRNAs in atherosclerosis were plotted. Bargraphs indicate number of abstracts a miRNA was mentioned at least twice in.
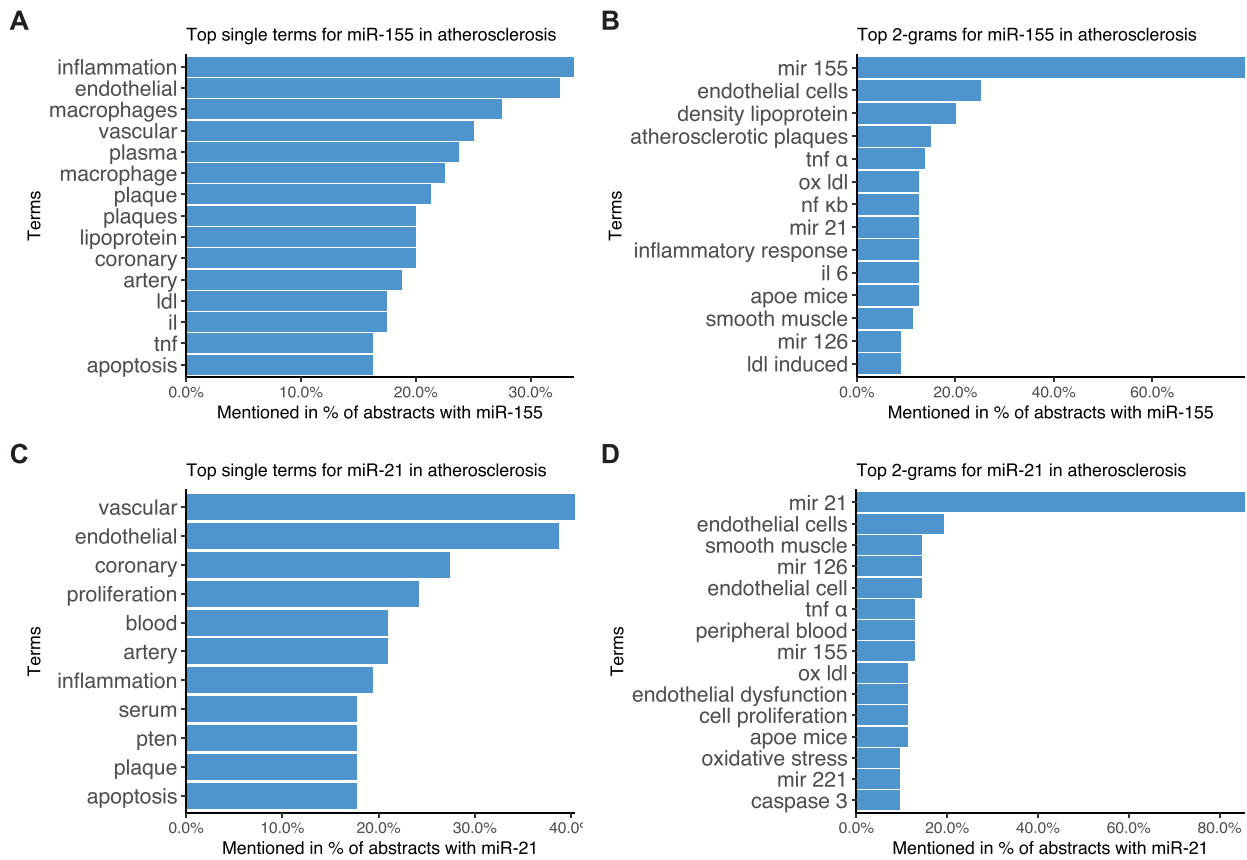
Figure 3. Term-associations of miR-155 and miR-21 in atherosclerosis. Most frequently single and two adjacent words associated with miR-155 (**A** and **B**) and miR-21 (**C** and **D**) in atherosclerosis were plotted, while common words and terms such as *a*, *is*, *significant*, or *western* were ignored. Additionally, field specific stop words such as *atherosclerosis* or *atherosclerotic* were added to aid readability. Bargraphs indicate number of abstracts mentioning single and two adjacent words relative to all abstracts mentioning the respective miRNA.

A similar reciprocal relationship exists between miR-155 and *NF-κB* (25,27).

Lastly, the association between miR-155 and *miR-21/miR-126* is explained by that fact that miR-155 and miR-21 are inversely expressed in atherosclerosis (36), whereas levels of miR-155 and miR-126 are jointly altered in atherosclerotic plaques (37). Furthermore, a mutual altered expression of all three miRNAs in atherosclerosis is described in at least two studies (38).

This demonstrates that miRetrieve's term-associations summarize the putative role of a miRNA in a focused and accurate manner. Additionally, miRetrieve reveals which miRNAs are conjointly differently expressed in a disease.

Term-associations of miR-21 in atherosclerosis

In atherosclerosis, miR-155 is related to terms of inflammation. However, as inflammation is one of the main drivers in atherosclerosis (39), these *inflammatory* signaling words might reflect a common mechanism, and thus be unspecific for miR-155. To investigate this possibility, terms frequently associated with miR-21, the fourth most frequently mentioned miRNA in atherosclerosis (Figure 2), are compared to terms associated with miR-155.

Opposed to miR-155, miR-21 is not as frequently associated with *inflammation* (35% of abstracts with miR-155, 18% of abstracts with miR-21), *macrophage* (20%, 9%) or *nf κb* (12%, 4%), while the frequency of *il 6* is similar (13%, 7%) (Figure 3C,D).

This suggests that miR-155 plays a more active role in inflammation than miR-21, which is confirmed by a manual search: although an absence of miR-21 in macrophages results in inflammation (40) and serum miR-21 correlates positively with atherosclerotic inflammation (41), most studies only associate miR-21 with inflammation, but do not demonstrate a clear mechanistic link (e.g. (42)).

In contrast, miR-21 frequently associates with *apoptosis (18%)*, *pten (16%)* and *mir 221 (11%)*. This hints that miR-21 regulates apoptosis in atherosclerosis and is connected to PTEN, while its expression might be associated with miR-221 (Figure 3C,D). Indeed, miR-21 targets *PTEN* (43) and correlates negatively with *PTEN* in patients (44). Amongst others, this prevents *apoptosis* (45), either by stimulating the PI3K pathway (46) or by increasing the expression of Bcl-2 (47). Additionally, miR-21 averts *apoptosis* by targeting PDCD4 (48) and MKK3, subsequently decreasing p38-CHOP and JNK signaling (40).

This illustrates that miRetrieve's term-associations are not only meaningful, but also characterize the role of a miRNA independent from disease state.

Targets of miR-155 and miR-21 in atherosclerosis

miRNAs exert their main function by translational repression of complimentary mRNA targets (1). Thus, the characterization of miR-155 and miR-21 in atherosclerosis is completed by viewing their validated targets using miRTarBase (5). As only abstracts reporting miRNAs in atherosclerosis are present, these validated targets are confined to miRNA–mRNA interactions in atherosclerosis.

According to miRTarBase, miR-155 targets >20 genes in atherosclerosis, whereas miR-21 targets 7 genes (Figure 4). Among these, miR-155 targets *ICAM1*, *VCAM1* (49), *NFKB1* and *NSD3* (50), which confirm miR-155's inflammatory role in atherosclerosis. In addition, *PTEN* is listed as a target of miR-21 and also encountered in miR-21's term associations (51).

However, miRNA listings in miRTarBase are incomplete: While miRTarBase lists only one publication each of miR-155 to target SOCS1 (34) and miR-21 to target PTEN (51), notably more studies confirm miR-155 to target SOCS1 (e.g. (26,27)) or miR-21 to target PTEN (e.g. (43,52)) in atherosclerosis. Moreover, miR-21 has been confirmed to target other proteins that are not listed, e.g. SMAD7 (53) or TIMP3 (54).

This discrepancy highlights the dependency of external databases on manual curation: Regular updates are necessary to cover a wide range of known, field specific miRNA–mRNA interactions without bias. As this is barely feasible in practice, these results must be interpreted with care.

In conclusion, miRetrieve visualizes field specific miRNA–mRNA interactions by using external databases such as miRTarBase. Although the full scope of known miRNA–mRNA interactions per field is not covered, this approach demonstrates the potential to provide insights into disease specific miRNA behavior.

Cross-field comparison of miRNAs in atherosclerosis to miRNAs in lung cancer

Most frequently mentioned miRNAs in lung cancer. As many miRNAs exists and miRNAs are believed to influence about any disease, it is hoped that a combination of different miRNAs might provide a signature distinctive for any disease. A unique signature could not only help in diagnosing a disease but also aid in understanding its individual pathophysiology. As miRetrieve's frequency distribution reflects the degree of investigation of miRNAs and the number of papers reporting their substantial role, it is interesting to compare which miRNAs are well investigated between diseases, such as atherosclerosis and lung cancer.

In lung cancer, the most frequently mentioned miRNAs are miR-21, miR-200 and miR-34 (Figure 5).

However, it is notable that 7 of 10 miRNAs are shared among the most frequently mentioned miRNAs in atherosclerosis and lung cancer, namely miR-21, miR-30, miR-29, miR-146, miR-125, miR-155 and miR-145 (Figures 2, 5). As the frequency distribution of miRNAs reflects the knowledge and their substantial role in a disease, these seven miRNAs are unlikely to explain the unique

pathophysiology or to serve as a specific miRNA signature in either disease.

This is an example of how miRetrieve can identify similarities and dissimilarities between disease driving miRNAs. This might aid in distinguishing miRNAs that are involved in general processes from miRNAs that contribute to a disease's unique pathophysiology.

Shared terms of miR-21 in atherosclerosis and lung cancer

It is hypothesized that one miRNA exerts different functions depending on environment (4). This raises the question if one miRNA plays different roles in atherosclerosis and lung cancer. As miR-21 is the most frequently mentioned miRNA in lung cancer and already characterized in atherosclerosis above, its role in both diseases shall be contrasted.

In atherosclerosis and lung cancer, miR-21 is associated with *apoptosis*, *pi3k akt* and *nf κb* (Figure 6A,B). Additionally, miR-21 is related to miR-155 and miR-126 in both fields (Figure 6B).

In fact, miR-21 is anti-*apoptotic* in lung cancer (55) by targeting PTEN (56), just as in atherosclerosis (see above). Additionally, miR-21 prevents apoptosis in lung cancer by targeting KIBRA (57) and the PI3K/Akt/NF-κB signaling pathway (58). However, miR-21 also induces apoptosis in lung cancer by targeting hMSH2 (59).

Regarding *nf κb*, miR-21 either suppresses (60) or increases (61) NF-κB signaling in both entities (see above for atherosclerosis).

Concerning the connection between miR-21, miR-126 and miR-155, miR-21 and miR-155 (62) as well as miR-21 and miR-126 (63) are commonly concomitantly aberrantly expressed in lung cancer, just as in atherosclerosis (see above).

This is another example that the term associations revealed by miRetrieve are accurate. Furthermore, miRetrieve uncovers general functions of miRNAs, which might help to sort miRNAs into larger functional classes. Additionally, miRetrieve reveals relevant miRNA–miRNA co-expressions, which might hint at regulatory commonalities independent from environment.

miRNAs as biomarker in atherosclerosis and lung cancer

miRNAs are highly stable in body fluids and thus hold potential to serve as biomarkers (64). The more specific a biomarker, the better it differentiates between diseases. Thus, potential biomarker miRNAs in atherosclerosis and lung cancer shall be contrasted and their specificity estimated.

Considering the 8 most frequent biomarker miRNAs in atherosclerosis and lung cancer, miR-155 is only listed in atherosclerosis, whereas miR-200 is only mentioned in lung cancer. Therefore, miR-155 and miR-200 might be more indicative biomarker of atherosclerosis and lung cancer, respectively. Furthermore, at least four miRNAs (miR-126, miR-21, miR-30, miR-125) hold potential to serve as biomarkers in both entities (Figure 7).

Manual validation confirms that miR-155 is a promising biomarker for atherosclerosis (65), whereas miR-200 is a
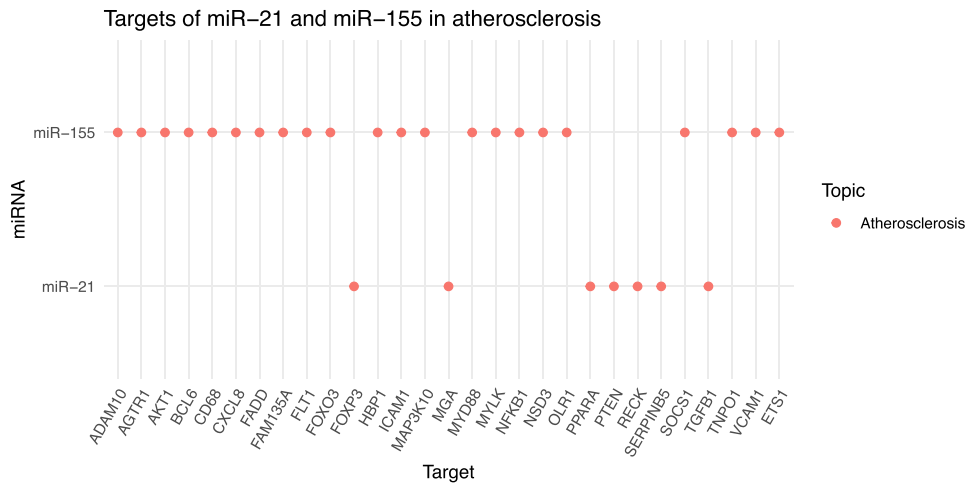
Figure 4. Validated targets of miR-155 and miR-21 in atherosclerosis according to miRTarBase (5). PubMed IDs of *atherosclerosis* abstracts containing miR-155 and miR-21 were matched to PubMed IDs of miRTarBase, and validated targets of miR-155 and miR-21 were displayed. Each dot indicates one study validating the miRNA–mRNA interaction.
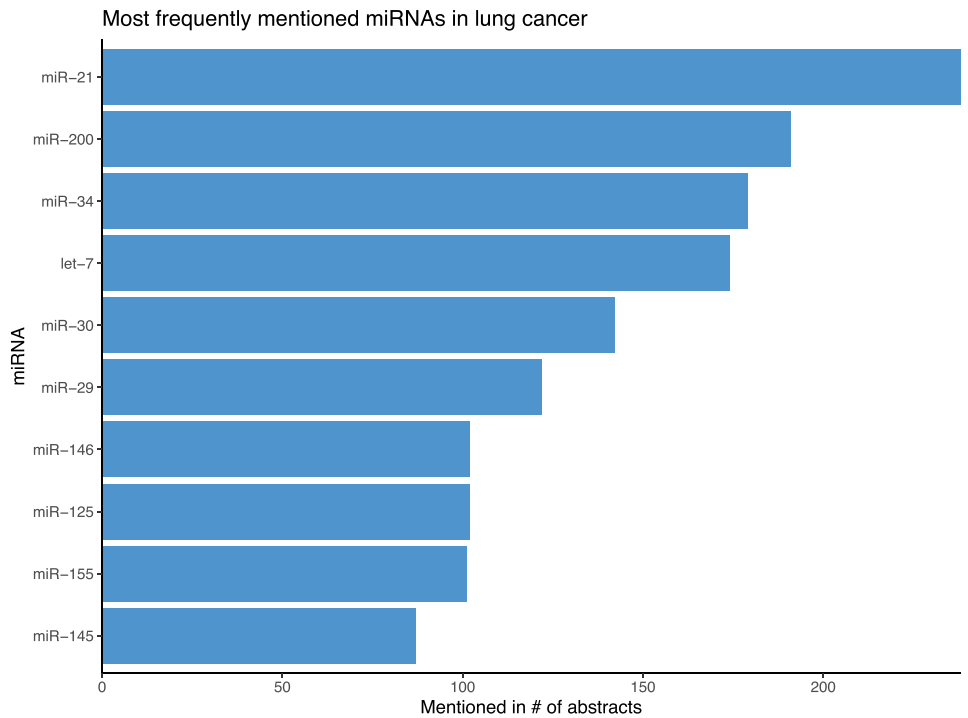

Figure 5. Most frequently mentioned miRNAs in lung cancer. *Lung cancer* abstracts were confined to research articles and miRNA names mentioned at least twice in an abstract were extracted. Ten most frequently mentioned miRNAs in lung cancer were plotted. Bargraphs indicate number of abstracts a miRNA was mentioned at least twice in.

well-confirmed biomarker for lung cancer (66). Of note, miRetrieve lists 16 and 27 abstracts of miR-155 and miR-200 as biomarker in atherosclerosis and lung cancer, respectively, whereas only 9 and 19 abstracts confirm these miRNAs as biomarker (Supplementary File S4), which corresponds to a recall between 50% and 60%.

Regarding miR-126, miR-21, miR-30 and miR-125, all four miRNAs are indeed established biomarkers for either atherosclerosis (e.g. miR-126: (67), miR-21: (41), miR-30:

(68), miR-125: (69)) or lung cancer (e.g. miR-126: (70), miR-21: (71), miR-30: (72), miR-125: (73)).

This demonstrates that miRetrieve identifies established biomarker miRNAs in a disease. Even though the abstracts listed by miRetrieve correspond to confirmed miRNAs as biomarker in only 50–60% of cases, this cross-disease comparison allows a first estimate on a miRNA's specificity or commonality as a biomarker, which can guide further studies.
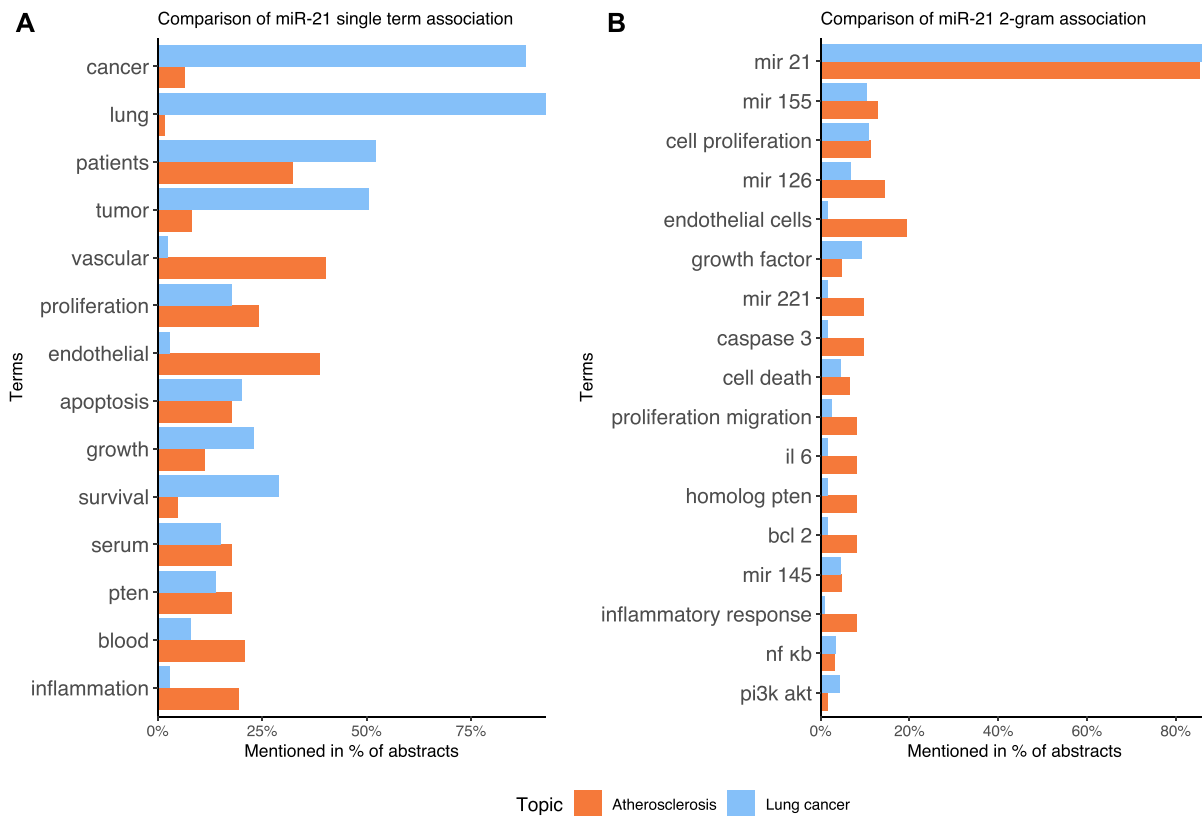
Figure 6. Shared terms of miR-21 in atherosclerosis and lung cancer. Most frequently single and two adjacent words associated with miR-21, shared in atherosclerosis and lung cancer, were juxtaposed, while common words and terms such as *a*, *is*, *significant* or *western* were ignored. Bargraphs indicate relative number of abstracts mentioning single and two adjacent words as compared to all abstracts mentioning miR-21 in *atherosclerosis* (orange) or *lung cancer* (blue), respectively.

DISCUSSION

Here, we introduced miRetrieve, a tool to summarize information on miRNAs from thousands of abstracts within a short amount of time. By demonstrating a typical miRetrieve workflow, we showed how miRetrieve effectively condenses the wealth of information on miRNAs, such as their degree of investigation, function in a disease, their larger functional class, or their suitability as a biomarker. Additionally, the accuracy of these results was confirmed by manual validation. This indicates that miRetrieve can not only help to understand the unique pathophysiology of a disease, but also help to improve our understanding of miRNAs in general.

One core function of miRetrieve is to extract differently spelled miRNA names and unify them into one format. Even though a wide variety of miRNA spellings is recognized, not every format can be accounted for, which might cause miRetrieve to miss miRNAs. Problematic is especially the extraction of miRNA names from chained patterns, such as *miR-146–155*, as these might conflict with the standard nomenclature (e.g. *miR-24–1*). This is aggravated by the fact that the possibilities of enumerating miRNAs with slashes, dashes and commas are almost endless. However, manual validation showed that miRetrieve extracted >99% of all miRNA names correctly. Another potential drawback is the extraction of miRNA

names from abstracts without proven biological function in the study at hand. Here, one approach is to extract miRNAs that are referenced more than once in an abstract. While this approach bears the risk of ignoring relevant miRNAs that are seldomly mentioned, it improves precision, as demonstrated above.

Once extracted, miRetrieve characterizes miRNAs by displaying the terms they are frequently associated with. Whereas these miRNA-term associations depend on a sufficient amount of provided abstracts, they are also suggestive and potentially incomplete. Yet, we demonstrated that miRetrieve's term-associations are not only meaningful but also accurate and reflect a miRNA's role independent from disease. Additionally, this can be extended to compare a miRNA's role across different diseases, which may aid in sorting miRNAs into larger functional classes. Lastly, we showed that these miRNA-term associations point at concurrently differently expressed miRNAs, which might help to uncover co-regulated miRNA-miRNA networks and stable miRNA signatures. Hence, miRetrieve's term-associations provide a reliable foundation to generate new hypotheses and guide further studies.

Next to associating miRNAs with terms, miRetrieve also dissects which miRNAs are frequently related to defined terms, which can be used, amongst others, to identify
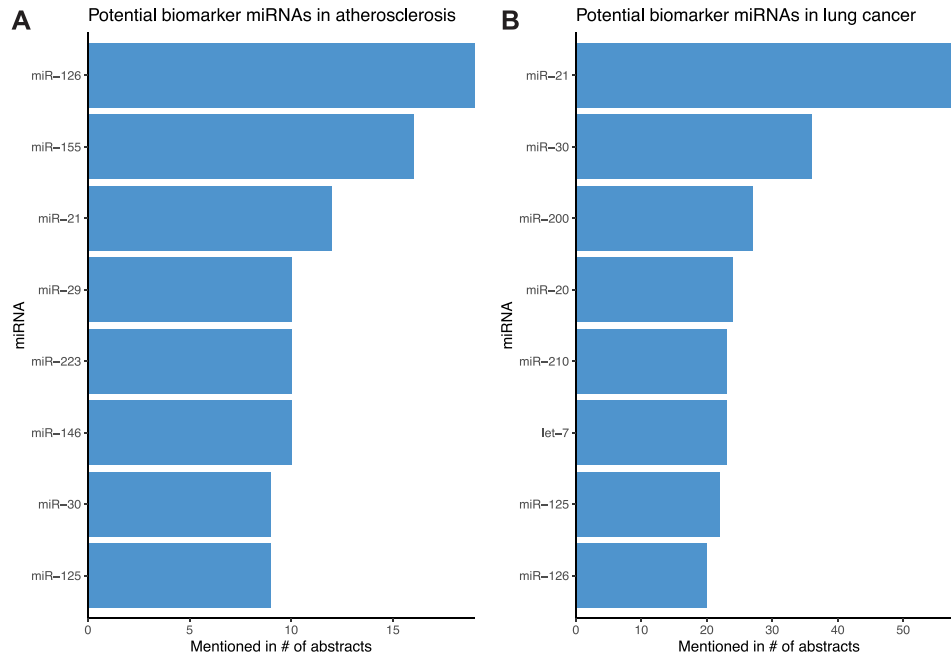
Figure 7. Potential biomarker miRNAs in atherosclerosis and lung cancer. Abstracts containing at least five marker words for biomarker (*bio-marker, biological marker, biomarker, body fluid, bodyfluid, circulating, diagnostic, exosomal, exosomes, extracellular vesicles, plasma, serum, urinary, urine*) in *atherosclerosis* (**A**) and *lung cancer* (**B**) were identified, and the eight most frequently reported miRNAs in these abstracts were visualized. Bargraphs indicate number of abstracts in *atherosclerosis* (A) or *lung cancer* (B) mentioning a miRNA as a potential biomarker.

potential biomarker miRNAs. As most biomarker studies investigate multiple miRNAs at once, they also mention miRNAs that fail to be suitable biomarkers, which explains miRetrieve's diminished accuracy when identifying biomarker miRNAs from abstracts. Furthermore, a miRNA's suitability as a biomarker depends on various factors, such as individual expression levels, the direction of expression alteration, or combination of several miRNAs to a signature. While these factors are not covered by miRetrieve, miRetrieve allows to quickly skim for candidate miRNAs that may serve as biomarker in a disease. This helps in estimating a miRNA's specificity as a biomarker, and can subsequently direct further studies.

All miRNA extractions, term- and class-associations are generated directly from text without lexico-syntactic rules or stochastic models. This facilitates miRetrieve's interpretability and allows miRetrieve to extract new miRNAs or recognize a differentiated scope of miRNA-associations independent from defined named entities, while the option to combine all analysis steps with rule-based tools such as miRSel (6), miRTex (7) and emiRIT (8) still remains.

Apart from characterizing miRNAs using generated associations, miRetrieve visualizes validated miRNA–mRNA interactions confined to the disease at hand by incorporating databases such as miRTarBase (5). However, up until now, these databases depend on manual curation and require extensive updates, which is why they do not cover the full scope of known miRNA–mRNA interactions. Nonetheless, visualizing miRNA–mRNA interactions for specific diseases may direct the investigation of other miRNA–mRNA interactions in similar fields more effectively than predicted miRNA–mRNA interactions that do not take environmental factors into consideration. Additionally, disease-specific miRNA–mRNA interactions may facilitate our understanding of what drives miRNA-target selectivity, and consequently improve miRNA-target predictions. Lastly, hope remains that this approach is reinforced further once validated miRNA–mRNA interactions are automatically extracted from text in the future.

CONCLUSION

In conclusion, miRetrieve provides accurate insights on miRNAs from huge amounts of text within a short amount of time. These insights aid in characterizing miRNAs in one or many fields, which may facilitate and accelerate the generation of new hypotheses on how miRNAs modify diseases, ultimately improving our understanding of miRNAs and diseases alike.

DATA AVAILABILITY

miRetrieve is an R package (9) released under a GPL-3 license and freely available from GitHub (https://github.com/JulFriedrich/miRetrieve) and CRAN.

The most important miRetrieve functions are available as a browser based Shiny application (10) under (https://miretrieve.shinyapps.io/miRetrieve/).

The raw abstracts for the use case are released as a GitHub tag at https://github.com/JulFriedrich/miRetrieve/releases/tag/v1.3.2, while Supplementary Data are available at https://github.com/JulFriedrich/miRetrieve-paper.

SUPPLEMENTARY DATA

Supplementary Data are available at NARGAB Online.

FUNDING

REFERENCES

1. Ha,M. and Kim,V.N. (2014) Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.*, **15**, 509–524.
2. Ambros,V. (2001) microRNAs: tiny regulators with great potential. *Cell*, **107**, 823–826.
3. Lu,Q., Wu,R., Zhao,M., Garcia-Gomez,A. and Ballestar,E. (2019) miRNAs as therapeutic targets in inflammatory disease. *Trends Pharmacol. Sci.*, **40**, 853–865.
4. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
5. Huang,H.Y., Lin,Y.C., Li,J., Huang,K.Y., Shrestha,S., Hong,H.C., Tang,Y., Chen,Y.G., Jin,C.N., Yu,Y. *et al.* (2020) miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res.*, **48**, D148–D154.
6. Naeem,H., Küffner,R., Csaba,G. and Zimmer,R. (2010) miRSel: automated extraction of associations between microRNAs and genes from the biomedical literature. *BMC Bioinform.*, **11**, 135.
7. Li,G., Ross,K.E., Arighi,C.N., Peng,Y., Wu,C.H. and Vijay-Shanker,K. (2015) miRTex: a text mining system for miRNA-Gene relation extraction. *PLoS Comput. Biol.*, **11**, e1004391.
8. Roychowdhury,D., Gupta,S., Qin,X., Arighi,C.N. and Vijay-Shanker,K.(2021) emiRIT: a text-mining based resource for microRNA information. *Database (Oxford)*, **2021**, baab031.
9. R Core Team (2019) In: R: A Language and Environment for Statistical Computing.
10. Chang,W., Cheng,J., Allaire,J., Xie,Y. and McPherson,J. (2019) In: shiny: Web Application Framework for R.
11. Griffiths-Jones,S. (2006) miRBase: the microRNA sequence database. *Methods Mol. Biol.*, **342**, 129–138.
12. Silge,J. and Robinson,D. (2016) tidytext: text mining and analysis using tidy data principles in R. *Open J.*, **1**, 37.
13. Wickham,H. (2016) In: *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, NY.
14. Pedersen,T.L. (2019) In: R package version 1.0.0 ed.
15. La Marca,V. and Fierabracci,A. (2017) Insights into the diagnostic potential of extracellular vesicles and their miRNA signature from liquid biopsy as early biomarkers of diabetic micro/macrovascular complications. *Int. J. Mol. Sci.*, **18**, 1974.
16. Mandraffino,G., Aragona,C.O., Cairo,V., Scuruchi,M., Lo Gullo,A., D'Ascola,A., Alibrandi,A., Loddo,S., Quartuccio,S., Morace,C. *et al.* (2017) Circulating progenitor cells in hypertensive subjects: effectiveness of a treatment with olmesartan in improving cell number and miR profile in addition to expected pharmacological effects. *PLoS One*, **12**, e0173030.
17. Donners,M.M., Wolfs,I.M., Stoger,L.J., van der Vorst,E.P., Pottgens,C.C., Heymans,S., Schroen,B., Gijbels,M.J. and de Winther,M.P. (2012) Hematopoietic miR155 deficiency enhances atherosclerosis and decreases plaque stability in hyperlipidemic mice. *PLoS One*, **7**, e35877.
18. Du,F., Yu,F., Wang,Y., Hui,Y., Carnevale,K., Fu,M., Lu,H. and Fan,D. (2014) MicroRNA-155 deficiency results in decreased macrophage inflammation and attenuated atherogenesis in apolipoprotein E-deficient mice. *Arterioscler. Thromb. Vasc. Biol.*, **34**, 759–767.
19. Johnson,C., Drummer,C.t., Virtue,A., Gao,T., Wu,S., Hernandez,M., Singh,L., Wang,H. and Yang,X.F. (2018) Increased expression of resistin in microrna-155-Deficient white adipose tissues may be a possible driver of metabolically healthy obesity transition to classical obesity. *Front. Physiol.*, **9**, 1297.
20. Li,X., Kong,D., Chen,H., Liu,S., Hu,H., Wu,T., Wang,J., Chen,W., Ning,Y., Li,Y. *et al.* (2016) miR-155 acts as an anti-inflammatory factor in atherosclerosis-associated foam cell formation by repressing calcium-regulated heat stable protein 1. *Sci. Rep.*, **6**, 21789.
21. Li,Y., Peng,Y., Yao,S., Chen,L., Li,S., Wang,M., Chen,S., Chen,X., Deng,F., Hu,W. *et al.* (2020) Association of miR-155 and angiotensin receptor type 1 polymorphisms with the risk of ischemic stroke in a chinese population. *DNA Cell Biol.*, **39**, 92–104.
22. Lv,J., Yang,L., Guo,R., Shi,Y., Zhang,Z. and Ye,J. (2017) Ox-LDL-Induced microrna-155 promotes autophagy in human endothelial cells via repressing the Rheb/mTOR pathway. *Cell. Physiol. Biochem.*, **43**, 1436–1448.
23. Ma,J., Yang,S., Ma,A., Pan,X., Wang,H., Li,N., Liu,S. and Wu,M. (2017) Expression of miRNA-155 in carotid atherosclerotic plaques of apolipoprotein e knockout (ApoE(-/-)) mice and the interventional effect of rapamycin. *Int. Immunopharmacol.*, **46**, 70–74.
24. Nazari-Jahantigh,M., Wei,Y., Noels,H., Akhtar,S., Zhou,Z., Koenen,R.R., Heyll,K., Gremse,F., Kiessling,F., Grommes,J. *et al.* (2012) MicroRNA-155 promotes atherosclerosis by repressing bcl6 in macrophages. *J. Clin. Invest.*, **122**, 4190–4202.
25. Park,M., Choi,S., Kim,S., Kim,J., Lee,D.K., Park,W., Kim,T., Jung,J., Hwang,J.Y., Won,M.H. *et al.* (2019) NF-kappaB-responsive miR-155 induces functional impairment of vascular smooth muscle cells by downregulating soluble guanylyl cyclase. *Exp. Mol. Med.*, **51**, 1–12.
26. Li,S., Sun,Y., Zhong,L., Xiao,Z., Yang,M., Chen,M., Wang,C., Xie,X. and Chen,X. (2018) The suppression of ox-LDL-induced inflammatory cytokine release and apoptosis of HCAECs by long non-coding RNA-MALAT1 via regulating microRNA-155/SOCS1 pathway. *Nutr. Metab. Cardiovasc. Dis.*, **28**, 1175–1187.
27. Zhang,H., Zhao,Z., Pang,X., Yang,J., Yu,H., Zhang,Y., Zhou,H. and Zhao,J. (2017) Genistein protects against ox-ldl-induced inflammation through microrna-155/SOCS1-Mediated repression of NF-kB signaling pathway in HUVECs. *Inflammation*, **40**, 1450–1459.
28. Yin,R., Zhu,X., Wang,J., Yang,S., Ma,A., Xiao,Q., Song,J. and Pan,X. (2019) MicroRNA-155 promotes the ox-LDL-induced activation of NLRP3 inflammasomes via the ERK1/2 pathway in THP-1 macrophages and aggravates atherosclerosis in ApoE-/- mice. *Ann Palliat Med*, **8**, 676–689.
29. He,S., Wu,C., Xiao,J., Li,D., Sun,Z. and Li,M. (2018) Endothelial extracellular vesicles modulate the macrophage phenotype: potential implications in atherosclerosis. *Scand. J. Immunol.*, **87**, e12648.
30. Zhang,R.N., Zheng,B., Li,L.M., Zhang,J., Zhang,X.H. and Wen,J.K. (2014) Tongxinluo inhibits vascular inflammation and neointimal hyperplasia through blockade of the positive feedback loop between miR-155 and TNF-alpha. *Am. J. Physiol. Heart Circ. Physiol.*, **307**, H552–H562.
31. Kazimierczyk,E., Eljaszewicz,A., Zembko,P., Tarasiuk,E., Rusak,M., Kulczynska-Przybik,A., Lukaszewicz-Zajac,M., Kaminski,K., Mroczko,B., Szmitkowski,M. *et al.* (2019) The relationships among monocyte subsets, miRNAs and inflammatory cytokines in patients with acute myocardial infarction. *Pharmacol. Rep.*, **71**, 73–81.
32. Zhu,M., Wei,Y., Geissler,C., Abschlag,K., Corbalan Campos,J., Hristov,M., Mollmann,J., Lehrke,M., Karshovska,E. and Schober,A. (2017) Hyperlipidemia-Induced microrna-155-5p improves beta-Cell function by targeting mafb. *Diabetes*, **66**, 3072–3084.
33. Ma,Y.L., Ma,Z.J., Wang,M., Liao,M.Y., Yao,R. and Liao,Y.H. (2015) MicroRNA-155 induces differentiation of RAW264.7 cells into dendritic-like cells. *Int. J. Clin. Exp. Pathol.*, **8**, 14050–14062.
34. Ye,J., Guo,R., Shi,Y., Qi,F., Guo,C. and Yang,L. (2016) miR-155 regulated inflammation response by the SOCS1-STAT3-PDCD4 axis in atherogenesis. *Mediators Inflamm.*, **2016**, 8060182.
35. Alexy,T., Rooney,K., Weber,M., Gray,W.D. and Searles,C.D. (2014) TNF-alpha alters the release and transfer of microparticle-encapsulated miRNAs from endothelial cells. *Physiol. Genomics*, **46**, 833–840.
36. Kalani,M., Hodjati,H., Sajedi Khanian,M. and Doroudchi,M. (2016) Lactobacillus acidophilus increases the Anti-apoptotic micro RNA-21 and decreases the Pro-inflammatory micro RNA-155 in the LPS-Treated human endothelial cells. *Probiotics Antimicrob. Proteins*, **8**, 61–72.
37. Leistner,D.M., Boeckel,J.N., Reis,S.M., Thome,C.E., De Rosa,R., Keller,T., Palapies,L., Fichtlscherer,S., Dimmeler,S. and Zeiher,A.M. (2016) Transcoronary gradients of vascular miRNAs and coronary atherosclerotic plaque characteristics. *Eur. Heart J.*, **37**, 1738–1749.
38. Park,M.Y., Herrmann,S.M., Saad,A., Widmer,R.J., Tang,H., Zhu,X.Y., Lerman,A., Textor,S.C. and Lerman,L.O. (2015)

Circulating and renal vein levels of microRNAs in patients with renal artery stenosis. *Nephrol. Dial. Transplant.*, **30**, 480–490.

39. Poston,R.N. (2019) Atherosclerosis: integration of its pathogenesis as a self-perpetuating propagating inflammation: a review. *Cardiovasc Endocrinol Metab*, **8**, 51–61.

40. Canfran-Duque,A., Rotllan,N., Zhang,X., Fernandez-Fuertes,M., Ramirez-Hidalgo,C., Araldi,E., Daimiel,L., Busto,R., Fernandez-Hernando,C. and Suarez,Y. (2017) Macrophage deficiency of miR-21 promotes apoptosis, plaque necrosis, and vascular inflammation during atherogenesis. *EMBO Mol. Med.*, **9**, 1244–1262.

41. Darabi,F., Aghaei,M., Movahedian,A., Elahifar,A., Pourmoghadas,A. and Sarrafzadegan,N. (2017) Association of serum microRNA-21 levels with visfatin, inflammation, and acute coronary syndromes. *Heart Vessels*, **32**, 549–557.

42. Alexandru,N., Andrei,E., Safciuc,F., Dragan,E., Balahura,A.M., Badila,E. and Georgescu,A. (2020) Intravenous administration of allogenic cell-derived microvesicles of healthy origins defend against atherosclerotic cardiovascular disease development by a direct action on endothelial progenitor cells. *Cells*, **9**, 423.

43. Zhu,J., Liu,B., Wang,Z., Wang,D., Ni,H., Zhang,L. and Wang,Y. (2019) Exosomes from nicotine-stimulated macrophages accelerate atherosclerosis through miR-21-3p/PTEN-mediated VSMC migration and proliferation. *Theranostics*, **9**, 6901–6919.

44. Nariman-Saleh-Fam,Z., Vahed,S.Z., Aghaee-Bakhtiari,S.H., Daraei,A., Saadatian,Z., Kafil,H.S., Yousefi,B., Eyvazi,S., Khaheshi,I., Parsa,S.A. *et al.* (2019) Expression pattern of miR-21, miR-25 and PTEN in peripheral blood mononuclear cells of patients with significant or insignificant coronary stenosis. *Gene*, **698**, 170–178.

45. Shen,Z. and She,Q. (2018) Association between the deletion allele of ins/del polymorphism (Rs145204276) in the promoter region of GAS5 with the risk of atherosclerosis. *Cell. Physiol. Biochem.*, **49**, 1431–1443.

46. Yang,Q., Yang,K. and Li,A.Y. (2015) Trimetazidine protects against hypoxia-reperfusion-induced cardiomyocyte apoptosis by increasing microRNA-21 expression. *Int. J. Clin. Exp. Pathol.*, **8**, 3735–3741.

47. Yang,Q., Yang,K. and Li,A. (2014) microRNA-21 protects against ischemia-reperfusion and hypoxia-reperfusion-induced cardiocyte apoptosis via the phosphatase and tensin homolog/Akt-dependent mechanism. *Mol Med Rep*, **9**, 2213–2220.

48. Shang,Y.Y., Fang,N.N., Wang,F., Wang,H., Wang,Z.H., Tang,M.X., Peng,J., Zhang,Y., Zhang,W. and Zhong,M. (2015) MicroRNA-21, induced by high glucose, modulates macrophage apoptosis via programmed cell death 4. *Mol Med Rep*, **12**, 463–469.

49. Zhu,N., Zhang,D., Chen,S., Liu,X., Lin,L., Huang,X., Guo,Z., Liu,J., Wang,Y., Yuan,W. *et al.* (2011) Endothelial enriched microRNAs regulate angiotensin II-induced endothelial inflammation and migration. *Atherosclerosis*, **215**, 286–293.

50. Huang,R.S., Hu,G.Q., Lin,B., Lin,Z.Y. and Sun,C.C. (2010) MicroRNA-155 silencing enhances inflammatory response and lipid uptake in oxidized low-density lipoprotein-stimulated human THP-1 macrophages. *J. Investig. Med.*, **58**, 961–967.

51. Ji,R., Cheng,Y., Yue,J., Yang,J., Liu,X., Chen,H., Dean,D.B. and Zhang,C. (2007) MicroRNA expression signature and antisense-mediated depletion reveal an essential role of MicroRNA in vascular neointimal lesion formation. *Circ. Res.*, **100**, 1579–1588.

52. Chen,L.Y., Wang,X., Qu,X.L., Pan,L.N., Wang,Z.Y., Lu,Y.H. and Hu,H.Y. (2019) Activation of the STAT3/microRNA-21 pathway participates in angiotensin II-induced angiogenesis. *J. Cell. Physiol.*, **234**, 19640–19654.

53. Wang,J., Gao,Y., Duan,L., Wei,S., Liu,J., Tian,L., Quan,J., Zhang,Q. and Yang,J. (2017) Metformin ameliorates skeletal muscle insulin resistance by inhibiting miR-21 expression in a high-fat dietary rat model. *Oncotarget*, **8**, 98029–98039.

54. Dai,J., Chen,W., Lin,Y., Wang,S., Guo,X. and Zhang,Q.Q. (2017) Exposure to concentrated ambient fine particulate matter induces vascular endothelial dysfunction via miR-21. *Int. J. Biol. Sci.*, **13**, 868–877.

55. Ono,M., Yamada,K., Avolio,F., Afzal,V., Bensaddek,D. and Lamond,A.I. (2015) Targeted knock-down of miR21 primary transcripts using snoMEN vectors induces apoptosis in human cancer cell lines. *PLoS One*, **10**, e0138668.

56. Li,X., Zang,A., Jia,Y., Zhang,J., Fan,W., Feng,J., Duan,M., Zhang,L., Huo,R., Jiao,J. *et al.* (2016) Triptolide reduces proliferation and enhances apoptosis of human non-small cell lung cancer cells through PTEN by targeting miR-21. *Mol Med Rep*, **13**, 2763–2768.

57. An,Y., Zhang,Q., Li,X., Wang,Z., Li,Y. and Tang,X. (2018) Upregulated microRNA miR-21 promotes the progression of lung adenocarcinoma through inhibition of KIBRA and the hippo signaling pathway. *Biomed. Pharmacother.*, **108**, 1845–1855.

58. Wang,T., Cai,Z., Hong,G., Zheng,G., Huang,Y., Zhang,S. and Dai,J. (2017) MicroRNA21 increases cell viability and suppresses cellular apoptosis in nonsmall cell lung cancer by regulating the PI3K/Akt signaling pathway. *Mol. Med. Rep.*, **16**, 6506–6511.

59. Song,Y., Zuo,Y., Qian,X.L., Chen,Z.P., Wang,S.K., Song,L. and Peng,L.P. (2017) Inhibition of microrna-21-5p promotes the radiation sensitivity of non-small cell lung cancer through HMSH2. *Cell. Physiol. Biochem.*, **43**, 1258–1272.

60. Gong,K., Guo,G., Gerber,D.E., Gao,B., Peyton,M., Huang,C., Minna,J.D., Hatanpaa,K.J., Kernstine,K., Cai,L. *et al.* (2018) TNF-driven adaptive response mediates resistance to EGFR inhibition in lung cancer. *J. Clin. Invest.*, **128**, 2500–2518.

61. Khaidakov,M. and Mehta,J.L. (2012) Oxidized LDL triggers pro-oncogenic signaling in human breast mammary epithelial cells partly via stimulation of mir-21. *PLoS One*, **7**, e46973.

62. Zheng,Y.Y., Fei,Y., Wang,Z., Chen,Y., Qiu,C. and Li,F.R. (2020) Tissue microRNAs in non-small cell lung cancer detected with a new kind of liquid bead array detection system. *J. Transl. Med.*, **18**, 108.

63. Cappellesso,R., Nicole,L., Caroccia,B., Guzzardo,V., Ventura,L., Fassan,M. and Fassina,A. (2016) Young investigator challenge: microrna-21/microrna-126 profiling as a novel tool for the diagnosis of malignant mesothelioma in pleural effusion cytology. *Cancer Cytopathol.*, **124**, 28–37.

64. Yao,Z.Y., Chen,W.B., Shao,S.S., Ma,S.Z., Yang,C.B., Li,M.Z., Zhao,J.J. and Gao,L. (2018) Role of exosome-associated microRNA in diagnostic and therapeutic applications to metabolic disorders. *J Zhejiang Univ Sci B*, **19**, 183–198.

65. Esplugas,R., Arenas,M., Serra,N., Belles,M., Bonet,M., Gascon,M., Vallve,J.C. and Linares,V. (2019) Effect of radiotherapy on the expression of cardiovascular disease-related miRNA-146a, -155, -221 and -222 in blood of women with breast cancer. *PLoS One*, **14**, e0217443.

66. Benjamin,H., Lebanony,D., Rosenwald,S., Cohen,L., Gibori,H., Barabash,N., Ashkenazi,K., Goren,E., Meiri,E., Morgenstern,S. *et al.* (2010) A diagnostic assay based on microRNA expression accurately identifies malignant pleural mesothelioma. *J. Mol. Diagn.*, **12**, 771–779.

67. Fan,J.L., Zhang,L. and Bo,X.H. (2020) MiR-126 on mice with coronary artery disease by targeting S1PR2. *Eur. Rev. Med. Pharmacol. Sci.*, **24**, 893–904.

68. Jimenez-Lucena,R., Camargo,A., Alcala-Diaz,J.F., Romero-Baldonado,C., Luque,R.M., van Ommen,B., Delgado-Lista,J., Ordovas,J.M., Perez-Martinez,P., Rangel-Zuniga,O.A. *et al.* (2018) A plasma circulating miRNAs profile predicts type 2 diabetes mellitus and prediabetes: from the CORDIOPREV study. *Exp. Mol. Med.*, **50**, 1–12.

69. Ormseth,M.J., Solus,J.F., Sheng,Q., Chen,S.C., Ye,F., Wu,Q., Oeser,A.M., Allen,R., Raggi,P., Vickers,K.C. *et al.* (2021) Plasma miRNAs improve the prediction of coronary atherosclerosis in patients with rheumatoid arthritis. *Clin. Rheumatol.*, **40**, 2211–2219.

70. Elfiky,A.M., Ahmed Mahmoud,A., Zeidan,H.M. and Mostafa Soliman,M. (2019) Association between circulating microRNA-126 expression level and tumour necrosis factor alpha in healthy smokers. *Biomarkers*, **24**, 469–477.

71. Abu-Duhier,F.M., Javid,J., Sughayer,M.A., Mir,R., Albalawi,T. and Alauddin,M.S. (2018) Clinical significance of circulatory miRNA-21 as an efficient non-invasive biomarker for the screening of lung cancer patients. *Asian Pac. J. Cancer Prev.*, **19**, 2607–2611.

72. Hojbjerg,J.A., Ebert,E.B.F., Clement,M.S., Winther-Larsen,A., Meldgaard,P. and Sorensen,B. (2019) Circulating miR-30b and miR-30c predict erlotinib response in EGFR-mutated non-small cell lung cancer patients. *Lung Cancer*, **135**, 92–96.

73. Shi,G.L., Zhang,X.Y., Chen,Y., Ma,S., Bai,W.Q. and Yin,Y.J. (2020) Prognostic significance of serum miR-22, miR-125b, and miR-15b in non-small cell lung cancer patients. *Clin. Lab.*, **66**, 6.