

Aus dem Bereich Klinische Bioinformatik
Klinische Medizin
der Medizinischen Fakultät
der Universität des Saarlandes, Homburg/Saar

Advancing miRNA Research: Computational Approaches for Single-Cell and Tissue-Resolved Analyses

Dissertation zur Erlangung des Grades eines Doktors
der Naturwissenschaften der Medizinischen Fakultät

der UNIVERSITÄT DES SAARLANDES
2025

vorgelegt von ANNIKA ENGEL
geb. am 01.09.1993 in Kaiserslautern, Deutschland

Tag der Promotion: 12. November 2025

Dekan: Univ.-Prof. Dr. med. dent. Matthias Hannig

Berichterstatter: Prof. Dr. Andreas Keller
Prof. Dr. Eckart Meese
Prof. Dr. Gunnar W. Klau

ABSTRACT

Since their discovery in 1993, microRNAs have been an active topic in molecular biology, a breakthrough that was honored with the 2024 Nobel Prize, reflecting their profound impact on the field. Despite the many years of research focused on microRNAs across species, their precise functional roles are still not fully understood. In particular, the dynamics behind timing and location of microRNA-mediated target repression or activation are yet to be discovered for most tissue and cell contexts. Moreover, these localized microRNA expression profiles are known to change throughout the lifespan of organisms. Recently, single-cell RNA sequencing has revealed age-modulated expression patterns such as waves of activation with exceptional detail by capturing cellular heterogeneity. Yet, the currently limited scalability and high costs involved with single-cell high-throughput microRNA protocols prevent large-scale, cell type-resolved application studies. As an alternative, fine-grained study designs which consider multiple tissues investigate microRNA expression heterogeneity with established bulk-sequencing protocols. Eventually, either approach results in large, multi-faceted datasets where computational methods are necessary to select promising microRNAs or identify behavior-driving sample properties, such as sex or age. To this end, this thesis presents a flexible computational framework for the downstream analysis of such complex microRNA datasets.

Three publications investigating individual application scenarios of microRNA functionality emphasize the customizability of the developed framework. In the first, which explored small non-coding RNAs in two mouse plasma fractions, nonnegative matrix factorization of the expression profiles was used to cluster samples based on their age. Second, in a case-control study investigating the effects of non-thermal plasma treatment, differential expression analysis revealed a set of microRNAs previously implicated in wound healing and tissue regeneration. Third, a clustering of time-series data in stem cell differentiation identified

increasing expression trajectories for genes related to cell-type differentiation. The adaptability of the computational framework was demonstrated by providing normalized count tables alongside detailed quality control metrics to optimize library preparation protocols for single-cell microRNA sequencing. Building upon this, the accessibility aspect of the framework was addressed with the development of a web-based platform to enable scientists world-wide to process and evaluate their sequencing runs, facilitating the rapid prototyping of single-cell microRNA preparation protocols.

Applying the computational framework to large-scale, multi-faceted datasets resulting from fine-grained bulk studies highlighted its scalability. Investigating an organ-resolved expression dataset (771 samples from 16 organs across ten time points) revealed both organ-specific and global microRNA profiles. Examining these recurrent expression patterns at multiple time points across the mouse lifespan uncovered dynamic expression patterns influenced by aging. A subsequent study focussed on the mouse brain (844 samples from 15 brain regions across seven time points), which is characterized by its substantial structural and functional heterogeneity. By leveraging all aspects of the developed computational framework, including embedding, differential expression analysis, and clustering of time-series data, the study identified brain region-specific and global aging signatures within a sex-specific dataset. Further, in a unique case-control study which involved accommodating mice at the International Space Station for 40 days, a profiling of single-cell messenger-RNA expression levels revealed the down-regulation of ribosomal protein genes. Both, the analysis of the single-cell RNA and the associated bulk microRNA dataset showed effects of spaceflight on the extracellular matrix and the immune system.

In the future, as more single-cell protocols become accessible and sequencing costs further decrease, fine-grained microRNA studies will be even more relevant. The computational framework presented in this dissertation provides a foundation for their analysis by offering customizable, adaptable, and scalable evaluation methods. Ultimately, these developments support the translation of findings to clinical applications, facilitate the development of intricate analysis methods, and therefore advance microRNA research.

ZUSAMMENFASSUNG

Seit ihrer Entdeckung im Jahr 1993 sind microRNAs ein zentrales Thema der Molekularbiologie. Ihre Bedeutung für das Fachgebiet wurde mit der Verleihung des Nobelpreises für Medizin im Jahr 2024 gewürdigt. Trotz jahrelanger Forschung an microRNAs sind ihre genauen und umfassenden Funktionen jedoch noch nicht vollständig verstanden. Insbesondere die Dynamik der zeitlichen und räumlichen Regulation durch microRNAs ist in den meisten Gewebe- und Zellkontexten unklar. Bekannt ist hingegen, dass sich diese lokalisierten Expressionsprofile im Laufe des Lebens eines Organismus verändern. Kürzlich hat die RNA-Sequenzierung einzelner Zellen altersbedingte Expressionsmuster, zum Beispiel Aktivierungswellen, mit außergewöhnlicher zellulärer Auflösung sichtbar gemacht. Derzeit verhindern jedoch die begrenzte Skalierbarkeit und die hohen Kosten solcher experimenteller Protokolle für microRNAs deren breiten Einsatz in Anwendungsstudien. Alternativ dazu ermöglichen Studiendesigns, die aus mehreren Einzelgeweben bestehen, die Untersuchung der Heterogenität der microRNA-Expression mittels etablierter Bulk-Sequenzierungstechniken. Beide Ansätze führen zu großen und vielfältigen Datensätzen, die computergestützte Methoden erfordern, um vielversprechende microRNA-Kandidaten auszuwählen oder regulatorisch relevante Merkmale wie Geschlecht oder Alter zu identifizieren. In dieser Dissertation wird ein computergestütztes Framework zur Analyse solcher komplexer microRNA-Datensätze vorgestellt.

Zunächst wird die Flexibilität des entwickelten Frameworks in drei Publikationen zu spezifischen Anwendungsfällen von microRNA-beeinflusster Funktionen demonstriert. In der ersten Studie wurden nichtkodierende RNAs in zwei Mausplasmen untersucht, wobei eine nichtnegative Matrixfaktorisierung der Expressionsprofile zur altersbasierten Clusterbildung der Proben verwendet wurde. In einer zweiten Fall-Kontroll-Studie, in der die Auswirkungen einer nicht-thermischen Plasmabehandlung betrachtet wurden, konnten durch diffe-

rentielle Expressionsanalysen microRNAs identifiziert werden, die bereits mit Wundheilung und Geweberegeneration in Verbindung gebracht wurden. Eine Clusteranalyse von Zeitreihendaten zur Stammzelldifferenzierung in der dritten Studie ergab ansteigende Expressionsmuster von Genen, die an der Zelltypdifferenzierung beteiligt sind. Die Anpassungsfähigkeit des Frameworks wurde anhand eines Einzelzell-microRNA-Datensatzes demonstriert für den eine normalisierte Expressionstabelle und detaillierte Metriken zur Qualitätskontrolle erstellt wurden. Darauf aufbauend wurde die Zugänglichkeit des Systems durch die Entwicklung einer webbasierten Plattform verbessert, die es Forschern weltweit ermöglicht, ihre Sequenzierungsdaten zu analysieren und damit die schnelle Prototypisierung von Einzelzellprotokollen für microRNAs unterstützt. Die weitere Anwendung des Frameworks auf große und komplexe Bulk-Datensätze, die aus mehreren Geweben bestehen, unterstreicht seine Skalierbarkeit. Die Untersuchung eines hochaufgelösten Expressionsdatensatzes (771 Proben aus 16 Organen über zehn Zeitpunkte) zeigte sowohl organspezifische als auch globale microRNA-Profile. Durch die Analyse von wiederkehrender Expressionsmuster über mehrere Zeitpunkte konnten dynamische, altersabhängige Expressionsmuster nachgewiesen werden. Eine Folgestudie mit Fokus auf das Mausgehirn (844 Proben aus 15 Hirnregionen zu sieben Zeitpunkten) wurde durchgeführt, da dieses Organ eine erhebliche strukturelle und funktionelle Heterogenität aufweist. Unter Verwendung aller Aspekte des entwickelten Frameworks, einschließlich Einbettungsmethoden, differentieller Expressionsanalyse und Clusteranalyse von Zeitreihendaten, wurden in einer geschlechtsspezifischen Untersuchung des Datensatzes sowohl regionsspezifische als auch globale Alterungssignaturen identifiziert. Darüber hinaus wurde in einer einzigartigen Fall-Kontroll-Studie an Mäusen, die für 40 Tage an Bord der Internationalen Raumstation untergebracht wurden, mittels einer Einzelzell-mRNA-Expressionsanalyse eine Herunterregulierung ribosomaler Proteingene nachgewiesen. Zusammen mit der begleitenden Bulk-Analyse von microRNAs zeigte sie Effekte des Weltraumfluges auf die extrazelluläre Matrix und das Immunsystem.

Mit der zunehmenden Verfügbarkeit von Einzelzellprotokollen und sinkenden Sequenzierungskosten werden detaillierte microRNA-Studien in Zukunft an

Relevanz gewinnen. Das in dieser Dissertation vorgestellte computergestützte Framework bietet eine Grundlage für deren Analyse, indem es anpassbare, flexible und skalierbare Analysemethoden bereitstellt. Diese Entwicklungen unterstützen letztlich den Transfer von Forschungsergebnissen in die klinische Anwendung, ermöglichen die Entwicklung neuartiger Analysemethoden und tragen somit zur Forschung auf dem Gebiet der microRNAs bei.

CONTENTS

List of figures	11
List of tables	11
List of abbreviations	12
Scientific publications	13
1 From molecular mechanisms to computational tools for applied miRNA research	17
1.1 Basic principles in molecular biology, RNA applications and profiling	17
1.1.1 Gene expression and regulation	17
1.1.2 Non-coding RNA and its role in gene regulation	20
1.1.3 RNA in therapeutics and aging	24
1.1.4 RNA profiling techniques	27
1.2 Mathematical foundations, methods and techniques	31
1.2.1 Data normalization and filtering	32
1.2.2 Statistical testing and correlation analysis	34
1.2.3 Embedding and variance analysis	47
1.2.4 Clustering methods	55
1.3 Development of a computational framework	63
1.3.1 Implementation	63
1.3.2 Reproducible pipelines	64
1.3.3 Accessibility through web-based tools	65
1.3.4 Related bioinformatics applications	66

2	Goals of the PhD thesis	69
3	Results	73
3.1	Ageing-associated small RNA cargo of extracellular vesicles	74
3.2	Skin treatment with non-thermal plasma modulates the immune system through miR-223-3p and its target genes	87
3.3	Paving the way to a neural fate – RNA signatures in naive and trans-differentiating mesenchymal stem cells	101
3.4	A novel approach to generate enzyme-free single cell suspensions from archived tissues for miRNA sequencing	120
3.5	SingmiR: a single-cell miRNA alignment and analysis tool	127
3.6	Experimental capture of miRNA targetomes: disease-specific 3'UTR library-based miRNA targetomics for Parkinson's disease	134
3.7	Characterizing expression changes in noncoding RNAs during aging and heterochronic parabiosis across mouse tissues	145
4	Discussion	169
	Bibliography	173
	Acknowledgments	223
	Curriculum Vitae	225

LIST OF FIGURES

Figure 1.1	Protein biosynthesis in eukaryotic cells	18
Figure 1.2	Microarray	28
Figure 1.3	Single-cell sequencing workflow	30
Figure 1.4	Differential expression analysis	42
Figure 1.5	Pearson correlation coefficient	44
Figure 1.6	Principal Component Analysis	49
Figure 1.7	Uniform Manifold Approximation and Projection	51
Figure 1.8	Principal Variance Component Analysis	54
Figure 1.9	Hierarchical clustering dendrogram	57
Figure 1.10	Iteration step for C-means algorithm	59
Figure 1.11	Statistics of web services at the Chair for Clinical Bioinformatics	66
Figure 2.1	Graphical abstract summary	69

LIST OF TABLES

Table 1.1	Decision table for a correlation method	46
-----------	---	----

LIST OF ABBREVIATIONS

BH	Benjamini-Hochberg procedure
DE	differential expression
DNA	deoxyribonucleic acid
EV	extracellular vesicle
FDR	False Discovery Rate
GEO	Gene Expression Omnibus
lncRNA	long non-coding RNA
miRNA	microRNA
mRNA	messenger-RNA
ncRNA	non-coding RNA
NGS	next-generation sequencing
PCA	Principal Component Analysis
pre-miRNA	precursor microRNA
pre-mRNA	precursor messenger RNA
pri-miRNA	primary microRNA
PVCA	Principal Variance Component Analysis
REML	restricted maximum likelihood
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
rpm	reads per million
rpmmm	reads per mapped million
rpmmm	reads per mapped million against miRNA
rRNA	ribosomal RNA
sncRNA	small non-coding RNA
tRNA	transfer RNA
UMAP	Uniform Manifold Approximation and Projection
UMI	unique molecular identifier
UTR	untranslated region
VCA	Variance Component Analysis

SCIENTIFIC PUBLICATIONS

This is a cumulative thesis based on the following peer-reviewed journal papers [1–7]. The publications included herein are identical to the published versions if not indicated otherwise. Equally, contributing first authors are denoted by a superscript asterisk (*) symbol and equally contributing last authors by a superscript ampersand (&) symbol.

1. Fabian Kern*, Thomas Kuhn*, Nicole Ludwig*, Martin Simon*, Laura Gröger, Natalie Fabis, Ernesto Aparicio-Puerta, Abdulrahman Salhab, Tobias Fehlmann, Oliver Hahn, **Annika Engel**, Viktoria Wagner, Marcus Koch, Katarzyna Winek, Hermona Soreq, Irina Nazarenko, Gregor Fuhrmann, Tony Wyss-Coray, Eckart Meese, Verena Keller&, Matthias W. Laschke&, Andreas Keller&. "Ageing-associated small RNA cargo of extracellular vesicles." In: *RNA biology* 20 (2023), pp. 482-494. url: <https://doi.org/10.1080/15476286.2023.2234713>.
2. **Annika Engel**, Nicole Ludwig, Friederike Grandke, Viktoria Wagner, Fabian Kern, Tobias Fehlmann, Georges P. Schmartz, Ernesto Aparicio-Puerta, DominicHenn, Barbara Walch-Rückheim, Matthias Hannig, Stefan Rupf, Eckart, Meese, Matthias W. Laschke&, Andreas Keller&. "Skin treatment with non-thermal plasma modulates the immune system through miR-223-3p and its target genes." In: *RNA Biology* 21.1 (2024), pp. 31-44. url: <https://doi.org/10.1080/15476286.2024.2361571>.
3. Caroline Diener, Konstantin Thüre, **Annika Engel**, Martin Hart, Andreas Keller, Eckart Meese, Ulrike Fischer. "Paving the way to a neural fate – RNA signatures in naive and trans-differentiating Mesenchymal Stem Cells." In: *European Journal of Cell Biology* 103 (2024), p. 151458. url: <https://doi.org/10.1016/j.ejcb.2024.151458>.

4. Stefan Scheuermann*, Sarah Hückler*, **Annika Engel**, Nicole Ludwig, Philipp Lebhardt, Jens Langejürgen, Stefan Kirsch. "A novel approach to generate enzyme-free single cell suspensions from archived tissues for miRNA sequencing." In: *SLAS Technology* 29.3 (2024), p. 100133. url: <https://doi.org/10.1016/j.slas.2024.100133>.
5. **Annika Engel**, Shusruto Rishik, Pascal Hirsch, Verena Keller, Tobias Fehlmann, Fabian Kern, Andreas Keller. "SingmiR: a single-cell miRNA alignment and analysis tool." In: *Nucleic acids research* 52.W1 (2024), W374-W380. url: <https://doi.org/10.1093/nar/gkae225>.
6. Martin Hart*, Fabian Kern*, Claudia Fecher-Trost, Lena Krammes, Ernesto Aparicio, **Annika Engel**, Pascal Hirsch, Viktoria Wagner, Verena Keller, Georges Schmartz, Stefanie Rheinheimer, Caroline Diener, Ulrike Fischer, Jens Mayer, Markus Meyer, Veit Flockerzi, Andreas Keller[&], Eckart Meese[&]. "Experimental capture of miRNA targetomes: disease-specific 3'UTR library-based miRNA targetomics for Parkinson's disease." In: *Experimental & molecular medicine* 56 (2024). url: <https://doi.org/10.1038/s12276-024-01202-5>.
7. Viktoria Wagner, Fabian Kern, Oliver Hahn, Nick Schaum, Nicole Ludwig, Tobias Fehlmann, **Annika Engel**, Dominic Henn, Shusruto Rishik, Alina Isakova, Michelle Tan, Rene Sit, Norma Neff, Martin Hart, Eckart Meese, Steve Quake, Tony Wyss-Coray[&], Andreas Keller[&]. "Characterizing expression changes in noncoding RNAs during aging and heterochronic parabiosis across mouse tissues." In: *Nature Biotechnology* (2023), pp. 1-10. url: <https://doi.org/10.1038/s41587-023-01751-6>.

PLANNED PUBLICATIONS

In the course of this work, additional publications are prepared, which, at the time of submitting this thesis, have not yet been accepted for publication but are within the publication process. Therefore, they are included as independent planned publications. It has to be mentioned that some manuscripts are still in progress and both the author list and the title may still change before their final publication. Equally, contributing first authors are denoted by a superscript asterisk (*) symbol and equally contributing last authors by a superscript ampersand (&) symbol.

- I **Annika Engel***, Viktoria Wagner*, Oliver Hahn, Aulden G. Foltz, Micaiah Atkins, Amila Beganovic, Ian H. Guldner, Nannan Lu, Verena Keller, Aryaman Saksena, Ulrike Fischer, Nicole Ludwig, Eckart Meese, Tony Wyss-Coray, Andreas Keller. "A spatio-temporal brain miRNA expression atlas identifies sex-independent age-related microglial driven miR-155-5p increase."
- II Friederike Grandke, Nicholas Schaum, Fabian Kern, Jeremy Amand, Tom Ambrosi, Teni Anbarchian, Karl Annusver, Reza Ashrafi, Kruti Calcuttawala, Charles Chan, Antoine de Morree, Angela Detweiler, Hala Dhowre, **Annika Engel**, Tobias Fehlmann, Virginia Ferguson, Matthias Flotho, Astrid Gillich, Michael Haney, Yan Hang, Pascal Hirsch, Taichi Isobe, Joanna Kalucka, Maria Kasper, Seung Kim, Thomas Kluiver, Davis Lee, Song Lee, Olivia Leventhaw, Ling Liu, Cati Logan, Maria Lugo-Fagundo, Luca Mannino, Maxim Markovic, Maurizio Morri, Shravani Mukherjee, Norma Neff, Patricia Nguyen, Robi Palovics, Weng Chuan Peng, Maksim Plikus, Ermalinda Porpiglia, Tom Rando, Nil Sanz, Ina Schiessl, Stephanie Schubert, Arun Sharma, Shaheen Sikander, Rahul Sinha, Daniel Staehli, Louis Stodieck, Aditi Swarup, Michelle Tan, Aris Tay, Stefan Veizades, Bruce Wang, Tobias William Wolff, Mike Wosczyzna, Albert Wu, Rose Yan, Viktoria Wagner, Andrew Yang, Macy Zardoneta, Weichen Zhao, Tony Wyss-Goray&, Stephen Quake&, Andreas Keller&. "Spaceflight and aging induce systemic effects on extracellular matrix and immune system."

III Friederike Grandke*, Shusruto Rishik*, Viktoria Wagner, **Annika Engel**,
Nicole Ludwig, Kruti Calcuttawala, Fabian Kern, Verena Keller, Marcin
Krawczyk, Louis Stodieck, Virginia Ferguson, Amanda Roberts, Eckart
Meese, Nicholas Schaum, Steven Quake, Tony Wyss-Coray, Andreas Keller.
“Age-dependent systemic changes in murine microRNA fingerprints after
spaceflight.”

FROM MOLECULAR MECHANISMS TO COMPUTATIONAL TOOLS FOR APPLIED MIRNA RESEARCH

This chapter establishes the foundational context for the bioinformatics approaches explored in this thesis. It begins with an introduction to molecular biology, providing the biological background for the computational analyses. The subsequent section introduces the mathematical foundations and methods that underlying the computational framework developed in this work, including statistical, embedding and clustering techniques. Finally, the last section focuses on the implementation and accessibility of the framework.

1.1 BASIC PRINCIPLES IN MOLECULAR BIOLOGY, RNA APPLICATIONS AND PROFILING

Molecular biology forms the cornerstone of understanding cellular processes. Deoxyribonucleic acid (DNA) mainly serves as the blueprint for cellular processes, encoding the information required to produce proteins, which carry out most functional roles in the cell. Focusing on eukaryotic cells, this section briefly introduces these key concepts, establishing the connection between molecular mechanisms and their computational exploration. By understanding the basics of gene expression and regulation, we can better contextualize the bioinformatics techniques used to analyze ribonucleic acid (RNA) data.

1.1.1 *Gene expression and regulation*

Genes are specific segments of DNA that encode the instructions for producing proteins by transcription and translation [8]. The transcribed messenger-RNA (mRNA) molecules serve as genetic blueprints controlling nearly every biological process, shaping the development, function, and adaptation of all living organ-

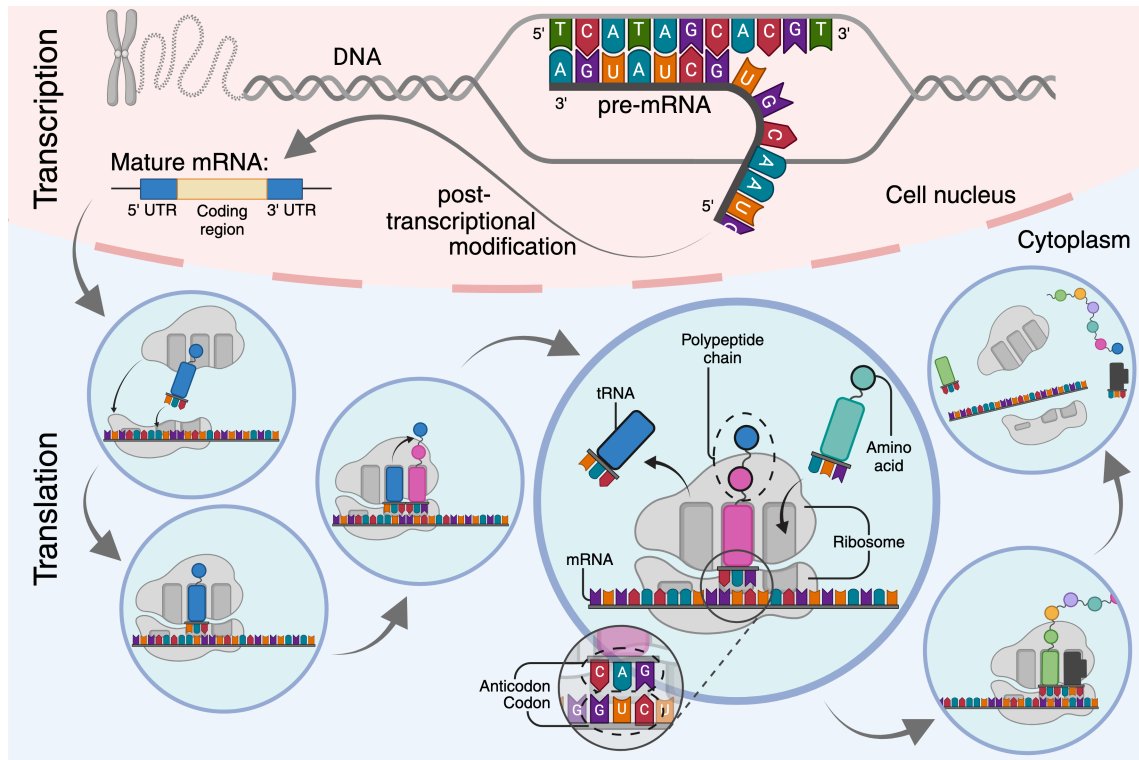


Figure 1.1: Protein biosynthesis in eukaryotic cells is presented. The main steps from the transcription to the translation into an amino acid chain are shown. Created in BioRender.com.

isms. The expression of genes, a tightly regulated process, involves the conversion of DNA-encoded information into functional products, such as proteins.

From DNA to protein - A general schematic illustration of eukaryotic protein biosynthesis is shown in Figure 1.1. Gene expression begins with the transcription in the cell nucleus, whereby the double-stranded DNA is read in 3' to 5' direction by complementary completion of the bases by a RNA polymerase II. The bases from the DNA adenine, guanine, cytosine, and thymine are supplemented by complementary bases from the RNA uracil, cytosine, guanine, and adenine, respectively. The so generated precursor messenger RNA (pre-mRNA) is processed into a mature mRNA by post-transcriptional modification and consists of two parts: the open reading frame in which the building instruction for the protein is decoded and a non-coding region called untranslated region (UTR) at both ends. The mRNA then leaves the cell nucleus into the cytoplasm, where translation is performed at the ribosomes. Here, the mRNA molecules serve as building

instructions for proteins. In a translation step, an amino acid sequence is created from the base sequence of the mRNA. Three consecutive bases always result in a base triplet, which is here denoted as the codon. This codon uniquely codes one of the 20 known amino acids. However, since the amino acid has no direct connection to the codon, an intermediate molecule, called transfer RNA (tRNA), is needed. It therefore possesses two distinct binding sites, the anticodon represents the corresponding match to the codon. The other one represents the amino acid binding site. By docking further tRNAs for the next base triplets in 5' to 3' direction to the mRNA, an amino acid chain is created. In order for the amino acids to be able to connect through a peptide bond, a ribosome is required, which in this case consists of a smaller and a larger subunit which both consist of a special type of RNAs called ribosomal RNAs (rRNAs) and ribosomal proteins. These amino acid chains form functional proteins when they are correctly folded and have undergone necessary modifications and assemblies. These processes do not necessarily happen immediately, but can be completed later at a different cellular location. The resulting proteins exhibit a specific three-dimensional structure that enables them to fulfill their respective cellular functions.

1.1.1.1 *Gene regulation*

Maintaining cellular identity requires precise regulation to respond to environmental stimuli and ensure proper development [9–12]. Regulatory elements, such as transcription factors and enhancers, modulate the onset and intensity of transcription. Epigenetic modifications, for example DNA methylation and histone acetylation, further impact gene accessibility and activity [13–23]. Post-transcriptional mechanisms, including RNA modifications and non-coding RNA (ncRNA) interactions, fine-tune mRNA stability and translation efficiency. Adding another layer of complexity to gene regulation, alternative splicing can generate isoforms with distinct functions.

The precise control of gene expression is vital for organismal health. Deregulation can lead to developmental abnormalities, diseases such as cancer and altered responses to environmental stressors [24–37]. Advances in genomics, transcriptomics, and epigenetics have expanded our understanding of these pro-

cesses, enabling researchers to map gene expression patterns across tissues, cell types, and developmental stages. Despite these efforts, the dynamic interplay between genotype and phenotype remains an intricate puzzle and requires further exploration to fully understand the molecular mechanisms driving life.

1.1.2 *Non-coding RNA and its role in gene regulation*

While genes serve as the templates for protein synthesis, they represent only a part of the transcriptome. Based on Frankish et al. [38], around 19,370 coding genes could be found in humans. Of all nucleotides in humans, only a minor fraction is considered to be coding, although around 85% can still be transcribed into ncRNA [39]. These ncRNA components, once dismissed as non-functional DNA, perform diverse regulatory functions that extend beyond coding regions and play a critical role in maintaining cellular stability [40–45].

NcRNAs are broadly categorized by size and function [40, 43]. Long non-coding RNAs (lncRNAs) often regulate gene activity at the transcriptional or post-transcriptional level. Small non-coding RNAs (sncRNAs) such as tRNA, rRNA, and microRNA (miRNA) guide or suppress specific genetic pathways and are involved in the translation process [42, 43, 46–54]. Additionally, lncRNAs may function as precursors for sncRNAs, underscoring their interconnected roles in regulating gene function [55].

The discovery of ncRNAs has reshaped our understanding of the genome. SncRNAs like miRNAs, for example, play a key role in gene silencing by binding to mRNAs and preventing their translation into proteins. Advances in molecular biology have revealed that ncRNAs influence a wide array of cellular processes, from development to disease [42, 55–58]. These regulatory elements are now recognized as integral to understanding how genes operate and how their deregulation contributes to conditions such as cancer and neurodegenerative diseases [59–67].

The study of ncRNAs highlights the complexity and precision of gene regulation. Far from being "junk" the non-coding genome is a dynamic and essential component of the genetic architecture, ensuring that genes fulfill their roles in

the intricate symphony of life. Among the various classes of ncRNAs, miRNAs stand out as potent regulators of gene expression.

1.1.2.1 *MicroRNA*

MiRNAs, a class of sncRNAs, are small RNA molecules which play an important key in fine-tuning gene expression. Discovered in 1993, miRNAs regulate genes by binding to mRNAs and influencing their stability or translation [68].

Subsequent research has shown that miRNAs are highly conserved across species, highlighting their evolutionary importance. With nearly 2,000 known human miRNAs and potentially more candidates, miRNAs regulate a significant portion of the protein-coding genome [69]. They are central to processes like development, differentiation, and cellular stress responses, often functioning as negative regulators by silencing target genes.

From DNA to microRNA - Analogous to the generation of mRNAs, there are special sites in the DNA that contain a blueprint for miRNAs. They are either read specifically for the production of miRNAs or as part of mRNA generation. The transcription of miRNAs works in the same way as for mRNAs. The corresponding site on the DNA is transcribed using a RNA polymerase II. The resulting RNA molecule is called primary microRNA (pri-miRNA) and has a hairpin structure, which plays an important role in the later recognition by enzymes [70, 71]. While still in the cell nucleus, the pri-miRNA is cut using the Drosha enzyme [72]. It retains the hairpin structure but is much shorter (about 70 nucleotides) and is now called precursor microRNA (pre-miRNA). Now the pre-miRNA is transported out of the cell nucleus into the cytoplasm with the help of another enzyme called Exportin-5 [73, 74]. Once there, the Dicer enzyme cuts the pre-miRNA further into a typical miRNA double strand consisting of about 22 nucleotides in length [75, 76]. One of the two strands is selected as the leading strand, which later becomes active, the other is disassembled. The miRNA guide strand is integrated into the RNA-induced silencing complex (RISC) protein complex [77, 78]. Here the miRNA acts like a key for mRNAs and waits to carry out its regulatory tasks.

The discovery of miRNAs has revolutionized our understanding of gene regulation by showing that genes are not merely static units of inheritance but

dynamic players in cellular networks. By acting as precise molecular switches, miRNAs exemplify the intricate mechanisms that maintain genetic balance and adapt to environmental changes. Their role in health and disease emphasizes the importance of genes not just as templates for proteins, but as hubs of regulatory activity.

The regulatory influence of miRNAs depends on their interactions with their so-called target genes. These interactions occur at defined binding sites, where miRNAs guide the suppression or fine-tuning of gene expression. Exploring the mechanisms of these interactions sheds light on the specificity and biological outcomes of miRNA activity.

1.1.2.2 *MicroRNA and its targets*

By influencing gene expression, miRNAs intervene in the production of proteins as post-transcriptional regulation [79]. This influence can occur by either blocking of translation, or the complete disassembly of the mRNA so that it does not even exist long enough to be translated. In rare cases in specific cellular conditions, miRNAs can also take on enhancing tasks. For example, miR-122 stabilizes mRNAs that are important for liver function and thus increases protein production [80].

After the miRNA has been integrated into RISC, it searches for a matching complementary mRNA. This means that the miRNA recognizes a mRNA as a match if its own seed sequence (a region about 6 to 8 nucleotides long) is complementary to that of the mRNA. The miRNA can then bind to the mRNA. However, miRNAs cannot bind everywhere. There are specific sites on the mRNA that are suitable for binding, these are located on the 3'UTR [81]. This is the end of the mRNA behind the coding sequence which itself is less suitable as it is much shorter and more protected. The other end of the mRNA, the 5'UTR, is shorter, it controls translation and is therefore usually covered by other structures and therefore less accessible. The 3'UTR is the evolutionarily conserved main site for regulation by miRNAs, its length makes it suitable for fine-tuned control of protein production. It has multiple binding sites for the same and other miRNAs to allow a fine and combined regulation. There are different ways of strengthening

of the resulting pairing. In perfect pairing, which occurs less frequently in animals or humans, the mRNA is completely degraded by the RISC complex [82–85]. More commonly, in partial pairing, the translation is blocked by the RISC. This means that the ribosome, which initiates the translation, can no longer bind properly to the mRNA. An individual miRNA may influence several mRNAs, while a single mRNA can be a target of numerous miRNAs. This forms a regulatory network [79, 86–88]. Mutations can cause a change in the binding sites, resulting in regulation not being carried out normally, which can play an important role in disease and aging.

Beyond the miRNA target interactions, their intercellular roles are facilitated by an active transportation between cells through extracellular carriers like vesicles. These transport mechanisms enable miRNAs to act as signaling molecules, influencing distant cells and tissues, which is particularly relevant in processes like aging and intercellular communication [89–92].

1.1.2.3 *Extracellular vesicle for microRNA transport*

Extracellular vesicles (EVs) are tiny, membrane-bound particles that are released from the cells into the surrounding tissue or blood. They play a crucial role in cellular communication. EVs are divided into three subtypes: exosomes, which are formed in intracellular endosomes and are then released by exocytosis, microvesicles, which are formed at the cell membrane and are released by being pinched off from the cell membrane, and apoptotic vesicles, which are formed during programmed cell death (apoptosis). In the context of this work, we only briefly present the functionality of exosomes and microvesicles.

EVs are used by cells to specifically exchange molecules such as miRNAs. In this way, large quantities of miRNAs can be sent to another cell at the same time [93, 94]. Therefore, the EVs are first loaded by the cell with specific miRNAs at a certain point in time. The vesicles are then released from the cell and enter the serum, which is the liquid component of the blood that remains after clotting, or other body fluids. During transportation, EVs protect the RNAs inside them from degradation by enzymes present in the serum. In the case of exosomes, they can then be transported through the serum to distant cells due to the blood

circulation. A cell can take up the EVs by its cell membrane swallowing the EVs or by fusion of the vesicle and cell membrane. Specific proteins among other molecules on the surface of the exosome membrane are recognized by special receptors on the target cell to initiate a connection. As a result, exosomes often travel longer distances in the body and the communication mechanism is precise and controlled. Microvesicles are not recognized by receptors due to the less pronounced surface molecules on the membrane. This means that any cell, with which the vesicle makes contact, can absorb it. Therefore, microvesicles mainly act in the local environment and occur more frequently in stress situations, such as inflammatory processes or acute wound care.

The transfer of miRNAs between cells is a highly specialized communication mechanism that transmits signals which promote certain behaviors in the target cells, such as cell growth, differentiation or immune response [95–98]. Systematic effects can arise when cells communicate with other cells in distant organs and regulate genes in them. It therefore plays a central role in many diseases and biological processes such as wound healing, the immune system or tumor regulation [99–102]. Tumor cells, for example, can suppress the immune system and prepare their environment for metastases.

Having discussed extracellular transport mechanisms for miRNAs and their intracellular and intercellular roles, we now focus on the transformative potential of miRNAs in research and clinical use.

1.1.3 *RNA in therapeutics and aging*

Individual miRNAs can regulate several mRNAs which, in turn, can be regulated by many other miRNAs, such that these interactions form a complex network. The same miRNA can have different and sometimes even opposing effects depending on the cell type, developmental stage or disease state and is therefore highly context-dependent [103–105]. Bioinformatic evaluations often link miRNAs and mRNAs, but experimental validation in the laboratory is usually time-consuming and often complicated [106, 107].

In addition to their central role in gene regulation, they are an essential factor in the developmental processes of stem cell differentiation, tissue and organ

development [108, 109]. Due to malfunction, miRNAs are associated with various diseases such as cancer, cardiovascular diseases and neurodegenerative diseases [110–113]. In addition to regulating mRNAs, they also regulate immune responses and, through their transport by EVs, they help cells to communicate systematically throughout the body.

MiRNAs nevertheless show promising potential for therapies. Their stability in EVs and their presence in body fluids make them ideal non-invasive biomarkers for disease [114–118]. Much is still unknown, but miRNAs harbor great potential to shed more light on the underlying mechanisms, for example, their highly conserved nature points to their important evolutionary function.

1.1.3.1 *Clinical treatment strategies and microRNA modulation*

Clinical treatments are therapeutic approaches or interventions that are directly applied in hospitals or clinics to prevent, diagnose, or treat diseases in patients. For miRNAs, while their role in biological processes and diseases is well established, most therapeutic applications involving miRNAs are still in the research or experimental stage [119–124]. Some promising examples are being tested in clinical trials, but very few have made it to actual use in hospitals or clinics. The targeted delivery to the relevant cells is challenging and side effects for other cells, tissues or organs may occur, thus, complicating therapeutic use [125–127].

To date, the most advanced therapeutic strategies using miRNAs involve miRNA mimics or miRNA inhibitors. These approaches aim to restore normal levels of miRNAs in cases where they are deregulated. For instance, MRX34, a synthetic mimic of tumor-suppressing miR-34, entered early-phase clinical trials for cancer treatment, but was discontinued due to immune-related side effects [128]. This highlights the potential of miRNA-based therapies while also emphasizing the challenges in translating them into clinical use. While some studies, such as the development of MRX34, have already explored the potential of miRNA-based therapeutics in early clinical trials, many other research efforts are focused on building the foundational knowledge necessary for future therapeutic integration. For instance, the study on the miR-23-27-24 cluster elucidates its roles in development and aging, highlighting potential targets for therapeutic

interventions [129]. Similarly, research on exosomal miR-21 demonstrates its impact on cancer progression and therapy resistance, offering insights for designing targeted treatments [130]. Furthermore, emerging strategies and challenges for incorporating miRNA modulation into cancer therapy are still be discussed [131–133]. Together, these studies exemplify the extensive work being done to advance miRNA research toward eventual clinical application.

1.1.3.2 *Differences in RNA signatures through aging*

Aging is a complex, natural biological process characterized by the gradual decline in the structure and function of cells, tissues, and organs over time. This decline increases the risk of diseases, reduces the body's ability to repair damage, and leads to a higher likelihood of death. Aging affects nearly all aspects of an organism's physiology, including the immune system, metabolism, and cellular repair mechanisms. The process of aging is driven by multiple molecular and cellular mechanisms, including damage to the DNA, proteins, and other biomolecules caused by reactive oxygen species and environmental stress which contributes to the decline in cellular function [134–137]. Additionally, senescence where cells arrive in a state of permanent cell cycle arrest in return to stress or damage, contributes to tissue dysfunction and aging [138]. Further, the telomeres, which are protective caps on the ends of chromosomes, get shorter, eventually leading to cellular aging and senescence [139–142]. Moreover, the roles of epigenetic modifications and DNA methylation changes in aging have been investigated. These methylation patterns can be used to estimate biological age and provide insights into age-related functional decline [143, 144].

Aging is accompanied by alterations in RNA expression profiles, which can affect various physiological processes. For instance, the miR-34a has been implicated in the regulation of genes involved in DNA repair and cellular senescence [145, 146]. Aging is often connected to chronic, low-grade inflammation, which is regulated by miRNAs like miR-146a that suppress inflammatory responses [147]. Understanding these changes in RNA expression with age can aid in developing interventions to mitigate age-related diseases and promote healthy aging. Therefore, researchers all over the world study the role of RNAs in aging and

use various methods, like expression profiling, RNA sequencing or microarrays, to identify age-related changes in RNA expression [148–158]. In addition, these changes have been investigated in sex-specific studies [159–162]. Notably, the emergence of single-cell RNA sequencing enabled studying expression changes while taking cellular heterogeneity into account [163–167]. By functional studies, like knocking out or over expressing specific miRNAs in model organisms, for example mice or *C. elegans*, helps to identify their role in aging pathways [168]. Incorporating clinical studies investigate miRNAs as biomarkers for aging-related diseases, such as cardiovascular disorders or neurodegenerative conditions [169].

In summary, miRNAs are critical regulators of aging processes, including cellular senescence, inflammation, and longevity pathways, making their modulation a promising therapeutic strategy and their expression valuable as biomarkers for disease. Continued research into miRNA functions and mechanisms will improve our understanding of their role in aging and enhance their application in clinical interventions [170].

Concluding this section with a discussion of miRNA challenges and impacts provides a broader perspective before transitioning to methodologies. Given the complexity of miRNA regulation and their diverse roles, robust profiling techniques are essential to gain a correct and proper understanding of their functions. These approaches range from bulk profiling methods to emerging single-cell technologies, providing the resolution necessary to decode the dynamic behavior of RNA molecules.

1.1.4 RNA profiling techniques

Quantifying the expression of RNAs, among them sncRNAs like miRNAs, is a pivotal task in molecular biology. A typical laboratory workflow starts with collecting the samples of interest, for example from bulk tissue or a specific cell type. For such bulk samples, the obtained expression profile represents an average over the cells in the processed sample. Subsequently, the RNA is extracted from the samples and possibly further purified. Eventually, a quality control step ensures the successful processing of the samples. From there, multiple techniques

can be utilized to quantify the RNAs. In the following, we briefly introduce the microarray technology [171–173] and next-generation sequencing (NGS) [174–176] while focussing on miRNA. These techniques are common in expression profiling, but not limited to this application. For instance, microarrays are also used to investigate DNA methylation [177].

1.1.4.1 Microarray

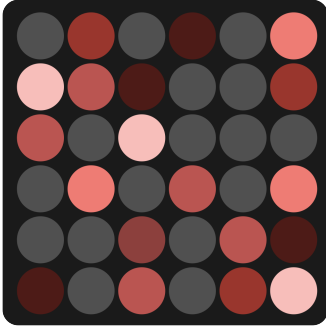


Figure 1.2: Exemplary microarray fluorescence.

MiRNA profiling via microarrays relies on the hybridization of the miRNA molecules of the sample to known miRNA probes on a solid substrate, the so-called chip. The probes are organized in spots which are laid out in a two-dimensional grid across the chip where each spot consists of numerous probes of a certain miRNA. To quantify the composition of miRNAs in the sample, each strand is ligated with a fluorescent dye and adapters to improve its binding capability to the probes. This so-called library is then applied to the microarray chip where the complementary strands of the miRNAs hybridize with the

corresponding probes. Note, the term *library size* denotes the number of unique miRNA in the library. After washing away free, unbound molecules, fluorescence detection measures the amount of remaining miRNAs as brightness per spot, as indicated in Figure 1.2. To convert these intensities into counts for each miRNA probe, sophisticated post-processing including localized background correction, aggregating multiple spots and normalization is necessary.

An advantage of the microarray techniques is that no miRNA amplification is required in the sample, therefore a potential amplification bias is avoided. Yet, the dynamic range of the detection leads to a diminishing accuracy for high expressed miRNAs caused by the saturation of the fluorescence. In contrast, low expressed molecules can be lost in the background signal. One major disadvantage is rooted in the process itself. Since it is based on the hybridization to previously selected probes, discovering novel miRNAs is challenging. While custom microarrays can be manufactured to incorporate candidates, explorative probing remains infeasible.

1.1.4.2 Next-generation sequencing

NGS is a widely used technology for miRNA screening and profiling since sequencers have become commercially available in the mid-2000s [178]. In contrast to the aforementioned microarray technique, NGS step-wise determines the sequence of nucleotides in the miRNAs. Starting from the purified sample, each miRNA is reverse transcribed and ligated with adapters which enable its binding to the experimental substrate. Additionally, a barcode is added to each complementary strand to later match it to the sample. This is followed by an optional amplification step to increase the number of complementary strands. In the case of synthesis sequencing, for example using the *Illumina* platform, the actual sequencing of the library is performed in a so-called flow-cells which consist of multiple fluid channels. In a first step, the complementary strands in the library are fixed inside these channels using the previously attached adapters. Here they are further amplified into clusters of the same kind. Then, a polymerase step-wise adds nucleotides to each strand whereby each type of nucleotide is augmented with a specific fluorescence marker. After each step, the type of each added nucleotide is determined via fluorescence detection, generating the reads. Following their detection, the fluorescence markers are removed from the strands and the flow-cell is flushed before the next cycle starts. Eventually, the sequencing stops after a pre-defined number of cycles which is called *sequencing length*. For miRNA, it typically ranges from 50 to 100 nucleotides which is sufficient to fully read miRNA including all attachments. Eventually, the sequencing results in sequences of measured nucleotides, called reads. They can be matched to the samples via the barcodes. When considering miRNAs, the *sequencing depth* denotes the number of reads per sample. Yet, multiple definitions for *sequencing depth* exist with regard to what is sequenced.

To actually perform the miRNA expression profiling, the reads need to be further processed by removing the parts of the reads which correspond to the platform-specific adapters using, for example, cutadapt [179, 180]. The cleaned reads are then mapped against a reference genome or a database such as mirBase [69] with tools such as bowtie [181]. In addition, the alignment, for example using

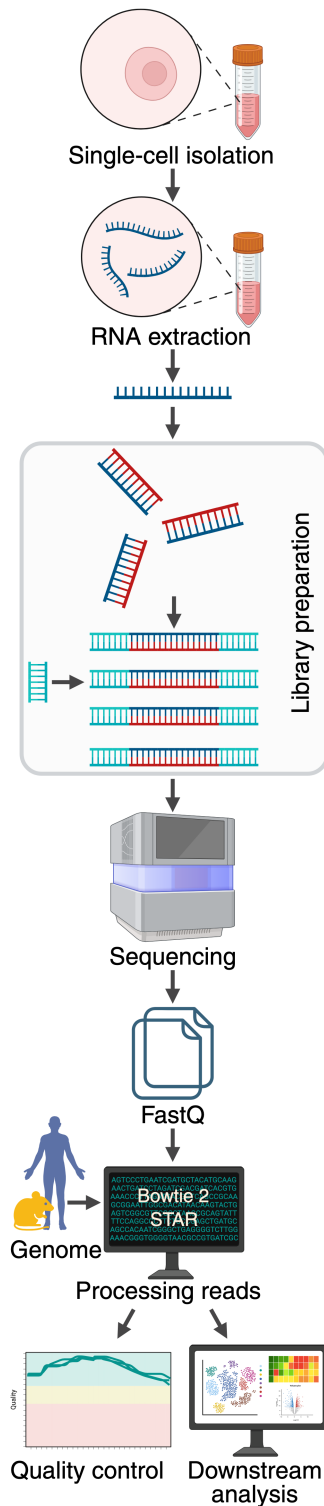


Figure 1.3: Single-cell sequencing workflow adjusted from [182]. Created in BioRender.com.

STAR [183], quantifies mismatches of reads with respect to the reference, for miRNA datasets, considering 0 or 1 mismatches are common. Finally, counting the detected miRNAs per sample results in a raw count matrix which can be further processed downstream. In the case of sncRNA, all these steps, including necessary data transformations in between them, have been consolidated into a single workflow by Fehlmann et al. [184].

In comparison to the microarray technique, NGSs performs a read-out of the library in question requiring minimal explicit assumptions about the contained miRNAs. Therefore, it is suitable to detect variations and novel miRNAs candidates without changing the experimental process [185, 186]. Moreover, since the resolution is at single-nucleotide level, re-evaluating reads at a later point in time with up-to-date reference databases can lead to further, previously hidden annotations.

1.1.4.3 Single-cell RNA sequencing

While bulk methods offer valuable insights into average gene expression across multiple cells, they often mask cellular heterogeneity. Overcoming this limitation, single-cell resolved RNA studies have gained increasing interest in the last decade [187–190].

To perform single-cell RNA studies, changes to established preparation protocol of bulk (tissue) samples are necessary [182, 191]. As outlined in Figure 1.3, the workflow starts by isolating single cells, for example via droplet-based methods or fluorescence activated cell sorting. Then, the RNA molecules are extracted per cell and transcribed to complementary strands and ligated with platform-dependent adapters. Additionally, a unique molecular identifier (UMI), to counter the amplification bias, and a barcode are added to each complementary strand. In contrast to bulk samples, these barcodes are split into a unique part for each sample and a second unique barcode within the sample for each individual cell. Since the amount of RNA per cell is

too low for synthesis sequencing, amplification of the complementary strands is necessary. The sequencing of the resulting library via NGS is performed via step-wise polymerase and the obtained reads are processed as presented in the above section.

In the recent years, studies explored cell-resolved expression heterogeneity on the miRNA level [192]. Yet, while commercial single-cell RNA workflows and platforms have emerged, to this date, none are available for miRNAs [193, 194]. Currently, the library preparation is still an active area of research where different protocols are evaluated and modified [4, 195]. In Chapter 3, we present a processing workflow for single-cell miRNA reads which can be easily utilized by scientist in the laboratory to quickly assess NGS results and use these insights to optimize their library preparation [5].

Through the expression profiling, we obtain numeric count values for any detected miRNAs, typically several hundred, per sample. Conducting similar profiling studies, we obtain even higher numbers of mRNAs per sample. In addition, each sample is annotated with attributes reflecting the experimental setup, for example age, sex, type of tissue, and treatment status. Combining the amount of miRNAs with the dimensionality given by the study setup leads to large, highly-complex datasets. Thus, manually selecting promising miRNAs and identifying behavior-driving attributes becomes infeasible and a need for data-driven methods using computational tools emerges.

1.2 MATHEMATICAL FOUNDATIONS, METHODS AND TECHNIQUES

A central part of this thesis is concerned with the rigorous interpretation of experimental data using a wide variety of analytical methods. In the publications forming Chapter 3, these methods and techniques are often presented as black boxes, where data is fed in and results are simply delivered. However, unveiling the mathematical foundations of these techniques is crucial for comprehending and assessing the obtained results. Therefore, this section provides an overview of key concepts, including statistical testing, correlation analysis, embedding techniques with variance analysis, and clustering methods. These methodologies

form the basis of the developed computational framework for uncovering patterns, relationships, and variability in complex datasets, providing robust and meaningful conclusions.

1.2.1 *Data normalization and filtering*

In the context of this thesis, we typically obtain a raw read count matrix originating from a transcriptomic readout. To draw meaningful conclusions, we perform a pre-processing consisting of three steps. The first step consists of normalization to make the samples comparable, afterwards we perform two quality control steps to remove sample and features with low quality [196–205].

1.2.1.1 *Normalization*

Depending on the specific case, we choose one of the following normalizations, which are common in the field: reads per million (rpm), reads per mapped million (rpmm), reads per mapped million against miRNA (rpmmm) (only in case of a miRNA dataset), quantile normalization and *DESeq2* normalization based on [184, 206]. The rpm-normalization divides for every sample the raw read counts of each feature by the total number of sequenced reads of this sample and multiplies the ratio by one million. This accommodates for different sequencing depths between the samples. Similar to the previous, the rpmm-normalization divides for every sample the raw read counts of each feature by the total number of reads of these samples which could be aligned to the genome, multiplied again by one million. This normalization also adjusts for the sequencing depth by considering the total number of genome-aligned reads, ensuring consistent comparisons across samples while reducing the bias from sequencing contaminants or non-species-specific reads. The rpmmm-normalization divides for every sample the raw read counts of each feature by the total number of reads of this sample which mapped only to miRNAs and multiplies the fraction by one million. This normalization method is only suitable for a miRNA dataset. The quantile normalization transforms the dataset in order to obtain an equal distribution of read counts for each sample. Therefore, we calculate the mean for every smallest count value of each sample, then for the second smallest, the third smallest until all values

are processed. Then we replace all smallest elements with their mean, every second-smallest elements with their mean and so on. This procedure removes technical biases, for example sequencing depth and batch effects while preserving relative expression levels. Additionally, it accounts well for differences in library size and corrects for compositional biases. Therefore, it is suited for datasets with homogenous expression patterns across the features. The *DESeq2* normalization uses a size factor based on the median ratio method. We first compare the geometric mean g_i for all features given for feature i by $g_i = \left(\prod_{j=1}^n X_{i,j} \right)^{\frac{1}{n}}$, if $X_{i,j} = 0$ holds, we remove feature i for every sample. The size factor s_j represents the median over $\frac{X_{i,j}}{g_i}$ for all features after removing the features containing at least one 0. For each feature i , the ratio reflects how much the expression level of that feature in sample j deviates from the overall geometric mean. The median across all features i ensures that the size factor is robust against outliers, for example if highly expressed features skew the results. To finally normalize the data, we divide for each sample j the feature values by the size factor s_j . This normalization removes differences according to the library size and corrects for compositional bias. For mRNA datasets using the *DESeq2* normalization or quantile normalization are best suited due to their compensation of compositional biases. Further, they stabilize the high variance within the datasets. For miRNAs, *rpmm* and *rpmm* are well suited, due to their low number of features and since their narrow dynamic range compared to mRNAs renders compositional biases are less pronounced. While the *rpm* normalization is also often used due to its simple and quick calculation, it may not be as robust for some properties within the datasets, like compositional bias.

Following the normalization, we perform a first quality control by discarding samples with a low read quantity followed by a filtering for features.

1.2.1.2 Quality control filtering criteria

Formally, let $X' \in \mathbb{R}^{s' \times n'}$ be the normalized data point matrix where s' describes the number of features and n' the number of samples. For example, the element $X'_{i,j}$ for $i \in 1, \dots, s'$ and $j \in 1, \dots, n'$ is the i -th feature of the j -th sample. Additionally, a metadata or annotation table M accompanies the dataset which consists of

one row for each sample and one column for each attribute of the samples, for example treatment, age, tissue or batch. Hence, we regard M as a table with size $n' \times (m + 1)$ where m denotes the number of attributes the dataset exhibits.

To ensure their experimental relevance, we filter the normalized data point matrix for samples with a high number of aligned reads against the genome. A common threshold in our analyses was 2 million reads. Most experimental designs ensure that the attributes are given in a balanced form where each attribute realization is more or less equally often present. This usually holds true after the sample filtering. In case too few samples for an attribute realization remain after the sample filtering, all samples belonging to this realization must be removed.

An additional filtering with respect to the features is performed by checking if enough samples exhibit a higher raw count than a preselected value. How many samples have to fulfill the criterion is often indicated by a percentage, typical values can be 10%, 50%, 75% and 100%, depending on the specific dataset. The preselected raw count threshold should be above the detected noise level of the data, in case of miRNAs this values is often set to 5 [184]. This filtering is often only applied to partitions of the dataset. The partitioning is achieved with respect to a chosen attribute of the samples and the filtering is performed individually for each part. The obtained features from every filtering are joined and serve as the new feature set for the final filtered and normalized data point matrix $X \in \mathbb{R}^{s \times n}$ where $s \leq s'$ and $n \leq n'$.

1.2.2 Statistical testing and correlation analysis

Statistical methods are essential tools in bioinformatics for drawing meaningful conclusions from experimental data. This section introduces fundamental hypothesis tests, including the *Shapiro-Wilk-Test*, *Student's t-test*, and *Wilcoxon rank-sum test*, alongside p -value adjustments to account for multiple comparisons. Additionally, the section explores correlation analysis, focusing on *Pearson correlation coefficient* and *Spearman's rank correlation coefficient*, to evaluate relationships among variables and their statistical significance. The presented descriptions are based on [207–211] and [207, 212–214] for the two parts respectively.

Many of the methods introduced below require normally distributed inputs. To push our data values closer to an approximation of a normal-distribution, to reduce skew within the dataset and to stabilize the variance, we apply a logarithmic-transform to the gene or miRNA expression matrix by computing $\log(X + \epsilon)$ with $\epsilon > 0$ (in our case $\epsilon = 1$) [215, 216].

1.2.2.1 Hypothesis testing

Comparing gene or miRNA expression across groups of samples, for example, in *case* versus *control* setups, relies heavily on hypothesis testing [217–225]. We can assess whether the observed dataset differences are statistically significant or due to random chance. Hypothesis testing starts with defining two mutually exclusive hypotheses: the null hypothesis (H_0) where we assume no effect, difference or relationship exists and the alternative hypothesis (H_1) where we assume that an effect, difference or relationship exists. The objective is to evaluate the evidence against H_0 in favor of H_1 based on a chosen significance level α , typically set to $\alpha \in \{0.05, 0.01, 0.001\}$. The result of a test is the p -value ranging between 0 and 1 which is defined as the probability of measuring a test statistic as extreme as or more extreme than the one calculated from the given dataset, under the assumption that H_0 is true. For a right-tailed test this means $p = P(T \geq T_{\text{observed}} | H_0)$, for a left-tailed $p = P(T \leq T_{\text{observed}} | H_0)$ and for the two-tailed $p = P(T \leq T_{\text{observed}} | H_0) + P(T \geq T_{\text{observed}} | H_0) = 2 \cdot P(T \geq |T_{\text{observed}}| | H_0)$. A small p -value ($p < \alpha$) provides evidence to reject H_0 , while a large p -value ($p \geq \alpha$) suggests insufficient evidence to reject H_0 . The p -value was originally obtained from look-up tables, based on precomputed null distributions. Modern implementations now either use mathematical approximations for the p -value or spline interpolation of less precomputed values from the null distribution.

Shapiro-Wilk-Test - Many methods require normally distributed data points. To test if one vector of the data point matrix exhibits a normal distribution with unknown mean μ and variance σ , the univariate *Shapiro-Wilk-Test* was developed [208]. In our case and for the further analysis, we are interested if a feature vector $X_{i,\cdot}$ is normally distributed. The hypotheses for this test are: H_0 that a feature vector $X_{i,\cdot}$ follows a normal distribution and H_1 that it does not follow a normal

distribution. The test calculates a statistic W based on the vector $X_{i,\cdot}$, which is sorted decreasingly and is given by

$$W = \frac{\left(\sum_{j=1}^n w_j X_{i,j}\right)^2}{\sum_{j=1}^n (X_{i,j} - \overline{X_{i,\cdot}})^2}$$

where $\overline{X_{i,\cdot}} = \frac{1}{n} \sum_{j=1}^n X_{i,j}$ and w_i are weights derived from the expected normal distribution. The numerator reflects how well the data points align with the normal distribution's expected structure, and the w_i -weights play a key role in this alignment. In case of a normal-distribution the numerator is an estimate of the total variance within the vector (like the denominator). Without these weights, the test statistic W would not properly reflect how the data point deviates from the expected structure of a normal distribution. The weights are designed to give more emphasis to the data points that are most informative about normality. This means the smallest and the largest values (tails of the distribution) tend to have larger weights because deviations in the tails are more indicative of non-normality. Since the weights depend only on the vector length, the n expected values and the covariance matrix of an order statistics of a standard normal distribution ($\mathcal{N}(0, 1)$), they can be pre-calculated and stored in look-up tables. The resulting values are the test statistic $W \in [0, 1]$ and the decisive parameter p -value. Since it uses a left-tailed test, the p -value is given by

$$p = P(W \leq W_{\text{observed}} | H_0)$$

to evaluate the evidence against H_0 . If $p < \alpha$ holds, we reject H_0 , which means $X_{i,\cdot}$ is not normally distributed. In contrast, if $p \geq \alpha$ holds, we fail to reject H_0 which means $X_{i,\cdot}$ is likely normally distributed. This statement is sufficient to proceed with methods requiring normally distributed data points. The test is sensitive to deviations from normality, which makes it effective for the task. The performance decreases for very small vector lengths or with data points heavily skewed by outliers. If the data points are highly discrete which means only a few unique values exist, the test may produce unreliable results because it assumes a

continuous distribution.

In this thesis, we often use tests to compare two attribute realizations m_A and m_B introduced by the metadata table M for each sample. We subset the data point matrix for the samples belonging to one realization each to obtain two matrices A and B . Therefore, $A \in \mathbb{R}^{s \times n_A}$ and $B \in \mathbb{R}^{s \times n_B}$ holds where n_A and n_B denote the number of samples fulfilling realization m_A and m_B , respectively. We remove constant rows (mainly rows containing only zeros) from our data point matrix. If the variance for a feature over all samples is constant, there is no distinguishing information included and many methods would fail for this feature.

The hypothesis testing framework is given by the null hypothesis (H_0) which stands for no difference between the groups for one feature i and the alternative hypothesis (H_1) claiming a significant difference between the two groups for one feature i . We assume that the feature vectors $a := A_{i,\cdot}$ and $b := B_{i,\cdot}$ are normally distributed since we applied *Shapiro-Wilk-Test* to the feature vectors across all samples. By Jensen [226] we conclude that if the feature vector across all samples is normally distributed, then the feature vector of a subset (apply for A and B) is also normally distributed.

Student's t-test - Here, we compare the means of the two feature vectors a and b under the assumption of normality. For H_0 , this implies that the means of the two groups are equal ($\mu_a = \mu_b$) and for H_1 that the two means are different ($\mu_a \neq \mu_b$). We compute the test statistic t with

$$t = \frac{\bar{a} - \bar{b}}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}}$$

where s_a^2 is the variance within a and n_a is the length of the vector a and for s_b^2 and n_b , respectively. The numerator measures the difference between the sample means, and the denominator adjusts this difference for the variability and sample sizes of the two groups. Since we allow that the two groups can have different variances, we consider an extension of the *Student's t-test* called *Welch's Test* to calculate the p -values. We obtain the degrees of freedom df by using the *Welch-Satterthwaite formula* which adjusts the df based on the relative variances and the

sample sizes of the groups. In the thesis, we use the two-sided test to calculate the p -value by

$$p = 2 \cdot P(t \geq |t_{\text{observed}}| | \text{df}) = 2 \cdot (1 - F_t(|t_{\text{observed}}| | \text{df}))$$

where t follows a t -distribution with df degrees of freedom and F_t is the cumulative distribution function of the t -distribution which can be determined analytically. The t -distribution is similar to the standard normal distribution, while for small degrees of freedom the distribution is wider with heavier tails (to account for uncertainty due to limited data). For $\text{df} \rightarrow \infty$ the t -distribution converges to the standard normal distribution. We compute the p -value by observing a t -statistic as extreme as t_{observed} in the right tail. If $p < \alpha$, we reject H_0 which indicates the means are significantly different. An advantage is that the two feature vectors can have different lengths and different variances which makes it widely usable in many real-world applications. The univariate test assumes normally distributed independent groups with large sample sizes and no significant outliers. This assumption is crucial because the test relies on the mean and the variance, which can be disproportionately affected by extreme values. Additionally, the data points are assumed to be continuous or mildly discrete.

Wilcoxon rank-sum test (Mann-Whitney U test) - Here, we compare the distributions, specifically the medians, of two feature vectors a and b without assuming normality. The hypothesis H_0 proposes that the distributions of the two feature vectors are identical and H_1 claims the distributions are different. We combine the two feature vectors and create a mapping table to rank the values. To do so, we sort the combined vector increasingly and rank it with integers starting by 1. In case of ties (identical values), we assign the average of those tied values. Using this mapping table we transform the two feature vectors a and b . We compute the test statistics W as the sum of ranks for one feature vector with

$$W_a = \sum_{i=1}^{n_A} R_i$$

where n_A denotes the number of samples in a and R_i the rank mapped to the sample i . By construction, it holds that $W_a + W_b = \frac{(n_A + n_B) \cdot (n_A + n_B + 1)}{2}$. If W_a is known, W_b can easily be computed by reformulating the equation above. Under the null hypothesis H_0 , the expected value $E(W)$ is given by

$$E(W_a) = \frac{W_a + W_b}{n_A + n_B} = \frac{n_A(n_A + n_B + 1)}{2}$$

and the variance σ_W^2 of the rank sum W by

$$\sigma_W^2 = \frac{n_A n_B \cdot (n_A + n_B + 1)}{12}.$$

Next, we standardize the test statistic W by

$$Z = \frac{W_a - E(W_a)}{\sqrt{\sigma_{W_a}^2}}.$$

The p -value is based on the standardized test statistic Z and is achieved using a two-tailed test by

$$p = 2 \cdot P(Z \geq |Z_{\text{observed}}|)$$

where $P(Z)$ is the cumulative probability from the standard normal distribution. If $p < \alpha$, we reject the null hypothesis H_0 which means that the distributions are significantly different. This univariate test is robust to non-normal data points (non-parametric) and outliers but is less sensitive to small differences in means.

1.2.2.2 Adjustments for multiple testing

When conducting multiple test as for instance by calculating p -values for every feature, the probability of false positives increases as well. Without adjustment, raw p -values only control the error rate per test. They do not account for the fact that multiple hypothesis tests are performed simultaneously. For example, if we do 20 tests with a significance level $\alpha = 0.05$ (which describes the probability of a false positive), we obtain a probability that at least one false positive is included as $1 - (1 - 0.05)^{20} \approx 0.64$ which means we have a 64% chance to obtain a false positive. The probability of at least one Type I error (false positive) in a family

of s simultaneous tests is called *family-wise error rate* FWER and can always be computed by $\text{FWER} = 1 - (1 - \alpha)^s$. To overcome this problem we adjust the p -values for this multiple testing. Adjustments control for the family-wise error rates or false discovery rates.

Bonferroni correction - To control the *family-wise error rate*, the significance level is rescaled by the number of tests s using the *Bonferroni correction*: $\frac{\alpha}{s}$. This changes the *family-wise error rate* to $\text{FWER}_{\text{Bonferroni}} = 1 - (1 - \frac{\alpha}{s})^s$ and the rejection criterion to $p < \frac{\alpha}{s}$. In practice, it is more convenient to transform the p -value instead of the significance level. Therefore, we adjust the p -values obtained for each feature by

$$p_{\text{adjusted}} = p \cdot s \geq p.$$

Considering the rejection criterion $p \cdot s < \alpha$, this is equivalent to reducing the significance level for each individual test which also induces a lower *family-wise error rate*. This approach is very conservative, it reduces Type I errors but raises Type II errors (false negatives).

Benjamini-Hochberg procedure (BH) - The share of Type I errors (false positives) within all rejected hypotheses, known as False Discovery Rate (FDR) and defined as $\text{FDR} = E \left(\frac{|\text{false positives}|}{|\text{true positives}| + |\text{false positives}|} \right)$, is controlled by the *BH* method.

For $R = 0$ where R corresponds to all rejected hypotheses (the denominator in the FDR formula) the FDR is set to 0. In contrast to the *Bonferroni correction*, we only consider the share of false positives within all positive values (rejected hypotheses). The p -values obtained for each feature are sorted increasingly $p_1 \leq p_2 \leq \dots \leq p_s$. For a given α , we assume i is the largest index fulfilling $p_i \leq \frac{i}{s} \cdot \alpha$. We reject all null hypotheses corresponding to $p_1, p_2, \dots, p_i$. To achieve the adjusted p -values, we calculate

$$p_i^{\text{adjusted}} = \min \left(\frac{s}{i} \cdot p_i, 1 \right)$$

for all features. With this we ensure that the FDR, the expected proportion, is below α . This approach balances the Type I and II errors.

Using the adjusted p -values, we check for each feature individually whether the adjusted p -value is smaller than the significance value. If this holds, we denote the feature as significant according to the underlying attribute comparison. The decision for a specific adjustment method depends on the desired outcome. If we want to achieve a strict control of false positives we use the *Bonferroni correction*. For large-scale exploratory studies, we use *BH*.

1.2.2.3 Differential expression (DE) analysis

The exemplary scenario of comparing the expression of two groups of samples serves as the basis for the DE analysis approach, which is a standard step in data-driven studies [224, 227–236]. For two attribute realizations m_A and m_B , we want to see if some features for the corresponding samples differ. Exemplarily, we consider the underlying comparison to be between the two realizations *case* versus *control*.

Fold change - First, we calculate the *fold change* FC describing, if the average expression of every feature across all samples between the realizations is deregulated. The FC is given by

$$FC = \frac{\frac{1}{n_A} \sum_{j=1}^{n_A} X_{.,j}}{\frac{1}{n_B} \sum_{j=1}^{n_B} X_{.,j}}.$$

If $FC \geq t_{\text{dereg}}$ holds for one feature i , then we say the feature i is upregulated. In contrast, if $FC \leq \frac{1}{t_{\text{dereg}}}$, we call it downregulated. Threshold t_{dereg} values like 1.5 or 2 are common. Using the $\log_2$ -transform, we can combine the inequalities to an overall deregulation statement with $|\log_2(FC)| \geq \log_2(t_{\text{dereg}})$ utilizing the logarithmic laws. The previously introduced p -values can be used to decide if the investigated comparison is significant by checking if $p_{\text{adjusted}} < \alpha$. A feature is considered significantly deregulated (up or down depending on FC) if it fulfills both criteria. The interplay between fold changes and the adjusted p -values is often displayed in a volcano plot which is a scatter plot showing the $\log_{10}$ -

transformed adjusted p -values against the $\log_2$ -transformed FCs.

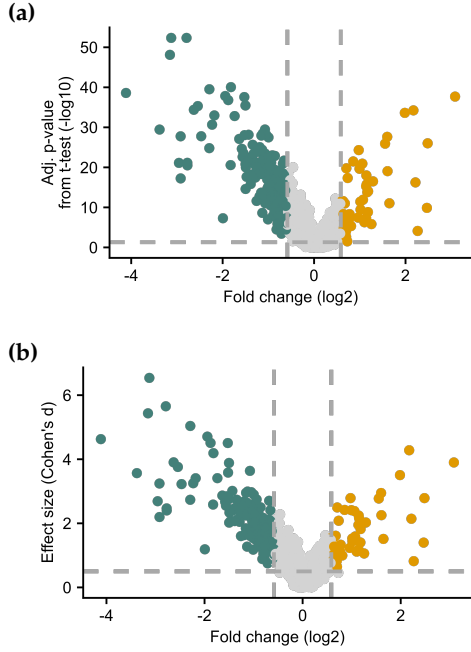


Figure 1.4: Visualization for a DE analysis. **(a)** Volcano plot. The significantly downregulated miRNAs are highlighted in green and in yellow the significantly upregulated miRNAs for $|\log_2(\text{FC})| \geq \log_2(1.5)$ and $\alpha = 0.05$. **(b)** Scatter plot. The downregulated miRNAs with a considerable effect are highlighted in green and in yellow the upregulated ones with a considerable effect for $|\log_2(\text{FC})| \geq \log_2(1.5)$ and $|d| \geq 0.5$.

Cohen's d - While hypothesis tests determine whether an effect exists (statistical significance), *Cohen's d* measures how large that effect is (practical significance). It is common to report both a p -value and *Cohen's d* to provide a complete picture of the results. The combination of adjusted p -values and the fold changes together with *Cohen's d* is used to detect promising feature candidates. The following is based on Cohen [237]. *Cohen's d* quantifies the difference between the two attribute realization means in terms of standard deviations by

$$d = \frac{\bar{a} - \bar{b}}{\sigma_{\text{pooled}}} \text{ where } \sigma_{\text{pooled}} = \left(\frac{(n_A - 1)\sigma_a^2 + (n_B - 1)\sigma_b^2}{n_A + n_B - 2} \right),$$

$\bar{a}$, $\bar{b}$ denote the means of feature vectors and σ_a^2 and σ_b^2 the variances of a and b , respectively. No effect can be measured, if $|d| = 0$, a small but noticeable difference, if $|d| = 0.2$. If $|d| = 0.5$, we call the difference moderate, so likely meaningful and a large effect can be seen for $|d| = 0.8$ which identifies a clearly noticeable difference. It is unit-less, making it comparable across different studies and datasets. However, it assumes the two input vectors to be normally distributed

and for large variances or small sample sizes the estimate can be less reliable. Like for the p -values, the magnitude of d just refers to the size of the effect but not which attribute realizations exhibit higher expression values.

An example of a volcano plot showing the adjusted p -value (*Student's t-test*, *BH*) against the fold change for each feature is depicted in Figure 1.4a. In green, the significantly downregulated miRNAs are highlighted and in yellow the significantly upregulated ones (for $|\log_2(\text{FC})| \geq \log_2(1.5)$ and $\alpha = 0.05$). An example of a scatter plot showing the absolute *Cohen's d* against the fold change for each feature is shown in Figure 1.4b. The downregulated miRNAs with a considerable effect are highlighted in green and in yellow the upregulated

one with a considerable effect (for $|\log_2(\text{FC})| \geq \log_2(1.5)$ and $d \geq 0.5$). The lines indicate the chosen thresholds, the comparison is based on two attribute realizations from the metadata table. The depicted data points were taken from the planned publication I and can be accessed online at GEO via the accession number GSE282205 (available at latest at publication).

1.2.2.4 *Correlation methods*

Understanding relationships between variables in high-dimensional datasets is crucial in bioinformatics analyses, a task often achieved using correlation methods [231, 238–246]. They offer an intuitive and quantitative way to assess pairwise relations together with their direction, which, in turn, can be used to identify patterns, redundancies and dependencies between variables. They serve as exploratory data analyses and enable further structured and focused downstream methodologies like feature selection and dimension reduction. General advantages of correlation analyses are that they are simple to understand and widely used. Furthermore, their simple calculations can be easily parallelized and therefore scale well with larger datasets. Additionally, they can be applied to a wide range of data types and tasks. In this thesis, we calculate correlations in different scenarios. First, we consider sample correlations which refer to the pairwise relations between every two samples. Next, the correlation value between a feature and an attribute from the metadata table, for example age, to determine if a feature is correlated with age. Lastly, we calculate the relation between a feature from one dataset to one from another, in our case this means to consider a miRNA from one dataset and a mRNA from another where the two datasets share identical samples.

However, there exists general limitations like the bivariate nature of correlations by presenting only the relation between exactly two variables each. Therefore, a complex interaction between three or more variables cannot be explained using bivariate correlation analysis. Since the calculated correlation is symmetric no statements regarding causality can be derived from the coefficients. Some methods can be sensitive to certain data characteristics like outliers, non-linearity, and skewed or non-uniform distributions. To be able to compare correlation coefficients from different pairs of variables, matching data values for both

variables are required. Here, we introduce the correlation methods exemplary for sample correlations.

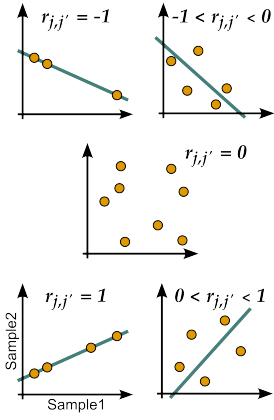


Figure 1.5: Exemplary scatter plot with correlation values $r_{j,j'}$.

Pearson correlation coefficients - Capturing linear relationships between variables in high-dimensional datasets is often achieved using *Pearson correlation coefficients*. The correlations values range from -1 (a perfect negative linear relationship) to 1 (a perfect positive linear relationship). A vanishing correlation coefficient means that there is no linear relationship. An exemplary, sketched overview of the different correlation values is depicted in Figure 1.5. For two samples j and j' the sample correlation is defined as

$$r_{j,j'} = \frac{\sum_{i=1}^s (X_{i,j} - \bar{X}_j)(X_{i,j'} - \bar{X}_{j'})}{\sqrt{\sum_{i=1}^s (X_{i,j} - \bar{X}_j)^2} \sqrt{\sum_{i=1}^s (X_{i,j'} - \bar{X}_{j'})^2}}$$

where $\bar{X}_j = \frac{1}{s} \sum_{i=1}^s X_{i,j}$ denotes the mean over all features of sample j and $\bar{X}_{j'}$ for sample j' analogously. The correlation coefficient $r_{j,j'}$ quantifies relationship of two variables in strength and direction. Still, it does not indicate whether the observed correlation is statistically significant or could have occurred by chance. Therefore, a hypothesis testing resulting in a p -value followed by an adjustment for multiple testing helps to assess statistical significance. The test hypothesis is given by H_0 which states that the correlation coefficient is $r = 0$, which means no linear relationship exists. In contrast, H_1 claims $r \neq 0$, indicating linear relation. The used test statistic is the t -statistic which we obtain by

$$t = r_{j,j'} \cdot \left(\frac{s-2}{1-r_{j,j'}^2} \right)^{\frac{1}{2}}.$$

We obtain the p -values from the t -distribution with $\text{df} = n - 2$, introduced in Section 1.2.2.1. Since we frequently calculate many correlations and p -values simultaneously, for example $\frac{n(n-1)}{2}$ in the case of sample correlations, an adjustment for multiple testing of the p -values is mandatory (see Section 1.2.2.2). We reject H_0 if $p_{\text{adjusted}} < \alpha$, which indicates a significant linear relationship. While the *Pearson correlation coefficient* assumes a linear relationship, it is used to validate

and quantify this assumption. Note, the method is sensitive to outliers which can distort the coefficients. Additionally, normally distributed data points are necessary for meaningful p-values in the significance testing.

Spearman's rank correlation coefficient - The strength and direction of a monotonic relationship between two variables can be measured using *Spearman's rank correlation coefficient*, which operates based on their ranks. Again, the obtained values range between -1 (a perfect negative monotonic relationship) and 1 (a perfect positive monotonic relationship). A value of 0 means that there is no monotonic relationship. It describes whether variables change simultaneously in a consistent direction, regardless of the form of the relationship, for example linear or non-linear trends like exponential or logarithmic behaviors. Consider one variable to increase exponentially while the second one increases linearly, then their relationship is not linear, but is still identified by *Spearman's rank correlation* and not by *Pearson correlation*. For two samples j and j' , the sample correlation is achieved by computing the ranks of the values of each sample across all features, then applying the *Pearson correlation* formula to the ranks. To rank the data, we need to create a mapping table by sorting the values and number them from 1 to s where s is the number of features. In case of ties we replace these ties by the averaged ranks for this tie to ensure consistency and robustness. Ranks are now computed by using the created mapping table to transform the original values. Applying the *Pearson correlation* results in the formula given by

$$\rho_{j,j'} = \frac{\sum_{i=1}^s (R_{i,j} - \bar{R}_j)(R_{i,j'} - \bar{R}_{j'})}{\sqrt{\sum_{i=1}^s (R_{i,j} - \bar{R}_j)^2} \cdot \sqrt{\sum_{i=1}^s (R_{i,j'} - \bar{R}_{j'})^2}}$$

where $R_{i,j}$ is the rank of feature i for sample j and $\bar{R}_j$ is the mean rank of all features for sample j , and it follows analogously for sample j' . Again, to guarantee statistical significance of the results, we use a hypothesis test to obtain p-values followed by a mandatory adjustment for multiple testings (see Section 1.2.2.1 and Section 1.2.2.2). The test hypothesis is given by H_0 assuming that the correlation coefficient is $\rho = 0$ which means no monotonic relationship. The alternative hypothesis H_1 is given by $\rho \neq 0$ indicating that a monotonic relationship exists.

Scenario	Recommended method
Linear relationship expected	<i>Pearson corr. coeff.</i>
Non-linear monotonic relationship	<i>Spearman's rank corr. coeff.</i>
Data points are normally distributed	<i>Pearson corr. coeff.</i>
Data points are non-normal or ordinal	<i>Spearman's rank corr. coeff.</i>
Outliers are present	<i>Spearman's rank corr. coeff.</i>
Exploring a dataset with unknown relationships	<i>Spearman's rank corr. coeff.</i>

Table 1.1: Decision-making guide for choosing between *Pearson correlation coefficient* and *Spearman's rank correlation coefficient*.

To obtain the p -value for small sizes of variable ($s \leq 30$), we use an exact but computationally expensive permutation-based method based on the distribution of the rank correlations coefficients. For larger variables ($s > 30$), we use an approximation of the t -distribution. The test statistic is given by

$$t = \rho_{j,j'} \cdot \left(\frac{s-2}{1-\rho_{j,j'}^2} \right)^{\frac{1}{2}}.$$

This formula is the same as the test statistic used in *Pearson correlation coefficient*, but it is applied to the *Spearman's rank correlation coefficient* instead of the *Pearson correlation coefficient*, since both only rely on the distributions of the two variables. We reject the H_0 if $p_{\text{adjusted}} < \alpha$, which indicates a significant monotonic relationship. The method is robust to outliers and non-normal distributions and is suitable for ordinal datasets where the actual values are not meaningful. Yet, by considering only the ranks, we lose the actual numerical amount of the relations and differences and therefore may fail to detect subtle linear trends.

Choosing the appropriate method for a dataset is dependent on the specific characteristics of the dataset. First, we need to inspect the relations within our dataset visually by utilizing scatter plots, check the distributions of the investigated variables by normality testing (*Shapiro-Wilk-Test*), Q-Q plots, histograms and check for outliers. If the relationships appear linear and the data points are normally distributed, the *Pearson correlation coefficient* is suitable. In contrast, if the relationships appear monotonic or the data points are not normally dis-

tributed *Spearman's rank correlation coefficient* is better suited. The Table 1.1 offers an overview regarding when to select which method. In doubt, the calculation of both correlation coefficients to compare and assess their consistency is necessary. For example a direct comparison of the correlation coefficients can be done. If both correlations yield the same coefficients ($|r - \rho| < 0.1$), the relationship is likely linear and *Pearson correlation coefficient* is sufficient to capture it. If the coefficients differ and the *Spearman's rank correlation coefficient* is high, it is likely, that there is a non-linear monotonic relationship, therefore the *Spearman's rank correlation coefficient* is the suitable approach. In contrast, if the *Pearson correlation coefficient* is high and the *Spearman's rank correlation coefficient* low, there might be subtle linear trend. Additionally, to consider the accompanying adjusted p -values together with the coefficients is universally recommended. If one of the correlation coefficients is not significant, the observed correlations are not statistically significant and may have occurred by chance.

1.2.3 Embedding and variance analysis

In high-dimensional datasets, the relationships between samples or features are often obscured by the complexity of the data structure. By projecting these datasets into lower-dimensional spaces using embedding methods, such as *Principal Component Analysis (PCA)* and *Uniform Manifold Approximation and Projection (UMAP)*, we reveal patterns and relationships while simultaneously reducing noise. These methods are widely used for dimensionality reduction and feature extraction in bioinformatics [247–263]. Complementing these approaches, *Principal Variance Component Analysis (PVCA)* leverages variance decomposition to explain variability in the dataset with experimental factors and thus sheds light onto the underlying drivers of the variation.

Principal Component Analysis - Simplifying complex datasets while preserving as much variation as possible is a key function of *PCA*, aiding interpretation and downstream analysis. Graphically speaking, in this representation, each feature is treated as a dimension in an s -dimensional space, and we rotate, shift and scale the feature space so that we end up with a new set of orthogonal axes (principal

components). The first few principal components capture most of the variance in the data.

The presented approach is adapted from Jolliffe [264]. To formulate the transformation, we first need a few definitions. The *mean centering* X_{centered} is given by the data point matrix $X \in \mathbb{R}^{s \times n}$ and the mean vector $\mu \in \mathbb{R}^s$. It consists of the mean of each row from X and therefore every feature

$$X_{\text{centered}} = X - \mu \text{ with } \mu = \frac{1}{n} \sum_{j=1}^n X_{\cdot j}$$

where $X_{\cdot j}$ denotes the j -th column of the matrix X (all feature values for sample j). Often an X_{scaled} is used where additionally each row of X_{centered} is divided by the standard deviation of this row (feature). The $s \times s$ sized *empirical covariance matrix* S is calculated for the given samples with

$$S = \frac{1}{n} X_{\text{centered}} X_{\text{centered}}^\top = \frac{1}{n} \sum_{j=1}^n (X_{\cdot j} - \mu_j)(X_{\cdot j} - \mu_j)^\top.$$

Note that the *covariance matrix* of X is defined by $S = \mathbb{E}(X - \mathbb{E}(X))(X - \mathbb{E}(X))^\top$. Since S is symmetric and positive semi-definite, because the quadratic form is based on a squared norm, we obtain the *eigen decomposition* by solving for the eigenvalues λ_i and the corresponding eigenvectors u_i of S :

$$S u_i = \lambda_i u_i$$

with $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_l \geq 0$. We assume that S is positive definite, this means $\lambda_l > 0$, and we define $U \in \mathbb{R}^{s \times l}$ as the matrix containing the eigenvectors u_i for $i = 1, \dots, l$, which we call *principal component axis*. Therefore, the span $(u_1, \dots, u_l)$ forms a basis of the $\mathbb{R}^l$. The projection of the original data points onto the *principal component axis* is obtained by

$$Y = U^\top X$$

where Y is the transformed data point matrix in the principal component space. We call the k -th row of Y the k -th principal component PC_k , containing the projections of all samples onto the k -th principal component axis. The proportion of the variance given by PC_k can be explained by:

$$\phi(PC_k) = \frac{\lambda_k}{\sum_{i=1}^s \lambda_i}.$$

To limit the number of principal component axis to the d axis which contain most of the variance in the data, we calculate

$$\frac{\sum_{i=1}^d \lambda_i}{\sum_{i=1}^l \lambda_i} \geq t \quad (1)$$

and choose the smallest $d \leq l$ which fulfills the inequality for a pre-selected t . Here, t represents the percentage of the original data variance which we want to capture in the transformed data.

For a typical graphical interpretation of the transformed data points, we choose $d = 2$. An example is shown in Figure 1.6, we use the same data points as in Figure 1.4. We use X_{scaled} and the percentages at the axis labels indicate the proportion of variance ϕ for the corresponding principal component. The data points are colored according to an attribute derived from the metadata table.

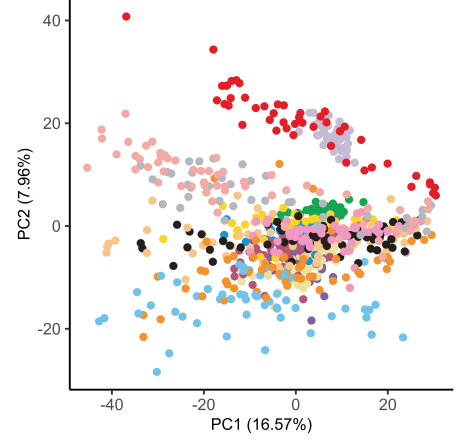


Figure 1.6: Graphical representation of a PCA for $d = 2$ colored by a metadata attribute. The percentages at the axis are the proportions of variance of each respective principal component PC_k .

Uniform Manifold Approximation and Projection - Preserving both local and global structures in high-dimensional data, *UMAP* is a nonlinear dimensionality reduction technique that excels in visualization and clustering. In comparison, *PCA* is solely a linear dimensionality reduction method and retains global relationships but may fail to capture local ones. The following approach is adapted from McInnes, Healy, Melville [265]. Manifold learning which consider that high-dimensional data is located on a lower-dimensional manifold which is embedded within the high-dimensional space is the foundation of *UMAP*. The topological structure represented by the graph of nearest neighbors capturing how points are connected locally and globally serves as a computational approximation of the manifold's structure. The embedding matrix Y is the equivalent of the data

point matrix X in the low-dimensional space. It has the dimensions $d \times n$, where d (typically 2 or 3) is the number of dimensions in the reduced space, and n corresponds to the number of samples.

First, we need to construct a weighted graph capturing the geometry of the manifold based on the given standardized (z-scored) data points X . Therefore, we calculate the *local neighborhood estimation* for each column (sample) of X by identifying its k -nearest neighbors based on the distances between the columns. To do so, we use the conditional probability $p_{j,j'}$, which describes the likelihood of a sample j being a neighbor of sample j' . Using the distance $d(X_{\cdot,j}, X_{\cdot,j'})$, typically the Euclidean or cosine distance, between samples j and j' , we formulate

$$p_{j,j'} = \exp \left(-\frac{d(X_{\cdot,j}, X_{\cdot,j'}) - \rho_j}{\sigma_j} \right)$$

where $\rho_j = \min_{l=1,\dots,n} d(X_{\cdot,j}, X_{\cdot,l})$ denotes the local connectivity adjustment (the distance to the nearest neighbor of sample j) and σ_j is a scaling parameter. The latter is optimized numerically per sample to balance the number of effective neighbors. The subtraction of ρ_j ensures that the closest neighbor always has a significant weight, regardless of the value of σ_j . The scaling factor σ_j then controls how quickly the probabilities vanish for distant neighbors. Driven by the pre-selected hyperparameter k , σ_j ensures that only for the k nearest neighbors, so the k samples with the highest $p_{j,j'}$ values, $p_{j,j'} > 0$ holds. In order to obtain a symmetric graph we replace the directed by undirected edges and combine $p_{j,j'}$ and $p_{j',j}$ via $p_{j,j'}^{\text{sym}} = p_{j,j'} + p_{j',j} - p_{j,j'} \cdot p_{j',j}$. The result is a symmetric, weighted graph called *nearest neighborhood graph* with vertices being the samples and undirected edges between all vertices. The edge weights are given by the $p_{j,j'}^{\text{sym}}$ which capture the pairwise local relationships between all samples.

Second, we optimize the low-dimensional embedding such that it forms a flattened representation of the manifold while preserving the local and global structure of the manifold. Therefore, we need to define a low-dimensional space by initializing a $d \times n$ sized representation of the data point matrix Y by random values, typically from a small Gaussian distribution. In this low-dimensional

space, we define the similarity or new edge weight of two samples j and j' by a modified form of a Student's t -distribution:

$$q_{j,j'} = \frac{1}{1 + a \|Y_{\cdot,j} - Y_{\cdot,j'}\|^{2b}}$$

where a controls the steepness of the similarity $q_{j,j'}$ for close neighbors and b determines how probabilities decrease for larger distances. The parameters a and b are derived numerically based on the pre-selected minimal distance parameter which serves as a hyperparameter. Both parameters control the embedding's structure, and they define how the *UMAP* embedding balances local and global structure by shaping the probability curve. The term $\|Y_{\cdot,j} - Y_{\cdot,j'}\|$ represents the distance between the two samples j and j' given by the Euclidean norm. To obtain Y , we minimize the loss function which is the difference between the high-dimensional probabilities $p_{j,j'}^{\text{sym}}$, originating from the original data, and the low-dimensional probabilities $q_{j,j'}$. Therefore, we write the *cross-entropy-based* loss function as

$$\text{loss} = - \sum_{j \neq j'} \left(p_{j,j'}^{\text{sym}} \cdot \log(q_{j,j'}) + (1 - p_{j,j'}^{\text{sym}}) \cdot \log(1 - q_{j,j'}) \right).$$

The necessary optimization can be realized using a gradient descent or a related optimization method. During each iteration step, the elements in Y are updated to better reflect the high-dimensionality relationships captured in $p_{j,j'}^{\text{sym}}$. Finally, the resulting matrix Y consists of the new coordinates for every sample.

An example of a *UMAP* is given in Figure 1.7, we use the same data points as in Figure 1.4. The hyperparameters are a minimal distance of 0.25 and local neighborhood size $k = 10$ while using the Euclidean distance as a metric in the high-dimensional space. These parameters were selected by visually assessing the results from a pool of linear combi-

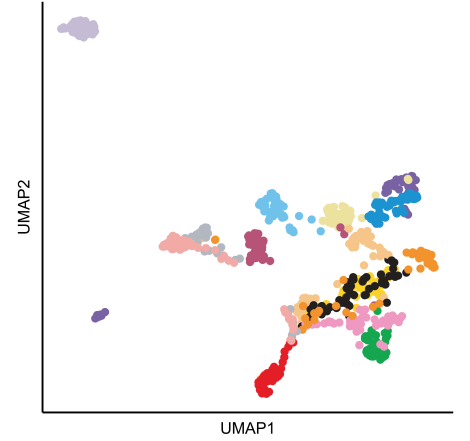


Figure 1.7: Graphical representation of a *UMAP* for $d = 2$ colored by a categorical metadata attribute.

nations of default parameters. The sample points are colored according to an attribute of the available metadata.

When comparing the results for the *PCA* and the *UMAP* applied to the same dataset depicted in Figure 1.6 and Figure 1.7, respectively, we notice that a linear reduction via *PCA* does not achieve a clear distinction between data points according to the color-coded property.

Principal Variance Component Analysis - Combining *PCA* and Variance Component Analysis (*VCA*), *PVCA* partitions the total variance across predefined attributes, allowing for the quantification and comparison of their contributions. In high-dimensional biological datasets, it is crucial to understand how much of the total variance is attributable to known experimental or biological factors, for example treatment, age, tissue or batch effects [112, 266–270]. The *PVCA* method decomposes the variance in high-dimensional datasets quantifies the relative impact of experimental factors. These insights can then guide downstream analyses.

The approach presented in the following sections is adapted from Li et al. [271]. As a first step, we start with a *PCA* of the data point matrix X . Using the formula introduced in Equation 1, we obtain the d principal components which explain a predefined proportion of the variance (typically 80% – 90%).

We proceed to the *VCA* which partitions the total variance in the dataset across given attributes by building a *linear mixed-effects model*. For the principal component i , we write

$$PC_i = \mu_i + \left(\sum_{k=1}^{m'} D_{j,k} \beta_{i,k} \right) + \varepsilon_{i,j}$$

with $\mu_i = \frac{1}{n} \sum_{j=1}^n Y_{i,j}$. The design matrix D with size $n \times (m' + 1)$ where $m' \geq m$ is based on the metadata table M and therefore possesses the same number of rows. To construct D , we keep all columns in M containing numbers and columns containing words are split in as many columns as the column has unique elements following *one-hot encoding*. The elements of the newly created columns are binary and equal to 1 if the sample has the specific attribute realization. For example, a column corresponding the attribute *treatment* exhibits two different realizations

case and *control*. We replace this attribute column by two columns called D_{case} and $D_{control}$ and assign 1 in D_{case} to the *case* samples and 0 otherwise. Analogous, we populate in the column $D_{control}$. The variable $\beta_{i,k}$ is a random effect representing the contribution of attribute k of the design matrix to the variability in the i -th principal component and $\varepsilon_{i,j}$ denotes the residual error, representing the variance which cannot be explained by the attributes.

The linear mixed-effects model can be solved by a statistical method typically the restricted maximum likelihood (REML) where a likelihood function needs to be maximized. This likelihood function does not depend on the $\beta_{i,k}$ or the $\varepsilon_{i,j}$ but only on their variances $\sigma_{i,k}^2$ and $\sigma_{i,\varepsilon}^2$, respectively, since we assume that their distributions are given as normal distributions with $\beta_k \sim \mathcal{N}(0, \sigma_{i,k}^2)$ and $\varepsilon_{i,j} \sim \mathcal{N}(0, \sigma_{i,\varepsilon}^2)$. The optimization of the likelihood function is achieved by standard numerical approaches, among those are *Quasi-Newton* methods and the *EM Algorithm*. For each principal component i , we estimate the variance proportion $v_{i,k}$ explained by each attribute k by

$$v_{i,k} = \frac{\sigma_{i,k}^2}{\left(\sum_{k=1}^{m'} \sigma_{i,k}^2\right) + \sigma_{i,\varepsilon}^2}.$$

We combine the proportions for one attribute k and each principal component by calculating

$$PV_k = \frac{\sum_{i=1}^d v_{i,k} \lambda_i}{\sum_{i=1}^d \lambda_i}$$

where λ_i denotes the eigenvalue corresponding to the principal component i . The proportion of variance of the residual which captures the variance not covered by the attributes in the metadata table can be calculated by $PV_\varepsilon = 1 - \sum_{k=1}^{m'} PV_k$. To map the proportion of variance PV_k back to the original attributes in M , we sum up the proportion of variances of all one-hot encoded columns corresponding to the attribute's realizations.

In real-world datasets, many effects are often driven by more than one attribute at the same time. Therefore, it is beneficial to consider the so-called *interaction terms*, which pairwise combine all attributes. The variation of proportion within

the dataset according to interaction terms can also be measured by the *PVCA*. Therefore, we need to reformulate the *linear mixed-effects model* as

$$PC_i = \mu_i + \left(\sum_{k=1}^{m'} D_{j,k} \beta_{i,k} \right) + \left(\sum_{k=1}^{m'} \sum_{l>k} D_{j,k} D_{j,l} \beta_{i,k,l} \right) + \varepsilon_{i,j}.$$

Solving using the aforementioned approach, yields the variances $\sigma_{i,k}^2$, $\sigma_{i,k,l}^2$ and $\sigma_{i,\varepsilon}^2$, which we use to calculate the variance proportions $v_{i,k,l}$ by

$$v_{i,k,l} = \frac{\sigma_{i,k,l}^2}{\left(\sum_{k=1}^{m'} \sigma_{i,k}^2 \right) + \left(\sum_{k=1}^{m'} \sum_{l>k} \sigma_{i,k,l}^2 \right) + \sigma_{i,\varepsilon}^2}.$$

This proportion represents the relative contribution of the interaction between the attributes k and l with respect to the total variance in the i -th principal component.

The proportion of variance for the two attributes k and l are given by

$$PV_{k,l} = \frac{\sum_{i=1}^d v_{i,k,l} \lambda_i}{\sum_{i=1}^d \lambda_i}.$$

Like before, we calculate the proportion of variance of the residual using PV_k and $PV_{k,l}$. All proportions of variance PV_k , $PV_{k,l}$ and PV_ε are often visualized in a bar plot or pie chart.

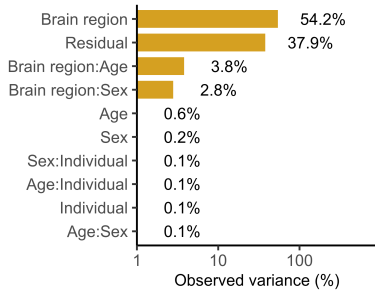


Figure 1.8: Graphical representation of a *PVCA* result including interaction terms.

An example of a *PVCA* result using the data points previously, introduced in Figure 1.4, is displayed in Figure 1.8, whereas vanishing proportions are omitted for clarity. In this example, we transformed the given data point matrix X in a pre-processing step with $X_{\log} = \log_{10}(X + \epsilon)$ whereas ϵ is a small positive constant added to avoid $\log(0)$. This transform is used to stabilize the variances, to compress the dynamic range of the expression value and to improve the distribution to approximate a normal distribution within the miRNA

dataset [215, 216]. As a threshold to calculate the d principal components, we chose $t = 0.9$.

1.2.4 Clustering methods

Clustering is a fundamental technique for exploring the structure in datasets of high-dimension and is widely applied [272–284]. Each cluster groups samples, features, or other entities based on their similarity, such that members within a cluster are more similar to each other than to members of other clusters, with similarity typically quantified using a distance metric. Depending on the methodology, clustering can produce either hard assignments (assigning each sample to exactly one cluster) or soft assignments (allowing partial membership across multiple clusters). The choice of the number of clusters c is a crucial parameter that directly influences the quality and interpretability of the clustering.

Hierarchical clustering - Grouping data points into a hierarchy based on their pairwise distances is achieved through the method of *hierarchical clustering*. It is particularly useful for exploring the underlying structure in high-dimensional data, such as gene or miRNA expression profiles or sample features where it identifies co-expressed features or related samples. The method's ability to visualize clusters in a dendrogram facilitates biological interpretation. In contrast to other clustering methods, initially specifying the number of clusters is not required. The presented method is based on Backhaus et al. [207] and Müllner [285].

Let us assume we aim to cluster the columns (samples) of the data point matrix $X \in \mathbb{R}^{s \times n}$. A row-wise clustering can be achieved by clustering the transpose of X . We start by computing the pairwise distance matrix D with $D_{j,j'} = d(X_{\cdot,j}, X_{\cdot,j'})$. Typically, the Euclidean distance is used for the distance function d and thus $d(X_{\cdot,j}, X_{\cdot,j'}) = \sqrt{\sum_{i=1}^s (X_{i,j} - X_{i,j'})^2}$. The main procedure in *hierarchical clustering* is the approach of agglomerative clustering where we start with each point, here column or sample, in a separate cluster. We then iteratively merge the two closest clusters according to the smallest distance value d until all clusters are merged. Hence, after merging two clusters, we need to find a strategy to calculate the distance between this new cluster and all remaining clusters. In the following, we present several strategies, called linkage criteria, to determine these distances:

Single Linkage - The distance between two clusters A and B is defined as the minimum pairwise distance between their elements:

$$d_{\text{single}}(A, B) = \min_{j \in A, j' \in B} d(X_{\cdot, j}, X_{\cdot, j'}).$$

Thus, we consider all possible combinations of two points, where one originates from each cluster, calculate their distance and use the minimum as the distance between the two clusters. This strategy produces elongated clusters as it prioritizes small pair-wise distances between points within the clusters. Additionally, this strategy tends to be sensitive to noise and outliers.

Complete Linkage - The distance between two clusters A and B is defined as the maximum pairwise distance between their elements:

$$d_{\text{complete}}(A, B) = \max_{j \in A, j' \in B} d(X_{\cdot, j}, X_{\cdot, j'}).$$

This ensures compact clusters by minimizing the maximum intra-cluster distance, but it is sensitive to outliers due to its reliance on maximum distances.

Average Linkage - The distance between two clusters A and B is the average of all pairwise distances between their elements:

$$d_{\text{average}}(A, B) = \frac{1}{|A| \cdot |B|} \sum_{j \in A} \sum_{j' \in B} d(X_{\cdot, j}, X_{\cdot, j'})$$

where $|A|$ denotes the number of elements in A and $|B|$ in B . This strategy circumvents the sensitivity to outliers compared to the single and complete linkage criteria.

Centroid Linkage - The distance between two clusters A and B is the Euclidean distance between their centroids (means):

$$d_{\text{centroid}}(A, B) = \|\mu_A - \mu_B\|_2, \text{ where } \mathbb{R}^s \ni \mu_A = \frac{1}{|A|} \sum_{j \in A} X_{\cdot, j}$$

and $\mu_B \in \mathbb{R}^s$ follows analogously. Computationally efficient, as it reduces the number of calculations. Instead of pairwise distance computations, we only compute the distance between the cluster centroids which makes this strategy

computationally more efficient. Yet, once a cluster is updated, its cluster centroid also needs to be updated.

Ward's Linkage - The distance between two clusters A and B is described by the increase in total *within-cluster variance* when the two clusters are merged into $A \cup B$:

$$d_{\text{Ward}}(A, B) = \Delta E = E(A \cup B) - (E(A) + E(B)).$$

Here, $E(A) = \sum_{j \in A} \|X_{.,j} - \mu_A\|_2^2$ denotes the within-cluster variance of cluster A where μ_A again is the centroid and $E(B)$ is the within-cluster variance of cluster B . This strategy tends to produce compact and spherical clusters, making it favorable for many datasets. Yet, the additional variance computations require additional computational effort.

The *hierarchical clustering* process is visualized in a dendrogram, where each leaf represents a sample or feature. The branches represent the sequence of cluster merges and the distance between a branch and the leaf corresponds to the distance values of the merged clusters. A cutoff of the dendrogram at a specific position yields a specific number of clusters together with their corresponding elements.

An example of a dendrogram calculated by a *hierarchical clustering* by the columns using *complete linkage* is depicted in Figure 1.9. In this example, the data point matrix X was manually designed and the values are either 6, 0 or -6 .

C-means algorithm - Partitioning the dataset into c non-overlapping, compact clusters is the goal of the *C-means algorithm*. This means that every data point is assigned to exact one cluster which is called *hard clustering*. The algorithm is a popular and efficient clustering method, widely used in bioinformatics for grouping genes or samples based on their similarity.

We present the algorithm based on Backhaus et al. [207] and focus on a clustering of the columns (samples). The algorithm needs to be provided with the

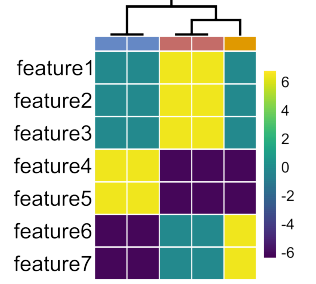


Figure 1.9: Graphical representation of the example dataset as heatmap. The dendrogram shows the result of a *hierarchical clustering* on the columns using *complete linkage*.

number of clusters c as a hyperparameter. We aim to minimize the within-cluster sum of squares J which serves as the objective function:

$$Z = \sum_{l=1}^c \sum_{j \in C_l} \|X_{.,j} - \mu_l\|_2^2$$

where C_l is the set of samples in cluster l and $\mathbb{R}^s \ni \mu_l = \frac{1}{|C_l|} \sum_{j \in C_l} X_{.,j}$. We thereby find a centroid for each cluster l such that the sum of the distances between the centroid and the elements of that cluster gets small. The minimization itself starts with a random initialization of c many centroids $\mu_1^{(I)}, \mu_2^{(I)}, \dots, \mu_c^{(I)} \in \mathbb{R}^s$ with $I = 0$. Two steps form the iteration. In the first one, we assign each sample j to the nearest centroid l if

$$\|X_{.,j} - \mu_l^{(I)}\|_2^2 \leq \|X_{.,j} - \mu_{l'}^{(I)}\|_2^2 \text{ for all } l' = 1, \dots, c.$$

Next, we update the cluster centroids with respect to their newly assigned samples

$$\mu_l^{(I+1)} = \frac{1}{|C_l|} \sum_{j \in C_l} X_{.,j}.$$

The iteration stops if we either reached a pre-selected number of iterations or if the centroids no longer change significantly.

In practice, the algorithm typically performs well for compact and spherical clusters, is simple and computational efficient. Yet, it is sensitive to the initialization values of the centroids and may converge to a local minimum. Since underlying distance measure is the Euclidean distances, the *C-means algorithm* can be suboptimal for non-spherical clusters, where points are not distributed symmetrically around the centroid.

A visualization of one iteration step of the algorithm for 16 manually created example data points and two features is depicted in Figure 1.10. On the left, the centroids within an iteration step are marked as circles, whereas the data points are denoted as gray squares. In the second image, we colored every data point according to the cluster centroid to which it is the closest. Note, the division lines only serve as a guideline in this example. In the third image, the centroid positions

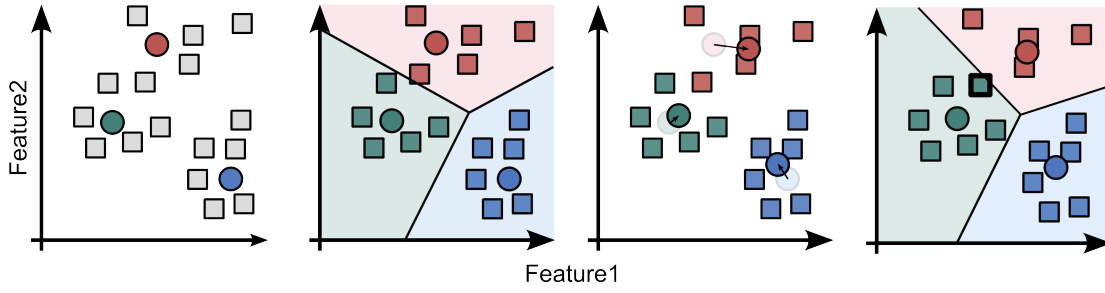


Figure 1.10: A visualization of one iteration step of the *C-means algorithm* for 16 samples and two features. Adjusted from Backhaus et al. [207].

are updated with respect to the cluster members. Again, we determine the nearest centroid for each data point and color them accordingly. As a result of this iteration step, one data point, marked with a thicker borderline, is transitioned from the red cluster to the green one.

Fuzzy C-means algorithm - Allowing each sample to correspond to multiple clusters with different degrees of membership is the key feature of the *Fuzzy C-means algorithm*. This is called *soft clustering* in contrast to the traditional *C-means algorithm*, which performs a hard clustering where each sample belongs to exactly one cluster. This property is especially useful when the dataset lacks hard boundaries between clusters, such as in gene or miRNA expression analysis, where samples might share characteristics across groups. Similar to the *C-means algorithm*, the number of clusters c has to be specified before the algorithm starts.

We introduce the algorithm according to Wierzchoń, Kłopotek [286] and again consider a clustering of the columns (samples). Here, we aim to minimize an objective function given by

$$Z = \sum_{l=1}^c \sum_{j=1}^n u_{l,j}^t \|X_{.,j} - \mu_l\|_2^2$$

where $\mu_l \in \mathbb{R}^s$ is the centroid of cluster l and $t > 1$ corresponds to the fuzziness parameter which controls the degree of fuzziness, typically set to $t = 2$ which yields a good balance between crispness and fuzziness of the clusters. For small values close to 1, for example $t = 1.1$, the clustering behaves like a hard clustering

($u_{l,j} \approx 1$ for the closest and $u_{l,j} \approx 0$ for all others), which is suitable for well-defined and distinct clusters. For a larger value $t > 2$ the membership values get more diffuse, allowing points to belong to multiple clusters, which is suitable if the clusters should overlap or be ambiguous. But too high values for t can lead to too fuzzy results which are hard to explain and to interpret. The variable $u_{l,j} \in [0, 1]$ denotes the degree of membership of sample j to cluster l . The matrix $U \in \mathbb{R}^{c \times n}$ is given by the $u_{l,j}$ and is called membership matrix. We start with setting $\sum_{l=1}^c u_{l,j}^{(I)} = 1$ for all samples j to obtain the membership matrix $U^{(I)}$ for iteration step $I = 0$. Next, we update (or initialize in case of $I = 0$) the centroids by computing

$$\mu_l^{(I+1)} = \frac{\sum_{j=1}^n u_{l,j}^t X_{.,j}}{\sum_{j=1}^n u_{l,j}^t}.$$

The numerator of the fraction contains a weighted sum of all samples with the membership degrees of cluster l and the corresponding sample. The denominator consists of the sum of the membership degrees of cluster l . Note that both the terms $u_{l,j}$ occurs to the power of t . We finish the iteration step by updating the membership degrees $u_{l,j}$ using

$$u_{l,j}^{(I+1)} = \frac{1}{\sum_{l'=1}^c \left(\frac{\|X_{.,j} - \mu_l^{(I+1)}\|_2}{\|X_{.,j} - \mu_{l'}^{(I+1)}\|_2} \right)^{\frac{2}{t-1}}}.$$

This ensures that $u_{l,j}$ decreases if the sample j is farther away from the centroid of cluster l and that $\sum_{l=1}^c u_{l,j} = 1$ holds for all samples. The iteration stops if a predefined iteration limit is reached, or the changes of U and of the centroids are below a predefined threshold, whereas the latter is more common. Finally, the membership degree $u_{l,j}$ can also be seen as a percentage how likely the sample j belongs to cluster l .

Similar to the *C-means algorithm*, fuzzy C-means variant is sensitive to initialization of the matrix U . Yet, it demands a greater computational effort caused by the additional updates for U .

In this work, we used the *fuzzy C-means algorithm* to analyze the time series of miRNA expression across tissues and ages. Therefore, we created time series

defined by all the combination of features (miRNA) and attribute realizations of the attribute *tissue*. Each time series consisted of different time points, representing miRNA expression at multiple ages from young to old. We applied the algorithm to cluster these time series according to their trajectory.

After obtaining the initial fuzzy clustering results, we post-processed them by first transforming the soft clustering into a hard clustering by assigning each miRNA-tissue combination to the cluster with the highest membership degree. This step served to remove ambiguity and overlap, ensuring that each combination contributed to only one cluster. Second, we removed miRNA-tissue combinations with a low membership degree to focus on confidently clustered time series. This step reduced noise and ensured that only well-clustered time series were included to avoid potential misinterpretations caused by uncertain assignments.

These post-processing steps were crucial for the subsequent analysis, where we focused on clusters with a high share of specific tissues or miRNA and clusters that showed distinct temporal trends (increasing or decreasing trajectories). This approach allowed us to concentrate on biologically meaningful and distinct clusters and improved the interpretability and robustness of our findings.

Determining the cluster number c is crucial in both *C-means* and *fuzzy C-means algorithms*, as it significantly influences the clustering results and their biological interpretability. Selecting the optimal c is a balance between algorithmic criteria and domain-specific relevance. We introduce common strategies based on Wierzchoń, Kłopotek [286] and Pal, Bezdek [287] which guide the choice of c to balance computational efficiency and meaningful partitioning of the data.

Elbow Method - This method evaluates the within-cluster sum of squares Z_c for multiple cluster sizes c :

$$Z_c = \sum_{l=1}^c \sum_{j=1}^n u_{l,j}^t \|X_{.,j} - \mu_l\|_2^2.$$

The values for Z_c decrease with increasing c . Plotting Z_c against c identifies an optimal number of clusters c' at the so-called *elbow point* where the decrease sharply slows, which indicates that the variance can be explained with c' clusters.

Mean Centroid Distances - Instead of within-cluster variance, we used a custom approach in this thesis to select the number of clusters by analyzing the mean distances between centroids MCD given by

$$\text{MCD}_c = \frac{1}{\binom{c}{2}} \sum_{l=1}^c \sum_{l'=l+1}^c \|\mu_l - \mu_{l'}\|_2$$

This metric emphasizes the separation between clusters, ensuring that clusters are not only compact but also distinct. We selected the c' at the "elbow point" of the plot between MCD and the corresponding c , where increasing the number of clusters yields a decrease in centroid separation.

Silhouette Analysis - This method measures how well samples fit within their assigned clusters compared to other clusters. Therefore, we calculate the *silhouette score* by

$$S(j) = \frac{b(j) - a(j)}{\max(a(j), b(j))}.$$

It ranges from -1 (poor clustering) to 1 (perfect clustering) for every sample j . Here

$$a(j) = \frac{1}{|A| - 1} \sum_{j' \in A, j' \neq j} \|X_{\cdot,j} - X_{\cdot,j'}\|_2$$

denotes the mean intra-cluster distance for sample j to every other sample j' in the same cluster A . The variable $b(j) = \min_{B \neq A} \frac{1}{|B|} \sum_{j' \in B} \|X_{\cdot,j} - X_{\cdot,j'}\|_2$ is the mean nearest-cluster distance for sample $j \in A$ to every other sample $j' \in B$ where $B \neq A$, while B is the nearest cluster. We determine the nearest cluster by the smallest mean distance $d(B, X_{\cdot,j})$, whereas the distance is defined by the average distance between the sample j and every sample $j' \in B$. The number of clusters with the highest average *silhouette score* across all samples is preferred.

Fuzzy Clustering Validity Indices - For soft clustering methods like the *fuzzy C-means algorithm*, we consider the *partition coefficient* P_{compact} which measures the cluster compactness with

$$P_{\text{compact}} = \frac{1}{n} \sum_{l=1}^c \sum_{j=1}^n u_{l,j}^2.$$

Higher values indicate better clustering but tend to decrease as the number of clusters increases. Additionally, we introduce the *partition entropy* P_{entropy} which indicates the fuzziness of the clustering by

$$P_{\text{entropy}} = -\frac{1}{n} \sum_{l=1}^c \sum_{j=1}^n u_{l,j} \cdot \log(u_{l,j}).$$

Here, lower values suggest more distinct clusters.

Biological or Domain-Specific Considerations - If known functional groups, temporal patterns, or tissue types exist, they can guide the selection of c' .

1.3 DEVELOPMENT OF A COMPUTATIONAL FRAMEWORK

This section explores practices for the development, application and accessibility of a computational framework designed to bridge theoretical methods with practical research applications. This framework integrates the mathematical principles and techniques from Section 1.2 into a cohesive pipeline, enabling reproducible analyses of large-scale datasets. Beyond implementation, this section puts our framework in the context of existing analysis platforms.

1.3.1 Implementation

The downstream analysis part of typical research studies considered here typically starts with an expression matrix or count table. Then, consecutively or simultaneously applying the methods introduced in Section 1.2 aims to answer pre-defined research questions, reveal patterns in large, multi-faceted datasets or identify angles for further, in-depth investigations.

We primarily implement the process from data point matrix X to results which are provided as matrices and visualizations using the programming languages *R* and *Python*. Both scripting languages have established themselves in the field of bioinformatics and data-science in general [288]. On the one hand, this can be attributed to their ease of use and comparably smooth learning curves. Compared to languages like *C* and its derivatives, *Python*, for example, does not require the user to carefully consider typing or perform manual memory management. While

this comes at the cost of reduced computational efficiency, it may drastically increase the accessibility and understandability of the resulting code. The lack of computational speed on the other hand can be tackled by interfacing with highly-optimized internal and external libraries [289–291].

Many of the analysis approaches and algorithms presented in Section 1.2 are available to the community in the form of packages, for example UMAP [265], PVCA [271, 292] and fuzzy c-means clustering [293, 294]. The packages, in turn, are provided via so-called repositories like *Bioconductor* [295] or *Bioconda* [296]. From this standpoint, we can consider the downstream analysis as a collection of self-written and external software which is tied together using custom routines. Considering that a typical analysis combines tens to hundreds of tools, a lack of structured organization among them severely hinders the reproducibility of studies and results. For example, uncertain software versions may prohibit performing the same analysis at a later point in time and changing the execution order may lead to different results. In particular, an ambiguous order of filtering steps can lead to the identification of different features. In the last years, the reproducibility of scientific results shifted into focus following a study by Baker [297], reporting that more than half of the researchers failed to reproduce at least one result of their peers.

1.3.2 Reproducible pipelines

To successfully recreate an analysis or computational method, we need to fulfill at least three criteria. First, the input data has to be available. This can be approached by providing the dataset to the public via repositories such as the *NCBI Gene Expression Omnibus (GEO)* [298, 299]. Second, the utilized functionalities and tools have to be precisely specified and available. Software environment management using *Conda* or containerization and virtualization techniques using *Docker* and *Singularity* create reproducible and shareable collections of tools with specified versions [300, 301].

Finally, the flow of execution of the analysis needs to be defined. Therefore, we consider the collection of routines to execute in the form of a directed acyclic graph. This graph is constructed from the inputs, outputs and function of each

execution step and is then executed via so-called workflow managers. In this thesis, we only utilize *Snakemake*, while *Nextflow* and *Galaxy* are also common in the community [302–304]. Besides being a comprehensive description of the execution order, the directed acyclic graph additionally enables to efficiently schedule single tasks in parallel if they do not interfere. In this way, scaling the workflow to fully utilize all computational resources is simplified. Additionally, workflow managers separate the functionality from the data, by defining both, the workflow itself and the environment in which it is executed using software environment managers.

1.3.3 Accessibility through web-based tools

While workflow pipelines themselves allow sharing computational methods with the research community, using them requires programming knowledge and the necessary computational resources to execute the pipeline. Web-based tools or services provide a solution for this twofold dilemma. On the one hand, scientist without expert programming knowledge are able to perform complex analyses using only a web browser while focussing on their respective research objective. The developers on the other hand retain control over the computational environment.

The web services which are developed in the context of this thesis follow a common design principle based on the *Django web framework*. In brief, the web application is split into *frontend* and *backend*. The frontend provides the users with a graphical interface which they use to upload data, fine-tune the desired computation via variables and observe the results. We implemented the interactions with the user in the form of so-called *wizards* which consist of a sequence of interfaces that simultaneously guide the user and provide immediate feedback regarding user choices. For example, settings for the computational method can be pre-selected or even prohibited based on previous user inputs or uploaded data. The backend receives the user input, performs cleaning and validation and starts a computational pipeline via a workflow manager. After the computation terminated successfully, the results are then provided to the user and visualized via the frontend. Using the pipelines to perform the actual bioinformatic func-

tionality simultaneously ensures a well-defined environment and allows easily re-using the functionality outside the web service for expert users and developers. In our applications, the frontend does not require any user-side computational workload besides validating file extensions. Yet, performing certain computations at the user-facing end of the web service can prove beneficial. For example, Hirsch et al. [305] demonstrated that calculating hashes of microbiome read files on the user machine makes uploading the full read files unnecessary and thereby reduced the uploaded data volume by more than two orders of magnitude, saving both time and resources. Additionally, user-side computations can be required in patient-related areas of research where uploading sensitive data to the backend is not desirable.

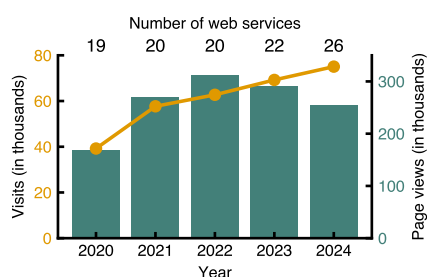


Figure 1.11: Visits (dots in yellow on the left side) and page views (bars in green on the right side) tracked with our tracking tool since Feb 10, 2020. On the top we provide the number of web services which are tracked in the respective year.

Web-based tools and databases serve as an important part of the bioinformatics research landscape [306–319]. Notably, various tools stay relevant throughout several years or decades [320–324] while steadily receiving updates [325–329]. Their importance and abundance can be assessed by considering the annual special-issue regarding web servers of *Nuclear Acids Research* [330]. The year 2024 marks the 22nd edition of this format where 248 proposals for tools led to 73 publications, one of them contributing to this thesis [5]. Apart from the momentary exposure at the time of publication, the long-term accessibility of these tools has been investigated

to formulate best practices [331, 332]. Figure 1.11 visualizes the number of web services which are provided by our own research group accompanied by the number of visits and page views per year. At the moment, we track 26 of our web services and in sum, we recorded more than 300 000 visits between 2020 and 2024. Overall, we offer 38 web services, some of them only for internal use of collaborators.

1.3.4 Related bioinformatics applications

In the last decade, various tools have emerged which investigate areas of miRNA research [333, 334]. A major effort has been conducted to predict targets of miR-

NAs [335, 336]. Some of these tools, for example *TargetScan* or *DIANA-microT* are accessible as web services [320, 337, 338]. Discovering novel miRNAs or detecting isomiRs is the major focus of another range of established tools [339–341]. Eventually, platforms like *tools4miRs* and *sRNAtoolbox* encapsulate a workflow starting from sequencing data files, perform the mapping and alignment, then provide downstream analysis tools and finally present the user with visualizations of their results [342, 343]. In the first case, *tools4miRs* interfaces with several external platforms such as *CPSS* and *mirTools*, while *sRNAtoolbox* relies on its own framework *sRNAde* [344, 345]. Notably, in both cases the downstream methodology is limited to a DE analysis and the associated hypothesis tests. While the DE analysis is a fundamental building block to analyze expression datasets, supplementary approaches, for example embedding, variance and cluster analysis, provide valuable insights into behavior-driving experimental factors and conditions. To this end, our proposed framework encompasses a wide range of data-driven methods to reveal relevant features and patterns in large, complex datasets up to single-cell resolution.

GOALS OF THE PHD THESIS

This thesis introduces a versatile computational framework designed for the exploration of complex microRNA (miRNA) datasets. It encompasses data-driven methods, including differential expression analysis, dimensionality reduction, and clustering to simultaneously address multi-faceted and time-resolved study setups. Within the publications that form this thesis, the framework is adapted to *applied* as well as *fundamental research* and contributes to *web-based tools* which simplify the use of intricate analytical methods. Each publication is briefly presented in this chapter accompanied by the core findings of the study. An overview of the distinct publications is presented in Figure 2.1. The projects are shown in two ways: on the right by arranging them on a timeline that indicates their publication dates and on the left by categorizing them according to their respective research focus. Additionally, the applied research projects are further organized according to the granularity and the time span of the respective study.

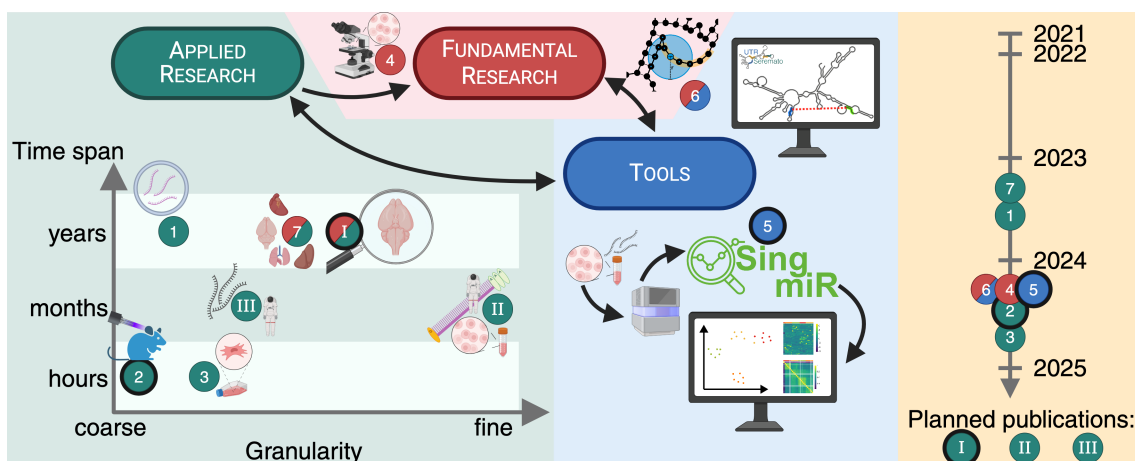


Figure 2.1: The graphical abstract shows the publications presented in this thesis. Publications where the author of this thesis contributed as first authorship are highlighted with a thick black border. Created in BioRender.com.

Starting in the realm of applied research, the flexibility of this framework is demonstrated through three distinct publications that investigate various applications of miRNA functionality. In the first study, small non-coding RNAs (sncRNAs) in two mouse plasma fractions were analyzed, where nonnegative matrix factorization enabled age-based clustering of samples [1]. The second study, a case-control experiment focusing on non-thermal plasma treatment, identified miRNAs associated with wound healing and tissue regeneration through differential expression analysis [2]. The third study applied a clustering method to a time-series dataset from stem cell differentiation, revealing upward trajectories in the expression of genes linked to cell-type differentiation [3].

Despite the valuable insights gained into post-transcriptional gene regulation, these studies were unable to capture the cellular heterogeneity inherent in complex tissues. To increase the granularity of analyzing miRNA heterogeneity to a cell-specific level, high-resolution methodologies are essential for uncovering more intricate patterns. Recent progress in single-cell ribonucleic acid (RNA)-sequencing supported by sophisticated bioinformatics tools has enabled the development of innovative analyses. Adapting these high-resolution approaches to study miRNA regulation is an emerging area of fundamental research. This thesis highlights the utility of such an approach in a pilot study that optimizes library preparation protocols and provides detailed quality control metrics alongside feature-normalized count tables [4]. Additionally, the development of a web-based tool integrating this bioinformatics pipeline with the presented computational framework allows for downstream analyses [5]. Besides its capabilities regarding the analytical methods, the defining characteristic of this tool is the unification of the process from raw sequencing dataset to the graphical representation of results in a unified user workflow. As the first platform of its kind, this tool offers researchers around the globe an accessible way for processing and analyzing single-cell miRNA sequencing datasets, thereby accelerating the development of sequencing protocols. Further exploring the usability aspect in combination with aspects of fundamental research, a second web-based tool was developed to visually indicate miRNA binding sites in the secondary structure of the 3'untranslated region (UTR) of a gene [6]. Additionally, a method was

developed to describe the coverage of such a binding site to investigate if the binding of a miRNA can be inhibited by the folding of the secondary structure of the messenger-RNA (mRNA).

As an alternative to cell-type-resolved studies, extensive tissue-resolved datasets can be employed to increase the granularity of exploring miRNA expression. In this thesis, an organ-specific expression dataset comprising 771 bulk samples from 16 organs across 10 time points, was analyzed [7]. Using the computational framework uncovered both organ-specific and global miRNA behaviors. By examining expression patterns over the mouse lifespan, dynamic changes influenced by aging were identified. A follow-up investigation further increased the granularity from an organ-specific to a tissue-specific level by focussing on the mouse brain, an organ notable for its high structural and functional diversity (Planned I). Consisting of 844 bulk samples from 15 regions from male and female mice across seven time points spread over the whole life span, this study established a time-, sex- and brain region-resolved atlas of miRNA expression patterns. Given the extent of the dataset, using the full range of methods in the developed computational framework was crucial to designate miRNAs for an in-depth analysis. Besides uncovering brain region-specific and global aging signatures, the granularity identified miR-155-5p as a promising therapeutic target.

Finally, a tissue-resolved, single-cell study for mRNA (Planned II) and a complementary bulk miRNA investigation (Planned III) were combined to form a comprehensive dataset studying the systemic effects of microgravity. This unique study design included housing mice aboard the International Space Station for 40 days, offering exceptional experimental insights. The single-cell dataset consists of 216 samples from 28 mouse organs and tissues which results in 280,745 cells. This part of the study indicated a down-regulation of ribosomal protein genes. The miRNA dataset is composed of 686 samples from 13 mouse organs and tissues. Benefiting from the granularity of the study, tissue specific influences of space flight could be identified in GAT, Spleen and SCAT.

Apart from the primary set of publications discussed before, two more application studies concerning miRNAs [346, 347] and two additional web services [305, 348] were published in the course of this thesis.

RESULTS

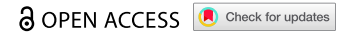
This cumulative thesis is based on 7 peer-reviewed publications whose published versions are included in full in this chapter.

3.1 AGEING-ASSOCIATED SMALL RNA CARGO OF EXTRACELLULAR VESICLES

RNA BIOLOGY
2023, VOL. 20, NO. 1, 482–494
<https://doi.org/10.1080/15476286.2023.2234713>



RESEARCH PAPER



Ageing-associated small RNA cargo of extracellular vesicles

Fabian Kern ^{a,b,#}, Thomas Kuhn ^{c,d,#}, Nicole Ludwig ^{e,f,#}, Martin Simon ^{g,h}, Laura Gröger ⁱ, Natalie Fabis ^j, Ernesto Aparicio-Puerta ^k, Abdulrahman Salhab ^l, Tobias Fehlmann ^m, Oliver Hahn ⁿ, Annika Engel ^o, Viktoria Wagner ^a, Marcus Koch ^p, Katarzyna Winek ^q, Hermona Soreq ^r, Irina Nazarenko ^s, Gregor Fuhrmann ^{c,d}, Tony Wyss-Coray ^t, Eckart Meese ^e, Verena Keller ^{u,v}, Matthias W. Laschke ^{w,x}, and Andreas Keller ^{a,b,n,x}

^aChair for Clinical Bioinformatics, Saarland Informatics Campus, Saarland University, Saarbrücken, Germany; ^bHelmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz-Centre for Infection Research (HZI), Department for Clinical Bioinformatics, Saarbrücken, Germany; ^cHelmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz-Centre for Infection Research (HZI), Biogenic Nanotherapeutics Group (BION), Saarbrücken, Germany; ^dDepartment of Pharmacy, Saarland University, Saarbrücken, Germany; ^eDepartment of Human Genetics, Saarland University, Homburg, Germany; ^fCenter for Human and Molecular Biology, Saarland University, Homburg, Germany; ^gMolecular Cell Biology and Microbiology, Wuppertal University, Wuppertal, Germany; ^hDepartment of Genetics and Epigenetics, Saarland University, Saarbrücken, Germany; ⁱDepartment of Neurology and Neurological Sciences, Stanford University, Stanford, USA; ^jINM – Leibniz Institute for New Materials, Saarbrücken, Germany; ^kThe Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel; ^lFaculty of Medicine, Institute for Infection Prevention and Control; Medical Center - University of Freiburg, Freiburg, Germany; ^mInstitute for Clinical and Experimental Surgery, Saarland University, Homburg, Germany; ⁿCenter for Bioinformatics, Saarland University, Saarbrücken, Germany

ABSTRACT

Previous work on murine models and humans demonstrated global as well as tissue-specific molecular ageing trajectories of RNAs. Extracellular vesicles (EVs) are membrane vesicles mediating the horizontal transfer of genetic information between different tissues. We sequenced small regulatory RNAs (sncRNAs) in two mouse plasma fractions at five time points across the lifespan from 2–18 months: (1) sncRNAs that are free-circulating (fc-RNA) and (2) sncRNAs bound outside or inside EVs (EV-RNA). Different sncRNA classes exhibit unique ageing patterns that vary between the fcRNA and EV-RNA fractions. While tRNAs showed the highest correlation with ageing in both fractions, rRNAs exhibited inverse correlation trajectories between the EV- and fc-fractions. For miRNAs, the EV-RNA fraction was exceptionally strongly associated with ageing, especially the miR-29 family in adipose tissues. Sequencing of sncRNAs and coding genes in fat tissue of an independent cohort of aged mice up to 27 months highlighted the pivotal role of miR-29a-3p and miR-29b-3p in ageing-related gene regulation that we validated in a third cohort by RT-qPCR.

ARTICLE HISTORY

Revised 6 June 2023
Accepted 9 June 2023

KEYWORDS

Noncoding RNA; tissue ageing; extracellular vesicles; plasma; miRNA

Introduction

Understanding and controlling the molecular hallmarks of age-related processes in higher organisms promises to greatly improve the quality of life [1]. For humans, ageing is frequently studied using easily accessible biospecimens such as blood, serum, or urine. Consequently, the scientific community generated models for a broad spectrum of molecular physiological and pathophysiological processes from different molecular types. For example, studies rely on long-lived individuals [2], serum proteomic profiling [3], small RNA patterns in blood cells [4,5], or the exploration of epigenetic control of ageing clocks [6]. Likewise, deeper profiles, such as gene expression fingerprints, are available for different tissues [7]. Murine models facilitate the analysis of such processes thanks to their restricted influence of genetic background and varying lifestyles compared to humans. In mouse models, the aged immune system drives senescence and

ageing of solid organs [8]. Further, organism-wide RNA-sequencing data of major organs and cell types across the mouse lifespan provide an important resource to study ageing [9,10]. The available data suggest complex ageing patterns, including both linear and non-linear effects that are either specific for organs or follow global organism-wide trajectories. Ageing and parabiosis-mediated rejuvenation suggest an almost universal loss of gene expression with age that is largely mimicked by heterochronic parabiosis: aged blood reduces global gene expression, and young blood restores it in select cell types [11]. In the same direction, Sahu and co-workers demonstrated that a beneficial effect of young blood on aged muscle regeneration was diminished when serum was depleted of extracellular vesicles (EVs), indicating the important role of EVs in ageing and rejuvenation [12]. In addition to blood, also young CSF has a beneficial effect by restoring oligodendrogenesis and memory in aged mice [13].

CONTACT Andreas Keller andreas.keller@ccb.uni-saarland.de Chair for Clinical Bioinformatics, Saarland Informatics Campus, Saarland University, Saarbrücken 66123, Germany

[#]These authors have equally contributed to the study and are listed in lexicographic order.

^{*}These authors contributed equally as senior authors to the study.

Supplemental data for this article can be accessed online at <https://doi.org/10.1080/15476286.2023.2234713>

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

These observations indicate a systemic and orchestrated exchange of information and molecules between organs. For example, extracellular vesicles (EVs) are membrane vesicles mediating the horizontal transfer of genetic information between different cell types. Specifically, EVs are postulated to play an important role [14,15] in, e.g. hypothalamic stem cells seem to control ageing through EV miRNAs [16]. Recently, targeted intervention of EV-mediated transfer of miRNAs from osteoclasts to chondrocytes was described as a promising method to slow or even inhibit osteoarthritis in mice [17]. Furthermore, several studies have addressed the relationship of EVs with ageing in a systematic manner [18–21]. Complementary studies have investigated the change in EV-bound noncoding RNAs depending on (treatment) interventions such as caloric restriction [22]. However, these are limited in their analysis scope by considering only one small RNA class at a time, often even to a subset of well-characterized representatives. Moreover, an inherent restriction is the limited sample volume, frequently leading to pooling of biosamples and blurring of fine-grained signals.

These and other issues complicate the analysis of EVs and their molecular cargos. Especially in the context of EVs in cancer, common pitfalls in purification have been summarized by Schekman and co-workers [23], with the correct nomenclature of EVs, purification and other aspects elucidated in great detail. Considering these inconsistencies, we use the term ‘extracellular vesicles’ (EVs) throughout the manuscript as recommended by the International Society for EVs. EV means the full fraction of vesicles up to 400 nm in diameter irrespective of their origin and biogenesis.

The main aim of our study was to provide a data resource of small non-coding RNAs included in EV cargo and freely circulating in plasma (fc-RNAs) in mice of different ages and to identify differences between the molecular information in these fractions associated with ageing that might advance our understanding of the systemic ageing process. Therefore, plasma fc-RNA and EV-RNA of individual mice were sequenced for noncoding RNA profiling and contrasted by computational approaches. To demonstrate the use of this resource, we conducted pathway and comparative analyses using original Tabula Muris senis (TMS) data [9,10] and performed sequencing of small RNAs from TMS as an independent cohort. Finally, we validated the core findings by RT-qPCR in a third cohort of aged mice (Figure 1a).

Results

Noncoding RNAs are modulated specifically upon ageing in EV- and fc-RNA samples

To uncover age-related dynamic processes and to model the information exchange involved, we sequenced both non-coding fc-RNAs and non-coding EV-RNAs from individual mice. The molecular profiles are available at five time points across the average lifespan between two and 18 months in two to four replicates per age group and biospecimen type (Supplementary Table S1). For the fc- and EV-RNA samples, we sequenced an average of 38 million reads per sample and mapped them to ten different noncoding RNA classes. Our

analysis covered a total of 80,688 different noncoding RNAs, with piRNAs, circRNAs, lincRNAs and miRNAs being the classes with the highest number of different features (Figure 1b). The first aspect of the analysis encompassed the distribution of molecules from the different RNA classes. While tRNA fragments were highly represented both in EV- and fc-fractions, piRNAs showed sharply lower levels in both specimen types (Figure 1c). However, varying amounts of circRNAs and rRNAs were predominantly observed in the EV- and fc-fractions. Notably, this analysis has a quantitative and RNA class-centric view but does not yet consider whether the representatives within the classes match across sample types. For example, only a small number of piRNAs were present in both the EV and fc fractions, even though the general abundance was high. Considering the sample type overlap for each class, the most significant difference was indeed observed between fc- and EV-bound piRNAs (Figure 1d). Similarly, we report large differences in the content of RNA molecules from snRNAs, snoRNAs, and scaRNAs. In contrast, detected tRNA fragments, lincRNAs, rRNAs and circRNAs are often shared between the two fractions. In summary, our data argue for type-specific expression patterns that differ significantly between noncoding RNA classes both in a quantitative and qualitative manner.

We thus asked whether unique ageing trajectories within and between noncoding RNA classes exist, either enclosed into/bound to EVs or freely circulating in plasma. One indicator is the proportion of variance in the RNA counts that can be explained by available sample covariates, i.e. either by age of the mice, the specimen type and donor mice identity, or linear combinations of such. Depending on the RNA class we observe varying results with respect to the separation in the fc- and EV-fraction in a 2-dimensional UMAP embedding, with a clear segregation in the scaRNAs (Supplementary Figure S1). Compared to the other RNA classes, tRNA fragments and miRNAs however showed the highest fraction of variance explained by age (Figure 1e). In comparison, the lowest variation with respect to age was observed for scaRNAs and rRNAs. Importantly, the individuality factor of each donor mouse used for this study was comparably small and independent of the RNA class. To uncover a potential relationship between each RNA class and mouse age, we used the expression at month 2 as baseline and modelled whether it increases or decreases over time for EV-RNA and fc-RNA separately but observed similar dynamics of change.

The largest age-related differences appear for rRNAs, where the overall amount increases for free circulating molecules with ageing but the EV loading of rRNAs decreases. This notable difference in the specimen types also explains the high proportion of variance attributed to the sample type annotation (Supplementary Figure S2). Our data further indicate a strong ageing signal in EV- and fc fractions, with varying strengths, again depending on the RNA class (Figure 1f). As our previous analyses emphasized the role of tRNA fragments, we investigated the expression profiles in an unbiased manner and performed a classification into three age groups (young, 2 months; middle aged, 6–8 months; old, 12–18 months). We modelled this classification task as an optimization problem through nonnegative matrix factorization, computing

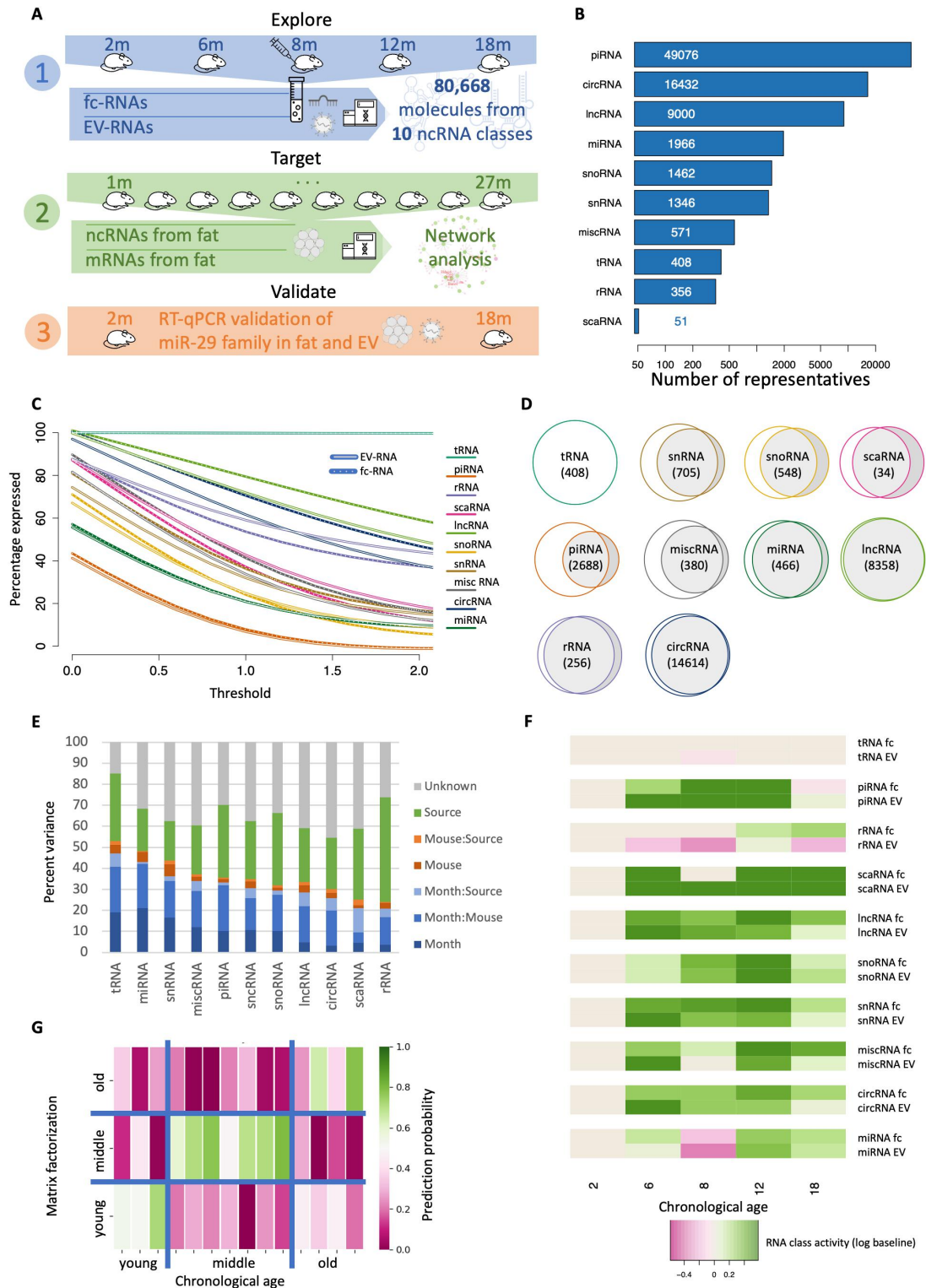


Figure 1. Distribution of sequencing reads into their mapped noncoding RNA classes and their relation to ageing across the mouse lifespan. (a) Study setup. We profiled EV- and fc fraction vesicle and plasma samples from mice in five age groups, sequenced 80,668 noncoding RNAs from 10 classes (1), sequenced fat tissue specimens from TMS as an independent cohort (2) and validated the key findings using RT-Qpcr in another independent cohort (3). (b) Overall distribution of molecules to the 10 noncoding RNA classes under investigation. (c) Fraction of representatives per RNA class (y-axis) exceeding the expression threshold (x-axis; normalized counts). The RNA class is presented as solid line in the foreground. Dashed with lines indicate the fc fraction and solid lines the EV fraction. (d) Overlap of

probabilities for each sample to belong to each of the three groups. We then assigned each sample to the age group with the highest probability. For both fc- and EVs, we computed varying prediction accuracies, once more with the best results obtained for tRNA fragments with a remarkable accuracy of 86% (Figure 1g).

Taken together, our analyses suggest that non-coding RNAs exhibit specific age trajectories, both qualitatively and quantitatively. Moreover, the data pinpointed substantial differences in the case of fc-RNAs and EV-RNAs, where the highest correlation with age was observed for tRNA fragments. This poses the question of whether loading of EVs follows biologically relevant environmental mechanisms. To potentially discover such patterns, we next performed a fine-granular and molecule-centric analysis.

Ageing patterns deviate within specific RNA classes

Of the 80,668 unique noncoding RNA molecules in *Mus musculus* included in our analysis, 23052 (28.6%) were stably present in the EV- and fc-fractions (Supplementary Table S2). We then computed the linear Pearson correlation as well as the nonlinear distance correlation for each of the investigated RNAs. While making conclusions on the correlations of single RNA features can be challenging in terms of type-I errors, comparing the different RNA classes and the fc- and EV-fractions globally can support an understanding of linear and non-linear ageing effects. We thus computed an estimate for each RNA whether it was linearly correlated with age, nonlinearly correlated with age, or not correlated with age at all for EV- and fc-fractions separately. Because of the large number of non-coding RNA features included in the study we are potentially facing an overplotting issue and thus computed a density estimation for the two correlation schemes. For both sample types, fc- (Figure 2a) and EV- (Figure 2b), the linear component was dominant, and only a few exceptions with nonlinear trajectories occurred. Those are characterized by a distance of at least 0.15 from a spline with eight degrees of freedom. The amplitude and frequency of nonlinear RNAs were both slightly enriched in EVs. Interestingly, we also observed a pattern towards a slightly negative correlation in EVs as compared to a positive correlation (average of 0.106) in the fc fraction. To test whether the average of 0.106 is the result of a random effect we performed permutation tests. Here, we reached an average correlation of -0.0003 , marking a statistically significant difference ($p < 10^{-16}$). Of note, most of the correlation values observed in our study are not significant. Generally, linear correlation coefficients above 0.5 and -0.5 in our study roughly corresponded to p-values with nominal significance at an alpha level of 0.05 (Supplementary Table S2).

Having observed noncoding RNAs that are either positively and negatively correlated with age in EV and fc fractions further called for exploring whether the up- and downregulated candidates show similar compositions in the two specimen types. In total, 27% and 22% increased and decreased with age in the EV fraction and fc fraction, respectively, slightly differing from what would be expected by a random distribution. However, 39% of the 23,052 expressed noncoding RNAs were negatively correlated with age in EVs but positively correlated in the fc fraction while only 12% presented the opposite behaviour, i.e. were negatively correlated with age in the fc- fraction and positively correlated with age in the EV-fraction (Figure 2c, Supplementary Table S2). To seek common patterns for the increasing and decreasing expression of non-coding RNAs, we clustered the expression in EV- and fc-fractions separately and extracted RNA clusters from the dendrogram. For each cluster, we then computed the average linear and nonlinear correlation with ageing and finally calculated the overlap of the sample types. Our analysis confirmed a strong decrease in the correlation with age in the EV fraction compared to the fc fraction (Figure 2d). The EV clusters are enriched in the lower left corner, indicating a significant trend towards a negative correlation with age in EVs. Furthermore, the data reveal an age-related loss in linear correlation compared to non-linear correlation. To validate the origin of these signals, we inspected all concordant and discordant noncoding RNAs and provide specific examples for markers clearly increasing and decreasing with age in both EV and fc fractions (*miR-466i-5p*, Figure 2e and *Gm16701*, Figure 2f, respectively), decreasing with age in the EV fraction but increasing in the fc fraction (*Gm20756*, Figure 2g), and finally increasing with age in the EV fraction but decreasing in the fc fraction (*miR-690*, Figure 2h). We further examined whether the patterns hold for all 10 noncoding RNA classes or if they are rather class specific. Here, the specificity of patterns for the different non-coding RNA classes was astonishing. For example, 94% of tRNA fragments increased with age in both the EV and fc fractions. Additionally, 54% of rRNAs decreased with age in the fc fraction but increased with age if EV-bound. Conversely, 42% of circRNAs increased with age in the fc fraction but decreased with age if EV-bound (in- or outside). Additionally, other RNA classes revealed distribution patterns significantly deviating from the 25% per group as expected by chance. Finally, 82% of miRNAs increased with age in the EV fraction (Figure 2i).

The miR-29 family controls ageing-related processes in fat tissues

In light of the regulatory role of miRNAs typically repressing gene expression [24] and further knowing that mRNA levels tend to decrease with age, we chose this particular class of noncoding RNAs to reveal further potential of our data

expressed RNAs in EVs and in the fc fraction as area proportional Venn diagrams. (e) Percent of total data variance attributed to different parameters, such as age (month), individual mice or specimen type (source). Columns are sorted according to decreasing fraction of variance attributed to age. (f) Relative expression of the different RNA classes per time point and sample type compared to the baseline (2 months). Green indicates higher expression, and purple indicates lower expression. The upper row per RNA class shows the fc fraction, and the lower row shows the EV fraction. (g) Prediction of age by nonnegative matrix factorization. The colour code represents the probability (trust) in the prediction, the x-axis represents the true age group, and the y-axis represents the predictions.

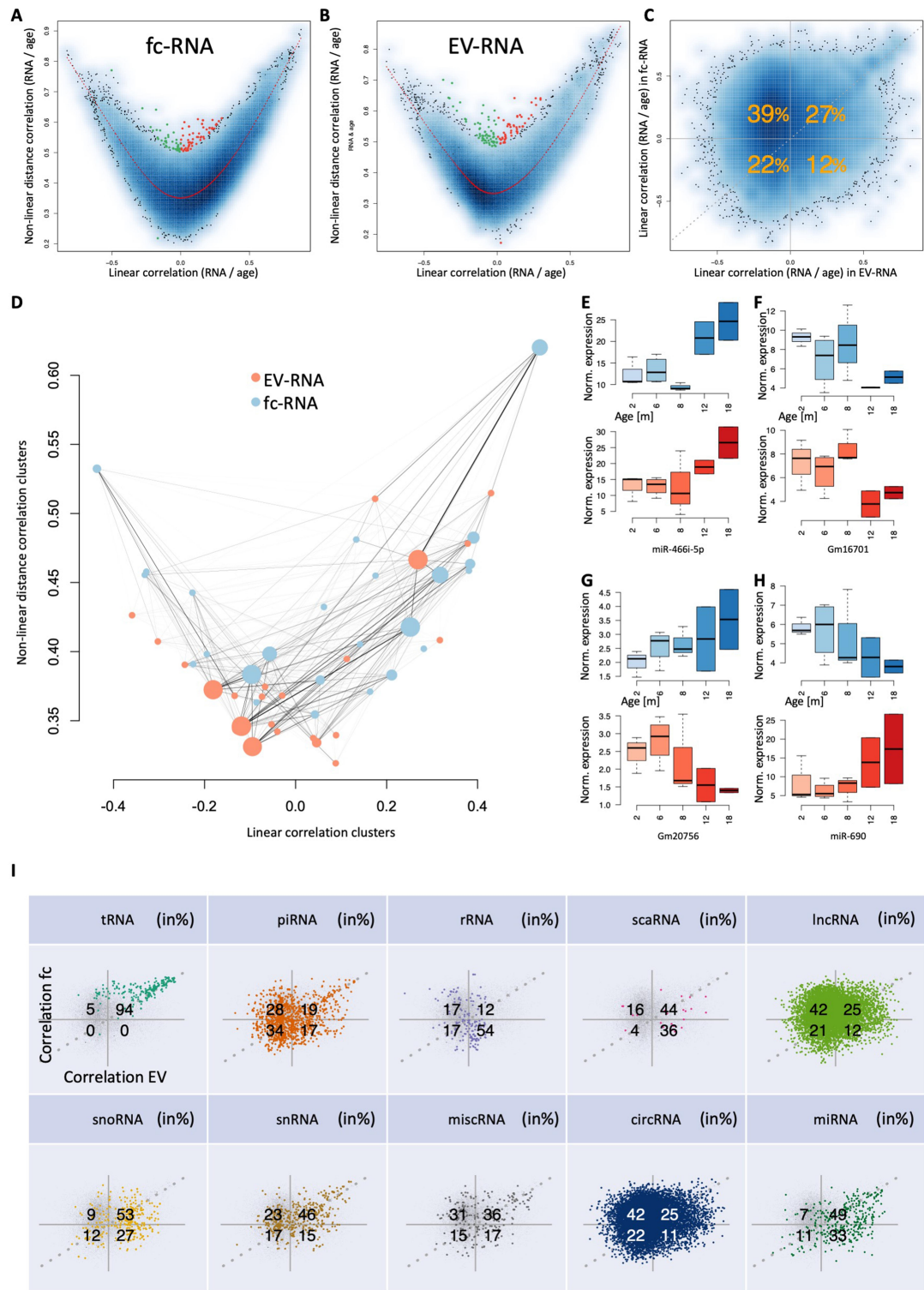


Figure 2. Correlation of noncoding RNAs with ageing in fc- and EV fractions. (a) for each noncoding RNA, the x-axis represents the Pearson correlation, and the y-axis represents the distance correlation with age in the fc fraction. The colour gradient in the background represents the density of non-coding RNA representatives with respective linear and non-linear correlation. The red line is a smoothed spline. The coloured dots (green, negatively; red, positively) are correlated with age in a predominantly nonlinear manner, i.e. those points are with a distance of at least 0.15 away from the spline. (b) the same information as in panel (a) but for EVs. (c) Scatter plot showing the Pearson correlation in EVs (x-axis) in relation to the Pearson correlation in the fc fraction (y-axis). Orange numbers represent the percentage

resource. We first asked whether the miRNAs increasing or decreasing with age in EV- and fc-fractions exhibit distinct functions in a pathway-specific manner. For each gene ontology category [25], we computed an enrichment score for the miRNAs in EVs and in the fc fraction. In more detail, we used the list of miRNAs sorted by their correlation with age to perform cut-off-free miRNA set enrichment analysis using miEAA [26,27] for all miRNAs instead of selecting only a subset of miRNAs. The gene ontology categories were extracted from the miEAA tool using annotations from the miRTarBase (Supplementary Table S3). A direct comparison provided strong evidence supporting the notion that miRNAs correlating with age in EVs are significantly enriched in biochemical categories compared to those in the fc fraction of plasma (Figure 3a). This finding argues further towards our initial hypothesis that EVs may specifically be loaded with noncoding RNAs that exert biological processes in remote sensing cells. To understand the nature of these processes, we compared the 16 categories that are at least three orders of magnitude more significant in EVs compared to the fc fraction to the two being at least three orders of magnitude more significant in the fc fraction compared to EVs. In the former, the strongest enrichment was found for protein heterodimerization activity, neural crest cell migration, negative regulation of inflammatory response, receptor internalization, positive regulation of neuroblast proliferation, the mitochondrial envelope, the positive regulation of DNA-templated transcription and the TORC2 complex. All categories were significant at an alpha level of 0.05 following Benjamini-Hochberg FDR adjustment in the EV fraction and none of the categories remained significant after adjustment for multiple testing in the fc fraction. As an example, we present the enrichment plots for the protein heterodimerization activity for both fractions (Figure 3b). Here, the original running sum curve for the EV fraction clearly exceeds the random background distribution while for the fc fraction random distributions reach the original one. The categories with higher significance in the fc fraction included cellular response to BMP stimulus ($p = 6 \times 10^{-5}$ vs. 0.11) and negative regulation of myotube differentiation ($p = 6 \times 10^{-7}$). Here, both categories were significant following adjustment for multiple testing for the fc fraction, in the EV fraction the negative regulation of myotube differentiation remained significant following adjustment for multiple testing. This case indeed indicates that pathways can be significant in both, the fc and EV fraction. Especially one category was highly significant in both: 'response to hypoxia' reached a p-value of 4.3×10^{-5} in the fc fraction and of 9.4×10^{-6} in the EV fraction (Figure 3c, Supplementary Table S3). Distinct pathways specifically

enriched in the EV fraction open the question of potential effects on gene regulation in different tissues.

The core hypothesis of our work stipulates specific loading of EVs, notably considering both possibilities, in- and outside-bound, with non-coding RNAs, first and foremost miRNAs, enabling the control of specific cellular functions and gene regulation in remote cells. To identify tissues most likely affected by the EV- and fc-RNA cargo, we next combined the miRNA data generated in this study with our previously established bulk- and single-cell murine tissue-ageing atlas, TMS [9,10]. In these studies, we reported both linear and nonlinear ageing trajectories in gene expression signals. Like the findings on noncoding RNAs observed here, the associated genes cluster with coherent biological functions, including extracellular matrix regulation, unfolded protein binding, mitochondrial function, and inflammatory and immune responses. The expression patterns are consistent across tissues, differing only in the amplitude and age of onset. In particular, fat tissues showed early ageing signals of biochemical pathways similar to those observed in the miRNA pathway analysis described above. It was previously shown that miRNAs target genes in a pathway-specific manner [28–30]. Thus, for miRNAs associated with age in the EV- or fc-RNA in the current work, we extracted the experimentally validated target genes from miRTarBase [31] and evaluated the correlation of these target genes with age in all tissues from TMS. Remarkably, the analysis was limited to miRNA-gene pairs with strong evidence of functional interactions, such as from reporter assays. In this context, the expected pattern is a negative correlation of target genes with age, where miRNAs show a positive correlation with age and *vice versa*. In particular, mesenteric fat, gonadal fat, the brain, white blood cells and brown fat fulfill this expectation (Figure 3d). This result is also in line with recent parabiose-mediated rejuvenation experiments, suggesting a loss of gene expression with age that is largely mimicked by rejuvenation. Likewise, the observed tissue-independent ageing patterns matched our previous results. While fat tissues generally showed the best concordance, other tissues, such as the lung or pancreas, did not. The target gene correlation for mesenteric and gonadal fat verified the increased correlation with age for miRNAs decreasing with age and *vice versa* (Figure 3e and 3f). This effect was more pronounced for gonadal fat in EV-miRNAs. Translating the miRNAs and genes to a target gene network identified eight core genes: *Notch1*, *Bace1*, *Hdac4*, *Igf1*, *Eln*, *Cav2*, *Insig1*, and *Scap* (Figure 3g, Supplementary Table S4), possibly reflecting physiological relevance for both signalling networks and epigenetic processes. All but *Notch1* are experimentally validated target genes of specific members of the

of points in each of the four quarters. The data suggest a shift to a negative correlation with age in EVs. Numbers are rounded to integers. (d) Noncoding RNAs in the fc- and EV fractions were clustered, and the resulting clusters were attributed to the average linear and nonlinear correlation. Solid lines represent a match of noncoding RNAs; the thicker the line is, the more noncoding RNAs match between an fc and EV clusters. The diameter of the points represents the cluster size. Most EV clusters accumulate in the lower left corner. (e-h) Examples of noncoding RNAs that increase or decrease with age. The x-axis represents the age, and the y-axis represents the expression of the selected noncoding RNA (orange, EV fraction; blue, fc). Colored boxes span the first to the third quartile, with the line inside the box representing the median value. The whiskers show the minimum and maximum values or values up to 1.5 times the interquartile range below or above the first or third quartile if outliers are present. (i) Confusion matrix scatter plots (see also panel (c)) split by the RNA classes.

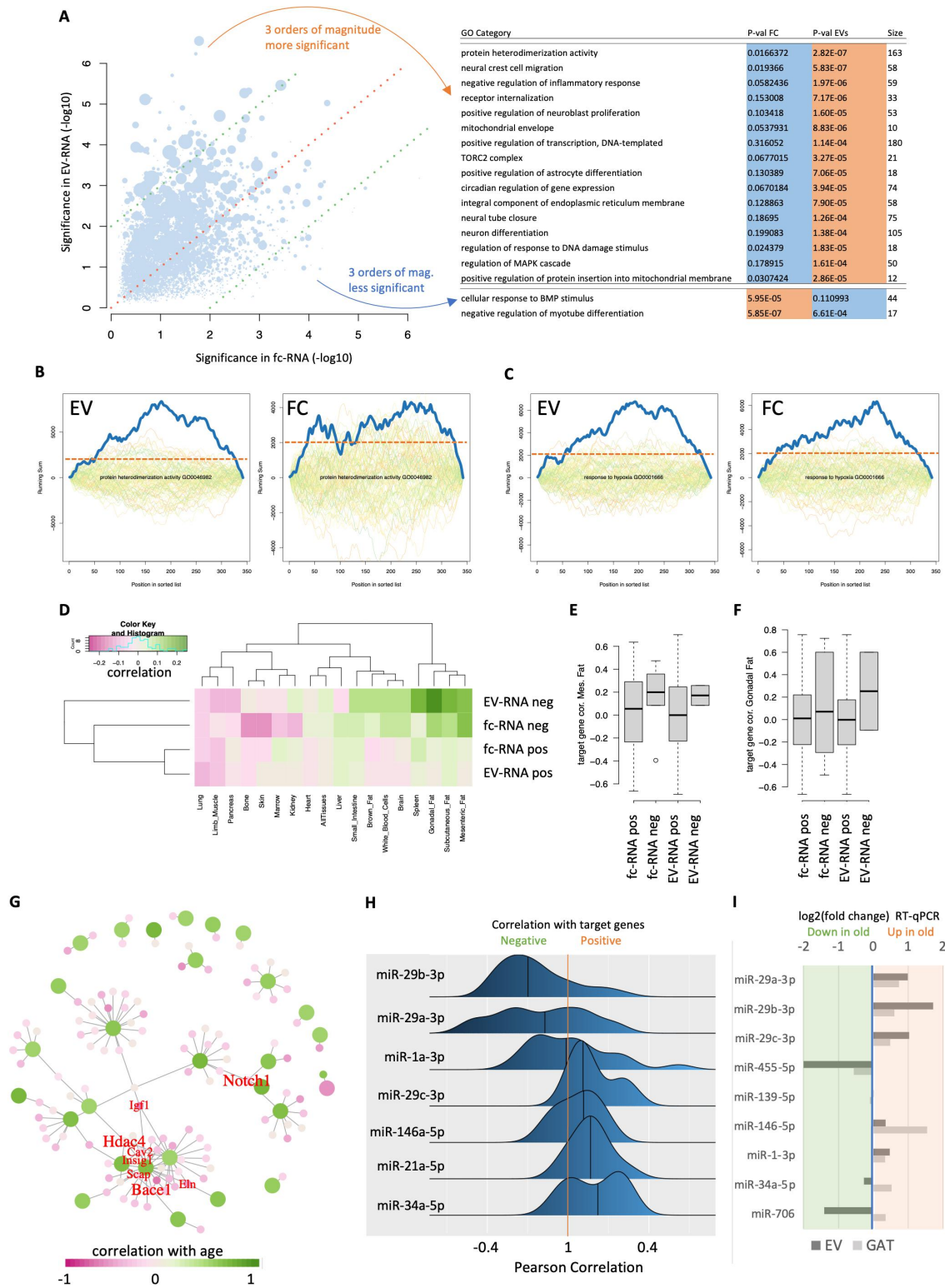


Figure 3. Pathway results and miRNA-target regulation networks. (a) miRNA pathway enrichment for age-related miRNAs in the fc fraction (x-axis) and EVs (y-axis). Each dot is one pathway, and the size represents the number of miRNAs associated with the pathway. The red dashed line is the bisector, and the green lines indicate two orders of magnitude higher significance in the fc fraction and in EVs. The pathways with at least a three orders of magnitude difference are listed on the right. (b) Enrichment plots for the category protein heterodimerization activity in the EV and fc fraction. The solid blue line denotes the enrichment for the miRNAs sorted

miR-29 family (miR-29a, -b or -c), pointing at an inherent regulatory role of this miRNA family in fat tissue. Among the gene nodes, *Nanog* could be distinguished by a positive correlation with age. This gene is targeted by mmu-miR-296-5p, which is negatively correlated with age.

Of note, the correlation presented in this analysis was indirect. The fc-RNA and EV-RNA fractions were sequenced from another cohort of mice and compared to the gene expression profiles in tissues. To test whether increased levels of these miRNAs have also been detected in gonadal fat tissue of aged mice, we used small RNA sequencing data from gonadal fat tissues from TMS [32]. Specifically, we examined identical tissue specimens that were used for sequencing the RNA expression atlas. For the microRNAs from the network (Figure 3h), we observed in most cases a mixed pattern of up- and down-regulation of the targets for those miRNAs. Nevertheless, members of the miR-29 family (namely, miR-29b-3p and miR-29a-3p) showed downregulation across all target genes in gonadal fat.

Due to a limited cohort size and potential challenges in using high-throughput screening based on NGS, we performed a further validation experiment. To this end, we investigated a third independent cohort of mice and analysed the relative expression of miR-29 family members along with other dysregulated miRNAs from the network (miR-455-5p, miR-139-5p, miR-146-5p, miR-1-3p, miR-34a-5p, and miR-706) in the EV-RNA of four young and two old mice using real-time quantitative PCR (RT-qPCR). The RT-qPCR results for the miR-29 family support the sequencing data and provide evidence for an increased abundance of miR-29a/b/c-3p in EV-RNA of old mice (Figure 3i). Additionally, in the gonadal fat samples of the same mice, a – however less pronounced – upregulation was confirmed upon ageing. miR-146-5p showed a similar trend and also miR-1-3p was slightly increased in the EV-fraction and GAT tissue. In contrast, miR-455-5p was down-regulated in the EV fraction and in GAT tissue. For miR-139-5p no dysregulation was observed and for miR-34a-5p and miR-706 the dysregulation in the EV-fraction and GAT tissue did not match. The strongest effects overall occurred in the miR-29 family.

In summary, the data presented here provide a valuable resource that can be used as a starting point to study the biology of circulating noncoding RNAs in the context of ageing in general. Especially in combination with published

tissue-specific gene expression atlases, these data enable the community to formulate novel hypotheses on distant cell-cell communication and affected tissues during ageing, which serves as a basis for functional studies in the future. At the same time, our analysis suggests a major role of fat in ageing processes, with EV-bound miRNAs of the miR-29 family performing important regulatory events.

Discussion

While our study presents intriguing new insights into the correlation of EV- and fc fractions and the molecular loading of EVs with noncoding RNAs in the context of ageing, it is important to mention the limitations and how they could be addressed. First, the purification of EVs is challenging and has many pitfalls [23]. Generally, the more purification steps that are applied, the less material is left, eventually requiring a pooling of samples. We decided to achieve the maximal possible purity while still leaving sufficient material for high-throughput sequencing of small RNAs, avoiding any pooling. As stated by Schekman et al. [23], healthy scepticism concerning the possible connection between EV-miRNAs and control of gene expression in target cells should remain until functional cell culture and animal studies are conducted with EVs purified by rigorous and quantitatively documented procedures, allowing depletion of lipoproteins and other non-EV contaminants and quantification and characterization of pure EVs. Similar concerns hold for the RNA molecules detected in the EV pellets, which may include RNAs that are inside the EVs, outside the EVs, or pelleted together with EVs within larger complexes. While NTA and cryo-EM do not replace purification, they verify the presence of vesicles in the samples used (Supplementary Figure S3). A second limitation comes down to the molecular measurement and annotation of the molecules. Having reached a high sequencing depth from low input volumes, the data were mapped to the standard reference databases. Whether read molecules, e.g. mapping to piRNAs represent functional piRNAs and not fragments or reads mapping to miRNAs annotated in miRBase represent functional miRNAs, is only partially known, calling for further functional validation experiments. Another challenge is the normalization of data, which in our case relies on global normalization. Nevertheless, differences between the fc-RNA and EV-RNA fractions as well as an additional targeted validation by RT-qPCR support the general high-throughput results. Nevertheless, we acknowledge that different normalization approaches can impact the results of respective studies relying on microarrays or sequencing. Additionally, we

with respect to the correlation with age. The coloured lines in the background denote the enrichment plots for random distributions. The more the blue line exceeds the background distribution the more significant the pathway is enriched. The horizontal dashed orange line represents a running sum of 2,000 to make the two curves better comparable to each other. (c) Same information as in panel (b) for the response to hypoxia. Here, both, the Ev and fc fraction exceed the random distributions significantly. (d) for miRNAs that are positively and negatively correlated with age in either the fc fraction or EVs, the average correlation with age of target genes from *Tabula muris senis* across 17 tissues is shown. As expected, miRNAs decreasing with age showed target genes increasing with age and vice versa. (e, f) for two fat tissues, the target gene correlation with age is detailed for the four groups shown as rows in (b). Gray boxes span the first to the third quartile, with the line inside the box representing the median value. The whiskers show the minimum and maximum values or values up to 1.5 times the interquartile range below or above the first or third quartile if outliers are present (shown as separate, black outlined dots). (g) Target network. Large green dots depict miRNAs, small pink dots represent genes, and lines delineate experimentally validated regulatory events between miRNAs and genes. The colour shading represents the correlation with age, and the hub genes targeted by at least three miRNAs are annotated in red. Relative font sizes represent the number of miRNAs targeting the respective gene. (h) for the core miRNAs from panel (g), the direct correlation of miRNAs to all target genes in fat tissues of the TMS cohort is shown. Distributions are sorted with respect to an increasing average correlation such that the most consistent downregulation of target genes is observed for miRNAs at the top. (i) RT-Qpcr results. The log2 fold-change based on the RT-Qpcr is presented for EV and GAT tissue. The miR-29 family members present a consistent up-regulation in both fractions.

want to highlight that we considered here only the non-coding RNA fraction from vesicles. But exosomes, known as crucial systemic signalling mediators, carry also mRNA as well as proteins. Especially when secreted by cancer cells the composition of exosomes varies depending on the disease, and these components can have an impact on the development and maintenance of the tumour microenvironment [33–35]. Applying multi-omics studies to ageing-associated vesicles is now crucial to get a comprehensive model on the action of the miRNAs reported herein and due to the complex EV micro-ecosystem.

In performing the study, we paid attention to limit confounding factors wherever possible. Because an influence of the sex on organ specific RNA signatures is known [9,10], we decided to only measure female mice in this study. Of note, the pathways comprising these sex DEGs in the previous RNA based studies largely differ from those describing ageing DEGs. We thus do not have evidence that sex differences influence the transcriptional ageing profiles from those studies. Nonetheless, the amount and cargo of EVs seems to be affected significantly by the sex overall [36–40]. In future targeted preclinical studies, the role of mouse sex on vesicle cargo type and amount should be further explored.

The primary aim of our study was to develop a resource for free circulating (fc-RNA) and extracellular vesicle-associated small RNAs (EV-RNA) in ageing. While we observe broad differences between RNA classes upon ageing, pathway and network analysis highlights a limited set of miRNAs with age-related regulatory activity. Missing miRNA target interactions (MTIs) in the regulatory core network (Figure 3e) await validation in future studies. Our results however do not only pinpoint a multifaceted ageing-factor depending on the RNA class but also on the organ. Strikingly, the results of EV-bound miRNAs suggest a major role of fat tissue in the process of ageing. The miR-29 family seems to play an essential role as a regulator of a tentative core ageing network and has been described in the context of ageing [41,42]. There is an increasing body of evidence associating fat tissue or processes therein with ageing in health and disease [43]. Age-related inhibition of adipogenesis and adipose tissue senescence leads to a decline in body fat in elderly individuals and is implicated in the development of metabolic diseases [44]. One hallmark of ageing is the so-called ‘inflammageing’, which describes a low-level chronic inflammation caused by cells taking on a senescence-associated secretory phenotype (SASP) and secreting proinflammatory molecules [45]. A recent single-cell sequencing study suggests the emergence of ‘ageing-dependent regulatory cells’ in fat tissue of mice with higher age that secrete, for example, the proinflammatory *Ccl6* as a major contributor to adipose cell senescence [46]. For some of the genes in the computed core networks, there is already evidence of a contribution to ageing or age-related diseases. The growth factor *IGF1*, for example, is a major player regulating ageing on a cellular and organism-wide level [47]. The histone deacetylase *HDAC4* has been shown to be polyubiquitinated and degraded during all types of senescence [48]; however, it seems to be overexpressed in ageing muscle [49]. Similarly, deregulation was shown for *CAV2* [50] and other genes reported in this study. The association of EV-RNAs is a likely contributor to these ageing

processes and is already known from cancer research. For example, miR-29a enclosed in tumour-derived exosomes has even been shown to directly bind to intracellular toll-like receptors in immune cells, generating a prometastatic inflammatory response [51]. Moreover, overexpression of miR-29a has been shown to repress adipogenesis in humans and mice due to repression of the glucocorticoid receptor as its target gene [52]. Our data, first and foremost the pathway enrichment analysis, provides evidence that miRNAs are not randomly associate with vesicles. It is hard to distinguish whether those miRNAs have synergistic or antagonistic effects. For both mechanisms examples have been published [53–55]. Of note, miRNAs with different (seed) sequences that nonetheless regulate similar gene sets or pathways exist, pinpointing synergistic or antagonistic effects [56]. Considering the set of pathways identified in our study we assume that the miRNAs in exosomes are synergistically targeting pathways but partially also exhibit antagonistic effects. Here, a systems biology analysis together with targeted validation experiments is required.

Together with the data presented herein, it is plausible to postulate that the miR-29 family might be one of the mediators for the inhibition of adipogenesis and the induction of a proinflammatory environment in ageing fat tissue. One interesting open task for future studies is to determine the cells of origin of the EVs loaded with miR-29 and clarify the regulatory functions of the miRNA in adipocytes. This finding adds to the knowledge on the importance of extracellular vesicles and the immune system interplay in ageing and immune diseases [57], together forming the complex cellular ecosystem.

Methods

Animals

We initially conducted the EV-RNA and fc-RNA isolation workflow on 18 female C57BL/6N mice, 14 of which resulted in a sufficient amount of biological material. Overall, we thus generated non-coding RNA-sequencing samples from mice at the age of 2 months ($n=3$; body weight (bw): 19–20g), 6 months ($n=4$; bw: 25–29g), 8 months ($n=3$; bw: 23–26g), 12 months ($n=2$; bw: 31g) and 18 months ($n=2$; bw: 34 & 41g). The age range was selected based on our previous results on age-related tissue-specific changes in gene expression that were already evident in mice 18months old (*Tabula muris senis* [9,10]. Leftover gonadal fat tissue samples from the TMS study were obtained as a second cohort for RNA sequencing. To assess the size distribution of vesicles and perform EM, an independent cohort of female C57BL/6N mice was used with an age of 2 ($n=4$; bw: 19–20g) and 18 months ($n=4$; bw: 29–41g). For validation experiments using RT-qPCR, a third independent cohort of female C57BL/6N mice was used (2 months: $n=4$, bw: 19–21g; 18 months: $n=2$, bw: 27–32g). The animals – excluding the existing specimens from TMS – were housed in groups on wood chips as bedding in the conventional animal facility of the Institute for Clinical & Experimental Surgery (Saarland University, Homburg/Saar, Germany). They had free access to tap water and standard pellet food (Altromin, Lage, Germany) and were maintained under

a controlled 12-h day/night cycle. This animal study was approved by the local State Office for Health and Consumer Protection and conducted in accordance with Directive 2010/63/EU and the NIH Guidelines for the Care and Use of Laboratory Animals (NIH Publication #85–23 Rev. 1985).

Blood sampling

The mice were anesthetized by an intraperitoneal injection of ketamine (100 mg/kg bw; Ursotamin®; Serumwerke Bernburg, Bernburg, Germany) and xylazine (12 mg/kg bw; Rompun®; Bayer, Leverkusen, Germany). Subsequently, they were fixed on a heating pad in the supine position. After midline laparotomy, a maximal volume of blood (~700–1000 µL) was taken from the vena cava and transferred into plasma tubes (Sarstedt, Nümbrecht, Germany). The blood samples were then centrifuged at 20°C and $10.000 \times g$ for 5 min to remove platelets, large vesicles, and cell debris; the resulting platelet-free plasma was stored at –80°C until further use. After blood collection, gonadal fat tissue of the mice in the third cohort was collected and snap-frozen.

Isolation of EVs

Two hundred microlitres of mouse plasma was transferred to a 1 mL open-top thickwall polypropylene ultracentrifugation tube (Beckman-Coulter, USA) and diluted with 800 µL of phosphate-buffered saline to prevent the tube from collapsing in the ultracentrifuge vacuum. Samples were centrifuged for 2 h at 4°C at $100,000 \times g$ using Type 50.4 Ti fixed-angle rotor (Beckman-Coulter, USA). Supernatants were carefully removed, and the EV-containing pellets were resuspended in 20 µL of phosphate-buffered saline. Samples were stored at –80°C until further analyses. The samples were characterized according to the MISEV2018 criteria [58], and NTA and EM were performed. Limitation of material (700 µL blood/animal/sample that resulted in approximately 250 µL plasma/sample) prevented application of other analytical approaches.

RNA extraction

EV-enriched pellets further referred to as EV fractions and EV-depleted plasma referred to as the free circulating (fc) fraction were used for RNA isolation. All samples (blood from one animal corresponding to one sample) were treated separately, and no samples were pooled. EV- and fc- total RNAs were isolated semi-automated using the miRNeasy Micro kit (Qiagen, Hilden, Germany) and Qiacube isolation robot according to the manufacturer's recommendations with the addition of 2 µL RNase-free glycogen (20 mg/mL, Invitrogen, Carlsbad, CA, USA) to facilitate RNA precipitation. For each sample, at least two replicated sequencing results were available (12- and 18-month replicates, all other time points in triplicate). Total RNA of gonadal fat tissue was isolated using a miRNeasy Mini kit and Tissue Lyser LT according to the manufacturer's protocol. The RNA concentrations of the EV- and fc-fractions were measured using a Qubit™ microRNA Assay Kit, and fat tissue was measured using a Nanodrop (Thermo Fisher Scientific, Waltham, MA, USA).

High-throughput RNA sequencing

Isolated EV-RNA and fc-RNA samples were analysed by Agilent small RNA chips, and 2 ng each (EV-RNA and fc RNA) was used for Illumina-compatible library preparation using the D-Plex Small RNA Kit (Diagenode, BE). The kits employ 3'-poly A tailing and template switch-based cDNA generation using unique molecular identifier (UMI)-tagged template switch oligos. After PCR amplification involving 13 cycles, libraries were purified from TBE-PAGEs. Illumina sequencing was carried out on a HiSeq2500 platform using the High Output mode for 96 cycles. Isolated RNA from gonadal fat tissue was sequenced using the MGISEq system with the standard MPS protocol as described previously [59].

Nanoparticle-tracking analysis

Each EV preparation and each EV-depleted sample were tested using NTA to estimate the number of particles in a sample. For that, 1 µL of plasma was diluted in 1199 µL, and 1 µL of the resuspended EV pellets was diluted in 999 µL of phosphate-buffered saline to achieve a final concentration between 20 and 120 particles/frame. Samples were then measured on NanoSight (Malvern, UK) at a camera level of 15. For each sample, three captures of 30 s were acquired. Videos were then analysed at a detection threshold of 5 using NTA 3.4 software.

Cryo-transmission electron microscopy

Three microlitres of each EV sample was transferred to a holey carbon film-coated copper grid (Plano S147–4), blotted for 2 s, and plunged into undercooled liquid ethane at –165°C (Gatan Cryoplunge3). The grid was then transferred to a cryo-TEM sample holder (Gatan model 914) under liquid nitrogen. Low-dose bright-field images were acquired at –170°C using a JEOL JEM-2100 LaB6 Transmission Electron Microscope and a Gatan Orius SC1000 CCD camera.

RT-Qpcr

Quantitative real-time PCR of EV-RNA and gonadal fat tissue RNA was used to validate age-related expression differences of selected miRNAs. Reverse transcription was performed using the miRCURY LNA RT Kit (Qiagen, Hilden, Germany) with 100 ng fat tissue RNA and EV-RNA equivalent to EVs from 20 µL plasma as input. qPCR was performed using the miRCURY LNA SYBR Green PCR Kit with miRCURY LNA miRNA PCR Assays specific for selected miRNAs in a 10 µL reaction volume. Uniform isolation and RT efficiency were checked using the manufacturer's recommended spike-in controls (Uni-Sp2, 4, 5 and 6). Expression differences were calculated using the $\Delta\Delta C_t$ method with miR-191a and let-7a (in conjunction) as endogenous controls [60].

Computational data analysis

The sample primary processing was performed with miRMaster [61] using standard parameters. The miRNAs

were mapped using Bowtie (version 1.2.3) and allowing up to 1 mismatch. As reference databases we used miRbase (version 22.1); Ensembl ncRNA (version 100), RNCentral piRNA (version 15), GtRNadb (version 18.1), NCBI RefSeq (bacteria & viruses; version 74) and NONCODE (version 5). As output, miRMaster generated a list with the expression of 80,668 RNAs from 10 RNA classes. The data were normalized to expression in one million reads and further processed with R (R 4.0.4 GUI 1.74 Catalina build (7936)). For quality control aspects, we compared the correlation of replicated samples to the correlation obtained for samples that were not replicated. In the first case, we reached an average correlation of 0.93 for the replicates and 0.80 between samples that were not replicates of each other ($p < 10^{-10}$). Venn diagrams were generated using the *eulerr* package from R. Mapping the fraction of variance to different parameters was performed using the principal variance component analysis (*pvca*) package. Splines were computed using the *smooth.spline* function with seven degrees of freedom. Colour palettes were generated using the *RColorBrewer* package. Smoothed scatter plots were computed using the *smoothScatter* function setting the point number to 500. Clustering was performed for the most highly expressed noncoding RNAs (at least 5 reads per million in at least one sample) using the scaled expression matrix (z-score of each feature). The clustering was performed with the *hclust* function using the Euclidean distance measure. Clusters were extracted by cutting the dendrograms at 1/1.25 of the maximal height. Heatmaps of target genes were computed using the *heatmap.2* function. Network visualization was performed using *iGraph*. As input for the network analysis, targets from miRTarBase [31] were used; however, they were restricted to strong evidence targets (i.e. experimentally validated). To compute the statistical concordance of RNAs correlated with ageing across sample types, a random background distribution with respect to positive and negative correlation was assumed. Briefly, a random distribution would mean that close to 25% of non-coding RNAs is consistently positively regulated in plasma and EVs, 25% is consistently negatively correlated with age and 25% in each are positively correlated in the one and negatively correlated in the other specimen type. Where applicable, p-values were corrected for multiple testing using the Benjamini Hochberg method with an alpha-level set to 5%.

Pathway analysis

For the pathway analysis we used miEAA 2.0. Precisely we performed a miRNA set enrichment analysis of the mature RNAs. To this end, we sorted the correlation value of the EV- and fc-fraction separately with the age and uploaded both sorted lists to miEAA. As organism we selected *mus musculus* and choose the gene ontology categories derived over the miRTarBase. We then selected a p-value threshold of 1 to force miEAA reporting of all nominal p-values, facilitating a direct comparison between the pathway results of the fc- and EV-fraction. We adjusted the p-values using the Benjamini-Hochberg method. As graphical output we present the enrichment plots. These plots describe the enrichment statistics for the input list as solid blue line.

In the background, the same distributions computed for random lists (corresponding to the result of non-parametric permutation tests) are shown.

Matrix factorization

We predicted the age of samples with respect to three age groups: ‘young’ (2 months), ‘middle’ (6–8 months) and ‘old’ (12–18 months). To this end, the expression patterns were split into 20 individual matrices for each of the 10 noncoding RNA classes and for plasma and EVs. We first normalized the given nonnegative Matrix D by dividing all elements by the maximum value in D . To obtain the probabilistic regarding the age groups, we decomposed the matrix D into two further matrices T and P , where P gives us the desired probabilities. T stands for the matrix of the typical age group vectors, i.e. in each entry of a column, there is a value representing all entries at this position of all samples belonging to this age group. The matrix P contains the probabilities of each sample to each age group respective to their typical vector in T . We formulated the non-negative matrix factorization as the optimization problem:

$$\begin{aligned} \text{Given : } D &\in \mathbb{R}^{m \times n} \\ \text{Unknown : } T &\in \mathbb{R}^{m \times 3}, P \in \mathbb{R}^{3 \times n} \\ \text{Solution : } \min_{T, P} & \|D - TP\|_F^2 \\ \text{s.t. } 0 &\leq T[i, s] \leq 1 \forall i \in \{1, \dots, m\}, \forall s \in \{1, 2, 3\} \\ 0 &\leq P[s, j] \leq 1 \forall s \in \{1, 2, 3\}, \forall j \in \{1, \dots, n\} \\ \sum_{s=1}^3 & P[s, j] = 1 \forall j \in \{1, \dots, n\} \end{aligned}$$

The first two constraints require all entries of the matrices T and P to lie between 0 and 1. Since we were interested in the percentage of a sample belonging to the three age groups, we also required all columns of the matrix P to sum up to 1 using a numerical solver [62].

We then classified each sample by choosing the index with the highest entry of each column in P and assigned the index as a label to each one. However, the rows of P were invariant to permutations. Here, this means that it is not clear which label corresponds to which age group. Furthermore, a measure of quality for the results was needed. Using the known age, we could construct a ground truth for each sample and calculated the classification accuracy for every permutation. The ground truth used was ‘young’ corresponding to two-month-old mice, ‘middle’ to six- and eight-month-old mice and ‘old’ to twelve- and 18-month-old mice. Finally, we chose the permutation labels that maximize the accuracy and obtain a measure of quality.

Acknowledgment

The authors would like to thank the The Michael J. Fox Foundation for Parkinson’s Research for their support with the Article Processing Charge. We acknowledge support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) and Saarland University within the ‘Open Access Publication Funding’ program.

Disclosure statement

Andreas Keller is member of the Scientific Advisory Board of Firalis.

Funding

This study was partially funded by the Regional Government of the Saarland and by Saarland University.

Author contributions

FK contributed to data analysis, data interpretation, and writing the manuscript. TK performed the vesicle extraction experiments. NL contributed to the study setup, the interpretation of the analysis and manuscript writing. MS contributed to the sequencing of the samples and to the study setup. LG contributed to the vesicle purification experiments. NF contributed to the sequencing experiments. AS contributed to the sequencing experiments. TF contributed to the analysis of primary sequencing data. OH contributed to the analysis of the tissue experiments from TMS. EA contributed to correcting and editing of the manuscript. AE performed the matrix factorization analysis. MK performed the cryo-EM analysis. KW and HM contributed to the tRNA fragment analysis and interpretation. GF contributed to the vesicle extraction and to the study setup. TW contributed to the analysis and interpretation of the data in the context of the tissue experiments from TMS. EM contributed to the study setup and the interpretation of the results. VK added to the interpretation of the data, writing, and correcting the manuscript. MWL performed the mouse experiments. IN contributed to manuscript writing and review of the methods used. AK contributed to the study setup, data interpretation, figure design, and manuscript writing.

Code availability

The primary data analysis was performed using miRMaster. The software is available as a web service. The other analyses were carried out using standard R packages that are freely available (see Methods Section).

Data availability statement

All sequencing data are freely available from the Gene Expression Omnibus under accession number GSE222857. All raw fastq files can be downloaded from the corresponding sequence read archive (SRA) entry. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE222857>

ORCID

Fabian Kern  <http://orcid.org/0000-0002-8223-3750>
 Thomas Kuhn  <http://orcid.org/0000-0003-2444-3204>
 Martin Simon  <http://orcid.org/0000-0002-0962-7788>
 Laura Gröger  <http://orcid.org/0000-0003-1043-3910>
 Ernesto Aparicio-Puerta  <http://orcid.org/0000-0002-3470-1425>
 Abdulrahman Salhab  <http://orcid.org/0000-0002-9649-6922>
 Tobias Fehlmann  <http://orcid.org/0000-0003-1967-2918>
 Oliver Hahn  <http://orcid.org/0000-0001-9440-1152>
 Annika Engel  <http://orcid.org/0000-0001-5570-3115>
 Marcus Koch  <http://orcid.org/0000-0002-1693-6829>
 Katarzyna Winek  <http://orcid.org/0000-0003-3085-9054>
 Hermona Soreq  <http://orcid.org/0000-0002-0955-526X>
 Irina Nazarenko  <http://orcid.org/0000-0002-2633-1161>
 Tony Wyss-Coray  <http://orcid.org/0000-0001-5893-0831>
 Matthias W. Laschke  <http://orcid.org/0000-0002-7847-8456>
 Andreas Keller  <http://orcid.org/0000-0002-5361-0895>

References

- [1] Wyss-Coray T. Ageing, neurodegeneration and brain rejuvenation. *Nature*. 2016;539(7628):180–186. doi: [10.1038/nature20411](https://doi.org/10.1038/nature20411)
- [2] Deelen J, Evans DS, Arking DE, et al. A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat Commun*. 2019;10(1):3669. doi: [10.1038/s41467-019-11558-2](https://doi.org/10.1038/s41467-019-11558-2)
- [3] Lehallier B, Gate D, Schaum N, et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nat Med*. 2019;25(12):1843–1850. doi: [10.1038/s41591-019-0673-2](https://doi.org/10.1038/s41591-019-0673-2)
- [4] Fehlmann T, Lehallier B, Schaum N, et al. Common diseases alter the physiological age-related blood microRNA profile. *Nat Commun*. 2020;11(1):5958. doi: [10.1038/s41467-020-19665-1](https://doi.org/10.1038/s41467-020-19665-1)
- [5] Kern F, Fehlmann T, Violich I, et al. Deep sequencing of sncRNAs reveals hallmarks and regulatory modules of the transcriptome during Parkinson's disease progression. *Nature Aging*. 2021;1(3):309–322. doi: [10.1038/s43587-021-00042-6](https://doi.org/10.1038/s43587-021-00042-6)
- [6] Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet*. 2018;19(6):371–384. doi: [10.1038/s41576-018-0004-3](https://doi.org/10.1038/s41576-018-0004-3)
- [7] Aramillio Irizar P, Schauble S, Esser D, et al. Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. *Nat Commun*. 2018;9(1):327. doi: [10.1038/s41467-017-02395-2](https://doi.org/10.1038/s41467-017-02395-2)
- [8] Yousefzadeh MJ, Flores RR, Zhu Y, et al. An aged immune system drives senescence and ageing of solid organs. *Nature*. 2021;594(7861):100–105. doi: [10.1038/s41586-021-03547-7](https://doi.org/10.1038/s41586-021-03547-7)
- [9] Schaum N, Lehallier B, Hahn O, et al. Ageing hallmarks exhibit organ-specific temporal signatures. *Nature*. 2020;583(7817):596–602. doi: [10.1038/s41586-020-2499-y](https://doi.org/10.1038/s41586-020-2499-y)
- [10] Tabula Muris C, Antony J, Baghel AS. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature*. 2020;583(7817):590–595. doi: [10.1038/s41586-020-2496-1](https://doi.org/10.1038/s41586-020-2496-1)
- [11] Palovics R, Keller A, Schaum N, et al. Molecular hallmarks of heterochronic parabiosis at single-cell resolution. *Nature*. 2022;603(7900):309–314. doi: [10.1038/s41586-022-04461-2](https://doi.org/10.1038/s41586-022-04461-2)
- [12] Sahu A, Clemens ZJ, Shinde SN, et al. Regulation of aged skeletal muscle regeneration by circulating extracellular vesicles. *Nat Aging*. 2021;1(12):1148–1161. doi: [10.1038/s43587-021-00143-2](https://doi.org/10.1038/s43587-021-00143-2)
- [13] Iram T, Kern F, Kaur A, et al. Young CSF restores oligodendrogenesis and memory in aged mice via Fgf17. *Nature*. 2022;605(7910):509–515. doi: [10.1038/s41586-022-04722-0](https://doi.org/10.1038/s41586-022-04722-0)
- [14] Pastuzyn ED, Day CE, Kearns RB, et al. The neuronal gene arc encodes a repurposed retrotransposon gag protein that mediates intercellular RNA transfer. *Cell*. 2018;172(1–2):275–88 e18. doi: [10.1016/j.cell.2017.12.024](https://doi.org/10.1016/j.cell.2017.12.024)
- [15] Tkach M, Thery C. Communication by extracellular vesicles: where we are and where we need to Go. *Cell*. 2016;164(6):1226–1232. doi: [10.1016/j.cell.2016.01.043](https://doi.org/10.1016/j.cell.2016.01.043)
- [16] Zhang Y, Kim MS, Jia B, et al. Hypothalamic stem cells control ageing speed partly through exosomal miRNAs. *Nature*. 2017;548(7665):52–57. doi: [10.1038/nature23282](https://doi.org/10.1038/nature23282)
- [17] Liu J, Wu X, Lu J, et al. Exosomal transfer of osteoclast-derived miRNAs to chondrocytes contributes to osteoarthritis progression. *Nat Aging*. 2021;1(4):368–384. doi: [10.1038/s43587-021-00050-6](https://doi.org/10.1038/s43587-021-00050-6)
- [18] DeCastro J, Littig J, Chou PP, et al. The microfluidic toolbox for analyzing exosome biomarkers of aging. *Molecules*. 2021;26(3):26. doi: [10.3390/molecules26030535](https://doi.org/10.3390/molecules26030535)
- [19] D'Anca M, Fenoglio C, Serpente M, et al. Exosome determinants of physiological aging and age-related neurodegenerative diseases. *Front Aging Neurosci*. 2019;11:232. doi: [10.3389/fnagi.2019.00232](https://doi.org/10.3389/fnagi.2019.00232)
- [20] Prattichizzo F, Micolucci L, Cricca M, et al. Exosome-based immunomodulation during aging: A nano-perspective on inflamm-aging. *Mech Ageing Dev*. 2017;168:44–53. doi: [10.1016/j.mad.2017.02.008](https://doi.org/10.1016/j.mad.2017.02.008)
- [21] Machida T, Tomofuji T, Ekuni D, et al. MicroRNAs in salivary exosome as potential biomarkers of aging. *Int J Mol Sci*. 2015;16(9):21294–21309. doi: [10.3390/ijms160921294](https://doi.org/10.3390/ijms160921294)

- [22] Lee EK, Jeong HO, Bang EJ, et al. The involvement of serum exosomal miR-500-3p and miR-770-3p in aging: modulation by calorie restriction. *Oncotarget*. 2018;9(5):5578–5587. doi: 10.18632/oncotarget.23651
- [23] Shurtleff MJ, Temoche-Diaz MM, Schekman R. Extracellular vesicles and cancer: caveat lector. *Ann Rev Cancer Biol*. 2018;2(1):395–411. doi: 10.1146/annurev-cancerbio-030617-050519
- [24] Bartel DP. Metazoan MicroRNAs. *Cell*. 2018;173(1):20–51. doi: 10.1016/j.cell.2018.03.006
- [25] Gene Ontology C, Douglass E, Good BM. The gene ontology resource: enriching a GOLD mine. *Nucleic Acids Res*. 2021;49(D1):D325–D34. doi: 10.1093/nar/gkaa1113
- [26] Kern F, Fehlmann T, Solomon J, et al. miEAA 2.0: integrating multi-species microRNA enrichment analysis and workflow management systems. *Nucleic Acids Res*. 2020;48(W1):W521–w8. doi: 10.1093/nar/gkaa309
- [27] Backes C, Khaleeq QT, Meese E, et al. miEAA: microRNA enrichment analysis and annotation. *Nucleic Acids Res*. 2016;44(W1):W110–6. doi: 10.1093/nar/gkw345
- [28] Kern F, Krammes L, Danz K, et al. Validation of human microRNA target pathways enables evaluation of target prediction tools. *Nucleic Acids Res*. 2021;49(1):127–144. doi: 10.1093/nar/gkaa1161
- [29] Kehl T, Kern F, Backes C, et al. miRpathdb 2.0: a novel release of the miRNA pathway dictionary database. *Nucleic Acids Res*. 2020;48(D1):D142–D7. doi: 10.1093/nar/gkz1022
- [30] Backes C, Kehl T, Stockel D, et al. miRpathdb: a new dictionary on microRNAs and target pathways. *Nucleic Acids Res*. 2017;45(D1):D90–D6. doi: 10.1093/nar/gkw926
- [31] Huang HY, Lin YC, Li J, et al. miRtarbase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res*. 2020;48:D148–D54. doi: 10.1093/nar/gkz896
- [32] Wagner V, Kern F, Hahn O, et al. Characterizing expression changes in noncoding RNAs during aging and heterochronic parabiosis across mouse tissues. *Nat Biotechnol*. 2023. doi:10.1038/s41587-023-01751-6.
- [33] Luo H-T, Zheng Y-Y, Tang J, et al. Dissecting the multi-omics atlas of the exosomes released by human lung adenocarcinoma stem-like cells. *NPJ Genom Med*. 2021;6(1):48. doi: 10.1038/s41525-021-00217-5
- [34] Valadi H, Ekström K, Bossios A, et al. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol*. 2007;9(6):654–659. doi: 10.1038/ncb1596
- [35] Zhou L, Lv T, Zhang Q, et al. The biology, function and clinical implications of exosomes in lung cancer. *Cancer Lett*. 2017;407:84–92. doi: 10.1016/j.canlet.2017.08.003
- [36] Rigamonti AE, Bollati V, Pergoli L, et al. Effects of an acute bout of exercise on circulating extracellular vesicles: tissue-, sex-, and BMI-related differences. *Int J Obes (Lond)*. 2020;44(5):1108–1118. doi: 10.1038/s41366-019-0460-7
- [37] Penna C, Femmino S, Alloati G, et al. Extracellular vesicles in comorbidities associated with ischaemic heart disease: focus on sex, an overlooked factor. *J Clin Med*. 2021;10(2):10. doi: 10.3390/jcm10020327
- [38] Kim Y, Perez-Gonzalez R, Miller C, et al. Sex differentially alters secretion of brain extracellular vesicles during aging: a potential mechanism for maintaining brain homeostasis. *Neurochem Res*. 2022;47(11):3428–3439. doi: 10.1007/s11064-022-03701-1
- [39] Pait MC, Kaye SD, Su Y, et al. Novel method for collecting hippocampal interstitial fluid extracellular vesicles (EV-ISF) reveals sex-dependent changes in microglial EV proteome in response to Abeta pathology. *bioRxiv*. 2023. doi: 10.1101/2023.03.10.532133
- [40] Chung DD, Mahnke AH, Pinson MR, et al. Sex differences in the transcriptome of extracellular vesicles secreted by fetal neural stem cells and effects of chronic alcohol exposure. *Biol Sex Differ*. 2023;14(1):19. doi: 10.1186/s13293-023-00503-0
- [41] Ugalde AP, Ramsay AJ, de la Rosa J, et al. Aging and chronic DNA damage response activate a regulatory pathway involving miR-29 and p53. *Embo J*. 2011;30(11):2219–2232. doi: 10.1038/emboj.2011.124
- [42] Swahari V, Nakamura A, Hollville E, et al. MiR-29 is an important driver of aging-related phenotypes. *bioRxiv*. 2022. doi: 10.1101/2022.11.29.518429
- [43] Liu Z, Wu KKL, Jiang X, et al. The role of adipose tissue senescence in obesity- and ageing-related metabolic disorders. *Clin Sci*. 2020;134(2):315–330. doi: 10.1042/CS20190966
- [44] Reyes-Farias M, Fos-Domenech J, Serra D, et al. White adipose tissue dysfunction in obesity and aging. *Biochem Pharmacol*. 2021;192:114723. doi: 10.1016/j.bcp.2021.114723
- [45] Fulop T, Larbi A, Pawelec G, et al. Immunology of aging: the birth of inflammaging. *Clin Rev Allergy Immunol*. 2021;64(2):109–122. doi: 10.1007/s12016-021-08899-6
- [46] Nguyen HP, Lin F, Yi D, et al. Aging-dependent regulatory cells emerge in subcutaneous fat to inhibit adipogenesis. *Dev Cell*. 2021;56(10):1437–51 e3. doi: 10.1016/j.devcel.2021.03.026
- [47] Anisimov VN, Bartke A. The key role of growth hormone–insulin–IGF-1 signaling in aging and cancer. *Crit Rev Oncol Hematol*. 2013;87(3):201–223. doi: 10.1016/j.critrevonc.2013.01.005
- [48] Yukioka N, Yamada E, Sasahara M, et al. Lysosomal enzyme activities in the cerebral microvessels in spontaneously and renal hypertensive rats. *Exp Mol Pathol*. 1988;49(1):111–117. doi: 10.1016/0014-4800(88)90025-1
- [49] Su J, Ekman C, Oskolkov N, et al. A novel atlas of gene expression in human skeletal muscle reveals molecular changes associated with aging. *Skelet Muscle*. 2015;5(1):35. doi: 10.1186/s13395-015-0059-1
- [50] Alsaqati M, Thomas RS, Kidd EJ. Proteins involved in endocytosis are upregulated by ageing in the normal human brain: implications for the development of alzheimer's disease. *J Gerontol A Biol Sci Med Sci*. 2018;73:289–298. doi: 10.1093/gerona/glx135
- [51] Fabbri M, Paone A, Calore F, et al. MicroRNAs bind to toll-like receptors to induce prometastatic inflammatory response. *Proc Natl Acad Sci USA*. 2012;109(31):E2110–6. doi: 10.1073/pnas.1209414109
- [52] Glantschnig C, Koenen M, Gil-Lozano M, et al. A miR-29a-driven negative feedback loop regulates peripheral glucocorticoid receptor signaling. *Faseb J*. 2019;33(5):5924–5941. doi: 10.1096/fj.201801385RR
- [53] Wang J, Jiang Q, Faleti OD, et al. Exosomal delivery of antagomirs targeting viral and cellular MicroRNAs synergistically inhibits cancer angiogenesis. *Mol Ther Nucleic Acids*. 2020;22:153–165. doi: 10.1016/j.omtn.2020.08.017
- [54] Padda J, Khalid K, Khedr A, et al. Exosome-derived microRNA: efficacy in cancer. *Cureus*. 2021;13:e17441. doi: 10.7759/cureus.17441
- [55] Hu G, Drescher K, Chen X. Exosomal miRNAs: biological properties and therapeutic potential. *Front Genet*. 2012;3:3. doi: 10.3389/fgene.2012.00056
- [56] Kehl T, Backes C, Kern F, et al. About miRNAs, miRNA seeds, target genes and target pathways. *Oncotarget*. 2017;8(63):107167–107175. doi: 10.18632/oncotarget.22363
- [57] Cheon SY, Lee JE. Extracellular vesicles and immune system in ageing and immune diseases. *Exp Neurobiol*. 2021;30(1):32–47. doi: 10.5607/en20059
- [58] Thery C, Witwer KW, Aikawa E, et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the international society for extracellular vesicles and update of the MISEV2014 guidelines. *J Extracell Vesicles*. 2018;7(1):1535750. doi: 10.1080/20013078.2018.1535750
- [59] Li Y, Fehlmann T, Borchering A, et al. CoolMPS: evaluation of antibody labeling based massively parallel non-coding RNA sequencing. *Nucleic Acids Res*. 2021;49(2):e10. doi: 10.1093/nar/gkaa1122
- [60] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2– $\Delta\Delta$ CT method. *Methods*. 2001;25(4):402–408. doi: 10.1006/meth.2001.1262
- [61] Fehlmann T, Backes C, Kahraman M, et al. Web-based NGS data analysis using miRmaster: a large-scale meta-analysis of human miRNAs. *Nucleic Acids Res*. 2017;45(15):8731–8744. doi: 10.1093/nar/gkx595
- [62] Lutsik P, Slawski M, Gasparoni G, et al. MeDeCom: discovery and quantification of latent components of heterogeneous methylomes. *Genome Biol*. 2017;18(1):55. doi: 10.1186/s13059-017-1182-6

3.2 SKIN TREATMENT WITH NON-THERMAL PLASMA MODULATES THE IMMUNE SYSTEM THROUGH MIR-223-3P AND ITS TARGET GENES

RNA BIOLOGY
2024, VOL. 21, NO. 1, 31–44
<https://doi.org/10.1080/15476286.2024.2361571>



RESEARCH PAPER

OPEN ACCESS

Skin treatment with non-thermal plasma modulates the immune system through miR-223-3p and its target genes

Annika Engel , Nicole Ludwig , Friederike Grandke , Viktoria Wagner , Fabian Kern , Tobias Fehlmann , Georges P. Schmartz , Ernesto Aparicio-Puerta , Dominic Henn , Barbara Walch-Rückheim , Matthias Hannig , Stefan Rupf , Eckart Meese , Matthias W. Laschke , and Andreas Keller

^aChair for Clinical Bioinformatics, Saarland University, Saarbrücken, Germany; ^bDepartment of Human Genetics, Saarland University, Homburg/Saar, Germany; ^cCenter for Human and Molecular Biology, Saarland University, Homburg/Saar, Germany; ^dDepartment of Clinical Bioinformatics (CLIB), Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research, Saarbrücken, Germany; ^eDepartment of Plastic Surgery, UT Southwestern Medical Center, Dallas, TX, USA; ^fCenter of Human and Molecular Biology (ZHMB), Virology & Immunology, Saarland University, Homburg/Saar, Germany; ^gClinic of Operative Dentistry, Periodontology and Preventive Dentistry, Saarland University, Homburg/Saar, Germany; ^hInstitute for Clinical and Experimental Surgery, Saarland University, Homburg/Saar, Germany

ABSTRACT

Non-thermal plasma, a partially ionized gas, holds significant potential for clinical applications, including wound-healing support, oral therapies, and anti-tumour treatments. While its applications showed promising outcomes, the underlying molecular mechanisms remain incompletely understood. We thus apply non-thermal plasma to mouse auricular skin and conducted non-coding RNA sequencing, as well as single-cell blood sequencing. In a time-series analysis (five timepoints spanning 2 hours), we compare the expression of microRNAs in the plasma-treated left ears to the unexposed right ears of the same mice as well as to the ears of unexposed control mice. Our findings indicate specific effects in the treated ears for a set of five miRNAs: mmu-miR-144-5p, mmu-miR-144-3p, mmu-miR-142a-5p, mmu-miR-223-3p, and mmu-miR-451a. Interestingly, mmu-miR-223-3p also exhibits an increase over time in the right non-treated ear of the exposed mice, suggesting systemic effects. Notably, this miRNA, along with mmu-miR-142a-5p and mmu-miR-144-3p, regulates genes and pathways associated with wound healing and tissue regeneration (namely ErbB, FoxO, Hippo, and PI3K-Akt signalling). This co-regulation is particularly remarkable considering the significant seed dissimilarities among the miRNAs. Finally, single-cell sequencing of PBMCs reveals the downregulation of 12 from 15 target genes in B-cells, Cd4+ and Cd8+ T-cells. Collectively, our data provide evidence for a systemic effect of non-thermal plasma.

ARTICLE HISTORY

Revised 14 May 2024
Accepted 24 May 2024

KEYWORDS

Non-thermal plasma; miRNA; single blood cell sequencing; wound healing; cell adhesion; miR-223-3p

Introduction

Non-thermal plasma (NTP), a partially ionized gas consisting of a mixture of ionized atoms, molecules, radicals, and photons, has been studied for clinical application in recent years. Technical applications include decontamination of sensitive instruments [1,2] due to its antibacterial effect [3–6]. Additionally, NTP treatment has been investigated in anti-tumour therapy [7–9]. Interestingly, NTP-induced apoptosis was shown in head and neck cancer (HNC) cells via accumulation of reactive oxygen species *in vitro* and inhibited growth of HNC tumours *in vivo* [10,11]. However, one of the most explored applications of NTP treatment is wound healing. Recently, Wang *et al.* demonstrated inhibited scar formation after surgery [12]. Additionally, Isbary *et al.* reported a reduction of the bacterial load in chronic infected wounds [13]. Preliminary *in vitro* experiments showed that epithelial cells display an increased proliferation when exposed directly to NTP and also when exposed indirectly via NTP-treated

culture medium [14–16], providing a potential mechanism underlying the improved wound healing processes. Considering stages of wound healing from inflammation to tissue formation and remodelling [17], NTP treatment supports wound recovery by stimulating the contribution of monocytes in the first stage [18]. The exact mode of action of NTP has not been fully identified and the knowledge on molecular processes underlying the improved wound healing and other positive effects of NTP are only partially understood. The aforementioned studies suggested the stimulation of the immune system as a systemic effect that might add to the local anti-microbial effects of NTP.

Beyond proteins, coding genes and metabolites, non-coding RNAs are well known to play an important role in processes related to the immune system. Among those, microRNAs (miRNAs) are a particular class of small non-coding RNAs, which are expressed in the course of a specific disease or in a cell-specific way [19–24]. The

CONTACT Andreas Keller andreas.keller@ccb.uni-saarland.de Chair for Clinical Bioinformatics, Saarland University, Campus E2.1, Saarbrücken 66123, Germany

*These authors contributed equally as senior authors.

Supplemental data for this article can be accessed online at <https://doi.org/10.1080/15476286.2024.2361571>.

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

mechanisms how miRNAs are transcribed and processed remain an active field of research [19] but even basic principles remain still incompletely understood [25]. Large-scale databases such as miDIP suggest however over 46 million interactions between miRNAs and genes [26]. Databases such as miRTarBase collect over 2 million of verified miRNA gene interactions [27]. Such comprehensive resources are partially created using curated functional evidence. As an example, the role of infections and how pathogens can hijack miRNAs to suppress innate immunity of the host are known [28], indicating their importance in immune responses. Performing microRNA studies should typically also include an investigation of the target effects to improve a lacking understanding of mechanism in the light of the suggested wide variety of roles for miRNAs [29].

We thus hypothesize that NTP treatment has (a) a local influence on miRNA profiles that (b) potentially extends to a systemic influence and (c) thereby alters gene expression in mediators such as blood cells. To test this hypothesis, we study the effect of NTP treatment on mouse auricular skin. Because of the known effect of NTP treatment in wound healing, we measure differences in miRNA expression in treated versus untreated skin samples using deep sequencing. By including five different timepoints up to 120 minutes after the exposure, we are able to analyse immediate and time-dependent changes in miRNA profiles. Because of potential systemic regulatory effects on target genes, we finally carry out single-cell RNA sequencing on peripheral blood mononuclear cells (PBMC)s.

Results

Non-thermal plasma treatment has a limited global impact on the miRNA repertoire

To investigate the effects of non-thermal plasma (NTP), we select mouse ears as a suitable model system and analysed four distinct subsets: the left and irradiated ear of treated mice (TL), the right ear of treated mice (TR), the left ear of untreated mice (UL), and the right ear of untreated mice (UR). For clarity, we refer to these four groups as the sample types throughout the following sections. To capture the time-dependent processes, we consider miRNA bulk samples from five timepoints after the treatment (Figure 1A). A detailed description of the study setup is given in the method section.

As our initial experimental readout, we choose small RNA deep sequencing. After stringent quality control and adapter trimming, we include an average of 30 million reads per sample (Figure 1B). On average 25 million reads align with the mouse reference genome. Notably, no significant difference was observed within the four groups in terms of quality filtering, mapping to the reference genome, or mapping to mouse miRNAs (Supplemental Figure S1A). Altogether, we obtain 1966 miRNAs of which 495 remain as stably expressed in the data set.

Performing a Principal Component Analysis (PCA) with a global clustering of the miRNA expression data including all samples, we find no discernible clustering patterns based on treatment or timepoints (Figure 1C), arguing against

a potential overarching molecular effect of NPT. To further explore the data and to identify less pronounced effects, we next conduct hierarchical clustering analysis for the top 100 miRNAs with the highest coefficient of variance (CV) across all samples (Figure 1D). Consistent with the PCA-based clustering results, we do not observe clear differentiation between treated and untreated samples, the different sample types, or the timepoints.

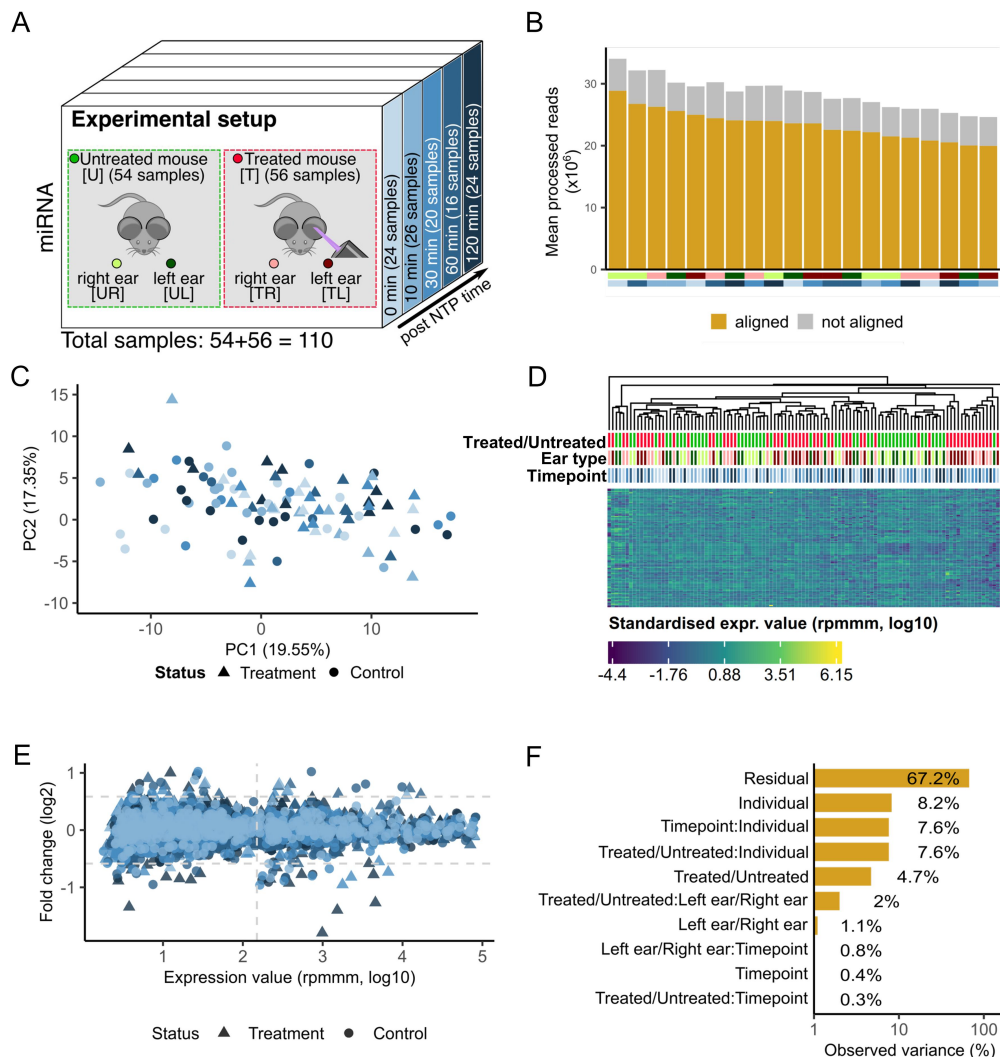
To check whether effects for single miRNAs exist at all, we calculate the fold changes (FCs) for all samples from treated and untreated mice separately across the timepoints (10 min vs. 0 min, 30 min vs. 0 min, 60 min vs. 0 min, 120 min vs. 0 min, Figure 1E). This analysis suggests a stronger effect in the treatment as compared to the control group, both, for miRNAs close to the background but also for highly abundant ones. Because of the large number of analyses (495 miRNAs, 4 timepoint comparisons for two groups) none of the markers in this global analysis remains significant following adjustment for multiple testing.

We thus ask for those factors yielding the overall largest impact on the expression levels using principal variance component analysis (PVCA). While PVCA is typically used to identify batch effects, we utilize it to uncover high-impact properties in our data set. We consider the attributes 'Treated/Untreated', 'Left ear/Right ear', 'Timepoints', and combinations of these attributes as potential factors affecting the expression levels. Additionally, we include the information that two samples always originate from the same mouse as a final attribute, 'Individual' (Figure 1F). The largest portion of the variance lies within the Residuals, indicating factors not explained by our experimental setup. We observe that the most substantial variance in the data set is associated with the 'Individual' property (8.2%), which is a common result pinpointing intra-individual variability. The combination of 'Individual' with 'Treated/Untreated', however, still significantly contributes to the overall data variance (7.6%). In a similar manner, the 'Timepoint' adds 7.6%. One notable result of the PVCA is that the left/right ear adds twice as much to the overall variance if considered in combination with the treatment.

In summary, our findings suggest that treatment with NTP has a limited influence on the global composition of miRNAs. However, it is important to note that this result does not imply the absence of changes in the abundance of individual miRNAs. In fact, the PVCA suggests significant differences between treated and untreated mice, especially if considered in the context of the timepoint and with respect to the left and right ear. This calls for a more focused analysis for each miRNA in the four sample types.

Comparison between treated and untreated mice yields significant upregulated miRNAs

For the four groups, we thus perform ear-wise comparisons within and between the treated and untreated cohorts and compare the results to each other (Figure 2A). Specifically, to identify miRNAs which are influenced by the treatment, we first compare all samples obtained from the treated to all samples from untreated mice since treatment accounts for



the highest variance in the PVCA analysis after excluding all properties related to the individual. Of note, we compute not only the fold changes (FC) and the adjusted p-values (t-test) for the comparison but also the effect sizes (Cohen's d). Our analysis reveals that three miRNAs mmu-miR-144-3p, mmu-miR-144-5p and mmu-miR-451a are more than 1.5-fold up-regulated in treated mice and have a high effect size (≥ 0.5) and a significant adjusted p-value (Figure 2B; upper part). When considering only the last timepoint (120 minutes) we calculate the overall largest effect sizes for a higher number of miRNAs

(Figure 2B; lower part). However, none of the miRNAs remain significant following adjustment. In the light of stronger effect sizes, the larger p-value may be explained by the smaller group sizes in this comparison ($n = 24$).

To make the comparison even more specific, we split the samples into the four sample types (TL, TR, UL, UR) and consider all combinations between treated and untreated with each side of the ear. Interpreting the results, it is important to keep in mind that NTP was only applied to the left ear. We start our analysis by comparing left and right ears first in the

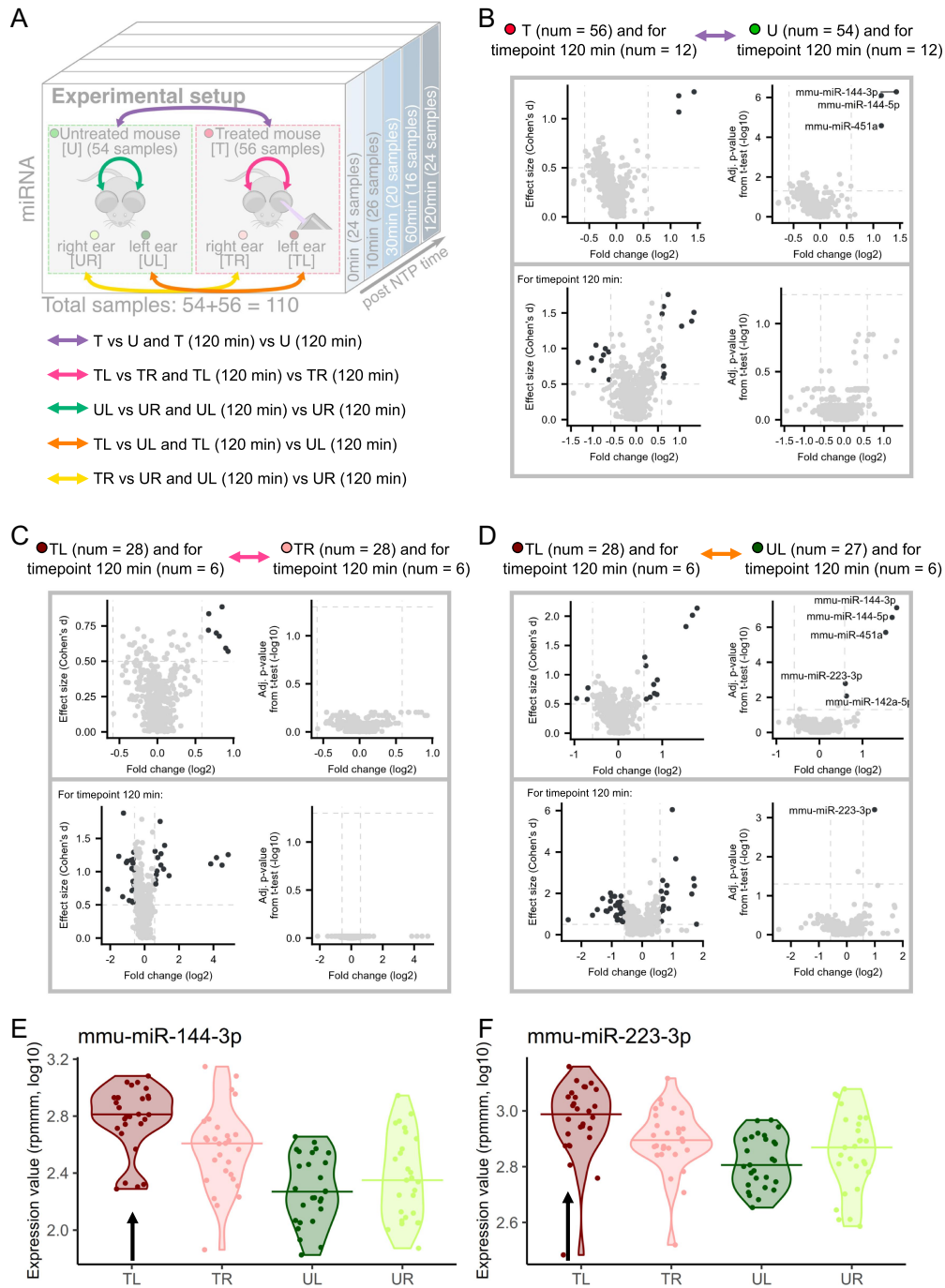


Figure 2. Group-based comparisons of miRNA expression. (A) the experimental setup with additional arrows indicating the comparisons we are investigating. (B) T versus U mice comparison. The effect size plots are located on the left side and Volcano plots on the right side of the different boxes with the comparisons from Figure 2a. The upper plots include all samples over all timepoints and the comparison for only the timepoint 120 minutes is depicted in the lower part of the box. Deregulated miRNAs with effect size ≥ 0.5 or p-value < 0.05 (t-test, $-\log_{10}$) are highlighted in dark grey. (C) TL versus TR mice comparison analogous to Figure 2b. (D) TL versus UL mice comparison analogous to Figure 2b. (E) Violin plot for mmu-miR-144-3p. Each violin corresponds to a sample type and the horizontal lines indicate the medians. The black arrow marks the violin corresponding to the treated ear. (F) Violin plot for mmu-miR-223-3p analogous to Figure 2e.

untreated (*Supplemental Figure S2A*) and then in the treated mice (*Figure 2C*). The results of the comparison of the left ear and the right ear within the untreated mice reveal no deregulated miRNAs, even if the latest timepoint is considered separately. In contrast, in the treated mice we identify upregulated miRNAs in the left ear compared to the right ear, even though they miss the significance level after adjustment for multiple testing. Focusing only on the 120 minutes timepoint confirms the result: we detect deregulated miRNAs, yet these do not prove significant following adjustment for multiple testing. We consequently investigate if the observed upregulation between treated and untreated samples can also be seen when comparing the left ears of treated versus the left ears of untreated mice (*Figure 2D*). In this comparison, we observe the same three miRNAs as above, i.e. mmu-miR-144-3p, mmu-miR-144-5p and mmu-miR-451a, to be of considerable effect size and significance level. In addition, two new candidates, mmu-miR-142a-5p and mmu-miR-223-3p, emerge for further investigation. When only looking at the 120 minutes timepoint, only mmu-miR-223-3p remains significantly upregulated.

Considering these results, we exemplarily compare the five normalized expression values for mmu-miR-144-3p and mmu-miR-223-3p across the four sample types directly to each other (*Figures 2E,2F*). The medians of the samples of the treated mice show a higher expression compared to the untreated ones. Since only the left ears were stimulated with NTP and we find no significant miRNAs between the right ears of the treated and right ears of the untreated mice (*Supplemental Figure S2B*), we suggest that the observed upregulation is caused by the application of NTP. Overall, this argues for the hypothesis of a systemic effect of the treatment.

To gain a better understanding of the underlying effects in the data set, we thus take a closer look at the dynamic variability dependent on the time after exposure to the NTP as a potential key factor.

Detailed view on significantly upregulated miRNAs

In the subsequent analysis, we focus our investigation on the five miRNAs that exhibit significant deregulation. When comparing samples within the same sample type but at the timepoints furthest apart from each other (*Figure 3A*, upper part), we observe a downregulation of miRNA expression in all the samples from the untreated mice and just the samples from the left ear from the untreated mice when considering the timepoint difference between 120 and 0 minutes. This downregulation suggests a general change in miRNA expression over time, potentially influenced by the anesthesia. Interestingly, a similar downregulation is also detected in the samples taken from the right ears of the NTP-treated mice. Furthermore, no significant deregulation is observed between the left and right ears, indicating a consistent behaviour of both ears. However, the downregulation over time does not reach statistical significance, suggesting a limited impact of the anaesthesia. In contrast, we observe a significant upregulation of all five miRNAs in the treated samples compared to the untreated samples (*Figure 3A*, lower

part). This change is consistent across all samples from the treated mice compared to all samples from the untreated mice. Moreover, this upregulation is also evident when comparing the left ears of the treated mice to the left ears of the untreated mice. However, we detect no significant difference in miRNA expression when comparing the right ears of the treated mice to the right ears of the untreated mice.

We thus calculate the Spearman correlation between the miRNA expression and the time following exposure to NTP (*Figures 3B,3C*). For both, treated and untreated samples, we observe mainly positively correlated miRNAs. In the sample-type-specific investigation, however, we find that the majority of miRNAs in the treated mice are positively correlated with the progression of time, while in the untreated mice they are negatively correlated. Notably, the five significant miRNAs from the previous analysis are all negatively correlated with time in the samples from untreated mice, even though not below our empirically determined threshold (-0.3). MiRNAs mmu-miR-223-3p and mmu-miR-142-5p present a positive correlation with time in the samples from treated mice ($\text{corr} = 0.36$ and $\text{corr} = 0.35$) and in the samples from left ears of treated mice ($\text{corr} = 0.61$ and $\text{corr} = 0.49$). Such a systemic effect, however, should become most obvious in the detailed time analysis and become more pronounced over time (*Figures 3D-3G*). We confirm the general trend of microRNAs being higher abundant in treated samples than untreated ones in the majority of timepoints. Resolving the expression over all timepoints does not indicate a distinct behaviour for mmu-miR-144-3p. In contrast, mmu-miR-223-3p, which is the only miRNA significantly upregulated for timepoint 120 minutes in the comparison between TL and UL, presents a monotonous increase from 0 minutes until 30 minutes and remains stable, highlighting it as a most likely systemic miRNA. Altogether, this narrow set of five miRNAs with potential effects opens the question on regulatory mechanisms affecting the expression of genes and pathways.

Upregulated miRNAs regulate wound healing and are enriched in immune and blood cells

To connect our miRNA findings to causal mechanisms, we perform a pathway analysis using the miRPathDB [30] resource (*Figure 4A*). Especially mmu-miR-142a-5p, mmu-miR-144-3p and mmu-miR-223-3p indicate a similar set of pathways connected to wound healing and tissue regeneration for example, the ErbB, FoxO, Hippo and PI3k-Akt signalling pathways [32–34]. The MAPK signalling pathway, which shows to be especially relevant in association with mmu-miR-144-3p, is known to be activated by NTP treatment [18]. In addition, the Adherens junction, Rap1 and Ras pathways exhibit effects regarding the cell adhesion and affiliations to epithelial tissue [35,36]. Furthermore, Ras and TGF-beta pathways are involved in cell-cycle control and apoptosis [37,38]. A co-regulation of miRNAs to similar sets of genes and thus of similar pathways is not surprising and well-known, especially if the seeds of the miRNAs are similar. But there are also examples of miRNAs with very different seeds regulating similar pathways [39]. We thus compute a multiple sequence alignment for the three miRNAs sharing similar target

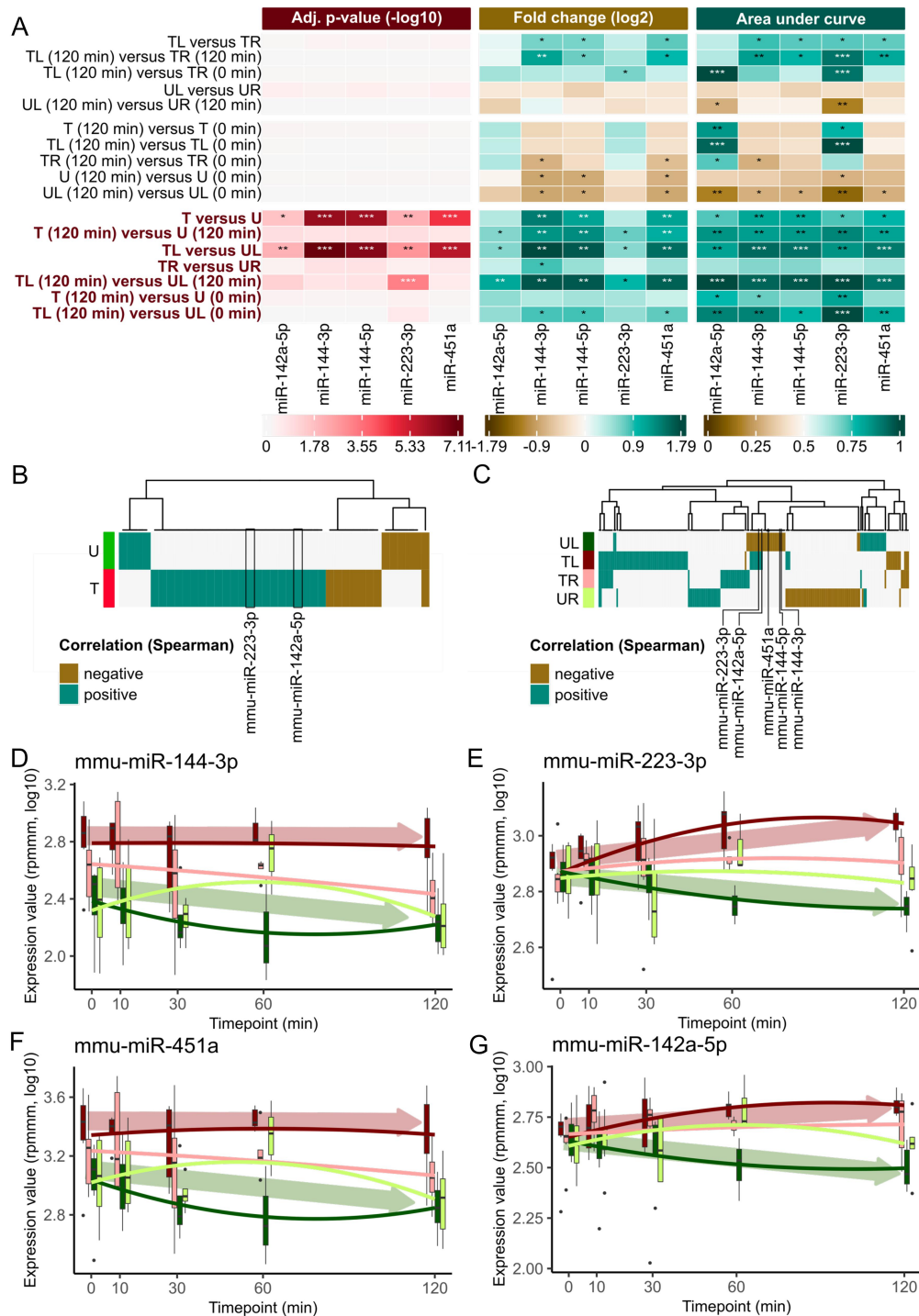


Figure 3. Group-based comparisons and time evolution of miRNA expression for mmu-miR-142a-5p, mmu-miR-144-3p, mmu-miR-144-5p, mmu-miR-223-3p and mmu-miR-451a. (A) Heatmaps of adjusted p-value from t-test ($-\log_{10}$), FC ($\log_2$) and AUC for manually selected comparisons. A p-value of under 0.05 is indicated by a *, of under 0.01 by ** and under 0.001 by ***. A deregulation of miRNAs by a factor of 1.5 is marked with a * and a factor of 2 with **. An AUC value between 0.7 and 0.8 or between 0.3 and 0.2 is marked by *, between 0.8 and 0.9 or between 0.2 and 0.1 is marked by ** and higher than 0.9 or lower than 0.1 is marked by ***. Promising comparisons between the treated and untreated mice are highlighted in red on the left side. (B) Heatmap plot of Spearman correlation between miRNA expressions and timepoints for the samples of the treated and untreated mice. Only miRNAs with a correlation lower than -0.3 or higher than 0.3 are depicted. Correlation values are highlighted as negative correlated if the correlation values are < -0.3 and positive correlated if the correlation values are > 0.3 . (C) Heatmap plot of Spearman correlation between miRNA expressions and timepoints for the samples of the four sample types analogous to Figure 3b. (D) Box plot over all five timepoints for mmu-miR-144-3p for each of the four sample types. The ends of the box denote the 25th percentile and the 75th percentile and the median is given as

pathways (Figure 4B). While miR-144-3p and miR-142a-5p indeed share some similarity in the relevant mature miRNA sequences, miR-223-3p has no seed similarity to the former. This again provides further evidence for a biologically relevant function. At the same time, it immediately poses the question how these miRNAs, and especially miR-223-3p, mediate a potential systemic effect.

To further improve our understanding of the causal mechanisms, we investigate the expression of the miRNAs across different cell types. Because of the homology of mouse miRNAs to humans and a significantly better data basis for human miRNAs in cell types, we perform this analysis on human cell lines. For four of the significantly upregulated miRNAs (miR-142(a)-5p, miR-144-3p, miR-223-3p and miR-451a) the computed alignments of their mouse and human sequences map in every position (Figure 4C). Accordingly, we are able to use a very comprehensive data set of miRNAs across human cell types [31] as a reference resource. We individually examine the expression of hsa-miR-144-3p, hsa-miR-223-3p, hsa-miR-451a and hsa-miR-142-5p at the cell type level (Figure 4D-G). We observe that these four miRNAs are common in blood (red coloured classes, e.g. red blood cells, plasma, and erythroblast-derived cells), immune (green coloured, e.g. CD15+ cells, neutrophil and B lymphocyte and CD56+ cells) and skin cells (pink coloured classes, e.g. hepatocyte, endothelial progenitor cells, conjunctival epithelial cells, Beta cell like derived cells and Endothelial cell retinal microvascular cells).

The significant abundance in skin cells is expected as this is the main specimen of the samples. The dysregulation of miRNAs mainly expressed in immune cells circulating in blood argues for a systemic effect. The principal occurrence of hsa-miR-144-3p and hsa-miR-451a in blood cells and the expression difference between the four experimental groups already visible at the earliest timepoint (0 minutes) (Figures 3D,3F) allows for the hypothesis that the miRNAs may show a direct response to the NTP treatment. In contrast, no difference for mmu-miR-223-3p and mmu-miR-142a-5p at 0 minutes exists (Figures 3E,3G) and the deregulation occurred delayed. This hints towards a dysregulation within the immune system occurring after an initial effect mediated by afore mentioned cell types within the treated tissue. To check our hypothesis of a systematic response to NTP treatment mainly within blood and especially circulating immune cells in blood, we take a closer look at our miRNA candidates and their occurrence in blood cells. For this purpose, we first take into account the data set presented by Juzenas *et al* [40]. This data set offers a comprehensive, cell-specific miRNA catalogue of selected peripheral blood mononuclear cell (PBMC) types (including CD4+ cells, CD8+ cells, B cells (CD19+ cells) and NK cells (CD56+ cells)). Since the data set again refers to human miRNAs, we consider the same four miRNAs as above with complete sequence homology between

human and mouse. We observe that hsa-miR-223-3p, hsa-miR-451 and especially hsa-miR-142-5p are strongly expressed in the cell types shown (NK cells, B cells, CD4+ cells and CD8+ cells) (Supplemental Figure S3A). Therefore, we decide to perform a further analysis to investigate the changes in PBMCs under NTP treatment. Here, single-cell RNA sequencing provides the best option to evaluate the changes over time for the different immune cell types.

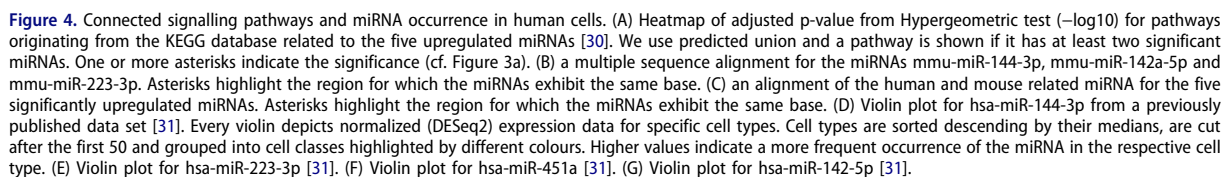
Single immune cell sequencing supports downregulation of targeted genes via miRNA overexpression

We thus perform single-cell sequencing of peripheral blood mononuclear cells (PBMCs) using the same experimental setup. Because the largest impact in our time-series experiment is present at the latest time point and potential regulatory mechanisms might take additional time, we perform the analyses for the latest time point in the study. For this experiment, we use a second cohort of mice and compare blood from two treated to two untreated mice. We sequence 7503 peripheral blood mononuclear cells (PBMCs) and after quality control end up with 5794 high-quality cells. The cell-type annotation highlights five different main cell types: B cells, CD4+ T cells, CD8+ T cells, plasmablasts and NK cells (Figures 5A,5B). Due to the fact that the single-cell method we use to sequence the PBMC data does not allow for quantifying mature miRNAs, we identify the target genes of the four miRNAs with the tool MirTarBase [27] to verify the previously seen effects. In a subsequent differential expression analysis between cells from treated and untreated mice, we find a downregulation for 12 of the 15 target genes linked to the deregulated miRNAs mmu-miR-223-3p and mmu-miR-451a (Figures 5C,5D). Most of the target genes show a deregulation in B cells (7 down and 3 upregulated), followed by CD4+ T cells (5 downregulated) and CD8+ T cells (1 downregulated). Mbnl1 that is targeted by miR-223-3p shows a downregulation both in CD4+ and CD8+ T cells and has been previously found to be involved in the differentiation of fibroblasts during wound healing [41]. This supports the assumption that a treatment with NTP influences both the level of the miRNA and their respective target genes.

Discussion & Conclusion

In this study, we systematically explored the effect of NTP treatment on mouse tissue samples. While the exact mode of NTP action has not been fully identified, previous studies showed that epithelial cells display an increased proliferation after NTP exposure and reported an inhibited scar formation in *in-vitro* experiments [12,15,16]. To examine the underlying molecular processes contributing to improved wound healing following NTP treatment, we analysed non-coding RNA from

a horizontal line. The solid lines show a polynomial of second degree fitted to the expression values using least-squares. Arrows indicate the overall trend of samples from treated (red) and untreated (green) mice. (E) Box plot for mmu-miR-223-3p analogous to Figure 3d. (F) Box plot for mmu-miR-451a analogous to Figure 3d. (G) Box plot for mmu-miR-142a-5p analogous to Figure 3d.



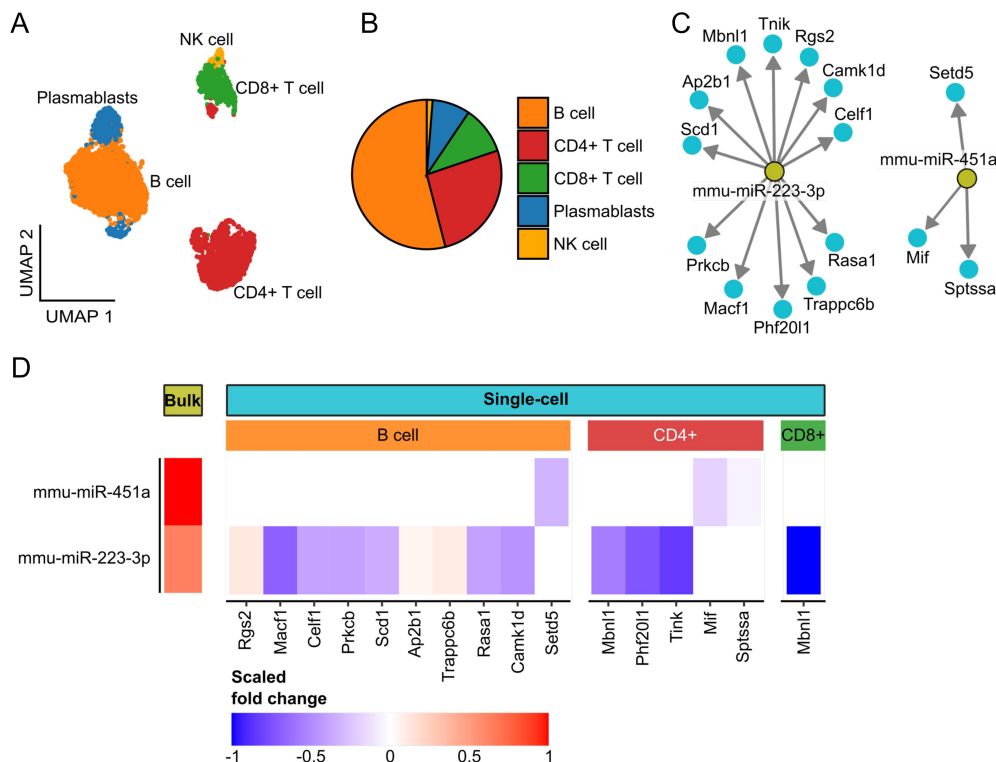


Figure 5. Single-cell quality assessment and associated gene miRNA expression changes. (A) UMAP of single-cell PBMC data coloured by cell type. (B) Cell type proportions in the data set (total: 5794 cells). (C) a graph network to display some of the target genes for miRNAs mmu-miR-223-3p and mmu-miR-451a provided by miRtarbase [27]. The target genes are selected by intersecting all target genes with the significant genes (p-value <0.05) from the single-cell data set. (D) Heatmap plot shows scaled FCs (scaled by the maximal absolute FC in the bulk and single-cell data, respectively) for the miRNAs (bulk data set) and some of their target genes (single-cell data set) from Figure 5c. Red yields an up- and blue a downregulation.

irradiated auricular skin and performed single-cell sequencing of blood.

Unsupervised hierarchical clustering and PCA as a first attempt to group the samples according to their known properties, e.g. treated/untreated, left/right ear or the timepoint, did not yield the anticipated result. Yet, a subsequent PVCA indicated a local influence by yielding promising comparisons between sample subsets, e.g. treated versus untreated left ears. From these comparisons, we obtained five significantly upregulated miRNAs (mmu-miR-144-5p, mmu-miR-144-3p, mmu-miR-142a-5p, mmu-miR-223-3p and mmu-miR-451a).

To reveal a potential systemic influence of the treatment, we investigated this set of miRNAs in a two-fold manner. First, we presented a relation between the upregulated miRNAs and pathways directly linked to wound healing and associated processes like tissue regeneration and cell cycle control. Additionally, mmu-miR-144-3p exhibited a connection to the MAPK signalling pathway which was reported to be induced by NTP treatment in human monocytes [18].

These findings fit to the second part of the investigation where we showed a correlation of miR-223-3p to the NTP treatment. This effect seemed not to be limited to the treated site but systemic. Several recent studies connected the deregulation of miR-223-3p to healing processes. Cheng *et al.* reported that mmu-miR-223-3p increases the muscle

regeneration in the early stage after an injury in mice [42]. They attributed the increase on the regulation of inflammation via the suppression of the target gene interleukin-6 (Il6). Effects concerning the inflammation regulation were also shown to suppress necroptosis in ischaemic hearts [43]. A study investigating the healing of human gingiva reported hsa-miR-223-3p as most upregulated miRNA [44]. In addition, Wu *et al.* presented an increase in hsa-miR-223-3p over the timespan of 21 days in patients with two kinds of fractures and accounted the effect to the regulation of cell viability and cell apoptosis [45].

Additionally, we were able to roughly sketch out the timing of the mmu-miR-223-3p upregulation induced by NTP, which reached the highest level after 30 minutes after treatment. Altogether, this suggests a connection to the inflammatory phase of wound healing occurring in the first few hours after an injury.

To support our findings, we prepared a second data set with single-cell PBMC data to collect more information based on immune cells in the blood. Especially for some target genes of mmu-miR-223-3p (provided by miTarBase [27]) we observed a downregulation in the differential expression analysis. This opposite behaviour compared to the upregulation of the miRNA supports the assumption of an effect of NTP treatment on a RNA-level. We detected a downregulation of the gene Mbnl1 that is targeted by miR-

223-3p in the blood. Although previous publications linked an increased expression of Mbnl1 in skin [41] and heart tissue [46] to an improved wound healing, this does not directly contradict a possible improvement of the wound healing due the downregulation of Mbnl1 in the blood after NTP treatment. Two factors that greatly influence the observed changes in gene expression are the used cells, as changes in the gene expression in blood often differ from the gene expression in other tissues, and the fact that previous publications were focused mainly on the effect in damaged tissue. Instead, we focused on the effect of NTP in healthy, non-injured mice.

Summarizing the bulk and the single-cell study, on the one hand, we found a set of upregulated miRNAs in a bulk experiment and a targeted gene in the subsequent single-cell analysis with reported effects in conjunction to wound healing processes. On the other hand, our analysis resulted in the connection to a signalling pathway, which was already reported in the context of NTP treatment. We conclude that NTP treatment induced some changes on miRNA-level, which might partially contribute to the supporting effects of NTP in wound healing, but further experiments are needed to confirm these conclusions and to decipher the exact mode of action.

Limitations and further work

In the bulk study, we were not able to observe a local difference between left and right ears nor a significant deregulation over time. This could be attributed to the low sample density, with only four to seven samples per timepoint and sample type. Additionally, many effects of wound healing occur immediately after the injury, yet further stages occur in the span of several days [17]. For example, mmu-miR-223-3p was only significantly upregulated for the last timepoint (120 minutes). This suggests that a further study would benefit from more samples per timepoint and a longer study period. However, we chose our experimental parameters optimizing the balance between sufficient samples per time point and minimizing the number of animals sacrificed for the study. The latest time point of 120 minutes was chosen to prevent possible narcosis damage of the mice caused by an extended sedation.

For the validation of the bulk results, we only focused on single-cell sequencing of blood. Due to the constraints of available blood volume from the experimental subjects, we omitted complementary measurements of microvesicles. Nonetheless, our previous studies supported the hypothesis that microvesicles are likely primary mediators of miRNA transport and signal transmission [47,48,50]. Furthermore, for a definitive exploration of the mechanisms contributed to a miRNA, e.g., mmu-miR-223-3p, an experimental setup could be considered where the development of the miRNA in question is artificially inhibited or prevented. Yet, such a tissue-specific knock-out remained out of the scope for this preliminary investigation. An additional experimental verification of the observed dysregulated genes, e.g. a qPCR analysis, would present a beneficial investigation but was left out considering the quality of our samples leveraging 10X Chromium technology. Although we showed that the miRNAs of interest are present in PBMCs, our single-cell data only represents

gene changes, as standardised high-throughput methods for the quantification of mature miRNAs are not yet established but are part of ongoing research [49,50]

Methods

Bulk study

Study setup

We carefully selected a suitable model system that would allow for controlled experiments in a well-defined environment. The mouse ear emerged as an ideal choice due to its physiological characteristics, including the absence of hair, accessibility, and the presence of a well-established vasculature. By focusing on uninjured auricular skin as our model system, we aimed to study the direct effects of NTP without inducing unnecessary stress on the mice. This particular choice of model system offered several advantages. First, it enabled us to analyse the molecular processes occurring in healthy skin that were stimulated by exposure to NTP. Second, the lack of injury eliminated confounding factors associated with wound healing processes, allowing us to isolate the specific effects of NTP on healthy skin. Additionally, the ease of sampling from the auricular skin simplified the experimental procedures. To capture the dynamic changes induced by NTP exposure, we obtained samples from the treated mice immediately after the end of NTP treatment (0 minutes), as well as at 10, 30, 60, and 120 minutes post-treatment. Besides the samples from the treated mice (irradiated and un-irradiated ears) we considered a equally sized control group.

Samples

All animal experiments were approved by the local authorities (Landesamt für Verbraucherschutz, Saarbrücken, Germany; permission number: 21/2018) and conducted in accordance with the European legislation on the protection of animals (Directive 2010/63/EU) and the National Institutes of Health (NIH) guidelines on the care and use of laboratory animals (NIH publication #85-23 Rev. 1985). The C57BL/6 wild-type mice used in this study were housed in groups on wood chips as bedding under a 12 h day/night cycle in the animal facility of the Institute for Clinical and Experimental Surgery (Saarland University, Homburg/Saar, Germany) with free access to water and standard pellet food (Altromin, Lage, Germany). To exclude age and gender as potentially confounding factors, we selected 3-month-old female mice (3.2 ± 0.3 months) for our study, which were assigned to either treatment or control group. All mice were anaesthetized by an intraperitoneal injection of ketamine (75 mg/kg body weight; Ursotamin®; Serumwerke Bernburg, Bernburg, Germany) and xylazine (15 mg/kg body weight; Rompun®; Bayer AG, Leverkusen, Germany). The left ear of the mice in the treatment group was irradiated with NTP for 10 minutes using an atmospheric pressure argon plasma jet kINPen Med [51] (Neoplas, Greifswald, Germany) with argon (Air Liquide, Düsseldorf, Germany) using a flow rate of 5 standard litres per minute (sL/m). The distance between the tip of the plasma

effluent and the mucosa was between 5 and 10 mm. The plasma jet was applied in meander-like motion to prevent heating of the treated tissue. The mice of the control group were equally anaesthetized but remained unexposed to NTP. Right and left ears of treatment and control mice were collected after 0, 10, 30, 60, and 120 minutes, immersed in RNAlater for 12 hours at 4°C and subsequently frozen at -80°C. RNA was isolated using miRNeasy Mini kit (Qiagen, Hilden, Germany) after manufacturer's recommendations. RNA quantity and quality was determined using Nanodrop (ThermoFisher Scientific, Waltham, MA, USA) and Bioanalyzer RNA Nano Chip Kit (Agilent Technologies, Santa Clara, CA, USA). Mean RNA Integrity Number (RIN) of the samples was 8.2 ± 0.9 .

Library preparation and sequencing

As our initial experimental readout, we chose small RNA deep sequencing which allows for high-resolution annotation of miRNAs. Small RNA sequencing libraries were prepared using MGIEasy Small RNA Library Prep Kit on the high-throughput MGI SP-960 sample prep system (MGI Tech, Shenzhen, China). According to the manufacturer's protocols, the 3'- and 5'-adapters were ligated to the RNA. Afterwards, an RT primer including sample-specific barcodes was utilized to perform reverse transcription (RT). Amplification of the cDNA was achieved using a polymerase chain reaction (PCR) with 21 cycles. PCR products were subsequently size selected via AMPure Beads XP (Beckman Coulter, Brea, CA, USA). Size and concentration of the purified PCR products were determined using an Agilent DNA 1000 Kit (Agilent Technologies, Santa Clara, CA, USA) and Qubit™ 1X dsDNA High Sensitivity (HS) (ThermoFisher Scientific, Waltham, MA, USA), respectively. Fifteen samples and one performance control sample were pooled in an equimolar fashion in one sequencing library, which was circularized using the MGI Easy circularization kit. In total, we prepared 8 libraries composed of a total of 120 samples which were sequenced on DNBSEQ-G400RS High-throughput Sequencing Set (Small RNA FCL SE50) by BGI in Wuhan, China.

Data analysis

The pipeline 'miRMaster 2.0' in standard settings [52] was used to process the sequencing data. Furthermore, only miRNAs were considered for further analysis and called expressed that exhibited more or equal than 5 raw reads in 100% of the samples of at least one sample type. We normalized the expression data to make the samples comparable. Therefore, we used the rpm-normalization integrated in the above-mentioned pipeline. This normalization considers the number of reads mapped to a miRNA and the total number of reads mapped to all miRNAs in units of million reads. In this work, we call this normalization rpmmm-normalization. For the further analysis, we used the filtered and normalized expression data, which consisted of 495 miRNAs. A total of 6 samples distributed over all sample types showed a vastly different behaviour in a preliminary analysis. We identified these

samples as outliers and removed both samples of the corresponding mice for further analysis. In total, we obtained 110 samples after outlier removal.

Data analysis was performed using R in version 4.2.2 with the following packages: data.table in version 1.14.6, tidyverse in version 1.3.2, effsize in version 0.8.1, pROC in version 1.18, cowplot in version 1.1.1, reshape2 in version 1.4.4, ComplexHeatmap [53] in version 2.14.0, viridisLite in version 0.4.1, lme4 in version 1.1, readxl in version 1.4.1 RhpBLASctl in version 0.21.

We always applied an unpaired two-tailed t-test. For the adjustment of p-values we used the Benjamini-Hochberg procedure. We used effect sizes to identify and specify quantifiable statements of an underlying effect between to subsets of a data set. To measure the effect size, which is done for every feature individually, we considered several measures such as fold change (FC), Cohen's d and the area under the receiver operator characteristics curve (AUC). The FC is derived by building a ratio between two expression values. If the ratio is 1, this means that the two values are equal. If the ratio is <1 (>1), we know that the part we divided by, exhibits the higher (lower) value. We say that the change between the two values is down (up) directed. We called a miRNA deregulated if it passed in one direction a specified fold change threshold and significant deregulated if its adjusted p-value is lower than 0.05. We defined a miRNA as upregulated if $FC \geq 1.5$ and downregulated if $FC \leq 1/1.5$. Whereas the fold change just represents the ratio between two values, Cohen's d uses a sophisticated ratio including the mean and the estimated variances of the two subsets calculated with the effsize package. We interpreted a feature as impactful if its Cohen's d value ≥ 0.5 . The third effect size we used was the AUC value. We took the expression values of the samples included in the comparison as a prediction and a binary value corresponding to the respective subset the sample belonged to as response. This allowed us to build an ROC curve. In this way, upregulated miRNAs achieved an AUC value close to 1 and downregulated ones close to 0. To determine the AUC value, we used the pROC package. All correlations were calculated using the Spearman rank correlation. For correlation analysis, only miRNAs exceeding the interval between -0.3 and 0.3 were marked as negatively or positively correlated. Principal Variance Component Analysis (PVCA) is a combination of principal component analysis (PCA) and variance component analysis (VCA). As a result, we obtained the proportion of variance for the available properties. The residual bar represented the remaining variance which is not covered by any of the given properties. The listed properties and their combinations presented their impact on the total variance in the data set. For clustering, we used hierarchal clustering with Euclidean distance and complete linkage. For the clustering of the expression data, we first standardized, which means the data were transformed to a mean of 0 with standard deviation of 1. To create the heatmaps, we used the ComplexHeatmap package. The data for the heatmap depicting pathways, originated from the tool miRPathDB [30], which outputs all significant KEGG pathways for the entered miRNAs. The values in the heatmap indicated the p-value in -log10-transform originating from the Hypergeometric test for the corresponding enrichment result.

Single-cell transcriptome analysis

Samples, library preparation and sequencing

The experimental setup for data collection was the same as for the bulk samples. As for the bulk samples, C57BL/6 mice from the Institute of Clinical and Experimental Surgery (Saarland University, Homburg/Saar, Germany) were used. All mice were anaesthetized as described above and, in the case of the treated mice, irradiated with NTP for 10 minutes. The plasma and the treatment itself were handled in the same way as for the bulk samples. Following a period of 120 minutes after irradiation for the treated mice and after the onset of anaesthesia for the untreated mice, blood (usually between 500 and 800 μ L) was collected from the vena cava.

PBMC were isolated from heparin blood of four mice using standard Ficoll gradient. Cells were counted in a haemocytometer using Trypan blue for live/dead staining. Viability was >95% for all samples. Single-cell transcriptome analysis was performed using Chromium Next GEM Single Cell 3' Kit v3.1 (10 $\times$ Genomics) with 5,000 cells input after the manufacturer's protocol. Generated single-cell libraries were sequenced by Novogene on an Illumina NovaSeq using PE150 sequencing.

Data analysis

Reads were aligned against the mm10 genome (refdata-gex-mm10-2020-A) using the 10 $\times$ Genomics Cell Ranger software (v.7.1.0) using a cut-off value of 200 UMIs.

For quality-control, SoupX [54] (v.1.6.1) was used to remove ambient cell-free mRNA contamination, cells with more than 10% mitochondrial reads, <200 features or >2,500 features were removed using Seurat [55] (v.4.3.0). As a third quality-control step, doublets were removed using DoubletFinder [56] (v.2.0.3) using automatically determined (using the recommended settings) parameters nExp and pK.

The samples were aligned using Seurat integration with default settings and dimensionality-reduction was performed using PCA and UMAP on the first 20 dimensions. The data was clustered using the FindNeighbors and FindClusters Functions of Seurat at 0.8 resolution.

For the cell type annotation, a differential expression analysis using Seurat's FindAllMarkers function (Wilcox) between each cluster and all other clusters was performed to identify marker-genes and the cells were annotated using known marker-genes. Clusters with the same cell type were merged.

Genes that were differentially expressed between treated and untreated mice were determined using the FindMarkers Function of Seurat with MAST [57]. Genes were selected if they had an adjusted p-value (Bonferroni correction) below 0.05.

Acknowledgments

We thank Stephanie Deutscher and Esther Maldener for excellent technical assistance.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

We acknowledge funding by Deutsche Forschungsgemeinschaft (DFG-project: Compute- und Storage Cluster with project number 469073465). For the Development of the software pipeline used to process the single-cell data, we are supported by funds of the European Health and Digital Executive Agency (HaDEA) (Project number 101057548-EPIVINF).

Author contributions

E.M., S.R., M.H., M.W.L. and A.K.: conception or design of the work. N. L., M.W.L., S.R., V.W. and B.W.: conducting the experiments and sequencing. A.E., F.G., T.F. and A.K.: data analysis and interpretation. A.E., N.L., F.G., A.K.: drafting the article. A.K., N.L., M.W.L., S.R., D.H., G.P.S., E.A., F.K. and A.E.: critical revision of the article. All authors read and approved the final manuscript.

Code availability

The code that supported the analysis within this paper is available at https://github.com/CCB-SB/2023_mouse_ear_non-thermal_plasma

Data availability statement

All sequencing data are available for download from NCBI's Gene Expression Omnibus [58] via GEO Series accession number GSE236802 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE236802>)

Ethical approval

All animal experiments were approved by the local authorities (Landesamt für Verbraucherschutz, Saarbrücken, Germany; permission number: 21/2018) and conducted in accordance with the European legislation on the protection of animals (Directive 2010/63/EU) and the National Institutes of Health (NIH) guidelines on the care and use of laboratory animals (NIH publication #85-23 Rev. 1985).

Extended Data

Supplemental Figure S1: (A) The percentage of reads, which are mapped to miRNAs. The bottom annotation indicates the sample type and timepoint.

Supplemental Figure S2: (A) TR versus UR mice comparison. The effect size plots are located on the left side and Volcano plots on the right side of the different boxes with the comparisons from Figure 2A. The upper plots include all samples over all timepoints and the comparison for only the timepoint 120 minutes is depicted in the lower part of the box. Deregulated miRNAs with effect size ≥ 0.5 or p-value <0.05 (t-test, -log10) are highlighted in dark grey. (B) UL versus UR mice comparison analogous to Supplementary Figure S2a.

Supplemental Figure S3: (A) Adjusted (layout changes and display only four cell types) plot from the web tool of Juzenas et al. [40]. This data set is offering a comprehensive, cell-specific miRNA catalogue of selected PBMC types.

ORCID

Annika Engel  <http://orcid.org/0000-0001-5570-3115>
 Nicole Ludwig  <http://orcid.org/0000-0003-4703-7567>
 Friederike Grandke  <http://orcid.org/0000-0001-6185-1472>
 Viktoria Wagner  <http://orcid.org/0000-0002-1957-0658>
 Fabian Kern  <http://orcid.org/0000-0002-8223-3750>
 Tobias Fehlmann  <http://orcid.org/0000-0003-1967-2918>
 Georges P. Schmartz  <http://orcid.org/0000-0002-9627-9223>
 Ernesto Aparicio-Puerta  <http://orcid.org/0000-0002-3470-1425>
 Dominic Henn  <http://orcid.org/0000-0003-4775-0151>
 Matthias W. Laschke  <http://orcid.org/0000-0002-7847-8456>
 Andreas Keller  <http://orcid.org/0000-0002-5361-0895>

References

- [1] Kalghatgi S, Kelly CM, Cerchar E, et al. Effects of non-thermal plasma on mammalian cells. *PLoS One*. 2011;6(1):e16270. doi: [10.1371/journal.pone.0016270](https://doi.org/10.1371/journal.pone.0016270)
- [2] O'Connor N, Cahill O, Daniels S, et al. Cold atmospheric pressure plasma and decontamination. Can it contribute to preventing hospital-acquired infections? *J Hosp Infect*. 2014;88(2):59–65. doi: [10.1016/j.jhin.2014.06.015](https://doi.org/10.1016/j.jhin.2014.06.015)
- [3] Koban I, Holtfreter B, Hübner N-O, et al. Antimicrobial efficacy of non-thermal plasma in comparison to chlorhexidine against dental biofilms on titanium discs in vitro - proof of principle experiment. *J Clin Periodontol*. 2011;38(10):956–965. doi: [10.1111/j.1600-051X.2011.01740.x](https://doi.org/10.1111/j.1600-051X.2011.01740.x)
- [4] Oehmigen K, Winter J, Hähnel M, et al. Estimation of possible mechanisms of *Escherichia coli* inactivation by plasma treated sodium chloride solution. *Plasma Process Polym*. 2011;8(10):904–913. doi: [10.1002/ppap.201000099](https://doi.org/10.1002/ppap.201000099)
- [5] Idlibi AN, Al-Marrawi F, Hannig M, et al. Destruction of oral biofilms formed in situ on machined titanium (Ti) surfaces by cold atmospheric plasma. *Biofouling*. 2013;29(4):369–379. doi: [10.1080/08927014.2013.775255](https://doi.org/10.1080/08927014.2013.775255)
- [6] Fallon M, Kennedy S, Daniels S, et al. Technologies to decontaminate bacterial biofilm on hospital surfaces: a potential new role for cold plasma? *J Med Microbiol*. 2022;71(10). doi: [10.1099/jmm.0.001582](https://doi.org/10.1099/jmm.0.001582)
- [7] Cheng X, Sherman J, Murphy W, et al. The effect of tuning cold plasma composition on glioblastoma cell viability. *PLoS One*. 2014;9(5):e98652. doi: [10.1371/journal.pone.0098652](https://doi.org/10.1371/journal.pone.0098652)
- [8] Vandamme M, Robert E, Lerondel S, et al. ROS implication in a new antitumor strategy based on non-thermal plasma. *Int J Cancer*. 2012;130(9):2185–2194. doi: [10.1002/ijc.26252](https://doi.org/10.1002/ijc.26252)
- [9] Kaushik NK, Kaushik N, Park D, et al. Altered antioxidant system stimulates dielectric barrier discharge plasma-induced cell death for solid tumor cell treatment. *PLoS One*. 2014;9(7):e103349. doi: [10.1371/journal.pone.0103349](https://doi.org/10.1371/journal.pone.0103349)
- [10] Kang SU, Cho J-H, Chang JW, et al. Nonthermal plasma induces head and neck cancer cell death: the potential involvement of mitogen-activated protein kinase-dependent mitochondrial reactive oxygen species. *Cell Death Dis*. 2014;5(2):e1056. doi: [10.1038/cddis.2014.33](https://doi.org/10.1038/cddis.2014.33)
- [11] Li X, Rui X, Li D, et al. Plasma oncology: Adjuvant therapy for head and neck cancer using cold atmospheric plasma. *Front Oncol*. 2022;12:994172. doi: [10.3389/fonc.2022.994172](https://doi.org/10.3389/fonc.2022.994172)
- [12] Wang XF, Fang Q-Q, Jia B, et al. Potential effect of non-thermal plasma for the inhibition of scar formation: a preliminary report. *Sci Rep*. 2020;10(1):1064. doi: [10.1038/s41598-020-57703-6](https://doi.org/10.1038/s41598-020-57703-6)
- [13] Isbary G, Morfill G, Schmidt HU, et al. A first prospective randomized controlled trial to decrease bacterial load using cold atmospheric argon plasma on chronic wounds in patients. *Br J Dermatol*. 2010;163(1):78–82. doi: [10.1111/j.1365-2133.2010.09744.x](https://doi.org/10.1111/j.1365-2133.2010.09744.x)
- [14] Barton A, Wende K, Bundscherer L, et al. Nonthermal plasma increases expression of wound healing related genes in a keratinocyte cell line. *Plasma Med*. 2013;3(1–2):125–136. doi: [10.1615/PlasmaMed.2014008540](https://doi.org/10.1615/PlasmaMed.2014008540)
- [15] Kalghatgi S, Friedman G, Fridman A, et al. Endothelial cell proliferation is enhanced by low dose non-thermal plasma through fibroblast growth factor-2 release. *Ann Biomed Eng*. 2010;38(3):748–757. doi: [10.1007/s10439-009-9868-x](https://doi.org/10.1007/s10439-009-9868-x)
- [16] Hasse S, Duong Tran T, Hahn O, et al. Induction of proliferation of basal epidermal keratinocytes by cold atmospheric-pressure plasma. *Clin Exp Dermatol*. 2016;41(2):202–209. doi: [10.1111/ced.12735](https://doi.org/10.1111/ced.12735)
- [17] Gurtner GC, Werner S, Barrandon Y, et al. Wound repair and regeneration. *Nature*. 2008;453(7193):314–321. doi: [10.1038/nature07039](https://doi.org/10.1038/nature07039)
- [18] Bundscherer L, Nagel S, Hasse S, et al. Non-thermal plasma treatment induces MAPK signaling in human monocytes. *Open Chem*. 2015;13(1). doi: [10.1515/chem-2015-0071](https://doi.org/10.1515/chem-2015-0071)
- [19] Bartel DP. Metazoan MicroRNAs. *Cell*. 2018;173(1):20–51. doi: [10.1016/j.cell.2018.03.006](https://doi.org/10.1016/j.cell.2018.03.006)
- [20] Keller A, Gröger L, Tschernig T, et al. iRNATissueAtlas2: an update to the human miRNA tissue atlas. *Nucleic Acids Res*. 2022;50(D1):D211–D221. doi: [10.1093/nar/gkab808](https://doi.org/10.1093/nar/gkab808)
- [21] Ludwig N, Leidinger P, Becker K, et al. Distribution of miRNA expression across human tissues. *Nucleic Acids Res*. 2016;44(8):3865–3877. doi: [10.1093/nar/gkw116](https://doi.org/10.1093/nar/gkw116)
- [22] Smith CM, Hutvagner G. A comparative analysis of single cell small RNA sequencing data reveals heterogeneous isomiR expression and regulation. *Sci Rep*. 2022;12(1):2834. doi: [10.1038/s41598-022-06876-3](https://doi.org/10.1038/s41598-022-06876-3)
- [23] Telonis AG, Lohrer P, Jing Y, et al. Beyond the one-locus-one-miRNA paradigm: microRNA isoforms enable deeper insights into breast cancer heterogeneity. *Nucleic Acids Res*. 2015;43(19):9158–9175. doi: [10.1093/nar/gkv922](https://doi.org/10.1093/nar/gkv922)
- [24] Fromm B, Høye E, Domanska D, et al. MirGeneDB 2.1: toward a complete sampling of all major animal phyla. *Nucleic Acids Res*. 2022;50(D1):D204–D210. doi: [10.1093/nar/gkab1101](https://doi.org/10.1093/nar/gkab1101)
- [25] Chu Y, Kilikevicius A, Liu J, et al. Argonaute binding within 3'-untranslated regions poorly predicts gene repression. *Nucleic Acids Res*. 2020;48:7439–7453. doi: [10.1093/nar/gkaa478](https://doi.org/10.1093/nar/gkaa478)
- [26] Hauschild AC, Pastrello C, Ekaputeri G, et al. MirDIP 5.2: tissue context annotation and novel microRNA curation. *Nucleic Acids Res*. 2023;51(D1):D217–D225. doi: [10.1093/nar/gkac1070](https://doi.org/10.1093/nar/gkac1070)
- [27] Huang HY, Lin YCD, Cui S, et al. miRTarBase update 2022: an informative resource for experimentally validated miRNA–target interactions. *Nucleic Acids Res*. 2022;50(D1):D222–D230. doi: [10.1093/nar/gkab1079](https://doi.org/10.1093/nar/gkab1079)
- [28] Trobaugh DW, Gardner CL, Sun C, et al. RNA viruses can hijack vertebrate microRNAs to suppress innate immunity. *Nature*. 2014;506(7487):245–248. doi: [10.1038/nature12869](https://doi.org/10.1038/nature12869)
- [29] Kilikevicius A, Meister G, Corey DR. Reexamining assumptions about miRNA-guided gene silencing. *Nucleic Acids Res*. 2022;50(2):617–634. doi: [10.1093/nar/gkab1256](https://doi.org/10.1093/nar/gkab1256)
- [30] Kehl T, Kern F, Backes C, et al. miPathDB 2.0: a novel release of the miRNA pathway dictionary database. *Nucleic Acids Res*. 2020;48(D1):D142–D147. doi: [10.1093/nar/gkz1022](https://doi.org/10.1093/nar/gkz1022)
- [31] Patil AH, Baran A, Brehm ZP, et al. A curated human cellular microRNAome based on 196 primary cell types. *Gigascience*. 2022;11. doi: [10.1093/gigascience/giac083](https://doi.org/10.1093/gigascience/giac083)
- [32] Wells A. EGF receptor. *Int J Biochem Cell Biol*. 1999;31(6):637–643. doi: [10.1016/S1357-2725\(99\)00015-1](https://doi.org/10.1016/S1357-2725(99)00015-1)
- [33] Hameedalddeen A, Liu J, Batres A, et al. FOXO1, TGF-β Regulation and wound healing. *Int J Mol Sci*. 2014;15(9):16257–16269. doi: [10.3390/ijms150916257](https://doi.org/10.3390/ijms150916257)
- [34] Xiao W, Tang H, Wu M, et al. Ozone oil promotes wound healing by increasing the migration of fibroblasts via PI3K/Akt/mTOR signaling pathway. *Biosci Rep*. 2017;37(6). doi: [10.1042/bsr20170658](https://doi.org/10.1042/bsr20170658)
- [35] Baum B, Georgiou M. Dynamics of adherens junctions in epithelial establishment, maintenance, and remodeling. *J Cell Bio*. 2011;192(6):907–917. doi: [10.1083/jcb.201009141](https://doi.org/10.1083/jcb.201009141)
- [36] Boettner B, Van Aelst L. Control of cell adhesion dynamics by Rap1 signaling. *Curr Opin Cell Biol*. 2009;21(5):684–693. doi: [10.1016/j.ceb.2009.06.004](https://doi.org/10.1016/j.ceb.2009.06.004)
- [37] Cox AD, Der CJ. The dark side of Ras: regulation of apoptosis. *Oncogene*. 2003;22(56):8999–9006. doi: [10.1038/sj.onc.1207111](https://doi.org/10.1038/sj.onc.1207111)
- [38] Li Y, Liu Y, Fu Y, et al. The triggering of apoptosis in macrophages by pristine graphene through the MAPK and TGF-beta signaling pathways. *Biomaterials*. 2012;33(2):402–411. doi: [10.1016/j.biomaterials.2011.09.091](https://doi.org/10.1016/j.biomaterials.2011.09.091)
- [39] Kehl T, Backes C, Kern F, et al. About miRNAs, miRNA seeds, target genes and target pathways. *Oncotarget*. 2017;8(63):107167–107175. doi: [10.18632/oncotarget.22363](https://doi.org/10.18632/oncotarget.22363)
- [40] Juzenas S, Venkatesh G, Hübenthal M, et al. A comprehensive, cell specific microRNA catalogue of human peripheral blood. *Nucleic Acids Res*. 2017;45(16):9290–9301. doi: [10.1093/nar/gkx706](https://doi.org/10.1093/nar/gkx706)

- [41] Davis J, Salomonis N, Ghearing N, et al. MBNL1-mediated regulation of differentiation RNAs promotes myofibroblast transformation and the fibrotic response. *Nat Commun.* **2015**;6(1):10084. doi: [10.1038/ncomms10084](https://doi.org/10.1038/ncomms10084)
- [42] Cheng N, Liu C, Li Y, et al. MicroRNA-223-3p promotes skeletal muscle regeneration by regulating inflammation in mice. *J Biol Chem.* **2020**;295(30):10212–10223. doi: [10.1074/jbc.RA119.012263](https://doi.org/10.1074/jbc.RA119.012263)
- [43] Qin D, Wang X, Li Y, et al. MicroRNA-223-5p and -3p cooperatively suppress necroptosis in Ischemic/Reperfused hearts *. *J Biol Chem.* **2016**;291(38):20247–20259. doi: [10.1074/jbc.M116.732735](https://doi.org/10.1074/jbc.M116.732735)
- [44] Wang Y, Tatakis DN. Integrative mRNA/miRNA expression analysis in healing human gingiva. *J Periodontol.* **2021**;92(6):863–874. doi: [10.1002/JPER.20-0397](https://doi.org/10.1002/JPER.20-0397)
- [45] Wang B, Wu W, Xu K, et al. MicroRNA-223-3p is involved in fracture healing by regulating fibroblast growth factor receptor 2. *Bioengineered.* **2021**;12(2):12040–12048. doi: [10.1080/21655979.2021.2002498](https://doi.org/10.1080/21655979.2021.2002498)
- [46] Bugg D, Bailey LRJ, Bretherton RC, et al. MBNL1 drives dynamic transitions between fibroblasts and myofibroblasts in cardiac wound healing. *Cell Stem Cell.* **2022**;29(3):419–433 e410. doi: [10.1016/j.stem.2022.01.012](https://doi.org/10.1016/j.stem.2022.01.012)
- [47] Wagner V, Kern F, Hahn O, et al. Characterizing expression changes in noncoding RNAs during aging and heterochronic parabiosis across mouse tissues. *Nat Biotechnol.* **2023**;42(1):109–118. doi: [10.1038/s41587-023-01751-6](https://doi.org/10.1038/s41587-023-01751-6)
- [48] Kern F, Kuhn T, Ludwig N, et al. Ageing-associated small RNA cargo of extracellular vesicles. *RNA Biol.* **2023**;20(1):482–494. doi: [10.1080/15476286.2023.2234713](https://doi.org/10.1080/15476286.2023.2234713)
- [49] Hübner SM, Fehlmann T, Werno C, et al. Single-cell microRNA sequencing method comparison and application to cell lines and circulating lung tumor cells. *Nat Commun.* **2021**;12(1):4316. doi: [10.1038/s41467-021-24611-w](https://doi.org/10.1038/s41467-021-24611-w)
- [50] Scheuermann S, Hübner S, Engel A, et al. A novel approach to generate enzyme-free single cell suspensions from archived tissues for miRNA sequencing. *SLAS Technol.* **2024**;100133:100133. doi: [10.1016/j.slats.2024.100133](https://doi.org/10.1016/j.slats.2024.100133)
- [51] Reuter S, von Woedtke T, Weltmann K-D. The kINPen—a review on physics and chemistry of the atmospheric pressure plasma jet and its applications. *J Phys D Appl Phys.* **2018**;51(23):233001. doi: [10.1088/1361-6463/aab3ad](https://doi.org/10.1088/1361-6463/aab3ad)
- [52] Fehlmann T, Kern F, Laham O, et al. miRMaster 2.0: multi-species non-coding RNA sequencing analyses at scale. *Nucleic Acids Res.* **2021**;49(W1):W397–W408. doi: [10.1093/nar/gkab268](https://doi.org/10.1093/nar/gkab268)
- [53] Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* **2016**;32(18):2847–2849. doi: [10.1093/bioinformatics/btw313](https://doi.org/10.1093/bioinformatics/btw313)
- [54] Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience.* **2020**;9(12). doi: [10.1093/gigascience/giaa151](https://doi.org/10.1093/gigascience/giaa151)
- [55] Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell.* **2021**;184(13):3573–3587 e3529. doi: [10.1016/j.cell.2021.04.048](https://doi.org/10.1016/j.cell.2021.04.048)
- [56] McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **2019**;8(4):329–337 e324. doi: [10.1016/j.cels.2019.03.003](https://doi.org/10.1016/j.cels.2019.03.003)
- [57] Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**;30(1):207–210. doi: [10.1093/nar/30.1.207](https://doi.org/10.1093/nar/30.1.207)
- [58] Finak G, McDavid A, Yajima M, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **2015**;16(1):278. doi: [10.1186/s13059-015-0844-5](https://doi.org/10.1186/s13059-015-0844-5)

3.3 PAVING THE WAY TO A NEURAL FATE – RNA SIGNATURES IN NAIVE AND TRANS-DIFFERENTIATING MESENCHYMAL STEM CELLS

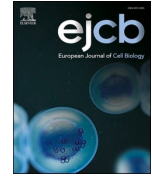
European Journal of Cell Biology 103 (2024) 151458



Contents lists available at ScienceDirect

European Journal of Cell Biology

journal homepage: www.elsevier.com/locate/ejcb



Paving the way to a neural fate – RNA signatures in naive and trans-differentiating mesenchymal stem cells

Caroline Diener^a, Konstantin Thüre^a, Annika Engel^{b,c}, Martin Hart^a, Andreas Keller^{b,c}, Eckart Meese^a, Ulrike Fischer^{a,*}

^a Saarland University (USAAR), Institute of Human Genetics, Homburg 66421, Germany

^b Saarland University (USAAR), Chair for Clinical Bioinformatics, Saarbrücken 66123, Germany

^c Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Center for Infection Research (HZI), Saarland University Campus, Saarbrücken 66123, Germany

ARTICLE INFO

Keywords:

Mesenchymal stem cells (MSCs)
Trans-differentiation
Neuron
Time-course
Transcriptome
miRnome

ABSTRACT

Mesenchymal Stem Cells (MSCs) derived from the embryonic mesoderm persist as a viable source of multipotent cells in adults and have a crucial role in tissue repair. One of the most promising aspects of MSCs is their ability to trans-differentiate into cell types outside of the mesodermal lineage, such as neurons. This characteristic positions MSCs as potential therapeutic tools for neurological disorders. However, the definition of a clear MSC signature is an ongoing topic of debate. Likewise, there is still a significant knowledge gap about functional alterations of MSCs during their transition to a neural fate. In this study, our focus is on the dynamic expression of RNA in MSCs as they undergo trans-differentiation compared to undifferentiated MSCs. To track and correlate changes in cellular signaling, we conducted high-throughput RNA expression profiling during the early time-course of human MSC neurogenic trans-differentiation. The expression of synapse maturation markers, including *NLGN2* and *NPTX1*, increased during the first 24 h. The expression of neuron differentiation markers, such as *GAP43* strongly increased during 48 h of trans-differentiation. Neural stem cell marker *NES* and neuron differentiation marker, including *TUBB3* and *ENO1*, were highly expressed in mesenchymal stem cells and remained so during trans-differentiation. Pathways analyses revealed early changes in MSCs signaling that can be linked to the acquisition of neuronal features. Furthermore, we identified microRNAs (miRNAs) as potential drivers of the cellular trans-differentiation process. We also determined potential risk factors related to the neural trans-differentiation process. These factors include the persistence of stemness features and the expression of factors involved in neurofunctional abnormalities and tumorigenic processes. In conclusion, our findings contribute valuable insights into the intricate landscape of MSCs during neural trans-differentiation. These insights can pave the way for the development of safer treatments of neurological disorders.

1. Introduction

Mesenchymal Stem Cells (MSCs) serve as a valuable and sustainable source of multipotent cells involved in the regeneration of mesodermal tissues (Hoang et al., 2022; Rohban and Pieber, 2017; Vasanathan et al., 2020). They differentiate into mesodermal cell types, such as osteoblasts, chondrocytes, and adipocytes (Hmadcha et al., 2020). Moreover, MSCs possess the ability to trans-differentiate into various cell types

from outside the mesodermal lineage, such as astrocytes, oligodendrocytes, and neurons. This trans-differentiation process can be induced by *in vitro* stimulation with specific differentiation media (Hermann et al., 2004; Hernandez et al., 2020; Kopen et al., 1999).

The ability of MSCs to develop into specific cell types combined with their ethical advantages over embryonic stem cells highlights their immense potential for therapeutic applications in personalized medicine (Feier et al., 2022; Harris, 2014; Hoang et al., 2022; Volarevic et al.,

Abbreviations: HBMSCs, human bone marrow-derived Mesenchymal Stem Cells; CCL2, C-C motif chemokine ligand 2; CNS, Central Nervous System; EM, extracellular matrix; FC, Fold Change; FDR, False Discovery Rate; GO, Gene Ontology; ISCT, International Society for Cellular Therapy; MiRNA, microRNA; HPO, Human Phenotype Ontology; NES, nestin; NPTX1, neuronal pentraxin 1; NSC, neural stem cell; RRNA, ribosomal RNA; TRNA, transfer RNA; R_q, Relative quantification; TUBB3, β-tubulin III; UC-MSC, umbilical cord-derived Mesenchymal Stem Cell.

* Corresponding author.

E-mail address: ulrike.fischer@uks.eu (U. Fischer).

<https://doi.org/10.1016/j.ejcb.2024.151458>

Received 23 April 2024; Received in revised form 18 September 2024; Accepted 21 September 2024

Available online 26 September 2024

0171-9335/© 2024 The Author(s).

Published by Elsevier GmbH. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

2018). The neural trans-differentiation capacity of MSCs has attracted significant interest for therapeutic interventions in neurological injuries and neurodegenerative diseases such as Parkinson's and Alzheimer's disease (Hernandez et al., 2020; Rahbaran et al., 2022). Ongoing clinical trials explore replacing damaged neuronal cells with trans-differentiated MSCs derived from the adult bone marrow (Choudhary et al., 2021; Hernandez et al., 2020; Hoang et al., 2022). To ensure safe use in clinical applications, it is crucial to have a comprehensive understanding on the processes that underlie MSC neural trans-differentiation (Lee et al., 2021; Lukomska et al., 2019; Musial-Wysocka et al., 2019). A thorough characterization could lead to more effective cell therapy approaches (Pittenger et al., 2019). However, the changes in cellular signaling that guide MSCs to their trans-differentiated state are still not fully understood (Choudhary et al., 2021; Hernandez et al., 2020; Lee et al., 2021).

Heterogeneous results have been reported for the molecular characteristics that define the MSC identity (Musial-Wysocka et al., 2019; Pittenger et al., 2019; Wang et al., 2021). A growing body of evidence could help to improve these criteria at the starting-point of the trans-differentiation process (Pittenger et al., 2019). Though, time-course RNA expression data are particularly useful to track and correlate changes in cellular signaling (Bar-Joseph et al., 2012; Diener et al., 2023a, 2020), only a few efforts have been made to decipher the transcriptional restructuring that directs MSCs towards their neuronal fate (Cortes-Medina et al., 2019; Khan et al., 2020), resulting in a disjointed picture of the relevant functional connections. Likewise, limited information is provided on factors that entail long-term therapeutic risks, including the expression of genes that bear tumorigenic potentials (Hernandez et al., 2020; Musial-Wysocka et al., 2019). Neuron-like morphologies can be observed as early as 48 hours after the initiation of the trans-differentiation process (Cortes-Medina et al., 2019; Miao et al., 2017). Based on this, we focused on this early time-window and generated time-resolved RNA expression data during the trans-differentiation of human bone marrow-derived MSCs (hBMSCs). We found relevant changes in cellular signaling that can be linked to the acquisition of neuronal features. Our study also uncovers potential issues that should be acknowledged when assessing the utility of MSC-derived neurons for therapeutic interventions. The transcripts and microRNAs with prominent time-course changes are both provided in an atlas format.

2. Materials and methods

2.1. Cultivation of human bone marrow derived mesenchymal stem cells (hBMSCs)

Human BMSCs (C-12974) were obtained from PromoCell GmbH (Heidelberg, Germany). The cell samples were derived from femoral head samples of a 66- (lot# 451Z012.3) and a 72-year-old (lot# 475Z010.3) female Caucasian donor, respectively. To strengthen the significance of the results, the RNA analyses were carried out on the cells of one donor and the protein-staining validation on the cells of the other donor, as indicated below. Corresponding cell charges were approved and certified by the supplier with immune staining and flow cytometric analyses. The cells were seeded at a density of 100,000/25 cm² flask with Mesenchymal Stem cell Growth Medium (PromoCell) and expanded for 1–2 passages before induction of the trans-differentiation.

2.2. Induction of neural trans-differentiation and collection of time-course samples

For neural trans-differentiation the BMSCs (lot# 451Z012.3) were seeded on fibronectin coated flasks and were allowed to reach a cell density of 60–80 % confluence. Once the appropriate confluence was reached, trans-differentiation induction was started by addition of Stem Cell Neurogenic Differentiation Medium (C-28015, PromoCell). Prior to this differentiation induction, the 0 h sample was collected.

Cellular samples were collected from 25 cm² culture flasks at 0, 3, 6, 9, 12, 24 and 48 h after addition of Neurogenic Differentiation Medium. Four independent replicates of the time-course experiment were performed resulting in a total of n=28 cell samples that were subject to subsequent RNA expression analyses. For cellular staining analyses, differentiation experiments were repeated with an independent biological replicate of cells (lot# 475Z010.3) and examined for representative time-points.

In a control experiment BMSCs were seeded on flasks with MSC Growth Medium 2 (C-28009, PromoCell). Cellular samples were collected at 0, 3, 6, 24 and 48 h after reaching 60 % confluence.

2.3. Staining of neural protein markers and fluorescence microscopy

Trans-differentiation experiments were repeated, to examine protein expression. Representative time-points were selected based on the information from time-course RNA expression data. For immune fluorescence analysis, BMSC were cultivated on fibronectin coated glass slides and allowed to reach 60–80 % confluence. Cells on glass slides for 0 h samples were methanol fixed when they reached 60–80 % confluence. Cells on glass slides for 9 h, 24 h, 48 h, 3 days and 7 days samples were differentiated by addition of Stem Cell Neurogenic Differentiation Medium and methanol fixed at the indicated time-points.

Cells on glass slides were permeabilized using 0.1 % Triton X100 for 15 minutes, blocked with 1 % goat serum for 30 minutes and incubated with primary antibodies: rabbit anti-NP-I 1:100 (Abcam, ab300404); rabbit anti-MCP1 1:50 (Abcam, ab214819); mouse anti-Nestin 1:200 (Abcam, ab18102); rabbit anti- β -III-Tubulin 1:200 (Abcam, ab18207) for one hour. After three PBS washing steps, secondary antibodies were added (goat anti-rabbit-Alexa Fluor 594 to detect NP-I and MCP1; goat anti-mouse-Alexa Fluor 594 and goat anti-rabbit-Alexa Fluor 488 to detect Nestin and β -III-Tubulin) 1:500. Nuclei were counterstained with DAPI.

2.4. Total RNA extraction, quantification and quality control

Cellular total RNA was extracted following the manufacturers' instructions from miRNeasy Mini Kit (Qiagen, Hilden, Germany). Concentrations of the isolated samples were determined by NanoDropTM 2000c Spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). RNA integrity was verified using an Agilent 2100 Bioanalyzer instrument with the RNA Nano Kit (Agilent Technologies, Santa Clara, CA, USA).

2.5. Analysis of time-resolved RNA expression profiles

miRNA and transcriptome profiles were determined based on the same total RNA samples. High-throughput miRNA and transcriptome expression analyses were conducted as detailed in a previous publication (Diener et al., 2020) using microarray systems from Agilent Technologies (miRNA: Complete Labeling and Hyb Kit with Human SurePrint G3 Unrestricted miRNA arrays (Release 21.0, G4872A); transcriptome: Low Input Quick Amp Labeling Kit and the Gene Expression Hybridization Kit with Human SurePrint G3 Gene Expression Microarrays (V3, G4851C)). For transcriptome analyses, cRNA was purified using RNeasy Mini Kit (Qiagen) and cRNA concentrations were determined with a NanoDropTM 2000c Spectrophotometer (Thermo Fisher Scientific Inc.).

2.6. Validation of time-resolved RNA expression patterns by quantitative reverse transcription polymerase chain reactions (RT-qPCRs)

The total RNA (500 ng) was reverse transcribed to cDNA using the QuantiTect RT Kit, (Qiagen GmbH, Hilden, Germany). For the qPCR analyses, 5 ng of the resulting cDNAs were used together with the SYBR[®] Green PCR Kit (Qiagen) and with Qiagen QuantiTect assays: *NES* (QT00235781), *NPTX1* (QT00083846), *NLGN2* (QT00007189), *TUBB3*

(QT00083713), *GAP43* (QT00023639), *ITGA5* (QT00080871), *PRSS35* (QT00204113), *ANLN* (QT00011585), *MMP13* (QT00001764). The qPCRs were run on a QuantStudio3™ Real-Time PCR System (Applied Biosystems™, Foster City, USA). Relative expression levels of the analyzed genes were determined with reference to *GAPDH* (Primer Assay: QT00079247) that served as endogenous control. Data evaluation was conducted using the Applied Biosystems™ Analysis Software, Relative Quantification Analysis Module, VERSION 4.3 (Thermo Fisher Scientific Inc.). Due to insufficient sample quantities, only three of the original four time-course replicates were included in the RT-qPCRs analyses. All analyses were conducted in duplicate technical replicates from each of the three biological replicates resulting in six analyses per tested gene and time-point.

In addition to RT-qPCR analysis of differentiated MSCs we investigated BMSCs with RT-qPCR during growth in MSC Maintenance Medium as described above.

2.7. Quantitative evaluation of miRNA expression data

For a quantitative evaluation of the miRNA expression data, a previously determined calibration curve (Diener et al., 2020) was applied to the accordingly processed data of the employed microarray analysis system. Using Avogadro's constant and assuming a total amount of 20 pg total RNA per cell (Monaco et al., 2012; Tang et al., 2019), absolute miRNA expression values [molecules/cell] were extrapolated.

2.8. Microarray processing and statistical comparisons

Microarray image analysis was conducted using Feature Extraction software (Agilent Technologies, Santa Clara, CA, USA). Absence of gene expression was specified by no detection throughout the microarray image analysis and excluded from further analysis. The raw microarray data were concatenated into one matrix using Python in version 3.7 with the package “NumPy” in version 1.16.4. Further, the data were background corrected, quantile normalized and log₂-transformed with R in version 3.5.1 with the packages “data.table” (v 1.12.0) and “Bioconductor-preprocesscore” (v 1.46.0). The processing was applied for the transcriptome and miRnome data, respectively. To determine medians, fold changes and t-tests, while obtaining the DEG analysis, R in version 4.0.3 was used with the package “data.table” in version 1.14.2. P-values were adjusted using the Benjamini-Hochberg method. An adjusted p-value of ≤ 0.05 was considered statistically significant. As for the pair-wise comparisons between the four replicated time-course experiments, correlation coefficients (Pearson) were determined by using GraphPad Prism software in version 10.1.1 (Graphpad Software, Inc.). Unless otherwise stated, high expression levels were defined by a log₂ expression >7.

2.9. Clustering analyses

For the trajectory analysis of the transcriptomic data the expression values were standardized via z-scores by transcript (mRNA). The clustering of the transcripts by their time trajectory was performed using the fuzzy c-means algorithm. The optimal number of clusters was selected by investigating the minimum centroid distance measure. The miRNA trajectory analysis was accomplished analogously. R in version 4.1.2 was used with the packages data.table (v 1.14.2) and Mfuzz (v 2.54.0), the last package was used for z-scoring and clustering.

2.10. Integration of known miRNA-target interactions

Lists of known miRNA-target relations were obtained from the miRTargetLink 2.0 online tool (Kern et al., 2021a) performing a search on the designated miRNAs under the criterion of a “strong validation” typus for the evidenced target interaction. Results were compared with the list of prominently altered mRNA transcripts.

2.11. Determination of functional interactions and pathway enrichment analyses

Functional interactions were analyzed by and exported from STRING database (version 11.5; <https://string-db.org/>) (Szklarczyk et al., 2021), excluding text-mining from the selection of active interaction sources and hiding the disconnected nodes. Pathway enrichment analyses were conducted by the embedded feature of corresponding STRING interaction networks (“Functional enrichments in your network”). In designated cases, subnetworks were represented for the enrichment of certain cellular processes.

3. Results

3.1. Experimental setup and quality control

Towards the establishment of neuron-like phenotypes, the cellular processes during the first two days of MSC trans-differentiation are yet to be deciphered. We conducted a detailed time-course RNA expression profiling (Fig. 1A) during the early trans-differentiation of hBMSCs. Cells were grown on fibronectin coated flasks and time-course samples were collected at 0, 3, 6, 9, 12, 24 and 48 h after addition of a standard commercial neurogenic differentiation medium. High quality of the RNA time-course samples was confirmed by RNA integrity numbers (RIN) ranging from 9.3 to 10.0. High-throughput RNA expression analyses were conducted for both the transcriptome and the miRnome.

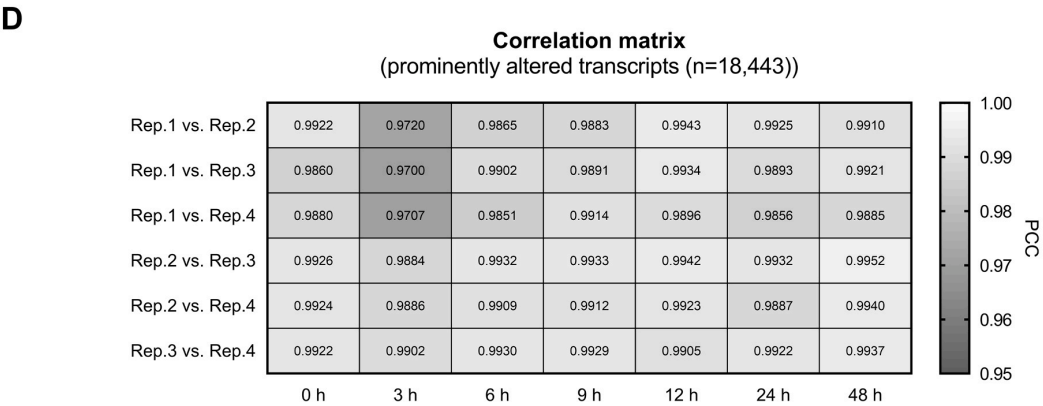
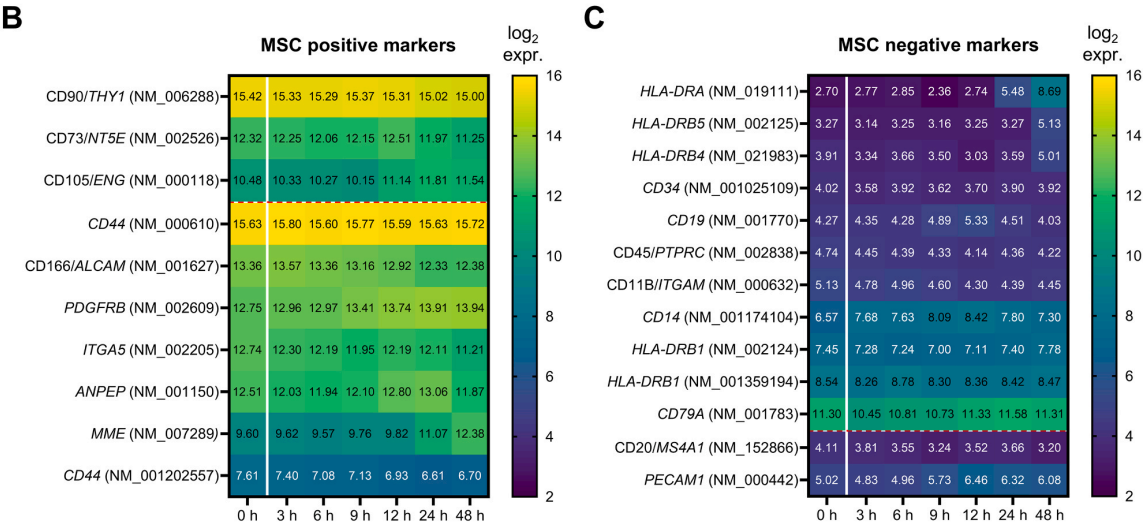
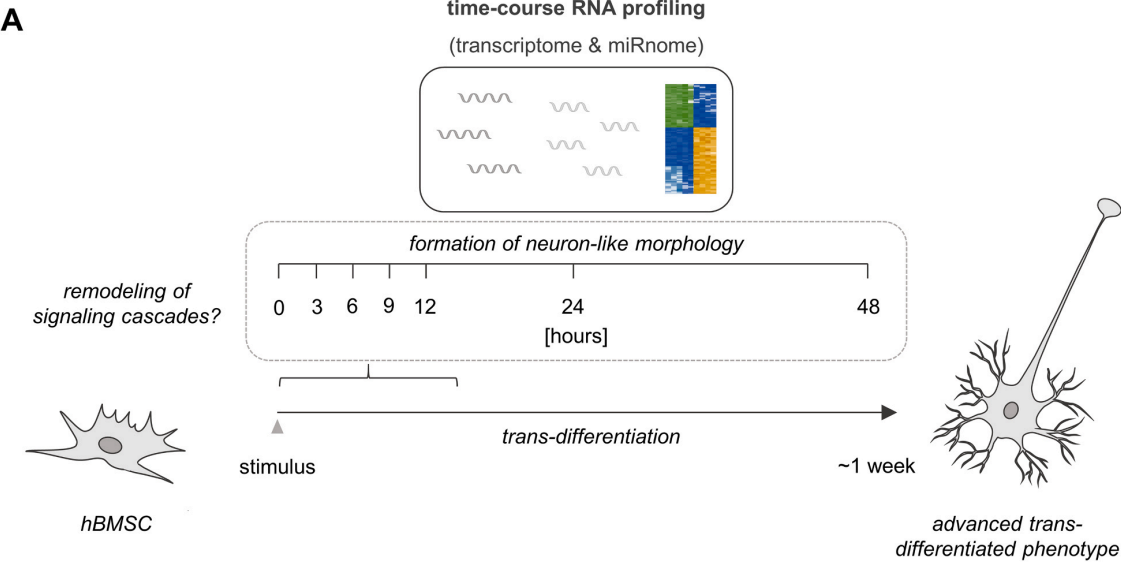
3.2. Transcriptome based characterization of the MSC identity

To contribute to a better definition of MSC signatures, we examined the expression of conventional MSC markers and of recently reported gene signatures in the analysis of our transcriptomics data (Fig. 1B and C). We first focused on the time-point before stimulation of the trans-differentiation process i.e., the 0 h time-point of our time-course analyses.

Our data confirmed high mRNA expression levels of the conventional MSC positive markers, CD90/*THY1*, CD73/*NT5E*, and CD105/*ENG*, as defined by the International Society for Cellular Therapy (ISCT) (Dominici et al., 2006). Corresponding median log₂ expression levels at the 0 h time-point ranged between 10.48 and 15.42. Antibody staining had confirmed these markers on the surfaces of 91 % of the cells (data provided by the PromoCell GmbH, Heidelberg, Germany).

Based on the definition by the ISCT, absence of CD11B or CD14, CD19 or CD79A, CD34, CD45, and HLA-DR surface molecules is also characteristic for the MSC identity (Dominici et al., 2006). Accordingly, we found comparatively low RNA expression levels of CD11B/*ITGAM* (median log₂ expression: 5.13), *CD14* (median log₂ expression: 6.57), *CD19* (median log₂ expression: 4.27), *CD34* (median log₂ expression: 4.02) and *CD45/PTPRC* (median log₂ expression: 4.74). We found high mRNA expression for *CD79A* with a median log₂ expression of 11.30. We detected comparatively low expression levels of most *HLA-DR* subtypes. The detected *HLA-DRA*, *HLA-DRB4*, *HLA-DRB5* transcripts showed a median log₂ expression of 4.02, 3.91 and 3.27, respectively. *HLA-DRB3* was not detected by our RNA expression analyses. Two mRNA isoforms of *HLA-DRB1* (NM_002124, NM_001359194), however, were found with high mRNA expression levels and median log₂ expressions of 7.45 and 8.54, respectively. As for the negative MSC markers CD14, CD34, CD45 CD19 and HLA-DR, protein expression was only found on the surfaces of 2–8 % of the cells (data provided by PromoCell).

Extending our MSC marker analysis to other recently suggested signature genes (Rohart et al., 2016), we confirmed high mRNA expression of *ALCAM/CD166* (median log₂ expression: 13.36), *ANPEP* (median log₂ expression: 12.51), *CD44* (median log₂ expression NM_001202557: 7.61; NM_001202557: 15.63), *ITGA5* (median log₂ expression: 12.74), *MME* (median log₂ expression: 9.60) and *PDGFRB* (median log₂ expression: 12.75). We detected comparatively low mRNA



(caption on next page)

Fig. 1. Overview of the experimental design, evaluation of MSC identity markers and time-course correlation of prominently altered transcripts. A: Trans-differentiation was *in vitro* stimulated by neurogenic differentiation medium to human BMSCs that were grown on the surface of fibronectin coated flasks. Cell samples for subsequent high-throughput mRNA and miRNA expression analyses were collected at early time-points (0, 3, 6, 9, 12, 24 and 48 h) after induction of the trans-differentiation process. The experiment was repeated in four replicates, resulting in a total of 28 RNA time-course samples. B, C: The time-course RNA signatures of MSC positive (B) and negative markers (C). Median $\log_2$ time-course expression patterns are shown for established and recently introduced MSC identity markers, respectively. Separation between these groups is indicated by the red dashed horizontal line.

D: Expression data of 18,443 transcripts with prominent time-course alterations were compared between the four replicated (rep.) experiments. Resulting Pearson's correlation coefficients (PCCs) are summarized in a correlation matrix.

expression levels of the recently reported negative markers (Oguma et al., 2022; Rohart et al., 2016) CD20/MS4A1 and PECAM1 (median $\log_2$ expression: 5.02).

3.3. Overall time-course expression changes

Extending our evaluation to further time-points of the early trans-differentiation process, a change in mRNA expression was evident for most of the MSC marker genes. Compared to the 0 h time-point, declining and increasing expression patterns were detectable in both, the groups of positive and the negative markers. The overall largest expression changes were observed at the end of the time-course analysis, i.e., at the 48 h time-point.

Specifically, out of the positive markers (Fig. 1B), six transcripts showed a declining expression pattern towards the end of the time-course (CD73/NT5E, CD90/THY1, ANPEP, CD44 (NM_001202557), CD166/ALCAM, ITGA5), while three showed an increasing expression pattern (CD105/ENG, MME, PDGFRB) and one transcript remained rather constant (CD44 (NM_000610)). The largest time-course expression changes were observed for ITGA5 and MME. Compared to the beginning of the time-course, median $\log_2$ fold changes ($\log_2$ FCs) of the corresponding mRNA transcripts were -1.53 and 2.78 , respectively.

Out of the MSC negative markers (Fig. 1C), six transcripts showed an increasing expression pattern (CD14, HLA-DRA, HLA-DRB1 (NM_002124), HLA-DRB4, HLA-DRB5, PECAM1), four showed a declining expression pattern towards the end of the time-course (CD11B/ITGAM, CD19, CD45/PTPRC, CD20/MS4A1) and three remained rather constant as compared to the beginning (HLA-DRB1 (NM_001359194), CD34, CD79A). Here, the largest time-course expression changes were observed for CD20/MS4A1 and HLA-DRA with median $\log_2$ FCs of the corresponding mRNA transcripts at -0.92 and 5.99 , respectively.

3.4. Grouping of transcripts in the early trans-differentiation process

We next identified transcripts with the most prominent time-course alterations and evaluated their functional relations. For each of the 35,369 detected transcripts, the maximum and minimum expression levels throughout the time-course were determined based on the median results of four independent time-course experiments. Comparing the corresponding time-points, prominent expression changes were defined by a median fold change (FC) criterion of ≥ 1.5 and their statistical significance ($p \leq 0.05$) after False Discovery Rate (FDR) correction. We identified 18,443 transcripts with prominent time-course alterations, comprising 9,009 transcripts with an overall decreased expression and 9,434 transcripts with an overall increased expression over the 48-hour time-period.

Comparing the expression data of the transcripts with prominent alterations revealed a very high concordance between the four replicated time-course experiments (Fig. 1D), demonstrating a rather uniform progress of the trans-differentiation process. For most of the comparisons the correlation coefficients were above a level of 0.985. A slightly reduced correlation was only observed for replicate 1 at the 3 h time-point, likely indicating some technical issues as compared to the other time-course replicates. However, corresponding Pearson's correlation coefficients were still in a high range between 0.970 and 0.972. This consistency allowed grouping the prominently altered transcripts

into distinct clusters, each with a specific expression pattern over the time-course.

3.5. Cellular processes associated with transcript clusters

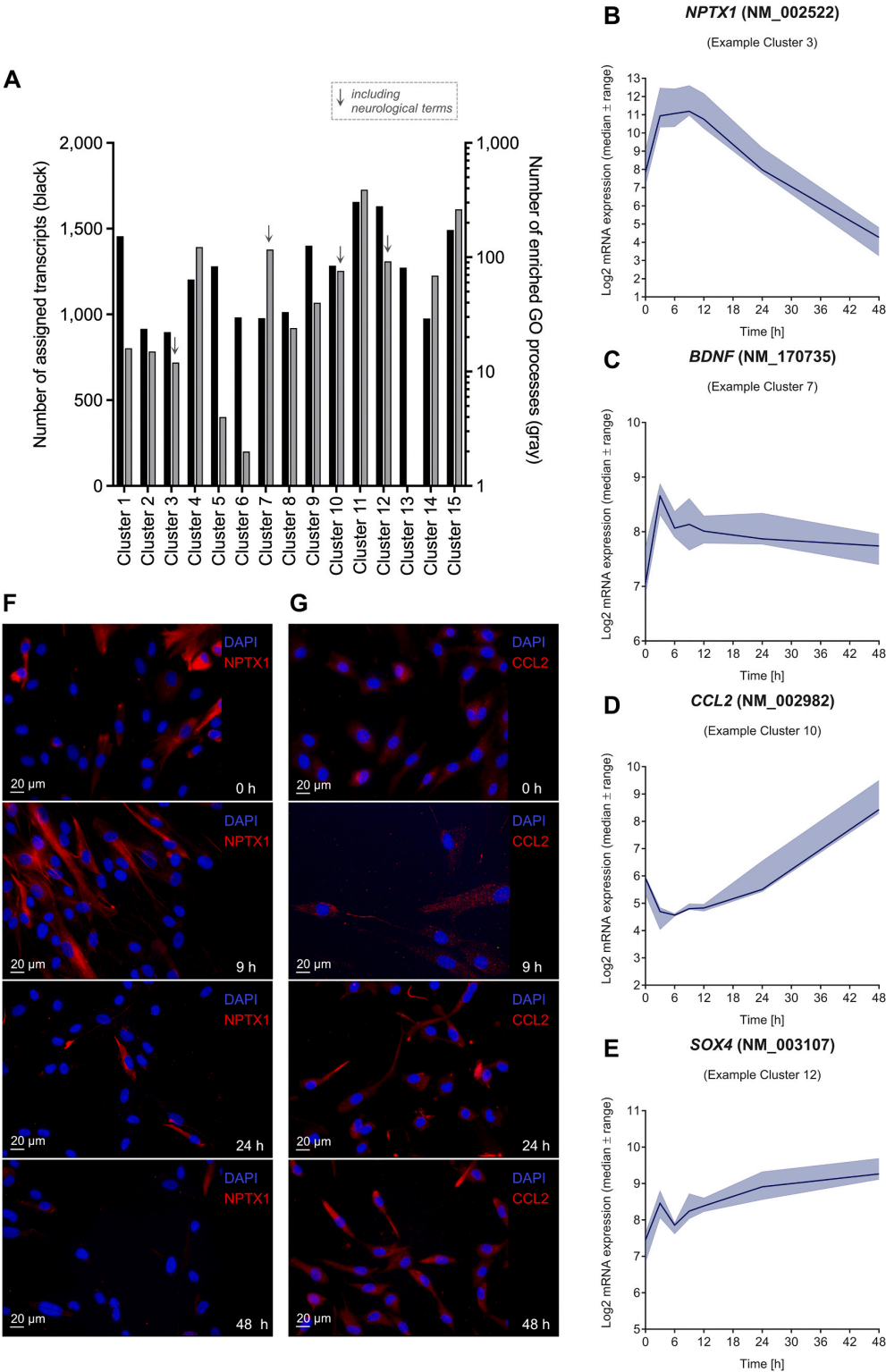
Categorization of the overall 18,443 prominently altered transcripts, identified 15 distinct clusters of expression trajectories (Supplementary Figure S1). Cluster 11 encompassed the highest number of transcripts, with a total of 1,656, while cluster 3 represented the smallest group with 897 transcripts (Fig. 2A left axis). To facilitate a comprehensive overview, the transcriptome and clustering data have been compiled in an atlas format and appended to the supplementary materials (Supplementary Table S1).

To unveil functional associations with cellular processes, enrichment analyses for each of the clusters were performed using the STRING 11.5 *in silico* tool (Szklarczyk et al., 2021). Considering "GO (Gene Ontology) biological processes" (Harris et al., 2004), the most significant terms were attributed to cluster 11 ($n=390$), whereas no significant relations to specific cellular processes were identified for Cluster 6. Supplementary table S2 summarizes the top 20 enriched GO processes for each cluster based on p-values (see also Supplementary table S3). Corresponding functional associations can be assigned to various head categories, including "cellular adhesion", "ribosome production and function", "transferRNA production and function", "mitosis and replication", "immune system and interferone signaling", "transcriptional activity", "metabolism", and "development and differentiation".

Upon searching for the terms "neuron" and "nervous" among the significantly enriched processes, specific functions were identified for four out of the 15 clusters, including the clusters 3, 7, 10, and 12 (Fig. 2A right axis). The expression trajectory of cluster 3, exemplified by the NPTX1 transcript, showed an early increase followed by a plateau-like phase and a sharp decrease for the remaining time-points (Fig. 2B). Cluster 7, represented by the BDNF transcript, exhibited an early increase followed by a very slight decrease at later time points (Fig. 2C). The expression trajectory of cluster 10, as represented by the CCL2 transcript, displayed an early decrease followed by a sharp increase for the remaining time points (Fig. 2D). Cluster 12, exemplified by the SOX4 transcript, represented a sharp increase, followed by a short decrease and a steady increase during the time course (Fig. 2E).

Besides the enriched GO processes, we identified enrichment of various "Monarch human Phenotype Ontology" (Monarch HPO) (Shefchek et al., 2020) terms denoting specific relations to neurological abnormality and Central Nervous System (CNS)-associated diseases (Supplementary Table S3). These terms were significantly enriched in clusters 12 and 15. Among the genes in cluster 12 that were associated with "Abnormality of the nervous system (HP:0000707)" (adj. $p=5.60 \times 10^{-3}$; $n=248$) were for example CYP1B1, MGP, and NTN1. Among the genes in cluster 15 that were associated with "Neurodevelopmental abnormality (HP:0012759)" (adj. $p=1.02 \times 10^{-2}$; $n=160$) were for example NEK2, BLM, and NALCN.

Next, we exemplarily show protein expression of the neuronal pentraxin 1 (NPTX1/NP-I) and C-C motif chemokine ligand 2 (CCL2/MCP1) during 48 h of trans-differentiation. For each time-point and each protein, expression was analyzed in 100 cells. Our cellular protein staining shows moderate fluorescence intensities of NPTX1 in about 25 % of the MSCs prior to the differentiation (Fig. 2F). After 9 h of trans-differentiation, the cells exhibited strong fluorescence intensities with



(caption on next page)

Fig. 2. Representation of prominently altered transcripts by 15 clusters of time-course expression trajectories. A: For the 18,443 prominently altered transcripts different shapes of expression trajectories are represented by 15 distinct clusters. The numbers of transcripts that have been assigned to each of the clusters are indicated by the black bars and the scale on the left axis. The numbers of significantly enriched Gene Ontology (GO) cellular processes per cluster are indicated by the gray bars and the scale on the right axis of the graph. Clusters that include functional connections to neurological processes are marked with arrows. B-E: Expression trajectories of the clusters that were enriched for neurological terms are exemplarily represented by *NPTX1*, *BDNF*, *CCL2* and *SOX4*. Log₂ expression data are shown as median results (line) with total ranges (filled areas) of the four replicated time-course experiments. F, G: Selected time-course expression patterns were confirmed by fluorescence microscopic analysis of neuronal pentraxin 1 (NPTX1) and C-C motif chemokine ligand 2 (CCL2) proteins. Characteristic time-points were chosen based on the corresponding time-course RNA expression data (see B, D). Representative fluorescence microscopic images were taken at a resolution of 40x. DAPI staining of cellular nuclei is depicted in blue and immune fluorescence staining of NPTX2 (F) or CCL2 (G) proteins is depicted in red, respectively.

a strong staining of cellular extensions. After 24 h and 48 h of trans-differentiation only about 8 % and 0.5 % of the cells, respectively, showed NPTX1 immune fluorescence. Protein expression analysis of CCL2 expression (Fig. 2G) showed weak fluorescence staining all cells at the 0 h time-point. After 9 h of trans-differentiation, almost 50 % of cells exhibited CCL2 immune fluorescence, albeit at low intensities. After 24 h and 48 h of trans-differentiation 75 % and 90 % of cells, respectively, showed enhanced CCL2 fluorescence staining. Overall, our immune fluorescence analyses are consistent with the expression changes observed by the RNA time-course analyses.

3.6. Cellular processes associated with most altered transcripts

In-depth examination of the top 100 transcripts with the most outstanding fold changes revealed high time-course correlations (Fig. 3A and B). For the specific comparisons, most correlation coefficients were above 0.94. Slightly reduced correlations were only observed for replicate 1 at the 3 h time-point, ranging between 0.858 and 0.890 for the top 100 decreasing and between 0.878 and 0.904 for the top 100 increasing transcripts, respectively.

An analysis of the top 100 decreasing transcripts revealed functional associations (Fig. 3C), such as with p53 signaling, including genes like *GTSE1*, *RRM2*, and *CCNE2*, and with cell cycle regulation, including *CDC45* and *MCM5*. Others of the most decreasing transcripts were linked to the maintenance of stemness, including *BIRC5*, *HMMR*, and *MYB*.

For the top 100 transcripts with increasing expressions (Fig. 3D), functional associations with prostaglandin signaling were evident, including genes such as *CHI3L1*, *PTGDS* and *PPARGC1A*. In addition, among the top transcripts with increasing expression were genes involved in complement signaling, including *C5AR1*, *CDKN1C* (p57) and *CIS/C1R* and genes involved in extracellular matrix remodeling, such as *MMP13*, *EFEMP1* and *FMOD*.

3.7. Time course expression of marker genes characteristic for the neural lineage differentiation and synapse maturation

To further depict the specificity of the MSC neurogenic trans-differentiation process, we evaluated the expression of characteristic marker genes for the neural lineage differentiation in context with our time-course experiments.

From the beginning of the time-course, high mRNA expression of the NSC marker nestin (*NES*) (median log₂ expression at 0 h: 10.78) indicated the neural differentiation capability of the analyzed MSCs (Minguell et al., 2005) (Fig. 4A). Likewise, high transcript levels of *MSI1* (median log₂ expression: 8.23) and *VIM* (median log₂ expression: 17.45) were observed at the 0 h time-point. Notably, the mRNAs of all the three NSC markers (Oikari et al., 2016a; Sun et al., 2008) showed rather constant decreases towards the end of the time-course, but still remained at high levels at the 48 h time-point (*NES*: 9.76; *MSI1*: 7.39; *VIM*: 17.18). For *NES* and *MSI1* these prominent decreases were statistically significant.

We next characterized the RNA expression of common neuron differentiation markers (Coccini et al., 2023; Fan et al., 2020; Jandial et al., 2008; McKenzie et al., 2018; Oikari et al., 2016a, 2016b; Shi et al., 2018; Sun et al., 2008) in context with our BMSC time-courses experiments

(Fig. 4B).

Amongst the neuron markers with the highest overall expression, *TUBB3* and *ENO1* showed distinct transcript expression already at the 0 h time-point, i.e. before stimulation of the trans-differentiation process (median log₂ expression *TUBB3* (NM_006086): 15.28; *ENO1* (NM_001428): 15.13). It is known that *TUBB3* is expressed in mesenchymal as well as neural lineages (Tondreau et al., 2004). With time progression a minor decline of the corresponding mRNAs was observed. Compared to the time-points of maximal and minimal expression, ten out of the overall 22 marker transcripts, displayed prominent time-course increases (*BTBD11* (NM_001018072), *DISP2* (NM_033510), *GAP43* (NM_002045), *MAPT* (NM_001123066), *MAPT* (NM_016835), *NEFM* (NM_005382), *RBFOX3* (NM_001350453), *SCN2A* (NM_021007), *SYT13* (NM_020826), *ZMAT4* (NM_024645)) and were, thus, listed in the atlas of Supplementary Table S1. Vice versa, six of the marker transcripts displayed prominent time-course decreases (*DLX2* (NM_004405), *ENO1* (NM_001428), *ENO2* (NM_001975), *GAD1* (NM_013445), *SERTM1* (NM_203451), *SYN1* (NM_006950)).

In addition to general neuron markers, we also characterized the expression patterns of common markers for synapse maturation (Cline, 2005; Coccini et al., 2023; Elias et al., 2008; Gomez de San Jose et al., 2022; Krueger et al., 2012; Kwon and Chapman, 2011; Pelkey et al., 2015; Varoqueaux et al., 2006) (Fig. 4C). Out of the 13 analyzed marker transcripts, six displayed a prominent increase in expression during the time-course (*DLG4* (NM_001365), *NLGN2* (NM_020795), *NLGN4X* (NM_001282145), *NPTXR* (NM_014293), *NRXN2* (NM_138732), *SYP* (NM_003179)). For example, *NLGN2* (NM_020795) continuously increased from a value of 9.60–11.93 during the initial 12 h of trans-differentiation. Afterwards the expression increase leveled off, reaching a maximum of 12.04 after 24 h, and ended with a final level of 11.71 (48 h). Another isoform of *NLGN2* (ENST00000575301) represented the transcript with the overall highest expression level and showed only minor changes with a short-term decrease from 12.92 to 12.14 at the 3 h time-point. Two synapse maturation markers showed prominent time-course decreases (*NPTX1* (NM_002522); *NPTX2* (NM_002523)). As described above and shown in Fig. 4C, the *NPTX1* transcript displayed a remarkable increase during the early 12 h of trans-differentiation (median log₂ expression *NPTX1* (NM_002522) 0 h: 7.85; 12 h: 11.19). With time progression, inversion of this trend towards an overall decline could be observed, resulting in a final median log₂ expression level of 4.27.

We exemplarily evaluated the time-course protein detection of nestin (*NES*) and β -tubulin III (*TUBB3*) in 100 cells (Fig. 4D–G and Supplementary Fig. S2 A–D). At the beginning of the time course (0 h) nestin expression was detected in 50 % of the cells and decreased remarkably until 48 h with only few cells with nestin expression at the ends of cell extensions. In contrast β -tubulin III expression at the beginning (0 h) was detectable in almost all cells. During the following 48 h β -tubulin III expression was detectable in 95 % of cells with strong expression in whole cell bodies and extensions. To allow for further assessment, we extended our observations of *NES* and *TUBB3* proteins to the advanced stages of the trans-differentiation (Fig. 4H and I and Supplementary Fig. S2E and F). Interestingly after 3 days cells with simultaneous nestin and β -tubulin III expression in dendrites were detected and after 7 days nestin expression increased predominantly at the ends of dendrites, while β -tubulin III expression decreased at the ends of those.

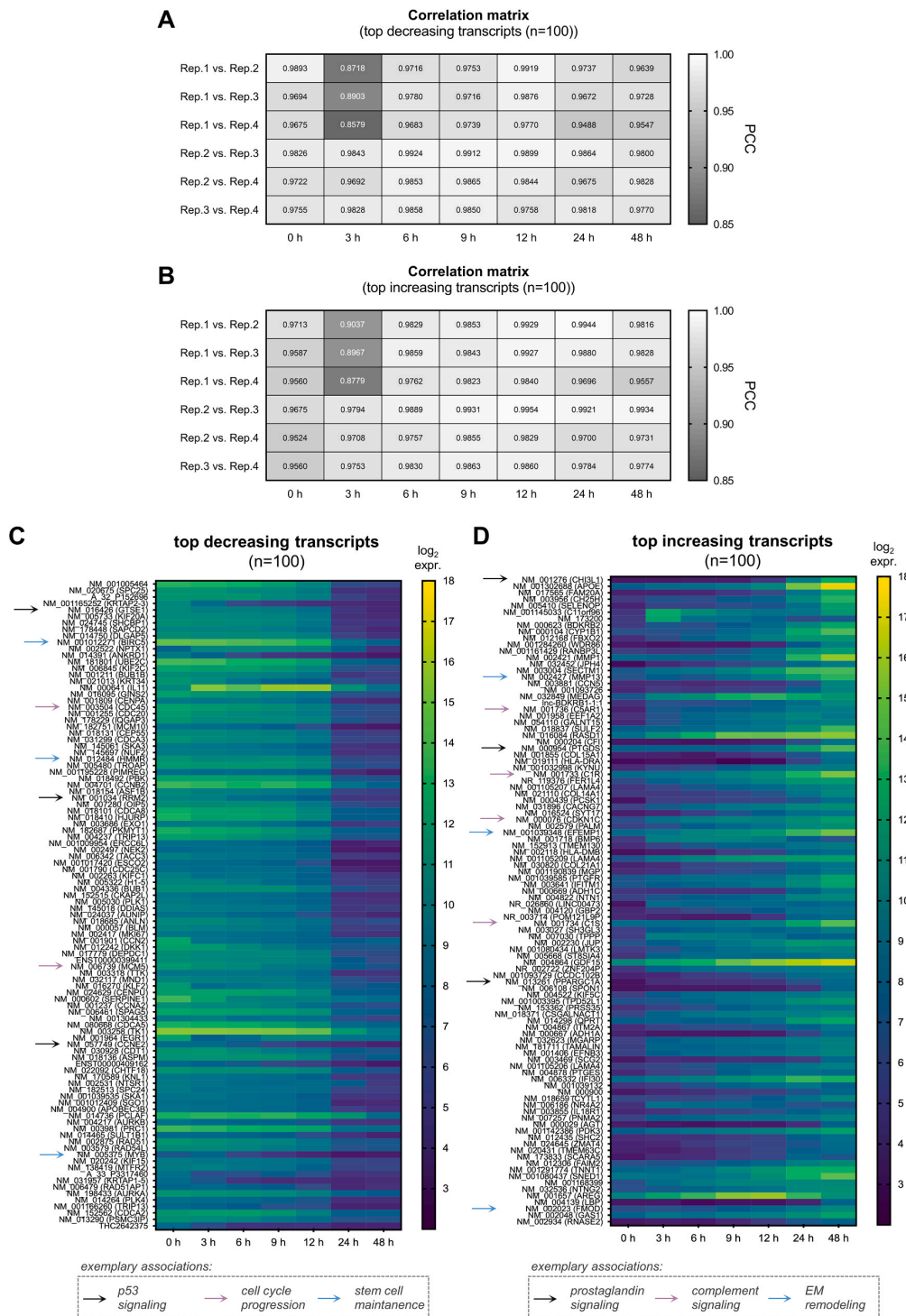


Fig. 3. Overview on the top hundreds of prominently altered RNA transcripts. A, B: Amongst the 18,443 prominently altered transcripts, the 100 with the highest fold changes were determined and separated according to decreasing (A) and increasing (B) expression levels. Corresponding expression data were compared between the four replicated (rep1.) experiments. Resulting PCCs are summarized in the correlation matrices. C, D: The top 100 transcripts are listed and represented with their median log₂ time-course expression. For the top 100 decreasing transcripts (C) associations with the p53 signaling, cell cycle progression and stem cell maintenance are marked by differently colored arrows as specified in the image. For the top 100 increasing transcripts (D) examples with functional associations to prostaglandin signaling, complement signaling and extracellular matrix remodeling are marked accordingly.

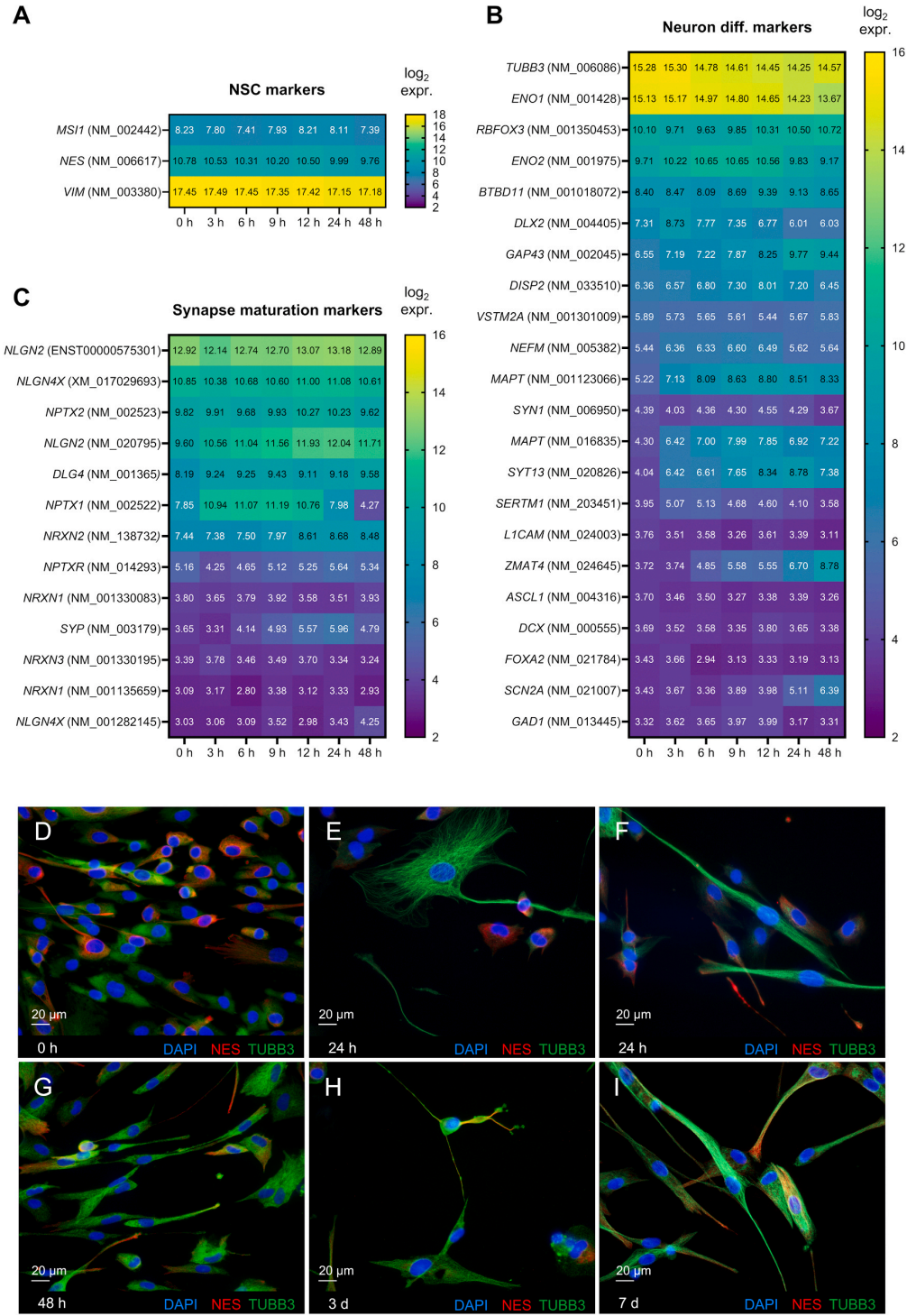


Fig. 4. Time-course evaluation of neural stem cell (NSC), neuron differentiation and synapse maturation markers. A-C: Time-course RNA expression data were evaluated for markers of common neural stem cells (NSC), neuron differentiation (diff.) and synapse maturation markers. Log₂ expression patterns of the corresponding transcripts are shown as median results of the four time-course experiments. D-I: Selected time-course expression patterns were confirmed by fluorescence microscopic analysis of nestin (NES) and class III β -tubulin (TUBB3) protein expression. Protein analyses were conducted at selected time-points during the early trans-differentiation process (0, 24 and 48 h; D-G) and after an extended period of 3 days (H) and 7 days (I), respectively. We show two exemplary pictures of the 24 h time-point to emphasize the early structural alteration and β -tubulin III expression within the cells. Representative fluorescence microscopic images were taken at a resolution of 40x. DAPI staining of cellular nuclei is depicted in blue. Immune fluorescence staining of NES and TUBB3 is depicted in red and green, respectively.

3.8. Time-course expression of astrocyte and oligodendrocyte markers

We further extended the evaluation of our hBMS time-courses data to various astrocyte (Fan et al., 2020; Jandial et al., 2008; McKenzie et al., 2018; Oikari et al., 2016b; Spurgat and Tang, 2022; Zhang et al., 2019) and oligodendrocyte differentiation markers (Huang et al., 2020; Jandial et al., 2008; Li et al., 2017; McKenzie et al., 2018; Oikari et al., 2016b; Zheng et al., 2018).

Nine out of the overall 18 astrocyte marker transcripts (Fig. 5A) displayed prominent time-course increases as specified above (*CLDN10* (NM_182848), *HES1* (NM_005524), *PRSS35* (NM_153362), *S100B* (NM_006272), *SLC1A2* (NM_004171), *STAT3* (NM_213662), *TPD52L1* (NM_001003395), *TPD52L1* (NM_001292026), *NDRG2* (NM_001354565)). A distinct time-course increase was observed for example for *PRSS35* (NM_153362). Starting at a median $\log_2$ expression value of 4.23, the expression quickly raised, but flattened towards the end of the time-course, resulting in final level of 9.65. Three of the astrocyte marker transcripts displayed prominent time-course decreases (*ALDH1L1* (NM_012190), *CD44* (NM_001202557), *ID1* (NM_002165)). For example, the *CD44* marker (NM_001202557), showed an initial median $\log_2$ expression of 7.61 (0 h) decreasing to a level of 6.70 (48 h). Another isoform of the *CD44* marker (NM_000610) represented the astrocyte marker with the overall highest expression level and showed a rather constant expression trajectory during the time-course, ranging between 15.59 and 15.80.

Amongst the oligodendrocyte differentiation markers (Fig. 5B), *PDGFRA* showed the overall highest expression starting with a median $\log_2$ expression of 11.56 (0 h) and displaying a minor increase to a level

of 14.61 (48 h). *PDGFRA* is involved in the growth and migration of mesenchymal cells as well as the signaling of oligodendrocyte progenitors (Funa and Sasahara, 2014). Six out of the overall 15 marker transcripts displayed prominent time-course increases as specified above (*CARNS1* (NM_001166222), *EGFR* (NM_201283), *PAIP2B* (NM_020459), *PDGFRA* (NM_006206), *QDPR* (NM_000320), *TMEM144* (NM_018342)). A remarkable increase in mRNA expression was observed, for example, for *TMEM144* (NM_018342), starting from a median $\log_2$ expression of 6.32 (0 h) and reaching a level of up to 7.94 (48 h). Four of the oligodendrocyte marker transcripts displayed prominent time-course decreases (*ANLN* (NM_018685), *CSPG4* (NM_001897), *EGFR* (NM_005228), *GALC* (ENST00000622264)). The *ANLN* (NM_018685) transcript, for example, showed a continuous mRNA decrease from a median $\log_2$ expression of 11.18 (0 h) to 6.12 (48 h) during the time-course.

3.9. Validation of time-resolved RNA expression patterns

Further validation of the microarray results was performed using RT-qPCR. *ITGA5* was chosen as a marker for MSC identity, *MMP13* as one of the top transcripts with increased expression, *NES* as a marker for NSCs, and *NLGN2* and *NPTX1* as markers for synapse maturation. Additionally, *TUBB3* and *GAP43* were analyzed as markers for neuronal differentiation, while the time-course mRNA expression patterns of *ANLN* and *PRSS35* were validated as markers for oligodendrocyte and astrocyte differentiation, respectively.

Re-analyzing our time-course RNA samples by an independent method, we found a striking similarity between the results of the RT-

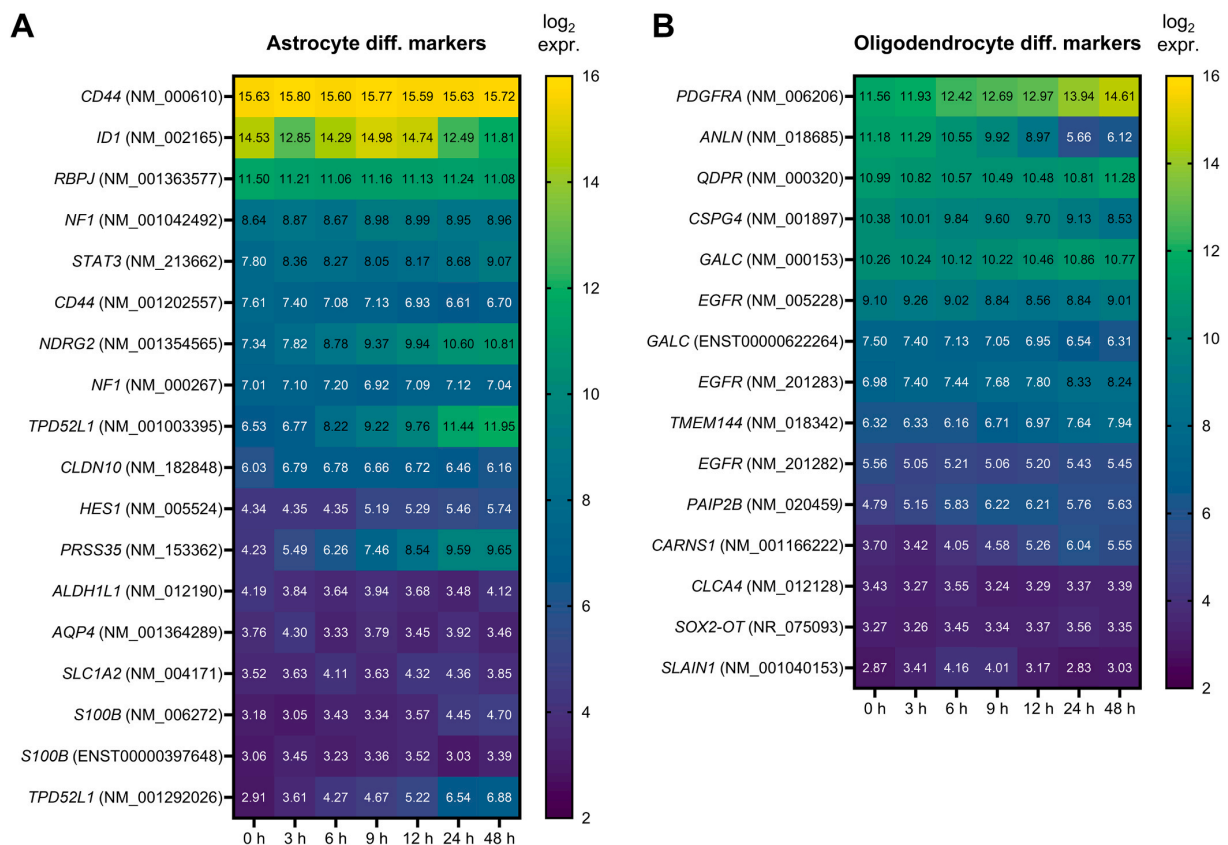


Fig. 5. Time-course evaluation of oligodendrocyte and astrocyte differentiation markers. Time-course RNA expression data were evaluated for common astrocyte (A) and oligodendrocyte (B) differentiation (diff.) markers. $\log_2$ expression patterns of the corresponding transcripts are shown as median results of the four time-course experiments.

qPCRs and the median results of the original microarray analyses. As shown in Fig. 6, the comparative analysis confirmed that both the ranges and shapes of the time-course patterns matched, confirming the validity of our microarray data.

As determined by RT-qPCR analyses, marker expression was different between trans-differentiated cells and control cells with mesenchymal maintenance medium as exemplified for expression of neuron differentiation marker *GAP43* and synapse maturation marker *NLGN2* and MSC marker *ITGA5* (Supplementary Fig. S3).

3.10. Time-course expression of microRNAs in the early trans-differentiation process

MicroRNAs (miRNAs) exhibit their function by inducing expression reduction of their specific target genes at the post-transcriptional level (Bartoszewski and Sikorski, 2018; Diener et al., 2023b; Yao, 2016). To investigate the potential impact of miRNA-mediated post-transcriptional regulation as part of the trans-differentiation process, we conducted a comprehensive analysis of the miRNome alongside the transcriptome at time-points 0, 3, 6, 9, 12, 24, and 48 h, following trans-differentiation induction.

We identified 91 prominently altered miRNAs exhibiting significant time-course expression changes with a fold change ≥ 1.5 . Among them, 46 demonstrated an overall decreased expression, while 45 displayed increased expression over the 48-hour period.

As for the mRNA transcripts we examined the dynamics of changes in the median $\log_2$ miRNA expression over the 48-hour timespan. High time-course consistency between the replicated experiments, comprising correlation coefficients of more than 0.960 (Supplementary Figure S4A), allowed grouping of the 91 prominently altered microRNAs into distinct clusters, each representing a specific expression pattern over the time-course. The various expression trajectories were categorized into five distinct clusters (Supplementary Fig. S4B). As for the transcriptome data, the miRNome expression and clustering data are provided in an atlas format in the supplementary materials (Supplementary Table S4).

Cluster 2 contained the highest number of miRNAs, with a total of 33, while clusters 4 and 5 were the smallest with 12 miRNAs each (Fig. 7A). Cluster 1, typified by a continuous decrease in expression following an initial latency of approximately 12 h, was exemplified by hsa-miR-15b-5p (Fig. 7B). Cluster 2, characterized by a continuous increase in expression after a certain latency period of 6 h, was exemplified by hsa-miR-27b-3p (Fig. 7C). Cluster 3, displaying a continuously decreasing expression trajectory, included several members of the miR-17 family, such as hsa-miR-18a-5p (Fig. 7D). Cluster 4 with a continuously increasing expression trajectory, encompassed various members of the let-7 miRNAs, including hsa-let-7c-5p (Fig. 7E). Cluster 5 was marked by an early decrease followed by a plateau-like phase at later time-points, featured 12 miRNAs including hsa-miR-503-5p (Fig. 7F).

To quantify the above-described expression changes, we utilized a calibration curve from a previous study (Diener et al., 2020), revealing that the most prominent alterations ranged within a magnitude of 10^2 – 10^3 molecules per cell during the 48-hour observation window. Notably, this range aligns well with findings in other cell types (Diener et al., 2020). Among the top miRNAs with the largest changes in molecular expression were for example hsa-let-7b-5p with +4,103 molecules/cell, hsa-miR-6089 with +2,311 molecules/cell, hsa-miR-34a-5p with +1,577 molecules/cell, hsa-miR-1260b with -2,303 molecules/cell, hsa-miR-221-3p with -520 molecules/cell and hsa-miR-15b-5p with -335 molecules/cell (Fig. 7G).

To gain insight into the regulatory roles of the prominently altered miRNAs, we searched for experimentally validated miRNA-target interactions using the *in-silico* tool miRTargetLink 2.0 (<https://ccb-compute.cs.uni-saarland.de/mirtargetlink2>) (Kern et al., 2021a). Since miRNAs often exert their regulatory effects by causing a decrease in the expression of their mRNA targets (Guo et al., 2010), we utilized the

time-course expression data to identify inversely expressed targets. We identified 203 miRNA-target interactions for the 45 miRNAs with increasing miRNA expression levels and 341 target interactions for the 46 miRNAs with decreasing expressions (see Supplementary Table S5). Subsequently, we constructed interaction networks for the corresponding proteins using the STRING 11.5 database (Szklarczyk et al., 2021) and highlighted sub-networks associated with the GO term "neuron differentiation" (Fig. 8A and B). Within the network for decreasing miRNAs, we identified central nodes such as *STAT3* and *VEGFA* mRNAs, both of which displayed increased expression during the observed time-course (Fig. 8A). The corresponding proteins have previously been linked to neural differentiation and MSC trans-differentiation (Snyder et al., 2011; Theis and Theiss, 2018; Wada et al., 2006). Within the networks for increasing miRNAs, we identified central nodes such as *MAPK1* and *MET*, both of which exhibited decreased expression over the time course (Fig. 8B). The expression of these appears to be restricted to specific stages of the neural differentiation process (Semprich et al., 2022; Zheng et al., 2013).

4. Discussion

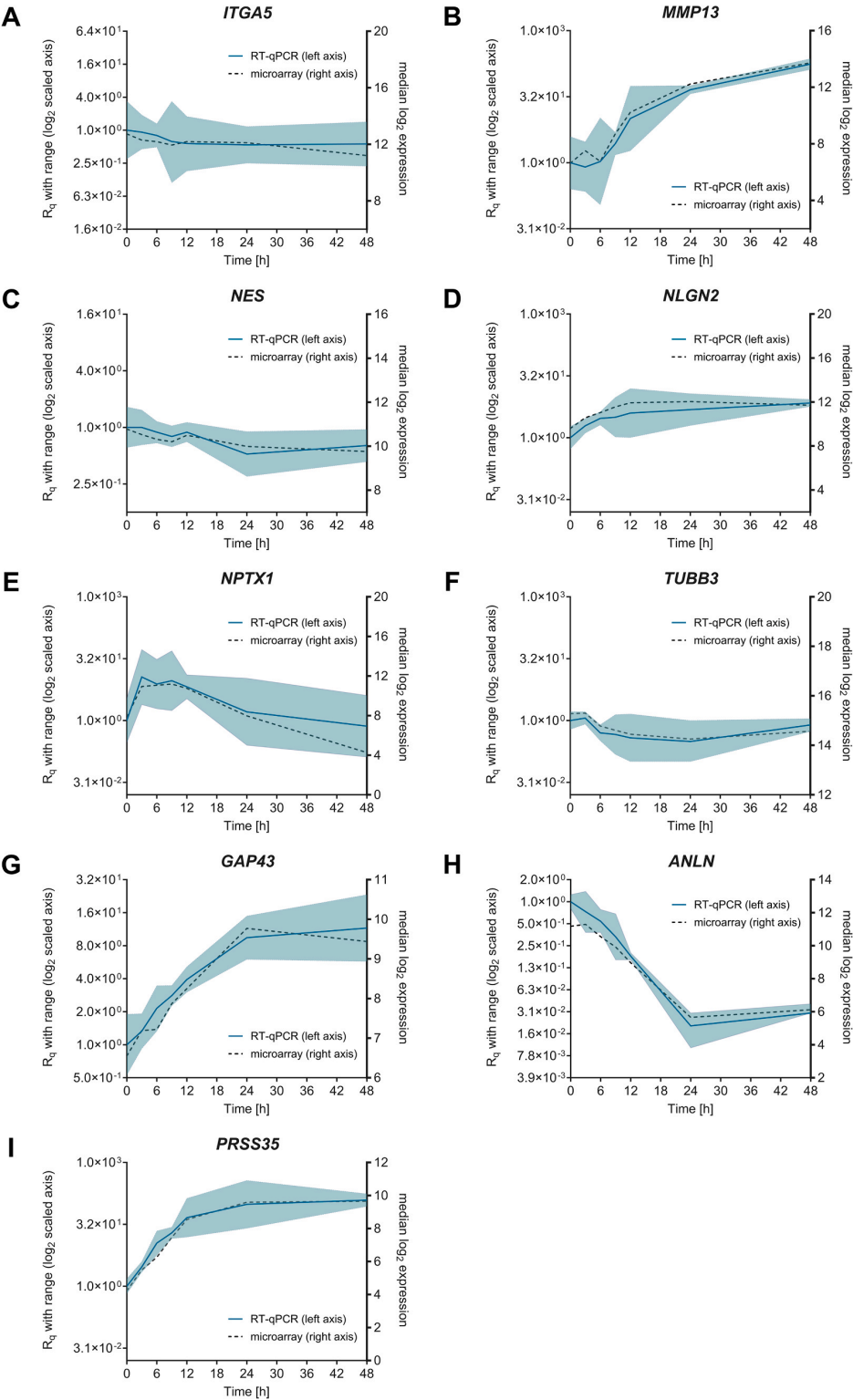
The ISCT defines MSC identity through the expression of specific surface markers (Dominici et al., 2006). However, the definition of clear signatures remains controversial due to heterogeneous results and is an ongoing process (Musial-Wysocka et al., 2019; Pittenger et al., 2019; Wang et al., 2021). Nonetheless, there is growing evidence that gene expression data can refine the criteria for a more precise definition of MSC identity (Pittenger et al., 2019). A major aim of our study was to contribute to this effort by better defining MSC signatures.

Most of the conventional and various recently introduced markers appear to efficiently indicate MSC identity. We found high RNA expression levels of various MSC positive markers including genes such as *CD44*, *THY1* (CD90), *NTSE* (CD73), *ENG* (CD105), *ALCAM* (CD166), *ITGA5*, *MME* and *PDGFRB*. Notably, *CD44* has also been considered as an intermediate marker for astrocyte differentiation from human pluripotent stem cells (Cai et al., 2012; Oikari et al., 2016b; Shaltouki et al., 2013). Low mRNA expression was detected for several MSC negative markers including *ITGAM* (CD11b), *CD14*, *CD19*, *MS4A1* (CD20), *CD34*, *PTPRC* (CD45), *PECAM1* and most *HLA-DR* isoforms. Consistent with our findings of an elevated expression of *HLA-DRB1*, an independent study also detected high *HLA-DRB1* mRNA expression in human BMSCs. The elevated expression was, however, not confirmed at the protein level (Oguma et al., 2022).

Our time-course transcriptomics data identified several markers like *ITGA5* or *HLA-DRA* that showed an altered expression shortly after the cellular stimulation. These markers appear to be well suited to distinguish between the original MSC identity and the emerging neurogenic trans-differentiation.

Our analysis of transcripts with significant time-course alterations aimed to decipher the transcriptional restructuring that drives MSCs towards a neuron-like phenotype (Cortes-Medina et al., 2019; Khan et al., 2020). Our data highlight functional connections to cellular processes, typically involved in neuronal differentiation.

Various transcripts that exhibit substantial decreases over the time-course, have been associated with p53 signaling and cell cycle progression. Exemplary genes, such as *GTSE1*, *RRM2* and *CCNE2*, are regarded as p53 negative regulators (Gorjala et al., 2016; He et al., 2017; Monte et al., 2003; Scolz et al., 2012). Examples involved in cell cycle regulation, particularly in S phase transition, included genes such as *CDC45* and *MCM5* (Chapouton et al., 2010; Huang et al., 2022; Quinn et al., 2001). Generally, an altered p53 activity and a delay in cell cycle have previously been linked to neurogenesis in the CNS (Farioli-Vecchioli and Tirone, 2015; Hardwick et al., 2015; Ruijtenberg and van den Heuvel, 2016; Xiong et al., 2020). Other transcripts displayed substantial increases and play critical roles in prostaglandin D2 (Augustyniak et al., 2017; Sakry et al., 2015; Zhou et al., 2015) and in



(caption on next page)

Fig. 6. Validation of time-resolved RNA expression patterns by RT-qPCRs. To confirm the microarray results, exemplary RT-qPCR were conducted to validate the mRNA expression patterns for nine representative genes: (A) *ITGA5* as a representative marker for the MSC identity, (B) *MMP13*, the transcript of which was amongst the top increasing mRNAs, (C) *NES* as NSC marker, (D, E) *NLGN2* and *NPTX1* as markers for synapse maturation, (F, G) *TUBB3* and *GAP43* as neuron differentiation markers, (H) *ANLN* as oligodendrocyte and (I) *PRSS35* as astrocyte differentiation marker. As the results of the RT-qPCR analyses, relative quantitation (R_q) of the corresponding mRNAs was conducted in relation to GAPDH housekeeping gene expression. R_q plots are shown on the left axis as median results (turquoise line) with the total range (turquoise shaded) of three replicated time-course experiments that were assayed in technical duplicates. They are illustrated in comparison to the median microarray results (black dotted line; right axis).

complement pathways (Coulthard et al., 2017; Furutachi et al., 2013; Walsh et al., 2017), both of which have been associated with neurogenic differentiation (Coulthard et al., 2017; Nango et al., 2020; Thomas et al., 2000). Notably, similar changes in both prostaglandin and complement signaling were also prevalent during the neurogenic differentiation of human induced pluripotent stem cells (iPSCs) (Augustyniak et al., 2017; Walsh et al., 2017), another type of stem cell with therapeutic potential that can be generated by the reprogramming of somatic cells (Thanaskody et al., 2022).

Alteration of additional processes with common links to neuronal development were identified as the result of our gene set enrichment analyses, including various head categories such as cellular adhesion (Migliorini et al., 2013), ribosomal and transfer RNA (rRNA, tRNA) functions (Fusco et al., 2021; Ramos and Fu, 2019), cell division (Farioli-Vecchioli and Tirone, 2015; McKinnon, 2013), immune signaling (Filiano et al., 2016; Morimoto and Nakajima, 2019), transcriptional activity (Hamby et al., 2008), and cellular metabolism (Fawal and Davy, 2018).

Additionally, there were expression changes in transcripts that have well-established roles in neural differentiation signaling. Examples include *NPTX1* that showed an early increase followed by a sharp decrease for the remaining time-points of our analysis. The specific expression pattern was also verified when staining the according cellular proteins and likely plays a crucial role in the early growth of synapse-like connections (Gomez de San Jose et al., 2022). Likewise, the expression of the transcription factor *SOX4* that showed an increasing expression in our time course analysis, is thought to be crucial for establishing neuronal identity (Bergsland et al., 2006). *CCL2* that also showed an increasing expression, both at the mRNA and protein levels, has former been demonstrated as a common regulator of neuronal functions (Hong et al., 2015). Moreover, the neurotrophic factor *BDNF* that showed altered expression pattern in our MSC trans-differentiation experiments has been associated with neurogenic differentiation. Evidence for a decisive role of BDNF stems from studies showing that BDNF supplementation enhances the efficiency of human neural precursor cell differentiation (Jiao et al., 2014) and that overexpression of the *BDNF* gene promotes the neurogenic trans-differentiation of rodent BMSCs (Liu et al., 2015).

Overall, the nature of the transcripts that are prominently altered in the first few hours following the induction of the trans-differentiation, clearly demonstrates a wide-ranging reorganization of cellular pathways. The according changes likely are the first essential steps towards for the acquisition of neuron-like features.

In addition to the evaluation of the transcriptomics, our data also imply that miRNAs act as potential drivers of the cellular trans-differentiation. As for the early dynamics of miRNAs during neurogenic MSC trans-differentiation, we found altered expression patterns of multiple miRNAs that have previously been associated with the differentiation of neurons. Specifically, hsa-let-7b-5p and hsa-miR-34a-5p that have been linked to neural stem cell differentiation and mature neuron signaling exhibit significant expression changes (Aranha et al., 2011; Kern et al., 2021b; Zhao et al., 2010). MiR-34a-5p has also been considered a potential driver of BMSC neurogenic trans-differentiation in rodent models (Liu et al., 2011). Hsa-miR-221-3p that showed a strong decreased expression during early BMSC trans-differentiation has been linked to the differentiation of neural crest cells (Greene and Tischler, 1976; Hamada et al., 2012). Its downregulation has also been shown to facilitate BMSC differentiation into osteoblasts (Gan et al.,

2020), suggesting an essential role in BMSC differentiation signaling.

Our clustering analysis identified further miRNAs that have previously been connected to neural development and differentiation. An increasing time-course trajectory was identified, for example, for hsa-miR-27b-3p that is an abundant neuronal miRNA, which negatively regulates pluripotency associated genes (Fuchs et al., 2014; Poon et al., 2016). Other examples include several members of the let-7 family that has been linked to the regulation of neural stem cell proliferation and differentiation (Roush and Slack, 2008; Zhao et al., 2010). The miR-17 family, which showed a decreasing time-course pattern in our experiments was previously found to have a reduced expression during progressive brain development (Mao et al., 2014; Xia et al., 2022). Hsa-miR-15b-5p, considered as a negative regulator of *BDNF*, and hsa-miR-503-5p, which likely inhibits neural lineages, also showed decreasing time-course expression patterns (Boone et al., 2017; He et al., 2021).

To date only a portion of the miRNAs that found their way into databases have been validated for their nature as true miRNAs (Alles et al., 2019; Diener et al., 2023b). Some candidates, particularly those with high miRNA numbers, have only been annotated based on recent RNA sequencing data and may not necessarily represent true miRNAs. However, this does not imply the absence of cellular functions for the corresponding molecules (Diener et al., 2024). Future analyses should explore the potential roles of poorly characterized miRNAs, like hsa-miR-1260b and hsa-miR-6089, which show significant expression changes in our time-course data. Nevertheless, our evaluation of experimentally validated miRNA-target interactions indicated multiple neuronal differentiation genes as potential targets of miRNAs in the context of neurogenic trans-differentiation. The finding of many shared miRNA-targets within the corresponding regulatory networks underscores the functional interaction of different miRNAs to enhance their regulatory efficiency (Bartoszewski and Sikorski, 2018; Diener et al., 2022, 2023b; Gebert and MacRae, 2019; Schmitz et al., 2014; Selbach et al., 2008).

MSCs' ability to trans-differentiate into neural cells makes them a promising source for treating neurological injuries and neurodegenerative diseases. However, given the critical nature of transplanting MSC-derived neurons into the CNS, safety assessments are crucial. Our time-course analyses suggest potential issues with the neural trans-differentiation process that warrant further investigation.

Prominent time-course alterations of several transcripts indicate associations with CNS abnormalities and diseases of the nervous system. Expressional increases of specific transcripts such as *CYP11B1*, *MGP* and *NTN1* (Alsubait et al., 2020; Mertsch et al., 2009; Ylivinkka et al., 2017) have previously been associated with the development of brain resident tumor entities. In addition, a reduced expression of *BLM* has been associated with genomic instability and a high predisposition to tumors (Cunniff et al., 2017; van Wietmarschen et al., 2018), while substantial reduction of *NALCN* mRNA has been associated with failures in neuronal excitability (Cochet-Bissuel et al., 2014). We also observed substantial increases in the expression of extracellular matrix (EM) modulators such as *MMP13*, *EFEMP1* and *FMOD* (Jan et al., 2016; Li et al., 2022; Livingston et al., 2020) during the early trans-differentiation process. While the peripheral secretion of growth factors and cytokines is likely advantageous for the therapeutic potential of MSCs (Han et al., 2022; Liang et al., 2021) the paracrine action of proteins including the EM modulators (Hu et al., 2012; Inoue et al., 2010; Sengupta et al., 2022) still raises concerns about their tumorigenic potential.

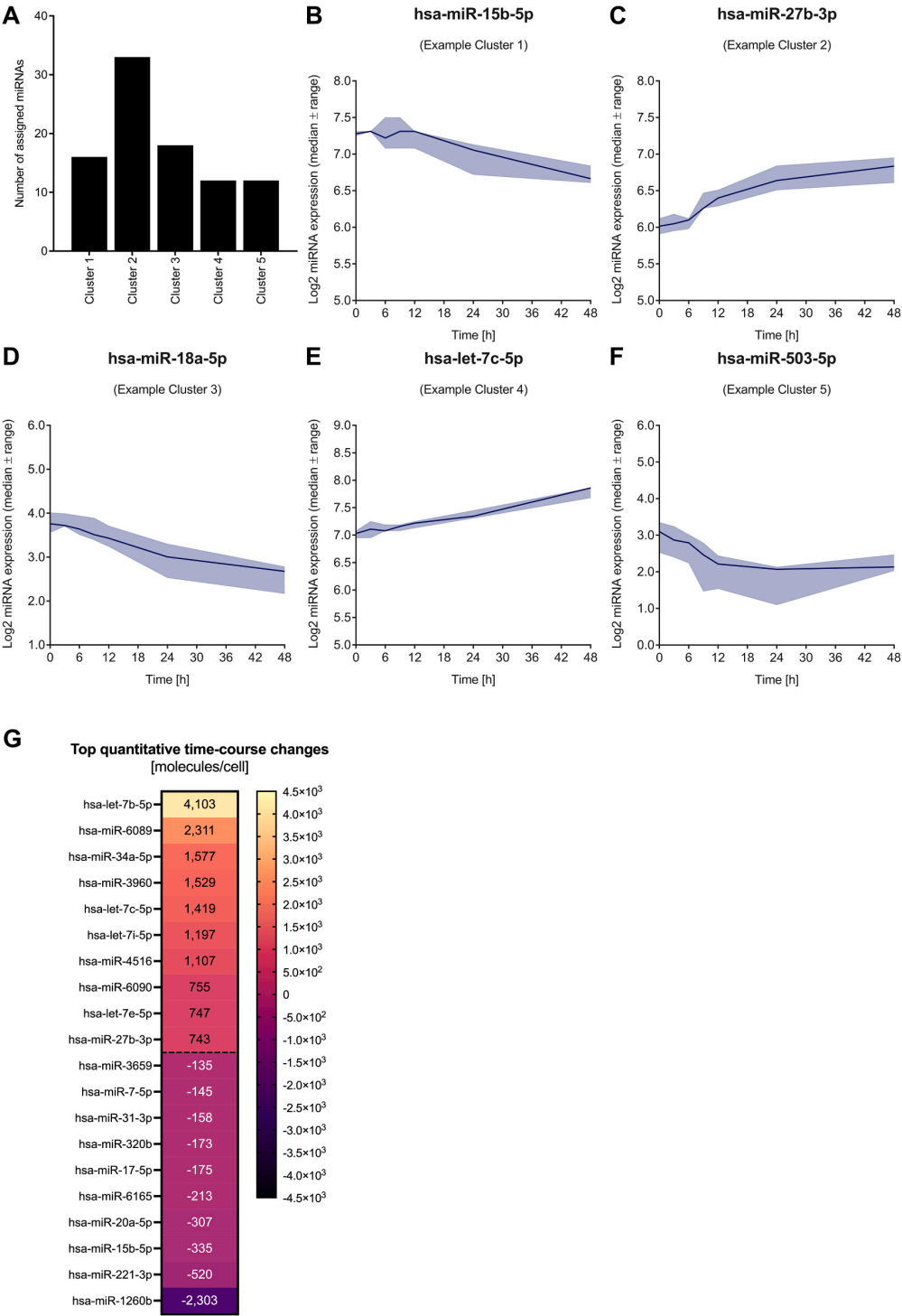


Fig. 7. Representation of prominently altered microRNAs by five clusters of time-course expression trajectories and quantitative overview on the highest miRNA expression changes. A: For 91 prominently altered microRNAs (miRNAs, miRs) the different shapes of time-course expression trajectories were represented by five distinct clusters. An overview on the number of miRNAs that have been assigned to each of the clusters is depicted. B-F: Representative expression trajectories are exemplarily illustrated by miR-15b-5p (B), miR-27b-3p (C), miR-18a-5p (D), let-7c-5p (E) and miR-503-5p (F), respectively. Log₂ expression data are shown as median results (line) with total ranges (filled areas) of the four replicated time-course experiments. G: As the result of quantitative evaluation, molecular changes are depicted for the most altered miRNAs. Maximum time-course changes are represented for the top 10 decreasing and increasing miRNAs, respectively. Separation between these groups is indicated by the black dashed horizontal line.

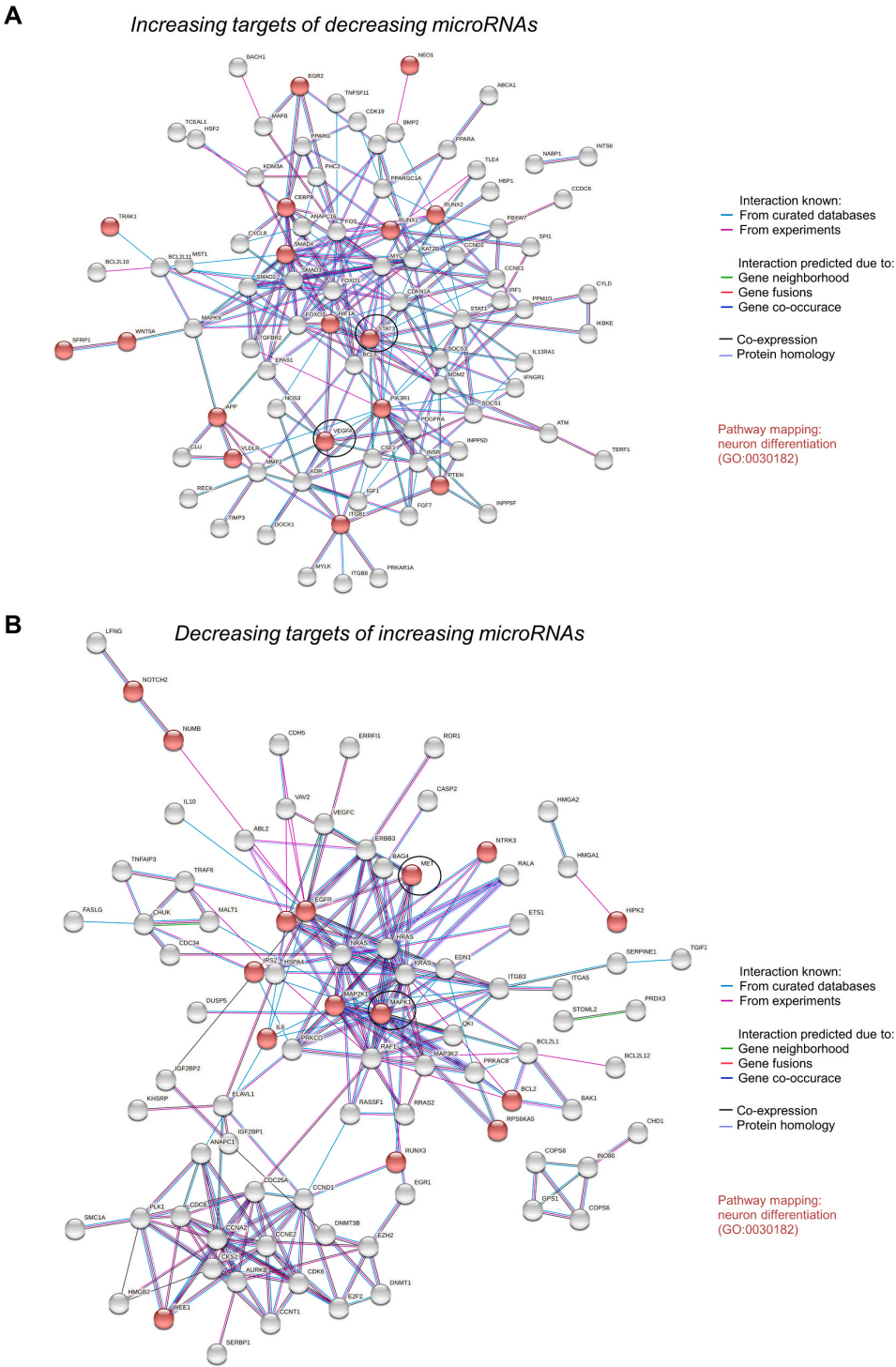


Fig. 8. Networks of miRNA-target interactions. Validated miRNA-target interactions were matched with time-resolved RNA expression data. Including the 91 prominently altered miRNAs, inverse directions of time-course expression patterns were assumed for their targeted transcripts. The represented networks, showing functional interactions between the resulting targets, were generated and exported from the STRING database. Networks are represented for the increasing targets of decreasing miRNAs (A) and the decreasing targets of increasing miRNAs (B), respectively. Targets that showed a connection to the GO term "neuron differentiation" are highlighted in red and exemplary targets at central nodes of the networks are identified by black circles.

In the CNS, neural stem cells (NSCs) originating from the embryonic ectoderm, give rise to both neurons and glia cells (Kennea and Mehmet, 2002; Tang et al., 2017). Similarly, MSCs have been described to possess the capability to trans-differentiate into these cell types (Urrutia et al., 2019). As for the expression dynamics of neural and glial signatures, we found ambiguous characteristics of neurons, astrocytes and oligodendrocytes upon the *in vitro* neurogenic trans-differentiation process. These different characteristics may be explained by a mixed cell population comprising neurons and glial cells. A similar scenario has previously been reported for umbilical cord derived MSCs (UC-MSCs). After stimulation of the UC-MSCs with neuronal conditioned medium for up to 12 days, expression of some glial proteins was observed in a portion of the trans-differentiated cells (Fu et al., 2004). Remarkably, our protein staining analyses revealed heterogeneity in the distribution of certain protein markers among the cells, indicating distinct cellular variability (Freeman et al., 2015). Future projects could benefit from more detailed insights into the composition of trans-differentiated cell populations, potentially using single-cell transcriptome or flow cytometric analyses.

It is also conceivable that the ambiguous marker characteristics reflect an uncompleted cell type differentiation. Various transcripts that are associated with the maintenance of stemness showed decreasing expression levels during the time-course, including genes such as *BIRC5* (Gil-Kulik et al., 2019), *HMMR* (Tilghman et al., 2014) and *MYB* (Malaterre et al., 2008). On the other hand, somatic stemness markers such as *KLF4* or *MYC* (Qin et al., 2011; Zaytseva et al., 2020; Zhao et al., 2017) exhibited an unexpected expression increase. Reduced expression of further stemness markers, as for example *NEK2*, has previously been linked to inefficient neural differentiation and the persistence of pluripotency (Spice et al., 2022). In addition, we were able to detect NSC specific genes throughout the trans-differentiation process, for example nestin that was detectable even up to 7 days, as validated by protein staining analyses. Overall, our findings and other reports on a later lineage conversion of terminally differentiated MSCs (Song and Tuan, 2004) imply the maintenance of a residual stem cell capacity. This should be acknowledged when assessing the utility of MSC-derived neurons for the transplantation into critical tissues like the CNS. Future studies should focus on a comprehensive assessment of persistent stem cell characteristics in the resulting neuron-like cells. Further transcriptomic analysis will be necessary during first hours after transplantation to decipher the directions of those cells in an *in vivo* environment. Potential strategies to mitigate tumorigenicity and stemness persistence will depend on results of transcriptomic analysis of cells isolated from *in vivo* experiments.

In summary, our time-resolved RNA profiling provides a comprehensive view on the transcriptomic landscape during the early MSC trans-differentiation process and highlights functional links to the acquisition of neuronal features. The assessment of experimentally validated target interactions together with the time-resolved RNA expression data indicate an essential role of microRNA mediated post-transcriptional regulations in the early MSC trans-differentiation. The combination of MSC identity markers shows that RNA signatures contribute to distinguish between the native MSC cell state and an emerging neurogenic trans-differentiation. Our data further indicate potential safety risks of MSC-derived neurons by showing an expression of common tumorigenic factors and the maintenance of stem cell characteristics.

Ethics approval and informed consent

Not applicable.

Author contributions

C.D., E.M. and U.F. designed the study; K.T., C.D. and U.F. performed the experiments; A.E., C.D., K.T., M.H., A.K. and U.F. analyzed and visualized the data; C.D., E.M., A.E., A.K., and U.F. wrote the

manuscript; all authors edited and approved the final manuscript.

Funding

This research was supported by the Hans-und-Ruth-Giessen-Stiftung (2021) and the Hedwig-Stalter-Stiftung (2022) to Diener C.

CRediT authorship contribution statement

Eckart Meese: Writing – original draft, Conceptualization. **Ulrike Fischer:** Writing – original draft, Formal analysis, Data curation, Conceptualization. **Andreas Keller:** Writing – review & editing, Formal analysis. **Caroline Diener:** Writing – original draft, Formal analysis, Data curation, Conceptualization. **Konstantin Thüre:** Formal analysis, Data curation. **Annika Engel:** Writing – original draft, Formal analysis. **Martin Hart:** Formal analysis.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors disclose that they did not use AI and AI-assisted technologies in the writing process.

Declaration of Competing Interest

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

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejcb.2024.151458.

References

- Alles, J., Fehlmann, T., Fischer, U., Backes, C., Galata, V., Minet, M., Hart, M., Abu-Halima, M., Grasser, F.A., Lenhof, H.P., Keller, A., Meese, E., 2019. An estimate of the total number of true human miRNAs. *Nucleic Acids Res* 47, 3353–3364. <https://doi.org/10.1093/nar/gkz097>.
- Alsubait, A., Aldossary, W., Rashid, M., Algamdi, A., Alrfaei, B.M., 2020. CYP1B1 gene: Implications in glaucoma and cancer. *J. Cancer* 11, 4652–4661. <https://doi.org/10.7150/jca.42669>.
- Aranha, M.M., Santos, D.M., Sola, S., Steer, C.J., Rodrigues, C.M., 2011. miR-34a regulates mouse neural stem cell differentiation. *PLoS One* 6, e21396. <https://doi.org/10.1371/journal.pone.0021396>.
- Augustyniak, J., Lenart, J., Zychowicz, M., Stepień, P.P., Buzanska, L., 2017. Mitochondrial biogenesis and neural differentiation of human iPSC is modulated by idebenone in a developmental stage-dependent manner. *Biogerontology* 18, 665–677. <https://doi.org/10.1007/s10522-017-9718-4>.
- Bar-Joseph, Z., Gitter, A., Simon, I., 2012. Studying and modelling dynamic biological processes using time-series gene expression data. *Nat. Rev. Genet* 13, 552–564. <https://doi.org/10.1038/nrg3244>.
- Bartoszewski, R., Sikorski, A.F., 2018. Editorial focus: entering into the non-coding RNA era. *Cell Mol. Biol. Lett.* 23, 45. <https://doi.org/10.1186/s11658-018-0111-3>.
- Bergsland, M., Werme, M., Malewicz, M., Perlmann, T., Muhr, J., 2006. The establishment of neuronal properties is controlled by Sox4 and Sox11. *Genes Dev.* 20, 3475–3486. <https://doi.org/10.1101/gad.403406>.
- Boone, D.K., Weisz, H.A., Bi, M., Falduto, M.T., Torres, K.E.O., Willey, H.E., Volsko, C.M., Kumar, A.M., Micci, M.A., Dewitt, D.S., Prough, D.S., Hellmich, H.L., 2017. Evidence linking microRNA suppression of essential prosurvival genes with hippocampal cell death after traumatic brain injury. *Sci. Rep.* 7, 6645. <https://doi.org/10.1038/s41598-017-06341-6>.
- Cai, N., Kurachi, M., Shibasaki, K., Okano-Uchida, T., Ishizaki, Y., 2012. CD44-positive cells are candidates for astrocyte precursor cells in developing mouse cerebellum. *Cerebellum* 11, 181–193. <https://doi.org/10.1007/s12311-011-0294-x>.
- Chapouton, P., Skupien, P., Hesl, B., Coolen, M., Moore, J.C., Madelaine, R., Kremmer, E., Faus-Kessler, T., Blader, P., Lawson, N.D., Bally-Cuif, L., 2010. Notch activity levels control the balance between quiescence and recruitment of adult neural stem cells. *J. Neurosci.* 30, 7961–7974. <https://doi.org/10.1523/JNEUROSCI.6170-09.2010>.
- Choudhary, P., Gupta, A., Singh, S., 2021. Therapeutic advancement in neuronal transdifferentiation of mesenchymal stromal cells for neurological disorders. *J. Mol. Neurosci.* 71, 889–901. <https://doi.org/10.1007/s12031-020-01714-5>.

- Cline, H., 2005. Synaptogenesis: a balancing act between excitation and inhibition. *Curr. Biol.* 15, R203–R205. <https://doi.org/10.1016/j.cub.2005.03.010>.
- Coccini, T., Schicchi, A., Locatelli, C.A., Caloni, F., Negri, S., Grignani, E., De Simone, U., 2023. Methylglyoxal-induced neurotoxic effects in primary neuronal-like cells transdifferentiated from human mesenchymal stem cells: Impact of low concentrations. *J. Appl. Toxicol.* 43, 1819–1839. <https://doi.org/10.1002/jat.4515>.
- Cochet-Bissuel, M., Lory, P., Monteil, A., 2014. The sodium leak channel, NALCN, in health and disease. *Front. Cell Neurosci.* 8, 132. <https://doi.org/10.3389/fncel.2014.00132>.
- Cortes-Medina, L.V., Pasantes-Morales, H., Aguilera-Castrejon, A., Picones, A., Lara-Figueroa, C.O., Luis, E., Montesinos, J.J., Cortes-Morales, V.A., De la Rosa Ruiz, M. P., Hernandez-Estevéz, E., Bonifaz, L.C., Alvarez-Perez, M.A., Ramos-Mandujano, G., 2019. Neuronal transdifferentiation potential of human mesenchymal stem cells from neonatal and adult sources by a small molecule cocktail. *Stem Cells Int* 2019, 7627148. <https://doi.org/10.1155/2019/7627148>.
- Coulthard, L.G., Hawsworth, O.A., Li, R., Balachandran, A., Lee, J.D., Sepehrband, F., Kurniawan, N., Jeanes, A., Simmons, D.G., Wolvetang, E., Woodruff, T.M., 2017. Complement C5aR1 signaling promotes polarization and proliferation of embryonic neural progenitor cells through PKC ζ . *J. Neurosci.* 37, 5395–5407. <https://doi.org/10.1523/JNEUROSCI.0525-17.2017>.
- Cunniff, C., Bassetti, J.A., Ellis, N.A., 2017. Bloom's syndrome: clinical spectrum, molecular pathogenesis, and cancer predisposition. *Mol. Syndr.* 8, 4–23. <https://doi.org/10.1159/000452082>.
- Diener, C., Hart, M., Kehl, T., Rheinheimer, S., Ludwig, N., Krammes, L., Pawusch, S., Lenhof, K., Tänzler, T., Schub, D., Sester, M., Walch-Rückheim, B., Keller, A., Lenhof, H.P., Meese, E., 2020. Quantitative and time-resolved miRNA pattern of early human T cell activation. *Nucleic Acids Res* 48, 10164–10183. <https://doi.org/10.1093/nar/gkaa788>.
- Diener, C., Hart, M., Kehl, T., Becker-Dorison, A., Tänzler, T., Schub, D., Krammes, L., Sester, M., Keller, A., Unger, M., Walch-Rückheim, B., Lenhof, H.P., Meese, E., 2023a. Time-resolved RNA signatures of CD4+ T cells in Parkinson's disease. *Cell Death Discov.* 9, 18. <https://doi.org/10.1038/s41420-023-01333-0>.
- Diener, C., Keller, A., Meese, E., 2022. Emerging concepts of miRNA therapeutics: from cells to clinic. *Trends Genet* 38, 613–626. <https://doi.org/10.1016/j.tig.2022.02.006>.
- Diener, C., Keller, A., Meese, E., 2023b. The miRNA-target interactions: An underestimated intricacy. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkad1142>.
- Diener, C., Keller, A., Meese, E., 2024. The miRNA-target interactions: an underestimated intricacy. *Nucleic Acids Res* 52, 1544–1557. <https://doi.org/10.1093/nar/gkad1142>.
- Dominici, M., Le Blanc, K., Mueller, I., Slaper-Cortenbach, I., Marini, F., Krause, D., Deans, R., Keating, A., Prockop, D., Horvitz, E., 2006. Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. *Cytotherapy* 8, 315–317. <https://doi.org/10.1080/14653240600855905>.
- Elias, G.M., Elias, L.A., Apostolides, P.F., Kriegstein, A.R., Nicoll, R.A., 2008. Differential trafficking of AMPA and NMDA receptors by SAP102 and PSD-95 underlies synapse development. *Proc. Natl. Acad. Sci. USA* 105, 20953–20958. <https://doi.org/10.1073/pnas.0811025106>.
- Fan, X., Fu, Y., Zhou, X., Sun, L., Yang, M., Wang, M., Chen, R., Wu, Q., Yong, J., Dong, J., Wen, L., Qiao, J., Wang, X., Tang, F., 2020. Single-cell transcriptome analysis reveals cell lineage specification in temporal-spatial patterns in human cortical development. *Sci. Adv.* 6, eaaz2978. <https://doi.org/10.1126/sciadv.aaz2978>.
- Farioli-Vecchioli, S., Tirone, F., 2015. Control of the cell cycle in adult neurogenesis and its relation with physical exercise. *Brain Plast.* 1, 41–54. <https://doi.org/10.3233/BPL-150013>.
- Fawal, M.A., Davy, A., 2018. Impact of metabolic pathways and epigenetics on neural stem cells. *Epigenet Insights* 11, 2516865718820946. <https://doi.org/10.1177/2516865718820946>.
- Feier, A.M., Portan, D., Manu, D.R., Kostopoulos, V., Kotrotsos, A., Strnad, G., Dobreanu, M., Salcudean, A., Batagor, T., 2022. Primary MSCs for personalized medicine: ethical challenges, isolation and biocompatibility evaluation of 3D electrospon and printed scaffolds. *Biomedicines* 10. <https://doi.org/10.3390/biomedicines10071563>.
- Filiano, A.J., Xu, Y., Tustison, N.J., Marsh, R.L., Baker, W., Smirnov, I., Overall, C.C., Gadani, S.P., Turner, S.D., Weng, Z., Peerzade, S.N., Chen, H., Lee, K.S., Scott, M.M., Beenhakker, M.P., Litvak, V., Kipnis, J., 2016. Unexpected role of interferon-gamma in regulating neuronal connectivity and social behaviour. *Nature* 535, 425–429. <https://doi.org/10.1038/nature18626>.
- Freeman, B.T., Jung, J.P., Ogle, B.M., 2015. Single-cell RNA-seq of bone marrow-derived mesenchymal stem cells reveals unique profiles of lineage priming. *PLoS One* 10, e0136199. <https://doi.org/10.1371/journal.pone.0136199>.
- Fu, Y.S., Shih, Y.T., Cheng, Y.C., Min, M.Y., 2004. Transformation of human umbilical mesenchymal cells into neurons in vitro. *J. Biomed. Sci.* 11, 652–660. <https://doi.org/10.1007/BF02256131>.
- Fuchs, H., Theuser, M., Wruck, W., Adjaye, J., 2014. miR-27 negatively regulates pluripotency-associated genes in human embryonal carcinoma cells. *PLoS One* 9, e111637. <https://doi.org/10.1371/journal.pone.0111637>.
- Funa, K., Sasahara, M., 2014. The roles of PDGF in development and during neurogenesis in the normal and diseased nervous system. *J. Neuroimmune Pharm.* 9, 168–181. <https://doi.org/10.1007/s11481-013-9479-z>.
- Furutachi, S., Matsumoto, A., Nakayama, K.I., Gotoh, Y., 2013. p57 controls adult neural stem cell quiescence and modulates the pace of lifelong neurogenesis. *EMBO J.* 32, 970–981. <https://doi.org/10.1038/emboj.2013.50>.
- Fusco, C.M., Desch, K., Dorrabaum, A.R., Wang, M., Staab, A., Chan, I.C.W., Vail, E., Villeri, V., Langer, J.D., Schuman, E.M., 2021. Neuronal ribosomes exhibit dynamic and context-dependent exchange of ribosomal proteins. *Nat. Commun.* 12, 6127. <https://doi.org/10.1038/s41467-021-26365-x>.
- Gan, K., Dong, G.H., Wang, N., Zhu, J.F., 2020. miR-221-3p and miR-222-3p downregulation promoted osteogenic differentiation of bone marrow mesenchyme stem cells through IGF-1/ERK pathway under high glucose condition. *Diabetes Res Clin. Pr.* 167, 108121. <https://doi.org/10.1016/j.diabres.2020.108121>.
- Gebert, L.F.R., MacRae, I.J., 2019. Regulation of microRNA function in animals. *Nat. Rev. Mol. Cell Biol.* 20, 21–37. <https://doi.org/10.1038/s41580-018-0045-7>.
- Gil-Kulik, P., Krzyzanowski, A., Dudzinska, E., Karwat, J., Chomik, P., Swistowska, M., Kondracka, A., Kwasniewska, A., Cioch, M., Jójczuk, M., Nogalski, A., Kocki, J., 2019. Potential involvement of BIRC5 in maintaining pluripotency and cell differentiation of human stem cells. *Oxid. Med. Cell Longev.* 2019, 8727925. <https://doi.org/10.1155/2019/8727925>.
- Gomez de San Jose, N., Massa, F., Halbgebauer, S., Oeckl, P., Steinacker, P., Otto, M., 2022. Neuronal pentraxins as biomarkers of synaptic activity: from physiological functions to pathological changes in neurodegeneration. *J. Neural Transm. (Vienna)* 129, 207–230. <https://doi.org/10.1007/s00702-021-02411-2>.
- Gorjala, P., Cairncross, J.G., Gary, R.K., 2016. p53-dependent up-regulation of CDKN1A and down-regulation of CCNE2 in response to beryllium. *Cell Prolif.* 49, 698–709. <https://doi.org/10.1111/cpr.12291>.
- Greene, L.A., Tischler, A.S., 1976. Establishment of a noradrenergic clonal line of rat adrenal pheochromocytoma cells which respond to nerve growth factor. *Proc. Natl. Acad. Sci. USA* 73, 2424–2428. <https://doi.org/10.1073/pnas.73.7.2424>.
- Guo, H., Ingolia, N.T., Weissman, J.S., Bartel, D.P., 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466, 835–840. <https://doi.org/10.1038/nature09267>.
- Hamada, N., Fujita, Y., Kojima, T., Kitamoto, A., Akao, Y., Nozawa, Y., Ito, M., 2012. MicroRNA expression profiling of NGF-treated PC12 cells revealed a critical role for miR-221 in neuronal differentiation. *Neurochem Int* 60, 743–750. <https://doi.org/10.1016/j.neuint.2012.03.010>.
- Hamby, M.E., Coskun, V., Sun, Y.E., 2008. Transcriptional regulation of neuronal differentiation: the epigenetic layer of complexity. *Biochim. Biophys. Acta* 1779, 432–437. <https://doi.org/10.1016/j.bbagr.2008.07.006>.
- Han, Y., Yang, J., Fang, J., Zhou, Y., Candi, E., Wang, J., Hua, D., Shao, C., Shi, Y., 2022. The secretion profile of mesenchymal stem cells and potential applications in treating human diseases. *Signal Transduct. Target Ther.* 7, 92. <https://doi.org/10.1038/s41392-022-00932-0>.
- Hardwick, L.J., Ali, F.R., Azzarelli, R., Philpott, A., 2015. Cell cycle regulation of proliferation versus differentiation in the central nervous system. *Cell Tissue Res* 359, 187–200. <https://doi.org/10.1007/s00441-014-1895-8>.
- Harris, D.T., 2014. Stem cell banking for regenerative and personalized medicine. *Biomedicines* 2, 50–79. <https://doi.org/10.3390/biomedicines2010050>.
- Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., Richter, J., Rubin, G.M., Blake, J.A., Bult, C., Dolan, M., Drabkin, H., Eppig, J.T., Hill, D.P., Ni, L., Ringwald, M., Balakrishnan, R., Cherry, J.M., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S., Fisk, D.G., Hirschman, J.E., Hong, E.L., Nash, R.S., Sethuraman, A., Theesfeld, C.L., Botstein, D., Dolinski, K., Feierbach, B., Berardini, T., Mundodi, S., Rhee, S.Y., Apweiler, R., Barrell, D., Camon, E., Dimmer, E., Lee, V., Chisholm, R., Gaudet, P., Kibbe, W., Kishore, R., Schwarz, E.M., Sternberg, P., Gwinn, M., Hannick, L., Wortman, J., Berriman, M., Wood, V., de la Cruz, N., Tonellato, P., Jaiswal, P., Seigfried, T., White, R., Gene Ontology, C., 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 32, D258–D261. <https://doi.org/10.1093/nar/gkh036>.
- He, Y., Cai, Y., Pai, P.M., Ren, X., Xia, Z., 2021. The causes and consequences of miR-503 dysregulation and its impact on cardiovascular disease and cancer. *Front. Pharm.* 12, 629611. <https://doi.org/10.3389/fphar.2021.629611>.
- He, Z., Hu, X., Liu, W., Dorrance, A., Garzon, R., Houghton, P.J., Shen, C., 2017. P53 suppresses ribonucleotide reductase via inhibiting mTORC1. *Oncotarget* 8, 41422–41431. <https://doi.org/10.18632/oncotarget.17440>.
- Hermann, A., Gastl, R., Liebau, S., Popa, M.O., Fiedler, J., Boehm, B.O., Maisel, M., Lerche, H., Schwarz, J., Brenner, R., Storch, A., 2004. Efficient generation of neural stem cell-like cells from adult human bone marrow stromal cells. *J. Cell Sci.* 117, 4411–4422. <https://doi.org/10.1242/jcs.01307>.
- Hernandez, R., Jimenez-Luna, C., Perales-Adan, J., Perazzoli, G., Melguizo, C., Prados, J., 2020. Differentiation of human mesenchymal stem cells towards neuronal lineage: clinical trials in nervous system disorders. *Biomol. Ther.* 28, 34–44. <https://doi.org/10.4062/biomolther.2019.065>.
- Hmadcha, A., Martin-Montalvo, A., Gauthier, B.R., Soria, B., Capilla-Gonzalez, V., 2020. Therapeutic potential of mesenchymal stem cells for cancer therapy. *Front. Bioeng. Biotechnol.* 8, 43. <https://doi.org/10.3389/fbioe.2020.00043>.
- Hoang, D.M., Pham, P.T., Bach, T.Q., Ngo, A.T.L., Nguyen, Q.T., Phan, T.T.K., Nguyen, G. H., Le, P.T.T., Hoang, V.T., Forsyth, N.R., Heke, M., Nguyen, L.T., 2022. Stem cell-based therapy for human diseases. *Signal Transduct. Target Ther.* 7, 272. <https://doi.org/10.1038/s41392-022-01134-4>.
- Hong, Y.R., Lee, H., Park, M.H., Lee, J.K., Lee, J.Y., Suh, H.D., Jeong, M.S., Bae, J.S., Jin, H.K., 2015. CCL2 induces neural stem cell proliferation and neuronal differentiation in Niemann-Pick type C mice. *J. Vet. Med. Sci.* 77, 693–699. <https://doi.org/10.1292/jvms.14-0352>.
- Hu, B., Nandhu, M.S., Sim, H., Agudelo-Garcia, P.A., Saldivar, J.C., Dolan, C.E., Mora, M. E., Nuovo, G.J., Cole, S.E., Viapiano, M.S., 2012. Fibulin-3 promotes glioma growth and resistance through a novel paracrine regulation of Notch signaling. *Cancer Res* 72, 3873–3885. <https://doi.org/10.1158/0008-5472.CAN-12-1060>.

- Huang, W., Bhaduri, A., Velmesshev, D., Wang, S., Wang, L., Rottkamp, C.A., Alvarez-Buylla, A., Rowitch, D.H., Kriegstein, A.R., 2020. Origins and proliferative states of human oligodendrocyte precursor cells. *e511 Cell* 182, 594–608. <https://doi.org/10.1016/j.cell.2020.06.027>.
- Huang, F., Chen, L., Guo, W., Huang, T., Cai, Y.D., 2022. Identification of human cell cycle phase markers based on single-cell RNA-Seq data by using machine learning methods. *Biomed. Res Int* 2022, 2516653. <https://doi.org/10.1155/2022/2516653>.
- Inoue, A., Takahashi, H., Harada, H., Kohno, S., Ohue, S., Kobayashi, K., Yano, H., Tanaka, J., Ohnishi, T., 2010. Cancer stem-like cells of glioblastoma characteristically express MMP-13 and display highly invasive activity. *Int J. Oncol.* 37, 1121–1131. <https://doi.org/10.3892/ijo.00000764>.
- Jan, A.T., Lee, E.J., Choi, I., 2016. Fibromodulin: a regulatory molecule maintaining cellular architecture for normal cellular function. *Int J. Biochem Cell Biol.* 80, 66–70. <https://doi.org/10.1016/j.biocel.2016.09.023>.
- Jandial, R., Singec, I., Ames, C.P., Snyder, E.Y., 2008. Genetic modification of neural stem cells. *Mol. Ther.* 16, 450–457. <https://doi.org/10.1038/sj.mt.6300402>.
- Jiao, Y., Palmgren, B., Novozhilova, E., Englund Johansson, U., Spieles-Engemann, A.L., Kale, A., Stupp, S.I., Olivius, P., 2014. BDNF increases survival and neuronal differentiation of human neural precursor cells cotransplanted with a nanofiber gel to the auditory nerve in a rat model of neuronal damage. *Biomed. Res Int* 2014, 356415. <https://doi.org/10.1155/2014/356415>.
- Kennea, N.L., Mehmet, H., 2002. Neural stem cells. *J. Pathol.* 197, 536–550. <https://doi.org/10.1002/path.1189>.
- Kern, F., Aparicio-Puerta, E., Li, Y., Fehlmann, T., Kehl, T., Wagner, V., Ray, K., Ludwig, N., Lenhof, H.P., Meese, E., Keller, A., 2021a. miRTargetLink 2.0-interactive miRNA target gene and target pathway networks. *Nucleic Acids Res* 49, W409–W416. <https://doi.org/10.1093/nar/gkab297>.
- Kern, F., Krammes, L., Danz, K., Diener, C., Kehl, T., Kuchler, O., Fehlmann, T., Kahraman, M., Rheinheimer, S., Aparicio-Puerta, E., Wagner, S., Ludwig, N., Backes, C., Lenhof, H.P., von Briesen, H., Hart, M., Keller, A., Meese, E., 2021b. Validation of human microRNA target pathways enables evaluation of target prediction tools. *Nucleic Acids Res* 49, 127–144. <https://doi.org/10.1093/nar/gkaa1161>.
- Khan, A.A., Huat, T.J., Al Mutery, A., El-Serafi, A.T., Kacem, H.H., Abdallah, S.H., Reza, M.F., Abdullah, J.M., Jaafar, H., 2020. Significant transcriptomic changes are associated with differentiation of bone marrow-derived mesenchymal stem cells into neural progenitor-like cells in the presence of bFGF and EGF. *Cell Biosci.* 10, 126. <https://doi.org/10.1186/s13578-020-00487-z>.
- Kopen, G.C., Prockop, D.J., Phinney, D.G., 1999. Marrow stromal cells migrate throughout forebrain and cerebellum, and they differentiate into astrocytes after injection into neonatal mouse brains. *Proc. Natl. Acad. Sci. USA* 96, 10711–10716. <https://doi.org/10.1073/pnas.96.19.10711>.
- Krueger, D.D., Tuffy, L.P., Papadopoulos, T., Brose, N., 2012. The role of neurexins and neuroligins in the formation, maturation, and function of vertebrate synapses. *Curr. Opin. Neurobiol.* 22, 412–422. <https://doi.org/10.1016/j.conb.2012.02.012>.
- Kwon, S.E., Chapman, E.R., 2011. Synaptophysin regulates the kinetics of synaptic vesicle endocytosis in central neurons. *Neuron* 70, 847–854. <https://doi.org/10.1016/j.neuron.2011.04.001>.
- Lee, D.Y., Lee, S.E., Kwon, D.H., Nithiyandam, S., Lee, M.H., Hwang, J.S., Basith, S., Ahn, J.H., Shin, T.H., Lee, G., 2021. Strategies to improve the quality and freshness of human bone marrow-derived mesenchymal stem cells for neurological diseases. *Stem Cells Int* 2021, 8444599. <https://doi.org/10.1155/2021/8444599>.
- Li, P., Li, H.X., Jiang, H.Y., Zhu, L., Wu, H.Y., Li, J.T., Lai, J.H., 2017. Expression of NG2 and platelet-derived growth factor receptor alpha in the developing neonatal rat brain. *Neural Regen. Res* 12, 1843–1852. <https://doi.org/10.4103/1673-5374.219045>.
- Li, S., Pritchard, D.M., Yu, L.G., 2022. Regulation and function of matrix metalloproteinase-13 in cancer progression and metastasis. *Cancers* 14. <https://doi.org/10.3390/cancers14133263>.
- Liang, W., Chen, X., Zhang, S., Fang, J., Chen, M., Xu, Y., Chen, X., 2021. Mesenchymal stem cells as a double-edged sword in tumor growth: focusing on MSC-derived cytokines. *Cell Mol. Biol. Lett.* 26, 3. <https://doi.org/10.1186/s11658-020-00246-5>.
- Liu, Q., Cheng, G., Wang, Z., Zhan, S., Xiong, B., Zhao, X., 2015. Bone marrow-derived mesenchymal stem cells differentiate into nerve-like cells in vitro after transfection with brain-derived neurotrophic factor gene. *Vitr. Cell Dev. Biol. Anim.* 51, 319–327. <https://doi.org/10.1007/s11626-015-9875-1>.
- Liu, Y., Jiang, X., Zhang, X., Chen, R., Sun, T., Fok, K.L., Dong, J., Tsang, L.L., Yi, S., Ruan, Y., Guo, J., Yu, M.K., Tian, Y., Chung, Y.W., Yang, M., Xu, W., Chung, C.M., Li, T., Chan, H.C., 2011. Dedifferentiation-reprogrammed mesenchymal stem cells with improved therapeutic potential. *Stem Cells* 29, 2077–2089. <https://doi.org/10.1002/stem.764>.
- Livingstone, I., Uversky, V.N., Furniss, D., Wiberg, A., 2020. The pathophysiological significance of fibulin-3. *Biomolecules* 10. <https://doi.org/10.3390/biom10091294>.
- Lukomska, B., Stanaszek, L., Zuba-Surma, E., Legosz, P., Sarzynska, S., Drela, K., 2019. Challenges and controversies in human mesenchymal stem cell therapy. *Stem Cells Int* 2019, 9628536. <https://doi.org/10.1155/2019/9628536>.
- Malaterre, J., Mantamadiotis, T., Dworkin, S., Lightowler, S., Yang, Q., Ransome, M.I., Turnley, A.M., Nichols, N.R., Emambokus, N.R., Frampton, J., Ramsay, R.G., 2008. c-Myb is required for neural progenitor cell proliferation and maintenance of the neural stem cell niche in adult brain. *Stem Cells* 26, 173–181. <https://doi.org/10.1634/stemcells.2007-0293>.
- Mao, S., Li, H., Sun, Q., Zen, K., Zhang, C.Y., Li, L., 2014. miR-17 regulates the proliferation and differentiation of the neural precursor cells during mouse corticogenesis. *FEBS J.* 281, 1144–1158. <https://doi.org/10.1111/febs.12680>.
- McKenzie, A.T., Wang, M., Hauberg, M.E., Fullard, J.F., Kozlenkov, A., Keenan, A., Hurd, Y.L., Dracheva, S., Casaccia, P., Roussos, P., Zhang, B., 2018. Brain cell type specific gene expression and co-expression network architectures. *Sci. Rep.* 8, 8868. <https://doi.org/10.1038/s41598-018-27293-5>.
- McKinnon, P.J., 2013. Maintaining genome stability in the nervous system. *Neurosci.* 16, 1523–1529. <https://doi.org/10.1038/nn.3537>.
- Mertsch, S., Schurgers, L.J., Weber, K., Paulus, W., Senner, V., 2009. Matrix gla protein (MGP): an overexpressed and migration-promoting mesenchymal component in glioblastoma. *BMC Cancer* 9, 302. <https://doi.org/10.1186/1471-2407-9-302>.
- Miao, Z., Sun, H., Xue, Y., 2017. Isolation and characterization of human chorionic membranes mesenchymal stem cells and their neural differentiation. *Tissue Eng. Regen. Med* 14, 143–151. <https://doi.org/10.1007/s13770-017-0025-6>.
- Migliorini, E., Ban, J., Greci, G., Andolfi, L., Pozzato, A., Tormen, M., Torre, V., Lazzarino, M., 2013. Nanomechanics controls neuronal precursors adhesion and differentiation. *Biotechnol. Bioeng.* 110, 2301–2310. <https://doi.org/10.1002/bit.24880>.
- Minguell, J.J., Fierro, F.A., Epunan, M.J., Erices, A.A., Sierralta, W.D., 2005. Nonstimulated human uncommitted mesenchymal stem cells express cell markers of mesenchymal and neural lineages. *Stem Cells Dev.* 14, 408–414. <https://doi.org/10.1089/scd.2005.14.408>.
- Monaco, E., Bionaz, M., Rodriguez-Zas, S., Hurley, W.L., Wheeler, M.B., 2012. Transcriptomics comparison between porcine adipose and bone marrow mesenchymal stem cells during in vitro osteogenic and adipogenic differentiation. *PLoS One* 7, e32481. <https://doi.org/10.1371/journal.pone.0032481>.
- Monte, M., Benetti, R., Buscemi, G., Sandy, P., Del Sal, G., Schneider, C., 2003. The cell cycle-regulated protein human GTSE-1 controls DNA damage-induced apoptosis by affecting p53 function. *J. Biol. Chem.* 278, 30356–30364. <https://doi.org/10.1074/jbc.M302902200>.
- Morimoto, K., Nakajima, K., 2019. Role of the Immune System in the Development of the Central Nervous System. *Front. Neurosci.* 13, 916. <https://doi.org/10.3389/fnins.2019.00916>.
- Musial-Wysocka, A., Kot, M., Majka, M., 2019. The pros and cons of mesenchymal stem cell-based therapies. *Cell Transpl.* 28, 801–812. <https://doi.org/10.1177/0963689719837897>.
- Nango, H., Kosuge, Y., Yoshimura, N., Miyagishi, H., Kanazawa, T., Hashizaki, K., Suzuki, T., Ishige, K., 2020. The molecular mechanisms underlying prostaglandin d (2)-induced neuritegenesis in motor neuron-like NSC-34 cells. *Cells* 9. <https://doi.org/10.3390/cells9040934>.
- Oguma, Y., Kuroda, Y., Wakao, S., Kushida, Y., Dezawa, M., 2022. Single-cell RNA sequencing reveals different signatures of mesenchymal stromal cell pluripotent-like and multipotent populations. *iScience* 25, 105395. <https://doi.org/10.1016/j.isci.2022.105395>.
- Oikari, L.E., Okolicsanyi, R.K., Griffiths, L.R., Haupt, L.M., 2016a. Data defining markers of human neural stem cell lineage potential. *Data Brief* 7, 206–215. <https://doi.org/10.1016/j.dib.2016.02.030>.
- Oikari, L.E., Okolicsanyi, R.K., Qin, A., Yu, C., Griffiths, L.R., Haupt, L.M., 2016b. Cell surface heparan sulfate proteoglycans as novel markers of human neural stem cell fate determination. *Stem Cell Res* 16, 92–104. <https://doi.org/10.1016/j.scr.2015.12.011>.
- Pelkey, K.A., Barksdale, E., Craig, M.T., Yuan, X., Sukumaran, M., Vargish, G.A., Mitchell, R.M., Wyeth, M.S., Petralia, R.S., Chittajallu, R., Karlsson, R.M., Cameron, H.A., Murata, Y., Colonnesse, M.T., Worley, P.F., McBain, C.J., 2015. Pentraxins coordinate excitatory synapse maturation and circuit integration of parvalbumin interneurons. *Neuron* 85, 1257–1272. <https://doi.org/10.1016/j.neuron.2015.02.020>.
- Pittenger, M.F., Discher, D.E., Peault, B.M., Phinney, D.G., Hare, J.M., Caplan, A.I., 2019. Mesenchymal stem cell perspective: cell biology to clinical progress. *NPJ Regen. Med* 4, 22. <https://doi.org/10.1038/s41536-019-0083-6>.
- Poon, V.Y., Gu, M., Ji, F., VanDongen, A.M., Fivaz, M., 2016. miR-27b shapes the presynaptic transcriptome and influences neurotransmission by silencing the polycomb group protein Bmi1. *BMC Genom.* 17, 777. <https://doi.org/10.1186/s12864-016-3139-7>.
- Qin, S., Liu, M., Niu, W., Zhang, C.L., 2011. Dysregulation of Kruppel-like factor 4 during brain development leads to hydrocephalus in mice. *Proc. Natl. Acad. Sci. USA* 108, 21117–21121. <https://doi.org/10.1073/pnas.1112351109>.
- Quinn, L.M., Herr, A., McGarry, T.J., Richardson, H., 2001. The Drosophila Geminin homolog: roles for Geminin in limiting DNA replication, in anaphase and in neurogenesis. *Genes Dev.* 15, 2741–2754. <https://doi.org/10.1101/gad.916201>.
- Rahbaran, M., Zekiy, A.O., Bahramali, M., Jahangir, M., Mardasi, M., Sakhaei, D., Thangavelu, L., Shomali, N., Zamani, M., Mohammadi, A., Rahnama, N., 2022. Therapeutic utility of mesenchymal stromal cell (MSC)-based approaches in chronic neurodegeneration: a glimpse into underlying mechanisms, current status, and prospects. *Cell Mol. Biol. Lett.* 27, 56. <https://doi.org/10.1186/s11658-022-00359-z>.
- Ramos, J., Fu, D., 2019. The emerging impact of tRNA modifications in the brain and nervous system. *Biochim Biophys. Acta Gene Regul. Mech.* 1862, 412–428. <https://doi.org/10.1016/j.bbaggm.2018.11.007>.
- Rohart, F., Mason, E.A., Matigian, N., Mosbergen, R., Korn, O., Chen, T., Butcher, S., Patel, J., Atkinson, K., Khosrotehrani, K., Fisk, N.M., Le Cao, K.A., Wells, C.A., 2016. A molecular classification of human mesenchymal stromal cells. *PeerJ* 4, e1845. <https://doi.org/10.7717/peerj.1845>.
- Rohban, R., Pieber, T.R., 2017. Mesenchymal stem and progenitor cells in regeneration: tissue specificity and regenerative potential. *Stem Cells Int* 2017, 5173732. <https://doi.org/10.1155/2017/5173732>.
- Roush, S., Slack, F.J., 2008. The let-7 family of microRNAs. *Trends Cell Biol.* 18, 505–516. <https://doi.org/10.1016/j.tcb.2008.07.007>.
- Ruijterberg, S., van den Heuvel, S., 2016. Coordinating cell proliferation and differentiation: antagonism between cell cycle regulators and cell type-specific gene

- expression. *Cell Cycle* 15, 196–212. <https://doi.org/10.1080/15384101.2015.1120925>.
- Sakry, D., Yigit, H., Dimou, L., Trotter, J., 2015. Oligodendrocyte precursor cells synthesize neuromodulatory factors. *PLoS One* 10, e0127222. <https://doi.org/10.1371/journal.pone.0127222>.
- Schmitz, U., Lai, X., Winter, F., Wolkenhauer, O., Vera, J., Gupta, S.K., 2014. Cooperative gene regulation by microRNA pairs and their identification using a computational workflow. *Nucleic Acids Res* 42, 7539–7552. <https://doi.org/10.1093/nar/gku465>.
- Scolz, M., Widlund, P.O., Piazza, S., Bublik, D.R., Reber, S., Peche, L.Y., Ciani, Y., Hubner, N., Isokane, M., Monte, M., Ellenberg, J., Hyman, A.A., Schneider, C., Bird, A.W., 2012. GTSE1 is a microtubule plus-end tracking protein that regulates EB1-dependent cell migration. *PLoS One* 7, e51259. <https://doi.org/10.1371/journal.pone.0051259>.
- Selbach, M., Schwanhauss, B., Thierfelder, N., Fang, Z., Khanin, R., Rajewsky, N., 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature* 455, 58–63. <https://doi.org/10.1038/nature07228>.
- Semprich, C.L., Davidson, L., Amorim Torres, A., Patel, H., Briscoe, J., Metz, V., Storey, K.G., 2022. ERK1/2 signalling dynamics promote neural differentiation by regulating chromatin accessibility and the polycomb repressive complex. *PLoS Biol.* 20, e3000221. <https://doi.org/10.1371/journal.pbio.3000221>.
- Sengupta, S., Mondal, M., Prasavi, K.R., Mukherjee, A., Magod, P., Urbach, S., Friedmann-Morvinski, D., Marin, P., Somasundaram, K., 2022. Differentiated glioma cell-derived fibromodulin activates integrin-dependent Notch signaling in endothelial cells to promote tumor angiogenesis and growth. *Elife* 11. <https://doi.org/10.7554/eLife.78972>.
- Shaltouki, A., Peng, J., Liu, Q., Rao, M.S., Zeng, X., 2013. Efficient generation of astrocytes from human pluripotent stem cells in defined conditions. *Stem Cells* 31, 941–952. <https://doi.org/10.1002/stem.1334>.
- Shefchek, K.A., Harris, N.L., Gargano, M., Matentzoglou, N., Unni, D., Brush, M., Keith, D., Conlin, T., Vasilovsky, N., Zhang, X.A., Balhoff, J.P., Babb, L., Bello, S.M., Blau, H., Bradford, Y., Carbon, S., Carmody, L., Chan, L.E., Cipriani, V., Cuzick, A., Della Rocca, M., Dunn, N., Essaid, S., Fey, P., Grove, C., Gouridine, J.P., Hamosh, A., Harris, M., Helbig, L., Hoatlin, M., Joachimiak, M., Jupp, S., Lett, K.B., Lewis, S.E., McNamara, C., Pendlington, Z.M., Pilgrim, C., Putman, T., Ravanmehr, V., Reese, J., Riggs, E., Robb, S., Roncaglia, P., Seager, J., Segerdell, E., Simulak, M., Storm, A.L., Thaxon, C., Thessen, A., Jacobsen, J.O.B., McMurtry, J.A., Groza, T., Kohler, S., Smedley, D., Robinson, P.N., Mungall, C.J., Haendel, M.A., Munoz-Torres, M.C., Osumi-Sutherland, D., 2020. The Monarch Initiative in 2019: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res* 48, D704–D715. <https://doi.org/10.1093/nar/gkz997>.
- Shi, Z., Geng, Y., Liu, J., Zhang, H., Zhou, L., Lin, Q., Yu, J., Zhang, K., Liu, J., Gao, X., Zhang, C., Yao, Y., Zhang, C., Sun, Y.E., 2018. Single-cell transcriptomics reveals gene signatures and alterations associated with aging in distinct neural stem/progenitor cell subpopulations. *Protein Cell* 9, 351–364. <https://doi.org/10.1007/s13238-017-0450-2>.
- Snyder, M., Huang, X.Y., Zhang, J.J., 2011. Stat3 is essential for neuronal differentiation through direct transcriptional regulation of the Sox6 gene. *FEBS Lett.* 585, 148–152. <https://doi.org/10.1016/j.febslet.2010.11.030>.
- Song, L., Tuan, R.S., 2004. Transdifferentiation potential of human mesenchymal stem cells derived from bone marrow. *FASEB J.* 18, 980–982. <https://doi.org/10.1096/fj.03-1100je>.
- Spice, D.M., Cooper, T.T., Lajoie, G.A., Kelly, G.M., 2022. Never in Mitosis Kinase 2 regulation of metabolism is required for neural differentiation. *Cell Signal* 100, 110484. <https://doi.org/10.1016/j.cellsig.2022.110484>.
- Spurgat, M.S., Tang, S.J., 2022. Single-Cell RNA-sequencing: astrocyte and microglial heterogeneity in health and disease. *Cells* 11. <https://doi.org/10.3390/cells11132021>.
- Sun, Y.M., Cooper, M., Finch, S., Lin, H.H., Chen, Z.F., Williams, B.P., Buckley, N.J., 2008. Rest-mediated regulation of extracellular matrix is crucial for neural development. *PLoS One* 3, e3656. <https://doi.org/10.1371/journal.pone.0003656>.
- Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N.T., Legeay, M., Fang, T., Bork, P., Jensen, L.J., von Mering, C., 2021. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* 49, D605–D612. <https://doi.org/10.1093/nar/gkaa1074>.
- Tang, S., Xie, Z., Wang, P., Li, J., Wang, S., Liu, W., Li, M., Wu, X., Su, H., Cen, S., Ye, G., Zheng, G., Wu, Y., Shen, H., 2019. LncRNA-OG promotes the osteogenic differentiation of bone marrow-derived mesenchymal stem cells under the regulation of hnRNPK. *Stem Cells* 37, 270–283. <https://doi.org/10.1002/stem.2937>.
- Tang, Y., Yu, P., Cheng, L., 2017. Current progress in the derivation and therapeutic application of neural stem cells. *Cell Death Dis.* 8, e3108. <https://doi.org/10.1038/cddis.2017.504>.
- Thanaskody, K., Jusop, A.S., Tye, G.J., Wan Kamarul Zaman, W.S., Dass, S.A., Nordin, F., 2022. MSCs vs. iPSCs: Potential in therapeutic applications. *Front Cell Dev. Biol.* 10, 1005926. <https://doi.org/10.3389/fcell.2022.1005926>.
- Theis, V., Theiss, C., 2018. VEGF - a stimulus for neuronal development and regeneration in the CNS and PNS. *Curr. Protein Pept. Sci.* 19, 589–597. <https://doi.org/10.2174/1389203719666180104113937>.
- Thomas, A., Gasque, P., Vaudry, D., Gonzalez, B., Fontaine, M., 2000. Expression of a complete and functional complement system by human neuronal cells in vitro. *Int Immunol.* 12, 1015–1023. <https://doi.org/10.1093/intimm/12.7.1015>.
- Tilghman, J., Wu, H., Sang, Y., Shi, X., Guerrero-Cazares, H., Quinones-Hinojosa, A., Eberhart, C.G., Latta, J., Ying, M., 2014. HMMR maintains the stemness and tumorigenicity of glioblastoma stem-like cells. *Cancer Res* 74, 3168–3179. <https://doi.org/10.1158/0008-5472.CAN-13-2103>.
- Tondreau, T., Lagneaux, L., Dejenef, M., Massy, M., Mortier, C., Delforge, A., Bron, D., 2004. Bone marrow-derived mesenchymal stem cells already express specific neural proteins before any differentiation. *Differentiation* 72, 319–326. <https://doi.org/10.1111/j.1432-0436.2004.07207003.x>.
- Urrutia, D.N., Caviedes, P., Mardones, R., Minguez, J.J., Vega-Letter, A.M., Jofre, C.M., 2019. Comparative study of the neural differentiation capacity of mesenchymal stromal cells from different tissue sources: an approach for their use in neural regeneration therapies. *PLoS One* 14, e0213032. <https://doi.org/10.1371/journal.pone.0213032>.
- Varoqueaux, F., Aramuni, G., Rawson, R.L., Mohrmann, R., Missler, M., Gottmann, K., Zhang, W., Sudhof, T.C., Brose, N., 2006. Neurotrophins determine synapse maturation and function. *Neuron* 51, 741–754. <https://doi.org/10.1016/j.neuron.2006.09.003>.
- Vasanthan, J., Gurusamy, N., Rajasingh, S., Sigamani, V., Kirankumar, S., Thomas, E.L., Rajasingh, J., 2020. Role of Human Mesenchymal Stem Cells in Regenerative Therapy. *Cells* 10. <https://doi.org/10.3390/cells10010054>.
- Volarevic, V., Markovic, B.S., Gazdic, M., Volarevic, A., Jovicic, N., Arsenijevic, N., Armstrong, L., Djonov, V., Lako, M., Stojkovic, M., 2018. Ethical and safety issues of stem cell-based therapy. *Int. J. Med. Sci.* 15, 36–45. <https://doi.org/10.7150/ijms.21666>.
- Wada, T., Haigh, J.J., Ema, M., Hitoshi, S., Chaddah, R., Rossant, J., Nagy, A., van der Kooy, D., 2006. Vascular endothelial growth factor directly inhibits primitive neural stem cell survival but promotes definitive neural stem cell survival. *J. Neurosci.* 26, 6803–6812. <https://doi.org/10.1523/JNEUROSCI.0526-06.2006>.
- Walsh, P., Truong, V., Hill, C., Stoflet, N.D., Baden, J., Low, W.C., Keirstead, S.A., Dutton, J.R., Parr, A.M., 2017. Defined culture conditions accelerate small-molecule-assisted neural induction for the production of neural progenitors from human-induced pluripotent stem cells. *Cell Transpl.* 26, 1890–1902. <https://doi.org/10.1177/0963689717737074>.
- Wang, Z., Chai, C., Wang, R., Feng, Y., Huang, L., Zhang, Y., Xiao, X., Yang, S., Zhang, Y., Zhang, X., 2021. Single-cell transcriptome atlas of human mesenchymal stem cells exploring cellular heterogeneity. *Clin. Transl. Med.* 11, e650. <https://doi.org/10.1002/ctm2.650>.
- van Wietmarschen, N., Merzouk, S., Halsema, N., Spierings, D.C.J., Guryev, V., Lansdorp, P.M., 2018. BLM helicase suppresses recombination at G-quadruplex motifs in transcribed genes. *Nat. Commun.* 9, 271. <https://doi.org/10.1038/s41467-017-02760-1>.
- Xia, X., Wang, Y., Zheng, J.C., 2022. The microRNA-17 ~ 92 family as a key regulator of neurogenesis and potential regenerative therapeutics of neurological disorders. *Stem Cell Rev. Rep.* 18, 401–411. <https://doi.org/10.1007/s12015-020-10050-5>.
- Xiong, Y., Zhang, Y., Xiong, S., Williams-Villalobo, A.E., 2020. A glance of p53 functions in brain development, neural stem cells, and brain cancer. *Biology* 9. <https://doi.org/10.3390/biology9090285>.
- Yao, S., 2016. MicroRNA biogenesis and their functions in regulating stem cell potency and differentiation. *Biol. Proced. Online* 18, 8. <https://doi.org/10.1186/s12575-016-0037-y>.
- Ylivinkka, I., Sihto, H., Tynnenen, O., Hu, Y., Laakso, A., Kivisaari, R., Laakkonen, P., Keski-Oja, J., Hyytiäinen, M., 2017. Motility of glioblastoma cells is driven by netrin-1 induced gain of stemness. *J. Exp. Clin. Cancer Res* 36, 9. <https://doi.org/10.1186/s13046-016-0482-0>.
- Zaytseva, O., Kim, N.H., Quinn, L.M., 2020. MYC in brain development and cancer. *Int. J. Mol. Sci.* 21. <https://doi.org/10.3390/ijms21207742>.
- Zhang, Z., Ma, Z., Zou, W., Guo, H., Liu, M., Ma, Y., Zhang, L., 2019. The appropriate marker for astrocytes: comparing the distribution and expression of three astrocytic markers in different mouse cerebral regions. *Biomed. Res Int* 2019, 9605265. <https://doi.org/10.1155/2019/9605265>.
- Zhao, W., Li, Y., Zhang, X., 2017. Stemness-related markers in cancer. *Cancer Transl. Med* 3, 87–95. <https://doi.org/10.4103/ctm.ctm.69.16>.
- Zhao, C., Sun, G., Li, S., Lang, M.F., Yang, S., Li, W., Shi, Y., 2010. MicroRNA let-7b regulates neural stem cell proliferation and differentiation by targeting nuclear receptor TLX signaling. *Proc. Natl. Acad. Sci. USA* 107, 1876–1881. <https://doi.org/10.1073/pnas.0908750107>.
- Zheng, B., Wang, C., He, L., Xu, X., Qu, J., Hu, J., Zhang, H., 2013. Neural differentiation of mesenchymal stem cells influences chemotactic responses to HGF. *J. Cell Physiol.* 228, 149–162. <https://doi.org/10.1002/jcp.24114>.
- Zheng, K., Wang, C., Yang, J., Huang, H., Zhao, X., Zhang, Z., Qiu, M., 2018. Molecular and genetic evidence for the PDGFRα-independent population of oligodendrocyte progenitor cells in the developing mouse brain. *J. Neurosci.* 38, 9505–9513. <https://doi.org/10.1523/JNEUROSCI.1510-18.2018>.
- Zhou, Y., He, C.H., Herzog, E.L., Peng, X., Lee, C.M., Nguyen, T.H., Gulati, M., Gochoico, B.R., Gahl, W.A., Slade, M.L., Lee, C.G., Elias, J.A., 2015. Chitinase 3-like-1 and its receptors in Hermansky-Pudlak syndrome-associated lung disease. *J. Clin. Invest* 125, 3178–3192. <https://doi.org/10.1172/JCI79792>.

3.4 A NOVEL APPROACH TO GENERATE ENZYME-FREE SINGLE CELL SUSPENSIONS FROM ARCHIVED TISSUES FOR MIRNA SEQUENCING

SLAS Technology 29 (2024) 100133



Contents lists available at ScienceDirect

SLAS Technology

journal homepage: www.elsevier.com/locate/slast

Technical Brief

A novel approach to generate enzyme-free single cell suspensions from archived tissues for miRNA sequencing

Stefan Scheuermann^{a,1}, Sarah Hückler^{b,1}, Annika Engel^c, Nicole Ludwig^d, Philipp Lebhardt^a, Jens Langejürgen^a, Stefan Kirsch^{b,*}

^a Clinical Health Technologies, Fraunhofer IPA, Mannheim, Germany

^b Biomarkers and innovative Technology Development, Division Personalized Tumor Therapy, Fraunhofer ITEM, Regensburg, Germany

^c Chair for Clinical Bioinformatics, Saarland University, Saarbrücken, Germany

^d Human Genetics, Saarland University, University Hospital, Saarbrücken, Germany

ARTICLE INFO

Keywords:

Personalized medicine
Molecular characterization
Micro RNAs (miRNA) sequencing
Single-cell analysis, Enzyme-free tissue dissociation

ABSTRACT

Obtaining high-quality omics data at the single-cell level from archived human tissue samples is crucial for gaining insights into cellular heterogeneity and pushing the field of personalized medicine forward. In this technical brief we present a comprehensive methodological framework for the efficient enzyme-free preparation of tissue-derived single cell suspensions and their conversion into single-cell miRNA sequencing libraries. The resulting data from this study have the potential to deepen our understanding of miRNA expression at the single-cell level and its relevance in the context of the examined tissues. The workflow encompasses tissue collection, RNALater immersion, storage, thawing, TissueGrinder-mediated dissociation, miRNA lysis, library preparation, sequencing, and data analysis. Quality control measures ensure reliable miRNA data, with specific attention to sample quality. The UMAP analysis reveals tissue-specific cell clustering, while miRNA diversity reflects tissue variations. The presented workflow effectively processes preserved tissues, extending opportunities for retrospective analysis and biobank utilization.

1. Introduction

In the era of personalized medicine, the molecular characterization of diseases is increasingly coming into focus, as the response of numerous therapies depends on the molecular genetic alterations of the individual disease [1]. In particular, addressing diseases with intricate and diverse patterns demands a thorough comprehension of intracellular processes to guide effective treatment strategies [2,3]. Single cell micro RNAs (miRNA) profiling can offer a valuable avenue for deciphering the complexities of these intracellular processes in order to enhance the understanding of individual disease [4]. miRNA are a type of small RNA molecules that play important roles in gene regulation [5]. They show tissue-specific expression patterns and are often dysregulated in diseases, making them valuable biomarkers. In the context of disease treatment, miRNA data can be used to: Identify new biomarkers for diagnosis and prognosis, study the effects of treatments on miRNA expression and develop new targeted therapies that target miRNAs

[6–8].

In the last decade, several technologies for single-cell profiling have been developed and used in various applications, revealing many new insights. High-throughput methods for single cell analysis such as flow or mass cytometry and single cell sequencing are ideal for comprehensive identification of single cells based on molecular information [9,10]. These methods have already begun to transform the understanding of complex tissues by enabling the identification of previously unknown cell types and states [11,12]. While single cell sequencing is clearly a powerful diagnostic method, there can be confounding factors in the tissue dissociation process into individual cells that can negatively impact the quality and reliability of the data. One factor is the lack of standardisation, which can lead to significant differences between different research groups and tissue types. Another major challenge is that incomplete disaggregation could bias results towards cell types that are easier to dissociate [13]. A recent study by Wu et al., in which single-cell RNA sequencing was performed using samples from mouse

* Corresponding author at: Fraunhofer ITEM-R, Head of Innovative Molecular Technologies & Biomarker Discovery, Division of Personalized Tumor Therapy, Am Biopark 9, 93053 Regensburg, Germany.

E-mail address: stefan.kirsch@item.fraunhofer.de (S. Kirsch).

¹ Shared first authorship

<https://doi.org/10.1016/j.slast.2024.100133>

Received 29 November 2023; Received in revised form 25 March 2024; Accepted 4 April 2024

Available online 6 April 2024

2472-6303/© 2024 The Author(s). Published by Elsevier Inc. on behalf of Society for Laboratory Automation and Screening. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

kidneys, found that endothelial cells and mesangial cells were under-represented in the scRNA-seq data [14]. Ultimately, long enzymatic digestion times have been shown to alter transcriptomic signatures and induce stress responses that affect cell classification. Solving these challenges would help drive the future of tissue mapping and disease diagnostics through the field of single-cell sequencing. Therefore, new approaches and technologies are urgently needed to ensure the reliability and wide acceptance of single-cell tissue analysis methods. Single cell sequencing methods mainly focus on mRNA, but for single cell miRNA sequencing only very few protocols are available [15–19]. Enzyme-free tissue dissociation is a promising new approach for miRNA-Seq from archived frozen biopsy samples. This method uses mechanical force to dissociate tissue into single cells without the use of enzymes [20,21]. While mRNA molecules are often subjected to degradation processes during long-term storage, miRNA molecules, owing to their compact size, exhibit robust resistance to such degradation. This characteristic allows miRNA profiles to retain a reflection of the native biological landscape, presenting an invaluable opportunity to investigate the molecular states as they existed prior to the archival process. Another reason for emphasizing miRNA analysis lies in the regulatory role of miRNAs in gene expression. By examining miRNA profiles, one not only gains insights into the preservation of the biological state but also indirectly access information about gene expression profiles. miRNAs serve as mediators of post-transcriptional gene regulation, and variations in their abundance can have profound effects on the gene expression landscape. In essence, the choice to prioritize miRNA analysis in archived tissue samples is driven by the unique advantages of miRNA stability over time and their pivotal role as regulators of gene expression. This would be particularly important for the transfer of single-cell diagnostics to human samples in the clinical setting.

This technical brief details the methods and procedures employed for the investigation of human tissue samples preserved in RNALater at -80°C , and processed for single-cell miRNA sequencing. The study aimed to assess a proof of concept workflow for an enzyme free tissue-to-single-cell conversion using a TissueGrinder and subsequent isolation of individual cells for miRNA-Seq analysis.

2. Material and methods

2.1. Human sample collection

Tissue samples from two human donors of the miRNATissueAtlas2 were used [22]. The bodies were obtained as donations for research and teaching purposes. The local institutional reviewer board (Ethikkommission der Ärztekammer des Saarlandes – Nr. 329/20) approved the study. The donor of the liver and spleen sample died on heart failure. The donor of the cortex sample also died on heart failure, but had the concomitant diagnosis of lung cancer. Upon arrival at the Department of Anatomy, between 8- and 48-h post-mortem, tissue samples were collected. Samples were immediately put in RNALater (Thermo Fisher Scientific) and stored at -80°C .

2.2. Enzyme-free tissue dissociation

The thawed tissue sample was placed in a petri dish filled with 1 ml PBS (Life Technologies) and cutted with a scalpel into smaller pieces. Every sample was distributed on two dissociation experiments. The tissue pieces were placed into the grinding gear of a TissueGrinder tube (Fast Forward Discoveries GmbH) and the grinding gear was filled with 800 μl PBS. For tissue dissociation of the spleen sample, the Tissue Grinder program spleen and a 70 μm filter or a modified version of the spleen program (50 % lower rpms at all steps) and a 40 μm filter were used. For the liver sample, the program liver or a modified version of the liver program (50 % lower rpms and 5 s shorter at all steps) and a 70 μm filter were used. For the cortex sample, a modified version of the spleen

program and a 100 μm filter was used. Next, the TissueGrinder tube was centrifuged for 5 min at 300 g. The grinding gear was opened, washed with 3–5 ml PBS and centrifuged again for 1 min. The cell count and the amount of dead cells of the resulting single cell suspension were determined using a Countess (Invitrogen).

2.3. Single cell isolation

A live-dead staining was performed using Trypan Blue Stain 0.4 % (Invitrogen). Cell suspensions were placed on AdcellTM diagnostic slides (Thermo Fisher Scientific). Single living cells were isolated under the microscope in 1 μl PBS using a micromanipulator (Patchman NP2) with pump (CellTram, both Eppendorf) and placed into 2 μl of lysis buffer (0.2 % Triton X-100 (Sigma-Aldrich)) and 4 U recombinant RNase inhibitor (Clontech Takara). Samples were stored at -80°C for up to six months.

2.4. Library preparation for miRNA sequencing

The preparation of the single cell miRNA sequencing libraries was performed as described in Hückler et al., 2021, Nature Comm. [15] for the SBN_CL protocol with minor modifications. First, 5.8S rRNA masking and 3' CL adapter ligation were performed. To prepare the Excess Adapter Removal mix, 5 μM RT primer (Biomers), 0.2 U 1 μl -1 RNase I and 5 mM DTT (both Biozym) were mixed and incubation at 37°C for 30 min, at 70°C for 20 min and forever at 8°C was performed. Then, 0.5 μl 5 U μl -1 Lambda Exonuclease and 0.5 μl 50 U μl -1 5' Deadenylase (both NEB) were added to the mix per sample. 2 μl of the Excess Adapter Removal mix were pipetted to every sample and incubated for 15 min at 30°C , for 15 min at 37°C and forever at 8°C . Ligation of the 5' CL adapter and reverse transcription were performed as described in the original SBN_CL protocol. For the first PCR amplification, 12.5 μl per sample of the following mix were prepared: 1 μM RP1 oligo (Biomers), 1x Phusion HF Buffer (Thermo Fisher Scientific), 0.15 mM dNTPs (Roche), 1.3 U RNase I and 5 mM DTT (both Biozym) and incubated for 30 min at 37°C , for 20 min at 70°C and forever at 8°C . Next, 1 U Phusion Hot Start II DNA Polymerase (Thermo Fisher Scientific) per sample were added. Then, 13 μl of the PCR1 mix were pipetted to every sample (total volume 30 μl) and the following PCR program was used: 30 s at 98°C , 13 cycles of 10 s at 98°C , 30 s at 60°C , 30 s at 72°C , followed by a final elongation of 5 min at 72°C and storage at 8°C . For the second PCR amplification, 23.25 μl per sample of the PCR2 mix were prepared: 0.8 μM RP1 oligo (Biomers), 1x Phusion HF Buffer (Thermo Fisher Scientific), 0.2 mM dNTPs (Roche), 2.35 U RNase I and 5 mM DTT (both Biozym) and incubated for 30 min at 37°C , for 20 min at 70°C and forever at 8°C . Next, 0.5 U Phusion Hot Start II DNA Polymerase (Thermo Fisher Scientific) per sample were added. Then, 23.5 μl PCR2 mix were pipetted in a fresh 0.2 ml PCR tube and 1 μl sample after PCR1 and 0.5 μl of an 100 μM individual index primer were added. For PCR2 the following program was used: 30 s at 98°C , 13 cycles of 10 s at 98°C , 30 s at 67°C , 30 s at 72°C , followed by a final elongation of 5 min at 72°C and storage at 8°C . Finally, Ampure XP Bead (Beckman Coulter) size selection was performed using bead to sample ratios of 1x and 1.6x.

The concentration of every miRNA library was determined using the Qubit High Sensitivity DNA kit (Life Technologies). If necessary, the samples were diluted to $1.8\text{ ng }\mu\text{l}^{-1}$ and the fragment length distribution was evaluated on a Bioanalyzer High Sensitivity dsDNA chip (Agilent Technologies).

2.5. Sequencing

The miRNA libraries were quantified using the KAPA Library Quantification Kit for Illumina (Roche). All libraries were pooled in equimolar concentration and sequenced on an Illumina MiSeq instrument. A final concentration of 18 pM library spiked with 5 % PhiX (Illumina) was sequenced (75 bp, single read) using a MiSeq Reagent Kit

V3 – 150 Cycles (Illumina).

2.6. Bioinformatic analysis

The bioinformatic analysis of the cell miRNA sequencing libraries were conducted following the methodology outlined in Hücker et al., 2021, published in Nature Communications [15]. Here, a concise summary of the described procedures is provided for clarity and reference. In the Bioinformatics analysis, Illumina BCL files were converted to FASTQ format using bcl2fastq (2.19.0.316). Reads with unique molecular identifiers (UMIs) had their UMIs incorporated into the header. After adapter and quality trimming with cutadapt (2.1042), reads were mapped to the human genome (GRCh38) using STAR (2.7.5b). RNA classes were assigned using featureCounts (1.5.2) based on annotations from GENCODE, miRBase, piRBase, and GtRNAdb. Unannotated regions were labeled "unmapped". Reads were hierarchically assigned to categories, with prioritization in cases of ambiguity. miRNA expression was quantified with mirquantify (1.0.1) from miRMaster (2.0) [23], and deduplication was performed with UMI-tools (1.0.0). Data were analyzed with and without subsampling using R and Seurat (4.4.0), and various visualizations and analysis were conducted for miRNA variability, clustering, and enrichment.

3. Results

In order to investigate the viability of converting tissues, initially stored in RNALater at -80°C for two years, into a single-cell suspension a TissueGrinder was used. Subsequently, the isolated individual cells were processed for analysis in miRNA lysis buffer, paving the way for miRNA-Seq investigations. This workflow (Fig. 1) outlines a comprehensive procedure for obtaining high-quality miRNA-Seq data from RNALater-stored tissues. Initially, following rapid autopsy, tissue samples undergo immediate immersion in RNALater, ensuring the stabilization of RNA. Subsequently, these RNALater-preserved tissues are stored at -80°C until required for processing. Upon thawing, tissues undergo effective dissociation into a single-cell suspension facilitated by the TissueGrinder. The TissueGrinder's grinding unit exerts controlled mechanical forces, ensuring gentle yet effective disruption of tissue structures, thereby maximizing cell yield and preserving cellular viability. This method is a key component and contributes to the overall success of the proof of concept workflow, providing a technological advancement in single-cell dissociation for subsequent miRNA-Seq analysis from RNALater-stored tissues. The generated single-cell population is then picked from the suspension using a micromanipulator. Following this,

single cells are lysed in a specialized miRNA lysis buffer, and the resultant lysates are stored at -80°C until they undergo library preparation. Libraries are prepared using a dedicated miRNA-Seq kit, and the generated sequences are subjected to next-generation sequencing (NGS). The resulting miRNA-Seq data are subjected to thorough analysis, enabling the identification of differentially expressed miRNAs and an exploration of their biological significance. This integrated workflow stands as a robust and reproducible approach for single-cell miRNA-Seq analysis from RNALater-stored tissues.

3.1. Sample overview

Table 1 presents a comprehensive summary of the samples included in the study, outlining the experimental protocols for tissue digestion (TG Protocols), cell viability assessments, cell yield data, and the specific cells selected for subsequent miRNA sequencing analysis. The tissues exhibited reduced rigidity compared to fresh specimens, necessitating a cautious and gentle dissociation approach. For the liver and spleen, the method resulted in a uniform single-cell suspension, although the liver exhibited a reduced proportion of viable cells, indicating increased susceptibility to freezing. Conversely, the cortex presented a more challenging scenario, with only a limited number of viable individual cells observed under the microscope, potentially attributed to the higher fat content in brain tissues, thereby increasing the complexity of the dissociation process.

3.2. Quality control miRNA sequencing

Quality control in miRNA sequencing is a critical step that ensures the reliability and accuracy of the obtained miRNA data. This process involves the assessment of sequence data, the removal of low-quality reads, and the validation of miRNA profiles, ultimately guaranteeing the robustness of downstream analysis. Beyond general quality control measures, our approach incorporates a bioinformatical evaluation of sample quality. Prior to downstream analysis, we assessed the distribution of base quality scores, identified and discarded low-quality reads, and validated miRNA profiles. This stringent quality control framework not only enhances the reliability of miRNA expression data but also addresses potential biases introduced during library preparation and sequencing. By implementing these measures, we ensure the generation of high-quality data, thereby fortifying the foundation for subsequent bioinformatic analysis and contributing to the overall robustness of the study's findings.

Fig. 2A presents the total count of miRNA Seq reads obtained per

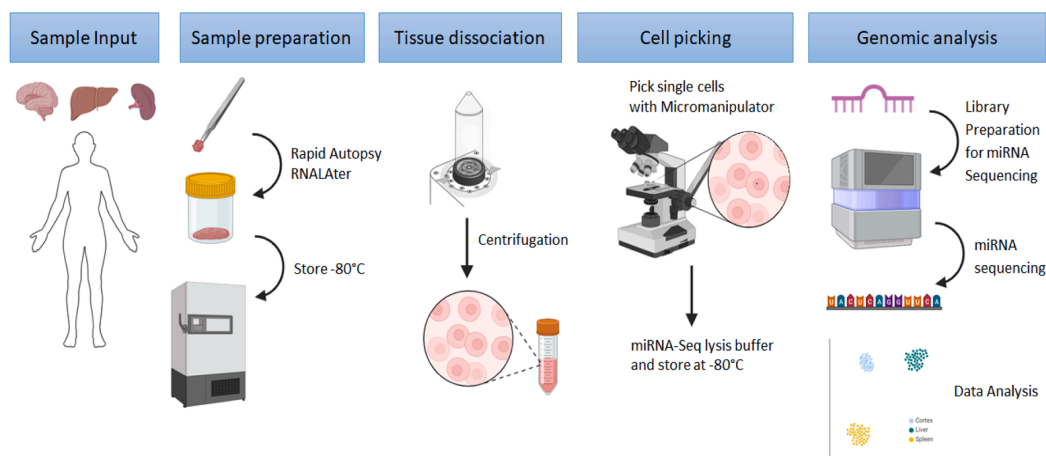


Fig. 1. To examine the feasibility of converting tissues, obtained from Saarbrücken University, preserved in RNALater at -80°C , into a single-cell suspension utilizing a TissueGrinder, with subsequent isolation of individual cells into miRNA lysis buffer for miRNA-Seq analysis. (Image Source: Created with BioRender.com).

Table 1

Sample overview, TG protocols, cell viability, cell yield, picked cells for miRNA sequencing. *In the case of the spleen samples, while our initial intent was to sequence a total of 20 individual cells, it is imperative to note that due to a technical error encountered during the library production process, we were only able to successfully sequence 10 of the intended cells.

Tissue type	RIN	Dissociation protocol	Filter [μ m]	Cell Yield [cells/ml]	Viability	Number of picked single cells [-]	Number of sequenced single cells [-]	Sample ID
Liver	3,9	Liver	70	3,99E+05	30 %	10	20	Liver SC1–10_Liver
		Liver Modified	70	4,80E+05	45 %	10		Liver SC11–20_Liver modified
Spleen	2,4	Spleen	70	2,58E+05	93 %	10	10*	Spleen SC1–10_Spleen
		Spleen modified	40	2,16E+06	90 %	10		Splee SC11–20_Spleen modified
Cortex	5	Spleen modified	100	6,45E+04	64 %	3	3	Cortex SC1–3_Spleen

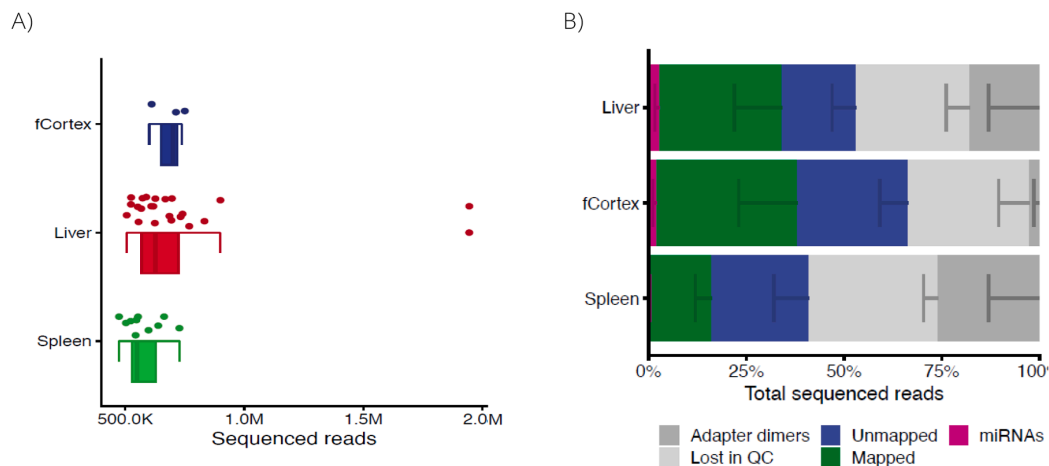


Fig. 2. Summarizing the results of single-cell miRNA sequencing data preprocessing. A) The chart shows the number of sequenced reads for each tissue sample, B) Number of reads lost in quality control (QC) and mapped to the reference genome. The chart also shows the percentage of reads that were adapter dimers and miRNAs.

individual cell, which averages at approximately 600,000 reads per cell. Notably, with the exception of two outliers observed in the liver samples, the distribution of read counts across the cells exhibits a consistent and uniform pattern, aligning with our anticipated expectations. Fig. 2B depicts the initial bioinformatic assessment conducted subsequent to sequencing. During this quality control phase, some reads are excluded due to their inadequate data quality, indicating a substantial level of inaccuracy in the determination of the DNA sequence at the specific position (indicated in light grey, and designated as "lost in QC"). An intriguing characteristic of miRNA libraries is the relatively elevated occurrence of "empty" reads, in which only the adapters were sequenced, without any accompanying insert. Purification methods are ineffectual in eliminating these reads, given that their size differs by only 20 base pairs from those containing an insert (illustrated in dark grey as "adapter dimers"). In our miRNA-Seq protocol, chemically modified adapters and a step to remove unligated 3' adapter are used to reduce the amount of adapter only reads, but even these measures are not able to completely avoid ligation of adapters to each other. Subsequent to the quality control phase, all retained reads are subjected to alignment against the human genome. Unmapped reads, depicted in blue, represent those that failed to align with the human genome. Conversely, green-shaded reads denote successful alignment with the human genome, yet they do not correspond to annotated miRNAs. The pink-shaded region, labeled as "miRNAs," illustrates the proportion of reads that align with annotated miRNAs within the human genome. It is evident from the figure that only a minute fraction of all reads (approximately 1–5 %) align with miRNAs, a pattern consistently observed in other single cell miRNA Sequencing studies [15]. The single-cell miRNA sequencing data were

successfully preprocessed, with a high percentage of reads mapped to the reference genome and a low percentage of adapter dimers and miRNAs. This data is now ready for further analysis to identify differentially expressed miRNAs between cell types.

3.3. Mapping of reads and number of miRNA per cell

Mapping of reads and determination of the number of miRNA per cell is a fundamental aspect of miRNA sequencing analysis. This step involves aligning sequencing reads to a reference genome or miRNA database to identify and quantify miRNA species within individual cells. The results provide insights into the miRNA expression landscape, enabling the characterization of tissue-specific miRNA profiles.

Within this section, a comprehensive breakdown of the reads that successfully aligned to the human genome (corresponding to the green and pink segments in the preceding image is provided (Fig. 3)). In this detailed view, annotations for the specific positions to which the reads mapped are presented. Notably, the pink portion represents the proportion of miRNA reads in relation to the overall reads, with larger proportions observed in liver and cortex samples compared to spleen samples. Additionally, it is evident that a substantial number of reads have mapped to mRNA and intergenic regions. These observed values are consistent with the findings from prior miRNA-Seq dataset, indicating a degree of similarity in the distribution of reads across these regions.

In Fig. 4A we conducted a UMAP analysis with the aim of simplifying the data's complexity. The resulting two-dimensional plot reveals only the values that exhibit the greatest dissimilarity among individual cells

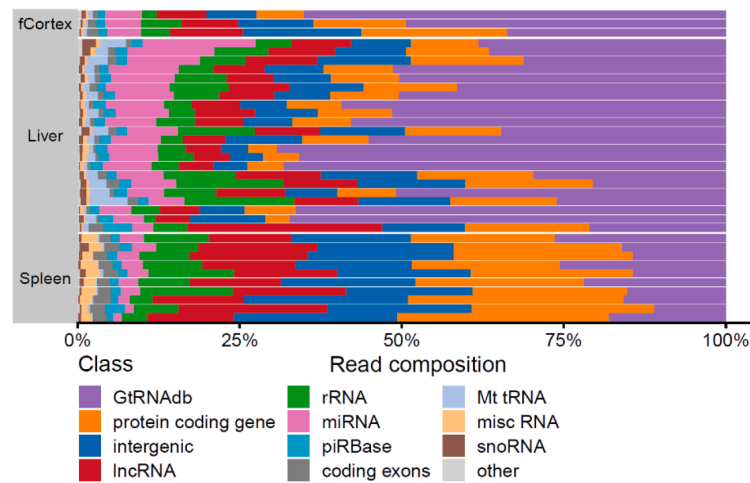


Fig. 3. Mapping of Illumina reads to the human genome.

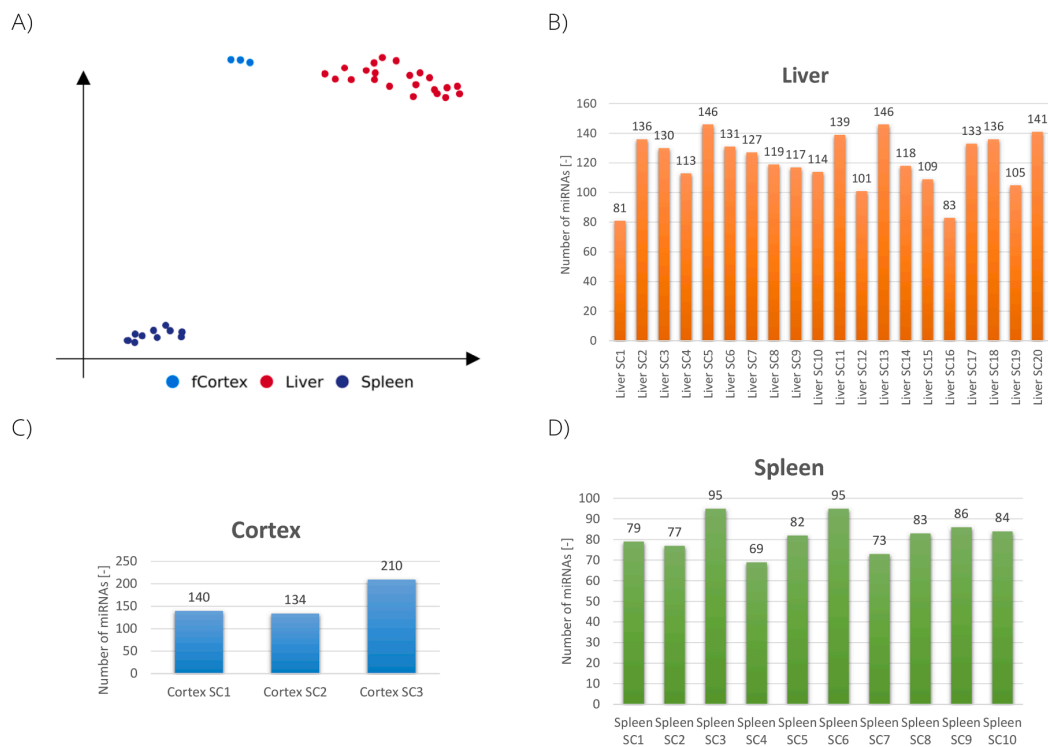


Fig. 4. In Fig. 4 A, depicts UMAP analysis to simplify data complexity, revealing the most dissimilar cell values in a two-dimensional plot. Cells from the same tissue consistently cluster together, indicating heightened similarity among them compared to cells from different tissues. Fig. 4B–D depicts the count of distinct miRNAs within individual cells, segmented by tissue type (rpmm; Cortex: th = 100, Liver: th = 100, Spleen: th = 500). A comprehensive overview of sequencing metrics for single-cell miRNA analysis can be found in the supplement.

within this dataset. The analysis distinctly illustrates that cells originating from the same tissue consistently cluster together, indicating a higher degree of similarity and shared characteristics compared to cells derived from distinct tissues. Within Fig. 4B–D, we provide an insight into the number of distinct miRNAs detected within individual cells, categorized by the three respective tissue types. For single liver and cortex cells, more different miRNAs could be detected compared to single spleen cells. Notably, these numerical findings align with the

miRNA diversity observed in other single-cell studies, underlining the consistency and congruity of our results in comparison to broader research in the field.

4. Discussion

The findings presented in this study have substantial implications for the preservation and analysis of tissues for miRNA research. Our

investigation revealed that tissues stored in RNALater at -80°C for an extended period of approximately two years maintained the potential for extracting single-cell suspensions, demonstrating robust cell viability suitable for single-cell miRNA analysis with the enzyme-free Tissue-Grinder technique.

Single-cell miRNA-seq necessitates the effective integration of two distinct techniques: the isolation of specific individual cells from cultures, tissues, or dissociated cell suspensions, and subsequently, the conversion of the limited cellular RNA into cDNA, followed by the high-throughput sequencing of cDNA libraries [24]. In deviation from earlier published research, this study provides the initial evidence supporting enzyme-free tissue dissociation followed by miRNA sequencing. Enzymes are commonly employed for tissue dissociation, requiring incubation at 37°C for variable durations depending on the tissue type. This temperature activates the cellular transcriptional machinery, potentially leading to alterations in gene expression in response to the dissociation process and other environmental stresses [25,26]. As opposed to enzyme-based approaches, we hypothesize that rapid and enzyme-free dissociation enables improved control over pre-analytic variables, facilitates standardization, and minimizes transcriptional drift [27,28]. Nevertheless, a comprehensive exploration of these aspects awaits further investigation. In this study, we successfully demonstrated the feasibility in principle, utilizing various archived tissue samples.

In case of tissue samples, preparation and isolation of single cells are additional very difficult steps in single-cell analysis [29]. The ability to effectively preserve tissue samples for this duration without compromising cell viability and miRNA integrity is a noteworthy achievement. This feature extends the window of opportunity for researchers to access archived tissue specimens and perform in-depth miRNA-Seq analysis. The detection of specific mutations or mutation spectra and/or perturbed gene expression profiles is crucial for the selection of modern drugs for personalised therapy [30]. This has significant implications for studies requiring long-term sample storage, retrospective analysis, and the utilization of biobanks with archived samples.

Importantly, the miRNA profiles obtained from these preserved tissues displayed a level of comparability with miRNA data from other cell cultures and research results performed with fresh cells [15]. This consistency underscores the reliability and validity of the preservation and isolation methods employed in our study. It indicates that the miRNA profiles derived from the preserved tissues are representative of the biological information they contain and are not significantly altered during the storage period.

5. Limitations

The technical brief faced several limitations that warrant acknowledgment. Firstly, the availability of tissue samples from healthy body donors through a rapid autopsy program inherently restricted the size and diversity of our sample pool. Additionally, the time frame between sample removal, processing and RNA stabilization remains crucial for preserving RNA integrity. Despite efforts to minimize time frames, the logistical challenges entail that not all preserved samples present the necessary quality for downstream analysis. In order to draw biologically meaningful conclusions, a significantly larger number of samples or cells from different healthy donors, comprising different age groups and genders, would have been required. Since we directly used a single cell suspension from disaggregated tissue for miRNA analysis without upstream fluorescence-activated cell sorting (FACS) to select different cell types, we could not analyze a sufficient number of single cells per cell type, which limited the statistical interpretability of our results. Addressing these limitations in future studies could provide a more comprehensive understanding of miRNA expression patterns in different cell types derived from healthy donor tissues, thereby advancing our knowledge of physiological processes of cellular behavior and cellular heterogeneity within organs.

6. Conclusion

In conclusion, our study demonstrates the feasibility of preserving tissues in RNALater at -80°C for an extended duration of approximately two years while maintaining the integrity of miRNA. This preservation method, coupled with the TissueGrinder technique, offers an effective means of converting tissues into single-cell suspensions for miRNA-Seq analysis. Moreover, our results emphasize the comparability of miRNA profiles obtained from these preserved tissues with those derived from other sources, confirming the utility and robustness of this preservation and isolation approach in advancing miRNA-related research. This approach holds the potential to benefit a wide range of scientific investigations and the long-term storage of invaluable tissue specimens. Single cell miRNA-seq could be used to identify specific cell types in which miRNAs are highly expressed. This information could be used to develop new biomarkers for diseases that are specific to certain cell types. Additionally, single cell RNA-seq could be used to study the effects of drugs and other treatments on miRNA expression in individual cells. This information could be used to develop new and more personalized treatments for diseases.

CRedit authorship contribution statement

Stefan Scheuermann: Conceptualization, Investigation, Methodology, Project administration, Resources, Writing – original draft, Writing – review & editing. **Sarah Hücker:** Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft. **Annika Engel:** Data curation, Formal analysis, Software, Visualization, Writing – original draft. **Nicole Ludwig:** Data curation, Formal analysis, Software, Visualization, Writing – original draft. **Philipp Lebhardt:** Conceptualization, Writing – original draft. **Jens Langejürgen:** Project administration, Resources, Writing – original draft. **Stefan Kirsch:** Conceptualization, Funding acquisition, Project administration, Resources, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Stefan Scheuermann reports a relationship with Fast Forward Discoveries GmbH that includes: employment and equity or stocks. Jens Langejürgen reports a relationship with Fast Forward Discoveries GmbH that includes: equity or stocks. Stefan Scheuermann has patent #WO2019145029A1 pending to Fast Forward Discoveries GmbH. S.S. and J.L. are co-founders of the company Fast Forward Discoveries GmbH, which commercializes the TissueGrinder technology. The other authors declare no competing interests. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank Dominik Wermers for his excellent technical assistance.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.slast.2024.100133](https://doi.org/10.1016/j.slast.2024.100133).

References

- [1] Strianese O, Rizzo F, Ciccarelli M, et al. Precision and personalized medicine: how genomic approach improves the management of cardiovascular and neurodegenerative disease. *Genes (Basel)* 2020;11. <https://doi.org/10.3390/genes11070747>.
- [2] Popova AA, Levkin PA. Precision medicine in oncology: in vitro drug sensitivity and resistance test (DSRT) for selection of personalized anticancer therapy. *Adv Therapeut* 2020;3. <https://doi.org/10.1002/adtp.201900100>.
- [3] Liebers N, Bruch P-M, Terzer T, et al. Ex vivo drug response profiling for response and outcome prediction in hematologic malignancies: the prospective non-interventional SMARTrial. *Nat Cancer* 2023. <https://doi.org/10.1038/s43018-023-00645-5>.
- [4] Naldini MM, Casirati G, Barcella M, et al. Longitudinal single-cell profiling of chemotherapy response in acute myeloid leukemia. *Nat Commun* 2023;14:1285. <https://doi.org/10.1038/s41467-023-36969-0>.
- [5] Dexheimer PJ, Cochella L. MicroRNAs: from mechanism to organism. *Front Cell Dev Biol* 2020;8:409. <https://doi.org/10.3389/fcell.2020.00409>.
- [6] Condrat C.E., Thompson D.C., Barbu M.G. et al. (2020) miRNAs as biomarkers in disease: latest findings regarding their role in diagnosis and prognosis. *Cells* 9. <https://doi.org/10.3390/cells9020276>.
- [7] He B, Zhao Z, Cai Q, et al. miRNA-based biomarkers, therapies, and resistance in cancer. *Int J Biol Sci* 2020;16:2628–47. <https://doi.org/10.7150/ijbs.47203>.
- [8] Gareev I, Beylerli O, Yang G, et al. The current state of MiRNAs as biomarkers and therapeutic tools. *Clin Exp Med* 2020;20:349–59. <https://doi.org/10.1007/s10238-020-00627-2>.
- [9] Bendall SC, Nolan GP. From single cells to deep phenotypes in cancer. *Nat Biotechnol* 2012;30:639–47. <https://doi.org/10.1038/nbt.2283>.
- [10] Wang Y, Navin NE. Advances and applications of single-cell sequencing technologies. *Mol Cell* 2015;58:598–609. <https://doi.org/10.1016/j.molcel.2015.05.005>.
- [11] Shekhar K, Lapan SW, Whitney IE, et al. Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell* 2016;166:1308–23. <https://doi.org/10.1016/j.cell.2016.07.054>. e30.
- [12] Villani A-C, Satija R, Reynolds G, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* 2017;356. <https://doi.org/10.1126/science.aah4573>.
- [13] Lombardo JA, Aliaghaei M, Nguyen QH, et al. Microfluidic platform accelerates tissue processing into single cells for molecular analysis and primary culture models. *Nat Commun* 2021;12:2858. <https://doi.org/10.1038/s41467-021-23238-1>.
- [14] Wu H, Kirita Y, Donnelly EL, et al. Advantages of single-nucleus over single-cell RNA sequencing of adult kidney: rare cell types and novel cell states revealed in fibrosis. *J Am Soc Nephrol* 2019;30:23–32. <https://doi.org/10.1681/ASN.2018090912>.
- [15] Hücker SM, Fehlmann T, Werno C, et al. Single-cell microRNA sequencing method comparison and application to cell lines and circulating lung tumor cells. *Nat Commun* 2021;12:4316. <https://doi.org/10.1038/s41467-021-24611-w>.
- [16] Faridani OR, Abdullayev I, Hagemann-Jensen M, et al. Single-cell sequencing of the small-RNA transcriptome. *Nat Biotechnol* 2016;34:1264–6. <https://doi.org/10.1038/nbt.3701>.
- [17] Hagemann-Jensen M, Abdullayev I, Sandberg R, et al. Small-seq for single-cell small-RNA sequencing. *Nat Protoc* 2018;13:2407–24. <https://doi.org/10.1038/s41596-018-0049-y>.
- [18] Wang N, Zheng J, Chen Z, et al. Single-cell microRNA-mRNA co-sequencing reveals non-genetic heterogeneity and mechanisms of microRNA regulation. *Nat Commun* 2019;10:95. <https://doi.org/10.1038/s41467-018-07981-6>.
- [19] Isakovaa A, Neff N, Quake SR. Single-cell quantification of a broad RNA spectrum reveals unique noncoding patterns associated with cell types and states. *PNAS* 2021;1–9. <https://doi.org/10.1073/pnas.2113568118>.
- [20] Soteriou D, Kubánková M, Schweitzer C, et al. Rapid single-cell physical phenotyping of mechanically dissociated tissue biopsies. *Nat Biomed Eng* 2023. <https://doi.org/10.1038/s41551-023-01015-3>.
- [21] Scheuermann S, Lehmann JM, Ramani Mohan R, et al. TissueGrinder, a novel technology for rapid generation of patient-derived single cell suspensions from solid tumors by mechanical tissue dissociation. *Front Med (Lausanne)* 2022;9:721639. <https://doi.org/10.3389/fmed.2022.721639>.
- [22] Keller A, Gröger L, Tschernig T, et al. miRNAAtlas2: an update to the human miRNA tissue atlas. *Nucl Acid Res* 2022;50:D211–21. <https://doi.org/10.1093/nar/gkab808>.
- [23] Fehlmann T, Kern F, Laham O, et al. miRMaster 2.0: multi-species non-coding RNA sequencing analyses at scale. *Nucl Acid Res* 2021;49:W397–408. <https://doi.org/10.1093/nar/gkab268>.
- [24] Saliba A-E, Westermann AJ, Gorski SA, et al. Single-cell RNA-seq: advances and future challenges. *Nucl Acid Res* 2014;42:8845–60. <https://doi.org/10.1093/nar/gku555>.
- [25] van den Brink SC, Sage F, Vértessy Á, et al. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat Method* 2017;14:935–6. <https://doi.org/10.1038/nmeth.4437>.
- [26] Denisenko E, Guo BB, Jones M, et al. Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genom Biol* 2020;21:130. <https://doi.org/10.1186/s13059-020-02048-6>.
- [27] El-Khoury V, Pierson S, Kaoma T, et al. Assessing cellular and circulating miRNA recovery: the impact of the RNA isolation method and the quantity of input material. *Sci Rep* 2016;6:19529. <https://doi.org/10.1038/srep19529>.
- [28] Becker N, Lockwood CM. Pre-analytical variables in miRNA analysis. *Clin Biochem* 2013;46:861–8. <https://doi.org/10.1016/j.clinbiochem.2013.02.015>.
- [29] Ståhlberg A, Kubista M. Technical aspects and recommendations for single-cell qPCR. *Mol Aspect Med* 2018;59:28–35. <https://doi.org/10.1016/j.mam.2017.07.004>.
- [30] Dietel M., Jöhrens K., Laffert M.V. et al. (2015) A 2015 update on predictive molecular pathology and its role in targeted cancer therapy: a review focussing on clinical relevance. *Cancer Gene Ther*:417–30.

3.5 SINGMIR: A SINGLE-CELL MIRNA ALIGNMENT AND ANALYSIS TOOL

Nucleic Acids Research, 2024, **52**, W374–W380<https://doi.org/10.1093/nar/gkae225>

Advance access publication date: 4 April 2024

Web Server issue



SingmiR: a single-cell miRNA alignment and analysis tool

Annika Engel¹, Shusruto Rishik¹, Pascal Hirsch¹, Verena Keller¹, Tobias Fehlmann¹, Fabian Kern^{1,2} and Andreas Keller^{1,2,*}

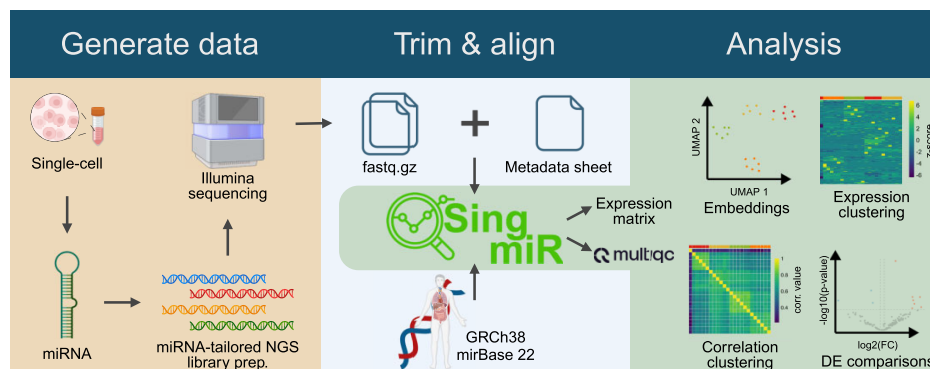
¹Chair for Clinical Bioinformatics, Saarland University, 66123 Saarbrücken, Germany²Department of Clinical Bioinformatics (CLIB), Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research, 66123 Saarbrücken, Germany

*To whom correspondence should be addressed. Tel: +49 681 30268611; Fax: +49 681 30268610; Email: andreas.keller@ccb.uni-saarland.de

Abstract

Single-cell RNA sequencing (RNA-seq) has revolutionized our understanding of cell biology, developmental and pathophysiological molecular processes, paving the way toward novel diagnostic and therapeutic approaches. However, most of the gene regulatory processes on the single-cell level are still unknown, including post-transcriptional control conferred by microRNAs (miRNAs). Like the established single-cell gene expression analysis, advanced computational expertise is required to comprehensively process newly emerging single-cell miRNA-seq datasets. A web server providing a workflow tailored for single-cell miRNA-seq data with a self-explanatory interface is currently not available. Here, we present SingmiR, enabling the rapid (pre-)processing and quantification of human miRNAs from noncoding single-cell samples. It performs read trimming for different library preparation protocols, generates automated quality control reports and provides feature-normalized count files. Numerous standard and advanced analyses such as dimension reduction, clustered feature heatmaps, sample correlation heatmaps and differential expression statistics are implemented. We aim to speed up the prototyping pipeline for biologists developing single-cell miRNA-seq protocols on small to medium-sized datasets. SingmiR is freely available to all users without the need for a login at <https://www.ccb.uni-saarland.de/singmir>.

Graphical abstract



Introduction

One of the best studied classes of noncoding RNAs is microRNAs (miRNAs), 20–25 nt long molecules that regulate potentially up to 60% of the coding genes found in humans, either degrading messenger RNA (mRNA) or repressing protein translation, mainly through 3′-UTR (untranslated region) binding (1,2). miRNAs play a key role in the regulation of cell states and are increasingly relevant biomarkers in new disease diagnostic and therapeutic approaches, e.g. the overexpression of miRNAs in lymphomas (3) or in cancer (4). Despite their importance and possible improvement to diagnostics, we are currently limited to quantifying miRNAs with bulk sequencing experiments only. For instance, we know the mRNA expression patterns of circulating tumor cells from single-

cell studies (5) but so far could not study in detail, through bulk sequencing alone, how the miRNome is shaping rare cell populations. Besides, miRNAs regulate genes and pathways (6), the analysis of which is typically accomplished with software tools (7,8) that are well established for bulk studies but so far are not tailored for single-cell sequencing data. To date, depth and availability of single-cell miRNA sequencing (miRNA-seq) datasets are lacking behind their mRNA counterparts, primarily due to persisting experimental challenges, which render the selective enrichment and subsequent sequencing of miRNA molecules an intricate affair. However, a few protocols exist that are continuously optimized for a better quantitative yield, while there is no commercial option currently on the market. Attempts to estimate miRNA

Received: February 19, 2024. Revised: March 8, 2024. Editorial Decision: March 17, 2024. Accepted: April 2, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

abundance in isolated cells based on primary miRNA expression in single-nucleus sequencing data have produced extremely sparse count matrices that show high variability (9). Even at the higher RNA quantities of standard bulk sequencing approaches, it can be challenging to quantify miRNA levels accurately. For example, the presence of other classes of noncoding RNA (10–12) and the formation of adapter dimers (13) are inevitable sources of bias. Existing technical issues are further complicated by the small RNA input quantities typically required for high-resolution single-cell libraries. The primary ideas to tackle these challenges are to remove excess adapters by combining digestion and size selection (14,15), reduction of adapter dimer formation by adapter chemical modification (16), mitigation of ligation bias by introducing degenerated bases to adapters (11) and polyadenylation (17). Performance optimization is therefore an ongoing effort, for which a recent benchmark provided new quantitative and qualitative grounds (18).

There exists a whole ecosystem of tools to analyze miRNA data: stand-alone tools such as miRDeep* (19) and web services such as miRMaster2.0 (20), sRNAbench/sRNAtoolbox (21), CBS-miRSeq (22) and others [a complete list is provided in miRMaster (20)]. However, currently no such tools exist for single-cell data analysis, which is projected to rapidly increase in the next few years due to newly developed protocols (18). The above-mentioned standard tools cannot be used right away for the single-cell analysis because they do not support the specific parameters used in such protocols, especially regarding the different adapter and barcode sequences as well as unique molecular identifier (UMI) layouts required.

We thus made our analysis pipeline available through a web server. We include the option to perform common comparative analyses, for instance embeddings by popular dimension reduction techniques, correlation and expression-based clustering, differential expression (DE) analyses and more. We aim to enable life science researchers planning to analyze and compare different single-cell miRNA-seq datasets with the necessary toolset, without requiring any computational or bioinformatics expertise.

Materials and methods

The computational workflow of SingmiR consists of two main stages. First, the alignment and trimming pipeline, which removes the adapters specific to the library preparation method used, aligns the reads to the human genome and quantifies miRNA abundances. Second, an optional analysis pipeline computes overview plots and statistics for the processed user dataset. A comprehensive submission interface guides the user through the necessary steps, such as providing data and specifying details for an optional in-depth analysis. The results page allows to download all results for both computational pipelines and displays multiple adjustable visualizations.

Alignment pipeline

SingmiR accepts inputs in the form of gzip compressed fastq files with the option to include a metadata file for downstream analysis. In the current implementation, each fastq file corresponds to one biological cell. Once uploaded, the data are extracted and the UMI sequence is added to the fastq file headers using a Python (version 3.12) script. It is

Table 1. Trimming parameters and metadata from the case study required for SingmiR

Protocol	3' Adapter	5' Adapter	Method	UMI length
SB	Sandberg	Sandberg	SB	8
SB_4N	4N	4N	SB	8
SB_CL	CleanTag	CleanTag	SB	8
SB_C3	Sandberg	C3	SB	6
SBN	Sandberg	Sandberg	SBN	8
SBN_4N	4N	4N	SBN	8
SBN_CL	CleanTag	CleanTag	SBN	8
CL	CleanTag	CleanTag	CleanTag	8
CL_16C	CleanTag	CleanTag	CleanTag	8
CL_4N	4N	4N	CleanTag	8
CL_Block	CleanTag	Block	CleanTag	6
CL_C3	CleanTag	C3	CleanTag	6
CL_Rand	Rand	CleanTag	CleanTag	6
CL_SB	Sandberg	Sandberg	CleanTag	8
CL_UMI6	CleanTag	CleanTag	UMI6	8
4N	4N	4N	4N	0
4N_C3	4N	C3	4N	0
4N_CL	CleanTag	CleanTag	4N	0
CATS	CATS	CATS	CATS	0

trimmed using cutadapt, version 2.10 (23), which utilizes the Illumina Universal adapter sequence, PCR primer sequence and truseq adapter sequences along with the adapter parameters uploaded with the fastq files. The details of the trimming are stated in Table 1. Fastqc, version 0.11.8 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) is used to perform quality metric checking for trimmed and raw reads. For the miRNA quantification, the cutadapt cleaned reads are mapped using bowtie, version 1.3.1 (24) against the human-derived miRNA from miRBase V22 (25) with 15-bp flanks. Fumi_tools, version 0.12.2 (https://ccb-github.cs.uni-saarland.de/tobias/fumi_tools) is used to deduplicate the resulting bam files and in-house scripts were used to produce count and normalized matrices. Due to the recent emergence of transfer RNA-derived fragments (tRF) as a noncoding RNA regulator and their structural similarity to miRNAs (26), tRFs were detected by mapping against transfer RNA with bowtie, subsequent deduplication with Fumi_tools and finally quantified using MINTmap 1.0 (27). The quantification for tRFs is made separately available as downloadable count matrix.

To quickly quantify all other classes of RNA, the cutadapt cleaned reads are also mapped using the STAR algorithm, version 2.5.3a (28) against an index of the human genome (GRCh38) using the parameters `-outSAMstrandField intronMotif -outFilterMultimapNmax 50 -outFilterScoreMinOverLread -outFilterMultimapScoreRange 0 -outFilterMatchNmin 18 -outFilterMatchNminOverLread 0 -outFilterMismatchNoverLmax 0.04 -alignIntronMax 1`. Next, the generated bam file is compared against the GRCh38 annotation in order to quantify the reads mapped to any gene in the human genome. The deduplicated and mapped reads are compared against annotations extracted from GENCODE v25, piRBase v1 and GtRNAdb v18.1, overall containing rRNA, Mt rRNA, snoRNA, snRNA, sRNA, scaRNA, scRNA, piRBase, misc RNA, ribozyme, coding exons, lncRNA, ncRNA and protein-coding genes, using the featureCounts [included in subread in version 1.5.2 (29)] using the options `-F SAF -O -M -f -fracOverlap 0 -s 0`. We perform a reads per million mapped miRNA normal-

ization to account for the differences in reads per file and coverage per miRNA. The trimming, mapping and feature count statistics are compiled into a MultiQC summary report in version 1.20.0 (30) and shared with the user. The rpmm-normalized matrix is used to produce further downstream analyses according to the user selection. Both the MultiQC and the expression matrix files (miRNAs, other RNAs and tRFs) are available for download.

Analysis pipeline

In addition to the raw data, a metadata sheet must be uploaded containing important sample parameters (one sample per row) and additional but optional descriptive information (column variables) for each sample. The user can select categories from the metadata sheet for an optional analysis of the miRNA features, as outlined in the following. To gain a deeper understanding of the dataset, we employ principal component analysis (PCA) and uniform manifold approximation and projection (UMAP) (31) as dimension reduction techniques. The resulting scatter plots are colored according to selected categories. In addition, a UMAP analysis is available for various preselected parameters to reveal higher order relationships between the samples and cells. We aim to discover batch effects and biologically relevant parameters by performing a principal variance component analysis (PVCA) on the selected categories. The residual category sums the variance in the data that cannot be associated with any of the categories provided in the metadata sheet.

A hierarchical clustering using the Euclidean distance and a complete linkage is performed on the standardized rpmm-normalized $\log_2$ -transformed expression values and on the sample correlation values calculated according to either Pearson or Spearman. When using the expression values, we performed clustering for different feature sets, all miRNAs, only expressed miRNA and top miRNAs determined by the highest coefficient of variation. We also provide the corresponding *P*-values calculated with the R function *cor.test* and adjusted with the Benjamini–Hochberg procedure, which controls the false discovery rate at an alpha level, together with a correlation plot indicating the significance of each value ($***P < 0.001$, $**P < 0.01$ and $*P < 0.05$). All results are presented in the form of a heatmap.

Performing DE analysis for all possible categories comes at increased computational costs. Therefore, DE analysis is performed in an interactive manner where the user selects a comparison, and the results are calculated on demand. This way, any comparison deemed valuable can be explored later in more detail. Besides the fold changes, we provide the *P*-values and adjusted *P*-values for *t*-tests and Wilcoxon signed-rank tests. The user can choose between the Benjamini–Hochberg procedure and the Bonferroni correction at a default alpha level of 0.05, the latter of which is known for its strong regulation of family-wise error levels. Additional measures include the effect size according to Cohen's *d* and the area under the receiver operator curve. Graphical representations in the form of volcano and scatter plots accompany the DE analysis in table form.

Web server implementation

The web server providing the front end and the underlying mechanics utilizes the Django Python web framework, version 2.1.7 (<https://djangoproject.com/>) inside Docker containers (<https://www.docker.com/>). Following data submission, we

use the task queue manager Celery, version 5.2.7 (<http://docs.celeryproject.org>) together with the in-memory data structure store Redis, version 5.0 (<https://redis.io/>) to efficiently handle concurrent tasks. Both the alignment and analysis employ a Snakemake pipeline, version 7.30.1 (32). Additionally, the front end of the website uses Bootstrap, version 5.1.3 (<https://getbootstrap.com/>) and Font Awesome, version 6.1.1 (<https://fontawesome.com/>) for design purposes, as well as jQuery, version 3.7.1 (<https://jquery.com/>).

Results

To test the capabilities of our web server, raw data from a previously published high-quality single-cell miRNA-seq dataset were re-analyzed with SingmiR (18). This study covers different sample types, of which we first process the samples from the second stage. In detail, these are 48 samples equal to 48 single-cell profiles obtained from the human breast cancer cell line MCF7, generated with eight different protocols (six samples each). To visualize expression profiles, we consider the results of the downstream analysis module. In addition to this, users have several quality measurements available through the MultiQC report. We restrict our re-analysis here to a single category also examined by the original study, i.e. comparing different versions of the experimental protocol. PCA (Figure 1A) and UMAP (Figure 1B) provide an initial overview of the sample variability and clustering. Color legends of UMAP plots can be modified by various preselected parameters; to exemplify this, we show a single parameter specification. We recognize a clustering for some of the protocols, for example for the protocol 'SBN_CL'. The PVCA indicates that the largest variance in the data can be found for the combination of the 5' adapter and the method (33.8%) variables, next to the combination of the 5' adapter with the UMI length (23%) (Figure 1C). A clustered heatmap of the $\log_2$ -transformed expression reveals miRNA clusters across the samples/cells. Separately clustered heatmaps for different feature sets, i.e. using all given miRNAs, only the expressed miRNAs or the top miRNAs based on the highest variance across all cells, respectively, are computed automatically. Annotation bars above the plot highlight the sample/cell clustering of the individual categories for the top 250 miRNAs (see the 'Materials and methods' section) (Figure 1D). To investigate cell similarity, we calculate the Pearson correlation values for all miRNAs between all cell combinations and display a row- and column-clustered heatmap (Figure 1E). We observe a strong correlation between samples of the protocols '4N', 'SBN_CL', 'SBN' and 'SB'. DE analysis can be performed on demand for any category of interest with at least two groups. We compare 'SBN_CL' against '4N_CL' using volcano plots for fold change against raw and adjusted *P*-values of a *t*-test (Figure 1F and G) and a Wilcoxon rank-sum test (Figure 1H and I). In addition, the effect size, which is calculated using Cohen's *d*, is plotted against the fold change (Figure 1J) and indicates a considerable upregulation of seven miRNAs for the protocol 'SBN_CL'. Yet only one significantly deregulated miRNA (hsa-miR-21-5p) remains for the adjusted *P*-value from the Wilcoxon rank-sum test and five significant deregulated miRNAs (hsa-miR-182-5p, hsa-miR-25-3p, hsa-miR-92a-3p, hsa-miR-183-5p and hsa-miR-21-5p) in the case of the *t*-test.

In the study by Hücker *et al.*, another high-quality dataset is presented, consisting of 48 samples from 8 different cell lines (6 samples each), which was obtained using the most promising protocol 'SBN_CL'. Thus, we also use this dataset to test

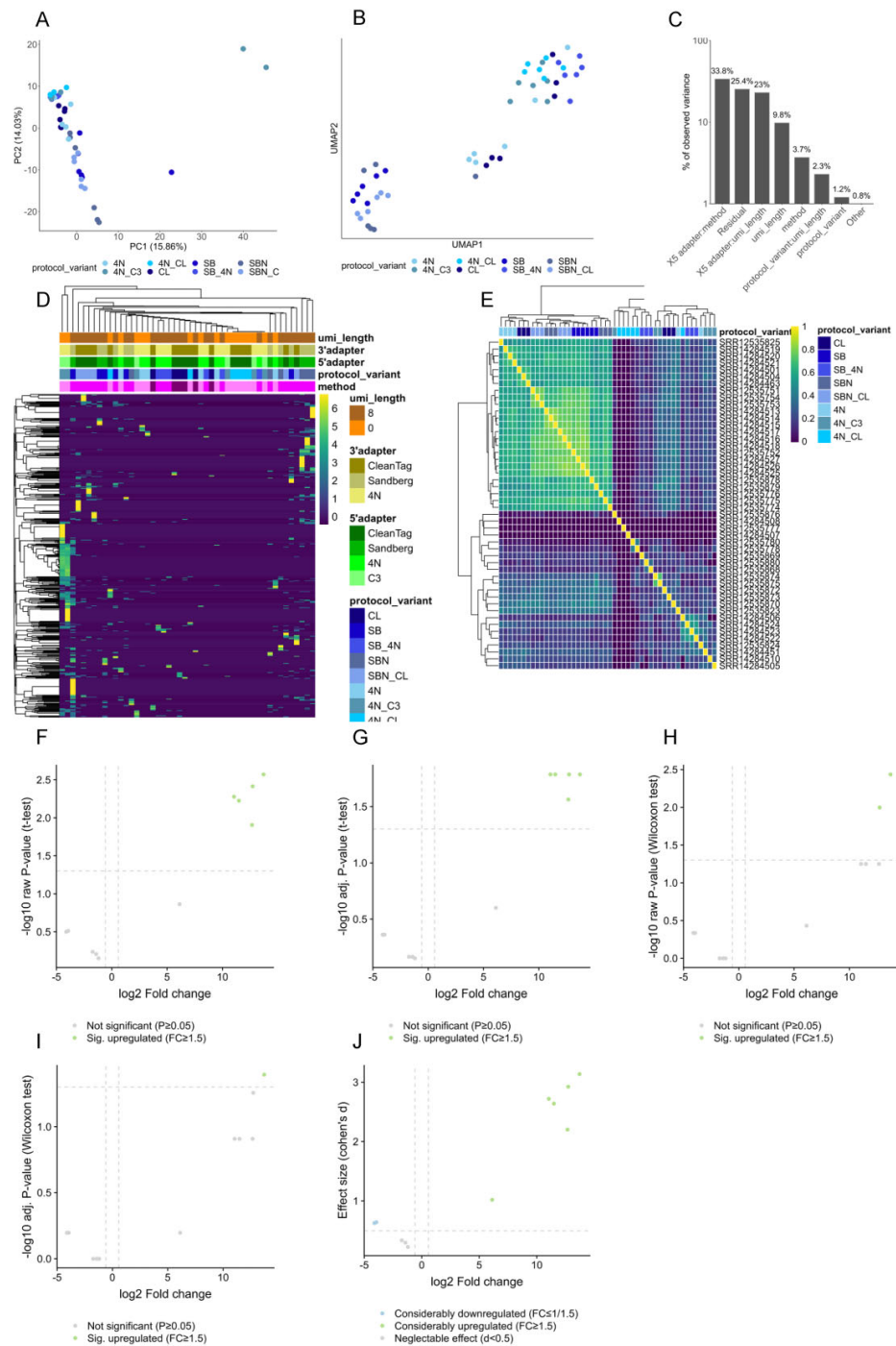


Figure 1. SingmiR results of the stage 2 dataset from (18) using the downstream analysis module. The comparison selected for the DE analysis is protocol 'SBN CL' versus '4N CL'. (A) Results of PCA. (B) Results of a UMAP analysis. (C) Results of the PVCA. (D) Results of the hierarchical clustering of the top 250 standardized log₂-transformed rpmmm-normalized miRNA values selected by the coefficient of variation. (E) Sample correlation calculated with Pearson and grouped by hierarchical clustering. (F) Volcano plot of the raw *P*-values from a *t*-test (used adjusting method is the Benjamini–Hochberg procedure). (G) Volcano plot of the adjusted *P*-values from a *t*-test (used adjusting method is the Benjamini–Hochberg procedure). (H) Volcano plot of the raw *P*-values from a Wilcoxon rank-sum test (used adjusting method is the Benjamini–Hochberg procedure). (I) Volcano plot of the adjusted *P*-values from a Wilcoxon rank-sum test (used adjusting method is the Benjamini–Hochberg procedure). (J) Scatter plot showing the effect size (Cohen's *d*) and the log₂-transformed fold change.

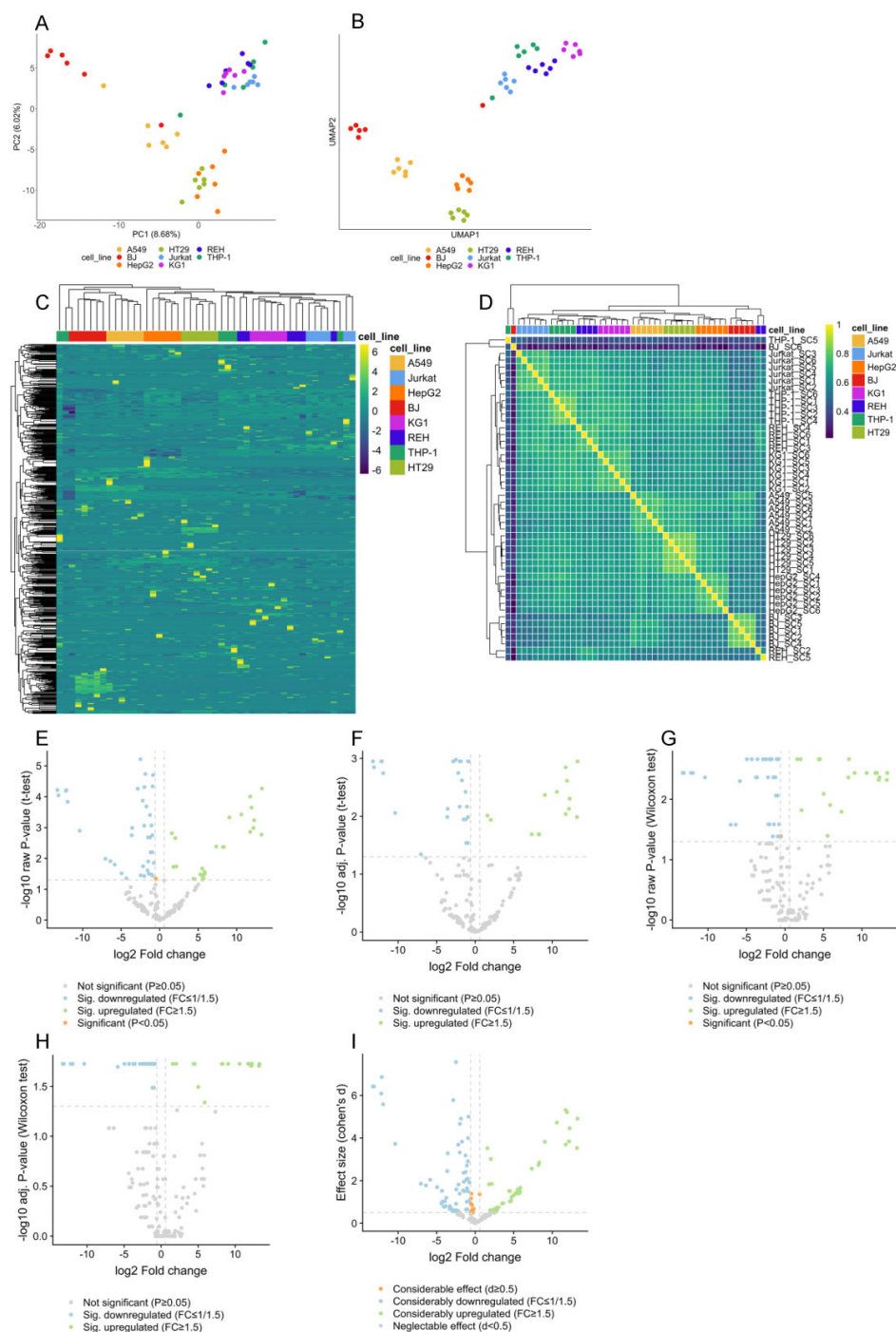


Figure 2. SingmiR results of the first stage 3 dataset from (18) using the downstream analysis module. The comparison selected for the DE analysis is cell line 'HT29' versus 'KG1'. (A) Results of PCA. (B) Results of a UMAP analysis. (C) Results of the hierarchical clustering of the expressed standardized $\log_2$ -transformed miRNAs selected by the coefficient of variation. (D) Sample correlation calculated with Pearson and grouped by hierarchical clustering. (E) Volcano plot of the raw P -values from a t -test. (F) Volcano plot of the adjusted P -values from a t -test (used adjusting method is the Benjamini-Hochberg procedure). (G) Volcano plot of the raw P -values from a Wilcoxon rank-sum test. (H) Volcano plot of the adjusted P -values from a Wilcoxon rank-sum test (used adjusting method is the Benjamini-Hochberg procedure). (I) Scatter plot showing the effect size (Cohen's d) and the $\log_2$ -transformed fold change.

our web server. The resulting embedding for PCA and UMAP shows a good separation of the samples by cell line origin (Figure 2A and B). This is also reflected in the clustering of the expression values and the correlation values (Figure 2C and D). As for this dataset there is only one categorical metadata variable with more than one level available, no PVCA can be performed. A DE analysis between the two cell lines 'HT29' and 'KG1' shows strong and significant fold changes, thus yielding numerous significantly deregulated miRNAs (Figure 2E–I). The results table for the comparison of the cell line 'HT29' to all other cell lines ('A549', 'BJ', 'HepG2', 'Jurkat', 'KG1', 'REH' and 'THP-1') is included in the supplementary material (Supplementary Table S1). By filtering for features that are significantly deregulated across all comparisons, we obtain three upregulated miRNAs (hsa-miRNA-200b-3p, hsa-miRNA-10a-5p and hsa-miRNA-141-3p). Remarkably, these miRNAs have been previously associated with human disease, namely in the context of colorectal and ovarian cancers (33–36). Consequently, it is plausible to observe a high expression predominantly in 'HT29', which had been derived from a human colon adenocarcinoma (37–40). While not pursued further in this work, the presented downstream analysis can serve as a starting point for further investigations.

Discussion

The current absence of a graphical user interface (GUI) based single-cell RNA-seq analysis pipeline for noncoding RNAs so far required researchers to have sufficient computational expertise to eventually generate reliable results. As a comprehensive best-practice and easy-to-use workflow, the here presented web server will hopefully enhance comparability between novel datasets and facilitate fast pilot studies where the analysis is currently conducted by the experimenter (41,42). After having demonstrated the capabilities of SingmiR with previously published datasets (18), the here presented web server is well positioned to serve as a useful tool for upcoming and larger single-cell miRNA-seq studies. Additionally, the development of techniques that aim to predict miRNA activity based on single-cell mRNA data (43) indicates interest in this field. Recent advancements in sequencing technology (44) and the need to study miRNA activity in complex tissues contribute to the increasing emergence of single-cell studies (45). Therefore, we anticipate that our server will not only simplify the analysis of initial pilot stage projects but also serve as a stepping stone to a greater understanding of the single-cell miRNA landscape. Therefore, we hope to provide a useful tool to enable benchmarking studies for single-cell miRNA-seq, as previously done for small RNA-seq methods (11). While we currently only support low-throughput sequencing data, generating up to a few hundred thousand reads for one cell replicate per fastq file, the expansion to high-throughput data marks a promising way forward, supporting novel high-throughput protocols once they have been established in the single-cell community. To further advance the development of SingmiR, we encourage the community to provide us feedback and to propose new features of interest.

Data availability

Data for the case study were made available by Hücker *et al.* on European Nucleotide Archive (<https://www.ebi.ac.uk/ena/browser/home>) under the project PRJNA659784.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

The graphical abstract was created with BioRender.com.

Author contributions: A.E. and S.R. designed and developed the SingmiR web server under the supervision of T.F., P.H., F.K. and A.K. A.E. and S.R. drafted the manuscript. T.F., P.H., V.K., F.K. and A.K. contributed to the manuscript writing.

Funding

Deutsche Forschungsgemeinschaft [469073465]; European Health and Digital Executive Agency [101057548-EPIVIN]. Funding for open access charge: Saarland University.

Conflict of interest statement

The authors have no relevant financial or nonfinancial interests to disclose.

References

- Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
- Di Leva,G., Garofalo,M. and Croce,C.M. (2014) MicroRNAs in cancer. *Annu. Rev. Pathol.*, **9**, 287–314.
- Peng,Y. and Croce,C.M. (2016) The role of microRNAs in human cancer. *Signal Transduct. Target. Ther.*, **1**, 15004.
- Rigg,E., Wang,J., Xue,Z., Lunavat,T.R., Liu,G., Hoang,T., Parajuli,H., Han,M., Bjerkvig,R., Nazarov,P.V., *et al.* (2023) Inhibition of extracellular vesicle-derived miR-146a-5p decreases progression of melanoma brain metastasis via Notch pathway dysregulation in astrocytes. *J. Extracell. Vesicles*, **12**, e12363.
- Negishi,R., Yamakawa,H., Kobayashi,T., Horikawa,M., Shimoyama,T., Koizumi,F., Sawada,T., Oboki,K., Omuro,Y., Funasaka,C., *et al.* (2022) Transcriptomic profiling of single circulating tumor cells provides insight into human metastatic gastric cancer. *Commun. Biol.*, **5**, 20.
- Zhu,M., Gao,J., Lin,X.J., Gong,Y.Y., Qi,Y.C., Ma,Y.L., Song,Y.X., Tan,W., Li,F.Y., Ye,M., *et al.* (2021) Novel roles of an intragenic G-quadruplex in controlling microRNA expression and cardiac function. *Nucleic Acids Res.*, **49**, 2522–2536.
- Aparicio-Puerta,E., Hirsch,P., Schmartz,G.P., Kern,F., Fehlmann,T. and Keller,A. (2023) miEAA 2023: updates, new functional microRNA sets and improved enrichment visualizations. *Nucleic Acids Res.*, **51**, W319–W325.
- Li,J., Han,X., Wan,Y., Zhang,S., Zhao,Y., Fan,R., Cui,Q. and Zhou,Y. (2018) TAM 2.0: tool for microRNA set analysis. *Nucleic Acids Res.*, **46**, W180–W185.
- Ji,J., Anwar,M., Petretto,E., Emanuelli,C. and Srivastava,P.K. (2022) PPMs: a framework to profile primary microRNAs from single-cell RNA-sequencing datasets. *Brief. Bioinform.*, **23**, bbac419.
- Dard-Dascot,C., Naquin,D., Aubenton-Carafa,Y., Alix,K., Thermes,C. and van Dijk,E. (2018) Systematic comparison of small RNA library preparation protocols for next-generation sequencing. *BMC Genomics*, **19**, 118.
- Giraldez,M.D., Spengler,R.M., Etheridge,A., Godoy,P.M., Barczak,A.J., Srinivasan,S., De Hoff,P.L., Tanriverdi,K., Courtwright,A., Lu,S., *et al.* (2018) Comprehensive multi-center assessment of small RNA-seq methods for quantitative miRNA profiling. *Nat. Biotechnol.*, **36**, 746–757.

12. Fuchs, R.T., Sun, Z., Zhuang, F. and Robb, G.B. (2015) Bias in ligation-based small RNA sequencing library construction is determined by adaptor and RNA structure. *PLoS One*, **10**, e0126049.
13. Hardigan, A.A., Roberts, B.S., Moore, D.E., Ramaker, R.C., Jones, A.L. and Myers, R.M. (2019) CRISPR/Cas9-targeted removal of unwanted sequences from small-RNA sequencing libraries. *Nucleic Acids Res.*, **47**, e84.
14. Faridani, O.R., Abdullayev, I., Hagemann-Jensen, M., Schell, J.P., Lanner, F. and Sandberg, R. (2016) Single-cell sequencing of the small-RNA transcriptome. *Nat. Biotechnol.*, **34**, 1264–1266.
15. Hagemann-Jensen, M., Abdullayev, I., Sandberg, R. and Faridani, O.R. (2018) Small-seq for single-cell small-RNA sequencing. *Nat. Protoc.*, **13**, 2407–2424.
16. Shore, S., Henderson, J.M., Lebedev, A., Salcedo, M.P., Zon, G., McCaffrey, A.P., Paul, N. and Hogrefe, R.I. (2016) Small RNA library preparation method for next-generation sequencing using chemical modifications to prevent adapter dimer formation. *PLoS One*, **11**, e0167009.
17. Turchinovich, A., Surowy, H., Serva, A., Zapatka, M., Lichter, P. and Burwinkel, B. (2014) Capture and amplification by tailing and switching (CATS). An ultrasensitive ligation-independent method for generation of DNA libraries for deep sequencing from picogram amounts of DNA and RNA. *RNA Biol.*, **11**, 817–828.
18. Hucker, S.M., Fehlmann, T., Werno, C., Weidele, K., Luke, F., Schlenska-Lange, A., Klein, C.A., Keller, A. and Kirsch, S. (2021) Single-cell microRNA sequencing method comparison and application to cell lines and circulating lung tumor cells. *Nat. Commun.*, **12**, 4316.
19. An, J., Lai, J., Lehman, M.L. and Nelson, C.C. (2013) miRDeep*: an integrated application tool for miRNA identification from RNA sequencing data. *Nucleic Acids Res.*, **41**, 727–737.
20. Fehlmann, T., Kern, F., Laham, O., Backes, C., Solomon, J., Hirsch, P., Volz, C., Müller, R. and Keller, A. (2021) miRMaster 2.0: multi-species non-coding RNA sequencing analyses at scale. *Nucleic Acids Res.*, **49**, W397–W408.
21. Aparicio-Puerta, E., Gomez-Martin, C., Giannoukakis, S., Medina, J.M., Scheepbouwer, C., Garcia-Moreno, A., Carmona-Saez, P., Fromm, B., Pegtel, M., Keller, A., *et al.* (2022) sRNAbench and sRNAtoolbox 2022 update: accurate miRNA and sncRNA profiling for model and non-model organisms. *Nucleic Acids Res.*, **50**, W710–W717.
22. Kesharwani, R.K., Chiesa, M., Bellazzi, R. and Colombo, G.I. (2019) CBS-miRSeq: a comprehensive tool for accurate and extensive analyses of microRNA-sequencing data. *Comput. Biol. Med.*, **110**, 234–243.
23. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.*, **17**, 10.
24. Langmead, B. (2010) Aligning short sequencing reads with Bowtie. *Curr. Protoc. Bioinformatics*, Chapter 11, Unit 11.17.
25. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. and Enright, A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.
26. Venkatesh, T., Suresh, P.S. and Tsutsumi, R. (2016) tRFs: miRNAs in disguise. *Gene*, **579**, 133–138.
27. Lohar, P., Telonis, A.G. and Rigoutsos, I. (2017) MINTmap: fast and exhaustive profiling of nuclear and mitochondrial tRNA fragments from short RNA-seq data. *Sci. Rep.*, **7**, 41184.
28. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
29. Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.
30. Ewels, P., Magnusson, M., Lundin, S. and Kaller, M. (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, **32**, 3047–3048.
31. McInnes, L., Healy, J., Saul, N. and Großberger, L. (2018) UMAP: uniform manifold approximation and projection. *J. Open Source Softw.*, **3**, 861.
32. Koster, J. and Rahmann, S. (2018) Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, **34**, 3600.
33. Collins, K.E., Wang, X., Klymenko, Y., Davis, N.B., Martinez, M.C., Zhang, C., So, K., Buechlein, A., Rusch, D.B., Creighton, C.J., *et al.* (2023) Transcriptomic analyses of ovarian clear-cell carcinoma with concurrent endometriosis. *Front. Endocrinol. (Lausanne)*, **14**, 1162786.
34. Rigoutsos, I., Lee, S.K., Nam, S.Y., Anfossi, S., Pasculli, B., Pichler, M., Jing, Y., Rodriguez-Aguayo, C., Telonis, A.G., Rossi, S., *et al.* (2017) N-BLR, a primate-specific non-coding transcript leads to colorectal cancer invasion and migration. *Genome Biol.*, **18**, 98.
35. Shadbad, M.A., Ghorbaninezhad, F., Hassani, H., Ahangar, N.K., Hosseinkhani, N., Derakhshani, A., Shekari, N., Brunetti, O., Silvestris, N. and Baradaran, B. (2022) A scoping review on the significance of programmed death-ligand 1-inhibiting microRNAs in non-small cell lung treatment: a single-cell RNA sequencing-based study. *Front. Med. (Lausanne)*, **9**, 1027758.
36. Zhang, N., Hu, X., Du, Y. and Du, J. (2021) The role of miRNAs in colorectal cancer progression and chemoradiotherapy. *Biomed. Pharmacother.*, **134**, 111099.
37. Cohen, E., Ophir, I. and Shaul, Y.B. (1999) Induced differentiation in HT29, a human colon adenocarcinoma cell line. *J. Cell Sci.*, **112**, 2657–2666.
38. Ye, F., Tang, H., Liu, Q., Xie, X., Wu, M., Liu, X., Chen, B. and Xie, X. (2014) miR-200b as a prognostic factor in breast cancer targets multiple members of RAB family. *J. Transl. Med.*, **12**, 17.
39. Wen, B., Zhu, R., Jin, H. and Zhao, K. (2021) Differential expression and role of miR-200 family in multiple tumors. *Anal. Biochem.*, **626**, 114243.
40. Liang, Z., Li, X., Liu, S., Li, C., Wang, X. and Xing, J. (2019) MiR-141-3p inhibits cell proliferation, migration and invasion by targeting TRAF5 in colorectal cancer. *Biochem. Biophys. Res. Commun.*, **514**, 699–705.
41. Heumos, L., Schaar, A.C., Lance, C., Litinetskaya, A., Drost, F., Zappia, L., Lucken, M.D., Strobl, D.C., Henao, J., Brunet, F., *et al.* (2023) Best practices for single-cell analysis across modalities. *Nat. Rev. Genet.*, **24**, 550–572.
42. Smith, C.M. and Hutvagner, G. (2022) A comparative analysis of single cell small RNA sequencing data reveals heterogeneous isomiR expression and regulation. *Sci. Rep.*, **12**, 2834.
43. Olgun, G., Gopalan, V. and Hannehalli, S. (2022) miRSCAPE—inferring miRNA expression from scRNA-seq data. *iScience*, **25**, 104962.
44. Qu, H.Q., Kao, C. and Hakonarson, H. (2024) Single-cell RNA sequencing technology landscape in 2023. *Stem Cells*, **42**, 1–12.
45. Zhang, J., Liu, L., Xu, T., Zhang, W., Zhao, C., Li, S., Li, J., Rao, N. and Le, T.D. (2021) Exploring cell-specific miRNA regulation with single-cell miRNA-mRNA co-sequencing data. *BMC Bioinformatics*, **22**, 578.

Received: February 19, 2024. Revised: March 8, 2024. Editorial Decision: March 17, 2024. Accepted: April 2, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

3.6 EXPERIMENTAL CAPTURE OF MIRNA TARGETOMES: DISEASE-SPECIFIC 3' UTR LIBRARY-BASED MIRNA TARGETOMICS FOR PARKINSON'S DISEASE


www.nature.com/emm

ARTICLE OPEN



Experimental capture of miRNA targetomes: disease-specific 3' UTR library-based miRNA targetomics for Parkinson's disease

Martin Hart^{1,6}, Fabian Kern^{2,3,6}, Claudia Fecher-Trost⁴, Lena Krammes¹, Ernesto Aparicio², Annika Engel², Pascal Hirsch², Viktoria Wagner², Verena Keller^{2,5}, Georges Pierre Schmartz², Stefanie Rheinheimer¹, Caroline Diener², Ulrike Fischer¹, Jens Mayer¹, Markus R. Meyer⁴, Veit Flockerzi², Andreas Keller^{2,3,6} and Eckart Meese^{1,6}

© The Author(s) 2024

The identification of targetomes remains a challenge given the pleiotropic effect of miRNAs, the limited effects of miRNAs on individual targets, and the sheer number of estimated miRNA–target gene interactions (MTIs), which is around 44,571,700. Currently, targetome identification for single miRNAs relies on computational evidence and functional studies covering smaller numbers of targets. To ensure that the targetome analysis could be experimentally verified by functional assays, we employed a systematic approach and explored the targetomes of four miRNAs (miR-129-5p, miR-129-1-3p, miR-133b, and miR-873-5p) by analyzing 410 predicted target genes, both of which were previously associated with Parkinson's disease (PD). After performing 13,536 transfections, we validated 442 of the 705 putative MTIs (62.7%) through dual luciferase reporter assays. These analyses increased the number of validated MTIs by at least 2.1-fold for miR-133b and by a maximum of 24.3-fold for miR-873-5p. Our study contributes to the experimental capture of miRNA targetomes by addressing i) the ratio of experimentally verified MTIs to predicted MTIs, ii) the sizes of disease-related miRNA targetomes, and iii) the density of MTI networks. A web service to support the analyses on the MTI level is available online (<https://ccb-web.cs.uni-saarland.de/utr-seremato>), and all the data have been added to the miRATBase database (<https://ccb-web.cs.uni-saarland.de/miratbase>).

Experimental & Molecular Medicine (2024) 56:935–945; <https://doi.org/10.1038/s12276-024-01202-5>

INTRODUCTION

miRNAs posttranscriptionally regulate the expression of target genes mainly by binding to the 3' untranslated region (3'UTR) and rarely to the 5'UTR or the open reading frame of their target genes^{1–3}. Depending on the complementarity of the miRNA seed region and the corresponding 3'UTR sequence of the target genes, posttranscriptional regulation can lead to mRNA destabilization, degradation, or inhibition of protein translation^{4–7}. Therefore, miRNAs regulate many target genes and are associated with a variety of cellular processes and different human diseases^{8–10}. Although hundreds to thousands of papers connect miRNAs to genes involved in diseases, the basis for this link, (i.e., evidence of a miRNA targeting a gene and affecting a given disease), is often missing or purely descriptive. Recently, miRNAs have been proposed to be powerful regulators of ribosome biogenesis, suggesting new applications for microRNA-mimic chemotherapeutics¹¹. Consequently, miRNAs have been identified as biomarkers and candidates for these and other therapeutic approaches^{12–15}. However, incorporating miRNAs in different therapeutic strategies requires the best possible knowledge of their target genes to avoid adverse side effects^{16,17}. Notably, the exact mechanism through which miRNAs target genes is not fully

understood. For example, studies suggest that argonaute binding within 3'-untranslated regions poorly predicts gene repression¹⁸.

Unfortunately, the full spectrum of miRNA–target gene interactions (MTIs) has not been fully elucidated due to the sheer number of more than 40 million MTIs¹⁹. This number also explains why knowledge about MTIs is mostly based on prediction algorithms that frequently yield heterogeneous and partially unspecific results²⁰. Approaches to providing experimental evidence for MTIs are frequently only descriptive and complicated by the pleiotropic effect of each miRNA and by rather limited effects on each single target²¹. Ideally, functional assays would take into account the different cellular backgrounds since MTIs are likely specific to tissues, cell types and disease states²². This, however, complicates the task of collecting further experimental evidence for MTIs since the potential MTIs would each have to be tested in specific cellular contexts. Although high-throughput methods such as CLIP-seq (cross-linking immunoprecipitation-high-throughput sequencing) or CLASH (cross-linking, ligation, and sequencing of hybrids) have the power to identify an enormous number of potential MTIs, they cannot discriminate between a functional MTI causing the downregulation of the target gene at the protein level and a short binding of the miRNA to its target

¹Human Genetics, Saarland University, 66421 Homburg, Germany. ²Clinical Bioinformatics, Saarland University, 66123 Saarbrücken, Germany. ³Helmholtz Institute for Pharmaceutical Research Saarland (HIPS)–Helmholtz Centre for Infection Research (HZI), Saarland University Campus, Saarbrücken, Germany. ⁴Department of Experimental and Clinical Pharmacology & Toxicology, Institute of Experimental and Clinical Pharmacology and Toxicology, Center for Molecular Signaling (PZMS), Saarland University, 66421 Homburg, Germany. ⁵Department for Internal Medicine II, Saarland University Hospital, 66421 Homburg, Germany. ⁶These authors contributed equally: Martin Hart, Fabian Kern, Andreas Keller, Eckart Meese. [✉]email: martin.hart@uks.eu

Received: 21 August 2023 Revised: 12 January 2024 Accepted: 30 January 2024
Published online: 1 April 2024

followed by the release of mRNA and miRNA from the RISC without any functional relevance. Here, we aimed to validate the complete target spectra of specific miRNAs under standardized conditions in a specific disease context. Since miRNAs have increasingly gained attention in Parkinson's disease (PD) research, we selected this disease, which is the second most common neurodegenerative disorder, to test our targetome capturing technique^{23–25}.

We calculated a maximum of 44,571,700 miRNA–target gene interactions (MTIs) based on 19,379 protein-coding human genes (GENCODE Release version 42) and an estimated number of 2300 human miRNAs²³. This calculation does not acknowledge the possibility of obtaining different transcripts from single genes or from different tissue- and cell type-specific MTIs. Since the first reviews on methods for miRNA target validation in 2008²¹, reporter assays are still considered the primary method for functionally analyzing the binding of a miRNA to a single target. However, the efficacy of this method is limited. Even in a simplified scenario with a single miRNA tested for all its predicted interactions to human protein-coding genes, our semiautomated high-throughput miRNA interaction reporter assay (HiTmiR) as it is used in this study enables at maximum the analysis of 120 3'UTR sequences within two weeks (assuming 20 target UTRs per 96-well plate, at an analysis time of 12 plates per week, and four replicates). Hence, the analysis of all 8,069 highly confidently predicted MTIs for miR-129-5p would take more than two years. To narrow the scope of this task, we limited our analysis to Parkinson's disease, as described above^{26,27} (Fig. 1a). Notably, we selected the miRNAs and the mRNAs based only on their association with PD without considering whether the mRNAs were predicted as targets of the chosen miRNAs. Since there are no publicly available human 3'UTR libraries that are enriched for genes belonging to a given disease phenotype and/or specific biological pathway, we generated a 3'UTR library that contained 410 potential miRNA target genes, all of which were reported to be involved in signaling pathways associated with PD.

In detail, these pathways included 14 PD- and dopamine-associated signaling pathways, including the signaling pathway dopaminergic synapse (KEGG, hsa04728), dopamine metabolic process (Gene Ontology, GO:0042417), and dopaminergic neurogenesis (WikiPathways, WP2855). We synthesized the 3'UTRs of the 410 genes as approx. 675 bp long fragments to account for length restrictions of the reporter gene plasmids, as addressed in our previous studies²³. The fragments were designed with a 30 bp overlap between consecutive fragments of a 3'UTR. The resulting library with 1,280 3'UTR sequences cloned in the reporter plasmid pMIR-RNL-TK (Supplementary Table 1) enables target screening of any miRNA. With this study, we aimed to validate the disease-specific targetomes of four miRNAs. Using a library of 3'UTRs specific to PD allowed us to assess the effect of disease-associated miRNAs on the regulation of key pathological players, enabling the elucidation of disease-associated miRNA–target–gene networks. The advantages of this method include flexibility in testing numerous disease-associated miRNAs; reproducibility, reliability and comparability of results due to the use of a standardized test method with identical 3'UTR sequences; and the possibility of further investigating the fundamental mechanisms of synergistic miR targeting in future studies.

MATERIALS AND METHODS

Cell lines

The cell lines HEK 293 T (ACC 635) and SH-SY5Y (ACC 209) were purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ). The authenticity of the cell lines was confirmed by short tandem repeat (STR) fingerprinting by the supplier. The cell lines were cultivated in DMEM (Life Technologies, Darmstadt, Germany) supplemented with penicillin (100 U/ml), streptomycin (100 µg/ml) and either 10% [v/v] FCS for 293T

cells or 20% [v/v] FCS for SH-SY5Y cells. Subculturing was performed two times a week for no more than three months after the start of culture.

miRNA expression plasmids

Sequence inserts representing miRNA precursor sequences with additional nucleotides up- and downstream of the precursor sequence for miR-129, miR-133b, and miR-873 were either synthesized by Eurofins Genomics (Ebersberg, Germany) or amplified by PCR using sequence-specific primers with human genomic DNA as a template and subsequently cloned and inserted into the expression plasmid pSG5 (Agilent Technologies, Santa Clara, California) using the EcoR I and BamH I restriction sites. The expression plasmid pSG5-miR-129 harbors the nucleotides 128207756–128208053[+] of human chromosome 7 (GRCh38/hg38), the pSG5-miR-133b nucleotides 52148873–52149091[+] of human chromosome 6 (GRCh38/hg38), and the pSG5-miR-873 nucleotides 2888786–28889058[-] of human chromosome 9 (GRCh38/hg38). Cloned sequences of the miRNA expression plasmids were verified by Sanger sequencing. Cloning of pSG5-miR-133b was performed as described in our previous study¹⁹.

3'UTR reporter plasmids

As positive controls for regulatory effects, sequence inserts containing two copies of the complementary sequence for the respective miRNAs as well as additional random flanking nucleotides with no self-complementary sites were generated in silico. Respective sequences were synthesized by Eurofins Genomics (Ebersberg, Germany) and cloned and inserted into the reporter plasmid pMIR-RNL-TK using the SpeI and SacI restriction sites for the positive controls of miR-129-5p, miR-129-3p, and miR-873-5p. Cloning of the pMIR-miR-133b positive control was performed employing the SpeI and NaeI restriction sites. Reporter plasmids of the PD-specific 3'UTR library containing approx. 675 bp long fragments of 3'UTR sequences of PD-associated target genes were synthesized and cloned and inserted into the reporter plasmid pMIR-RNL-TK employing SpeI and SacI restriction sites. The synthesis of 28 3'UTR sequence fragments failed. The remaining 1,280 3'UTR sequences were cloned and inserted into the reporter plasmid pMIR-RNL-TK, and the correct cloning of all reporter constructs was verified by Sanger sequencing. A complete list of all 3'UTR sequences of the PD-specific 3'UTR library, including the respective NM accession numbers, is given in Supplementary Table 1.

Northern blot analysis

Ectopic miRNA expression and correct miRNA processing in 293 T cells were verified by Northern blotting. To do so, 293 T cells were seeded at 2.5×10^5 cells per well in a 6-well plate. After 24 h, the cells were transfected with 2 µg of empty pSG5 plasmid or 2 µg of miRNA expression plasmid using Polyfect transfection reagent according to the manufacturer's instructions (Qiagen, Hilden, Germany). After an additional 48 h, the cells were lysed using QiAzol Lysis Reagent (Qiagen, Hilden, Germany). Total RNA was isolated using a miRNeasy Mini Kit following the manufacturer's protocol (Qiagen, Hilden, Germany). Northern blotting was performed employing radiolabeled DNA probes specific for miR-129-5p (5'-CTTTTTCGGTCTGGGCTTGCCTGTCTC3'), miR-129-1-3p (5'-AAGCCCTTACCCCAAAAAGTATCCTGTCTC3'), and miR-873-5p (5'-GCAG-GAAGCTGTGAGTCTCTCTGTCTC3') as described previously¹⁹.

Library-based miRNA target gene reporter-assay (LiMTaR)

To construct the PD-specific 3'UTR library, we selected 416 target genes from 14 PD-associated pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG)²⁸, Gene Ontology (GO)²⁹, WikiPathways³⁰ and Reactome³¹ databases. The 3'UTR sequences of the respective target genes were split into approximately 675 bp fragments, with an overlap of 30 bp between consecutive fragments. The respective sequences were synthesized and cloned and inserted into pMIR-RNL-TK. Twenty-eight 3'UTR sequence fragments were excluded due to critical sequence motifs that prevented correct synthesis. In total, we generated a PD-specific 3'UTR library that included 1,280 reporter plasmids containing 3'UTR sequences of the 410 PD-associated genes. To test the impact of the miRNAs miR-129-5p, miR-129-1-3p, miR-133b and miR-873-5p on PD-related genes, we selected the corresponding 3'UTR reporter plasmids harboring the respective binding site(s) for each miRNA from the 3'UTR library. To verify the validity of the HiTmiR-Assay, each 96-well plate contained 2 positive controls (sensor reporter plasmids harboring 2 sequences complementary to the respective miRNA) at various positions to exclude positioning effects. The results of these positive controls are depicted in Supplementary Fig. 1. High-throughput analysis of reporter constructs of our

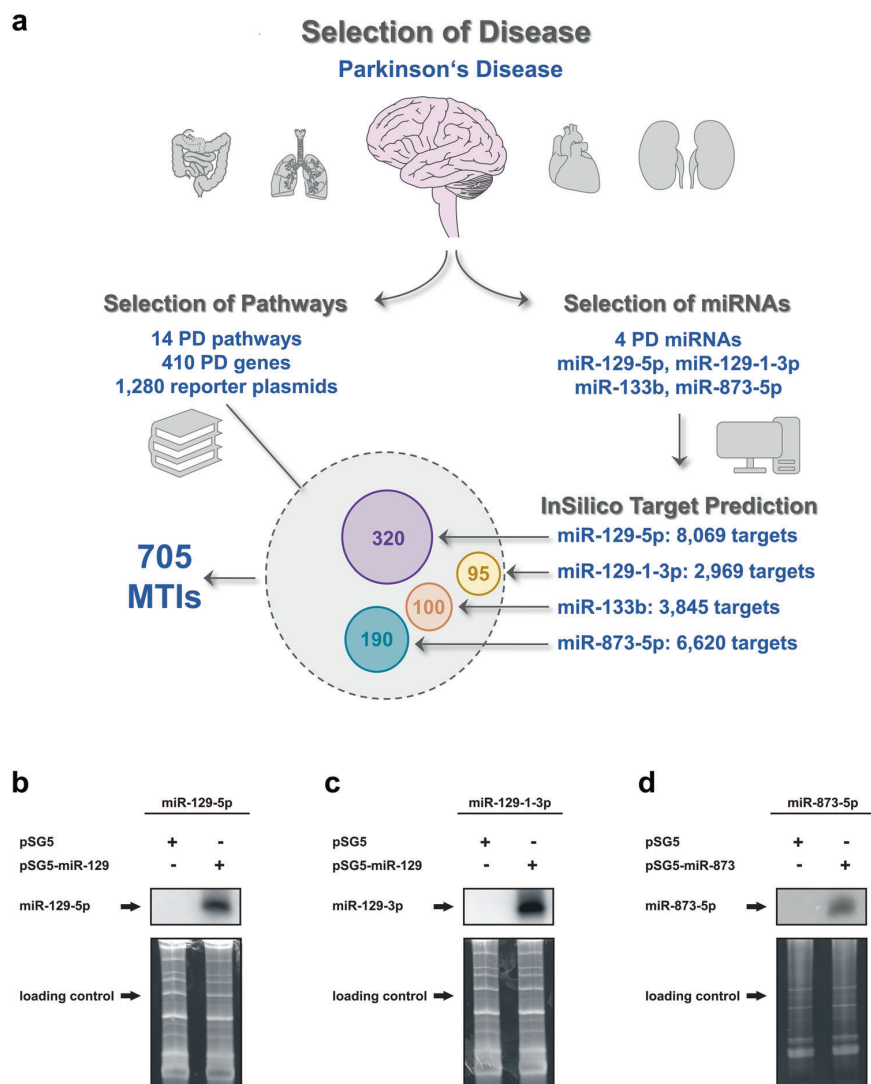


Fig. 1 Library-based miRNA target gene reporter assay. **a** Selection of MTIs. The MTIs were defined in a staged concept. Using PD as the disease phenotype, we collected data on 14 PD pathways involving 410 genes. From these genes, 1280 3'UTR reporter constructs were derived considering length constraints. We selected four relevant PD miRNAs unrelated to the analyzed genes. Via computational prediction, for the miR-129-5p 320 3'UTR reporter construct, 95 3'UTR constructs were used for miR-129-1-3p, 100 3'UTR constructs were used for miR-133b, and 190 3'UTR constructs were identified for miR-873-5p harboring at least one canonical binding site on the respective miRNA. In total, 705 MTIs for the four chosen miRNAs were analyzed. The 3'UTRs of the target genes harboring more than one exclusive miRNA binding site are shown in Supplementary Table 7, and the overlaps are visualized in Supplementary Fig. 1. **b** Northern blot analysis of miR-129-5p. 293 T cells were transfected with miRNA expression plasmids containing the sequence of miR-129-5p, miR-129-1-3p, or miR-873-5p. After 48 h, RNA was isolated, and Northern blot analysis was performed with specific probes against the indicated miRNAs. **c** Northern blot analysis of miR-129-1-3p. 293 T cells were transfected with a miRNA expression plasmid containing the sequence of miR-129-1-3p. After 48 h, RNA was isolated, and Northern blot analysis was performed with specific probes against miR-129-1-3p. **d** Northern blot analysis of miR-873-5p. 293 T cells were transfected with a miRNA expression plasmid containing the sequence of miR-873-5p. After 48 h, RNA was isolated, and Northern blot analysis was performed with specific probes against miR-873-5p.

PD-specific 3'UTR library was performed by a liquid handling system as described by our group previously²³. In brief, 293T cells were seeded at 3.2×10^4 cells per well in a 96-well plate using the liquid handling system epMotion® 5075 (Eppendorf, Hamburg, Germany). After 24 h, the cells were transfected with 50 ng/well of either the reporter plasmid pMIR-RNL-TK, with or without an insert, and 200 ng/well of the miRNA expression plasmid containing either the respective miRNA or no insert with PolyFect transfection reagent (Qiagen, Hilden, Germany). After an additional 48 h, the cells were lysed in passive lysis buffer (Promega, Madison, WI, USA). Luciferase substrates from the Dual-Luciferase® Reporter Assay System

(Promega, Madison, WI, USA) were added to the cell lysates, and luciferase activity was measured using a GloMax Navigator microplate luminometer (Promega, Madison, WI, USA). High-throughput dual-luciferase assays were performed four times in technical duplicates. Statistical significance was calculated by Welch's *t* test in GraphPad Prism 9.

Quantitative real-time PCR

Ectopic expression of miR-129-5p in SH-SY5Y cells was verified by quantitative real-time PCR (Supplementary Fig. 2). SH-SY5Y cells were

seeded at 4.5×10^5 cells per well in a 6-well plate. After 24 h, the cells were transfected with either Allstars Negative Control (ANC; Qiagen, Hilden, Germany) or the miR-129-5p miScript miRNA Mimic (MIMAT0000242, sequence: 5'CUUUUUGCGGUCUGGGCUUGC3'; Qiagen, Hilden, Germany) using HiPerFect Transfection Reagent (Qiagen, Hilden, Germany). After an additional 48 h, the cells were lysed using QiAzol Lysis Reagent (Qiagen, Hilden, Germany). Total RNA was isolated using miRNeasy Mini Kit following the manufacturer's protocol (Qiagen, Hilden, Germany). Total RNA (150 ng) was reverse transcribed using miScript II RT Kit (Qiagen, Hilden, Germany). qPCR was performed using the miScript Primer Assay (Qiagen, Hilden, Germany) with primers specific for miR-129-5p via the StepOnePlus Real-Time PCR System (Applied Biosystems, Foster City, United States). RNU6B (Qiagen, Hilden, Germany) served as an endogenous control. The statistical significance of differences among three independent replicates was calculated by Student's t test in GraphPad Prism 9.

Western blot analysis

For Western blot analysis, SH-SY5Y cells were transfected with the miR-129-5p mimic as described above. After 48 h, the cells were harvested, lysed in 2x sample buffer (130 mM Tris/HCl, 6% [v/v] SDS, 10% [v/v] 3-mercapto-1,2-propanediol, and 10% [v/v] glycerol) and sonicated three times for three seconds each. Total protein extract (10 µg) was electrophoresed in a 4–15% TGX gel (Bio-Rad Laboratories, Inc., Hercules, California, USA). The proteins were electroblotted onto nitrocellulose membranes (Whatman, GE Healthcare, Freiburg, Germany). Unspecific antibody binding was blocked by preincubation of the nitrocellulose membrane in 5% TBS milk with 0.1% Tween 20 for 30 min. SNCA was detected by a polyclonal rabbit antibody (#2642), COMT by a monoclonal rabbit antibody (#14368), CLOCK by a monoclonal rabbit antibody (#5157), and AKT3 by a monoclonal rabbit antibody (#14982), all of which were purchased from Cell Signaling Technology (Danvers, MA, USA). α-Tubulin served as an endogenous control and was detected by a monoclonal rabbit antibody (#2125 Cell Signaling Technology, Danvers, MA, USA). A secondary anti-rabbit antibody was purchased from Sigma–Aldrich (A0545; Sigma Aldrich, Munich, Germany).

Identification of miRNA binding sites in 3'UTR sequences in silico and secondary structure matching

In this analysis, we included miR-7-5p and -34a-5p from our previous study, which were also tested with our standardized high-throughput miRNA interaction reporter assay (HiTmiR) in the context of PD²³. For miR-129-5p, miR-129-1-3p, miR-133b, miR-873-5p, miR-7-5p and miR-34a-5p, we aligned the 3'UTR sequences of the target genes with the seed sequences of the respective miRNAs, focusing on the 8-mer, 7-mer-m8, 7-mer- A1, and 6-mer binding sites. Next, we searched for exact hits of these canonical binding sites in the 3'UTR sequences for each gene. For each 3'UTR sequence, we computed the secondary structure using RNAfold³². Here, we computed the minimum free energy and partition function; we did not allow GU pairs at the end of the helices and avoided isolated base pairs. We matched the dot-bracket notation of the folded 3' UTR sequence for each binding site and computed the percentage of bound bases within the local secondary structure (identified by “(” or “)”). We then related the percentage of bases bound within each binding site to the corresponding reduction in relative light units (RLUs) that was measured for each individual MTI.

Calculation of the coverage score of binding sites in the 3'UTRs

Using the Graph Modeling Language representation obtained from ViennaRNA (version 2.5.1), we determined a coverage value that provides information on how many bases are found in the 2D local area around the seed region. For each base in a seed sequence, a value that represents how many bases are located within a circle with a radius (r) around the seed base is assigned; this process is performed by using the Euclidean distance. To obtain the coverage value, we summed the values of the individual bases of the seed and divided this value by the number of bases in the seed to normalize seeds of different lengths. If more than one seed region was found, we always considered the seed with the smallest coverage value, corresponding to the most accessible site, for further analysis. For the calculation of correlations, we used Pearson correlation.

Formally, the coverage score is defined as follows: Let $B_r(x) := \{x \in R^2 \mid ||x||_2 \leq r\}$ be the 2-dimensional ball of radius $r \in \mathbb{R}_{\geq 0}$ and

$$X_A(x) := \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases}$$

the indicator function. The RNA secondary structure of a UTR sequence is given as a graph G with $G = (V, E)$, where V denotes the vertices and E the edges. In our case, the vertices correspond to the bases and the edges to the molecular compounds and to the hydrogen bonds. There is an edge $e_i \in E$ between two vertices $v_i, v_{i+1} \in V$ since the two bases are next to each other in the RNA secondary structure, so the edge e_i corresponds to a molecular compound. There is an edge $e_k \in E$ with $k \geq |V|$ if there is a hydrogen bond between two vertices $v_i, v_j \in V$ and $i < j$. Let $S_l \subseteq V$ be the set of vertices contained in the l -th seed region of a miRNA.

The coverage score for this miRNA, the seed region S_l and a fixed $r \in \mathbb{R}_{\geq 0}$ were defined as follows:

$$\frac{\sum_{s_l \in S_l} \sum_{v \in V} X_{B_r(s_l)}(v)}{|S_l|}$$

where $\sum_{v \in V} X_{B_r(s_l)}(v)$ gives the number of bases in a 2-dimensional ball around the seed bases s_l . Therefore, the coverage score yields the sum of the number of bases in the radius of the seed bases normalized by the number of bases in the seed region.

The calculations were carried out with python with the packages pandas (version 3.10.8), numpy (1.24.2), matplotlib (3.7.0), networkX (3.0), pillow (9.4.0) and scipy (1.10.1). The package scipy with the included interpolation method and the package Pillow were used to obtain line art figures.

RESULTS

Identification of 705 MTIs between the 410 genes and 4 PD-miRNAs

To demonstrate the potential of the PD-specific 3'UTR library, we explored the complex miRNA–target interactions of four well-known PD-related miRNAs (miR-129-5p, miR-129-1-3p, miR-133b, and miR-873-5p)^{23,33–35} (Fig. 1a). Notably, we intentionally selected the miRNAs independently of the 3'UTR library to obtain unbiased insights. The relation to PD was the only common selection criterion between the miRNAs and the genes. We first cloned the miRNAs into the expression plasmid pSG5 and verified the constructs by Sanger sequencing. We next transfected the miRNA expression plasmids into 293 T cells. While Northern blot validation for miR-133b has been previously described¹⁹, we demonstrated the correct miRNA processing and ectopic expression of miR-129-5p, miR-129-1-3p, and miR-873-5p in 293T cells (Fig. 1b–d). Because the miRNAs and genes were selected independently of each other, we identified potential targets of the four miRNAs in the 3'UTR library of the PD-associated genes. In silico consensus target prediction revealed 8069 potential target genes for miR-129-5p (Supplementary Table 2), 3844 for miR-129-1-3p (Supplementary Table 3), 3845 for miR-133b (Supplementary Table 4), and 6620 for miR-873-5p (Supplementary Table 5). Out of the PD-specific 3'UTR library, we subsequently extracted those plasmids that harbored a canonical miRNA binding site for the respective miRNA by sequence alignment of the 3'UTR sequences with the canonical binding sites (8-mer, 7-mer-m8, 7-mer- A1, 6-mer). As a result, we identified 320 3'UTR subregions with binding sites for miR-129-5p, 95 3'UTR subregions for miR-129-1-3p, 100 3'UTR subregions for miR-133b, and 190 3'UTR subregions for miR-873-5p (Supplementary Table 6). As mentioned, we split the 3'UTRs into 1280 overlapping fragments to account for length restrictions and obtain optimal results. In sum, 705 potential MTIs remained and were tested by our semiautomated reporter assay: 140 genes predicted for miR-129-5p were represented by 320 3'UTR reporter constructs, 60 genes predicted for miR-129-1-3p were represented by 95 3'UTR reporter constructs, 69 genes predicted for miR-133b were represented by 100 3'UTR reporter constructs, and 102 genes predicted for miR-873-5p were represented by 190 3'UTR reporter constructs (Fig. 1a).

The target gene 3'UTR reporter constructs harboring more than one exclusive miRNA binding site for one of the four tested miRNAs are summarized in Supplementary Table 7, and the overlap between the target gene 3'UTR constructs for these miRNAs is visualized in Supplementary Fig. 3. We identified 22

target gene 3'UTR vectors with binding sites for miR-873-5p and miR-133b, 6 for miR-133b and miR-129-1-3p, 25 for miR-133b and miR-129-5p, 5 for miR-129-1-3p and miR-129-5p, 14 for miR-873-5p and miR-129-1-3p, and 43 for miR-873-5p and miR-129-5p. A few of the target gene 3'UTR constructs harbored binding sites for 3 different miRNAs: 4 constructs harbored binding sites for miR-873-5p, miR-133b and miR-129-1-3p; 1 construct harbored binding sites for miR-133b, miR-129-1-3p and miR-129-5p; 3 constructs harbored binding sites for miR-873-5p, miR-133b and miR-129-5p; and 1 construct harbored binding sites for miR-873-5p, miR-129-1-3p, and miR-129-5p.

The semiautomated reporter assay validated up to 80% of the MTIs for miR-129-1-3p

To experimentally validate the MTIs between the in silico-identified UTRs of the library and the four miRNAs, we performed our semiautomated HiTmlR assays for all 705 predicted MTIs. Following cotransfection of a miRNA with the reporter plasmid, the assay indicated functional binding of the tested miRNAs to the respective 3'UTR as a reduction in RLUs. We distributed the 705 MTIs along with positive and negative controls on 35.25 96-well plates. Each dual luciferase assay was then performed four times, resulting in $35.25 \times 4 \times 96 = 13,536$ single transfections (Fig. 2a–d). The validated MTIs are shown in the upper left quadrant of a volcano plot (Fig. 2g). In line with the original miRATBase criteria, we required a mean reduction to <90% of the relative light units (RLU) compared to the empty miRNA expression vector. As a second criterion, we required a significant nominal *p* value at an alpha level of 0.05. Notably, the results were presented as the maximal reduction and the maximal significance of all 3'UTR sequences representing that gene. Interestingly, the four miRNAs varied significantly with respect to the number of targeted 3'UTRs and the target effect size (Fig. 2f, g). Out of a total of 320 predicted and tested 3'UTR constructs (i.e., predicted and tested MTIs), 215 were suppressed by miR-129-5p (i.e., verified MTIs). Out of 95 predicted and tested 3'UTR constructs (i.e., predicted and tested MTIs), 76 were suppressed by miR-129-1-3p (i.e., verified MTIs). Out of 190 predicted and tested 3'UTR constructs (i.e., predicted and tested MTIs), 96 were suppressed by miR-873-5p (i.e., verified MTIs). Out of 100 predicted and tested 3'UTR constructs (i.e., predicted and tested MTIs), 55 were suppressed by miR-133 (i.e., verified MTIs). Among these, the validation rate was highest for miR-129-1-3p (80%), while the lowest rate was 50.5% for miR-873-5p.

To understand whether these patterns persist when focusing on stronger reductions, we repeated the consideration with a reduction to less than 70% of the original intensity. Among the 320 3'UTR subregions with binding sites for miR-129-5p, we observed 50 respective reporter plasmids. The original expression of the target gene COX4I2 (cytochrome c oxidase subunit 4I2) was reduced by 6.4% by miR-129-5p, representing the most powerful targeting effect ($p < 10^{-6}$). For miR-129-1-3p, we detected a highly significant reduction in RLUs ($p < 0.001$) for most reporter plasmids. Of the 95 3'UTR sequences for that miRNA, 21 reporter plasmids exhibited a significant RLU reduction to less than 70% of their original expression. The expression of KCNJ9 (potassium inwardly rectifying channel subfamily J member 9) was reduced the most (39.2% reduction). For miR-133b (10 UTRs) and miR-873-5p (9 UTRs) less robust targeting effects were observed. miR-133b decreased the original expression of the target gene KCNJ9 by 26.0%, which was the lowest overall reduction. For miR-873-5p, 45.7% of the RLU of the 3'UTR of the target gene dopamine beta-hydroxylase_2 (DBH_2) remained. The lower number of target genes of miR-133b and miR-873-5p compared to miR-129-5p and miR-129-1-3p is indicative of the differences in the functional abilities of the respective miRNAs. Remarkably, we also identified few genes with a significant upregulation in the target validation (see also the “right” arm in the volcano plot, Fig. 2e), contrasting the expected patterns of reduced activity. The targeting of the

PPP2R5D gene (protein phosphatase 2 regulatory subunit B'delta) by miR-133b increased its activity by 130%, indicating a significantly increased expression level ($p = 0.008$). There is increasing evidence that miRNAs can also act as positive regulators of target genes. Recent studies have reported that miRNAs can positively regulate their target genes by binding to the promoter sequences of their respective target genes in the nucleus^{36,37}. However, the mechanisms by which a miRNA induces the transcription of a target gene are not fully understood.

The astonishing variability of the results motivated a more detailed look at the positive and negative binding events for MTIs. In the following section, we address the questions of whether and to what extent the binding site, (i.e., the number or the composition of the complementary nucleotides between UTR sequences and miRNAs), contributes to the variation that we found for the tested miRNAs. For the standardized validation experiments, we considered the 3'UTRs based on the binding site type (6-mer, 7-mer-A1, 7-mer-m8, 8-mer) and the number of miRNA binding sites. Generally, we expect a greater reduction in the RLU for 7-mer and 8-mer binding sites than for 6-mer binding sites. For this analysis, we additionally included data on the targets miR-7-5p and -34a-5p from our previous study; these data were also obtained with our standardized high-throughput miRNA interaction reporter assay (HiTmlR)²³. For miR-129-5p, the most prominent reduction was detected for the 7-mer-m8 binding site (Fig. 3a). Surprisingly, we found 7-mer and 8-mer binding sites within the 3'UTR sequence of this miRNA that did not reduce the RLU of the target gene. Similarly, we found a limited influence of the binding site type on the RLU for miR-129-1-3p (Fig. 3b). In this case, a 6-mer binding site yielded the most reduced RLU in the target 3'UTR. For miR-133b, we detected a significant decrease in the RLU at the 8-mer sites compared to the 6-mer sites (Fig. 3c). For miR-873-5p, we did not detect a significant difference between the target site types. The 7-mer-A1 target site yielded the strongest RLU reduction (Fig. 3d). After reanalyzing the data from our previous study on miR-7-5p (Fig. 3e) and miR-34a-5p (Fig. 3f), we found a significant negative correlation between the number of binding nucleotides and the reduction in the RLU. In particular, the presence of an 8-mer binding site resulted in a significant reduction compared to that of the other binding sites. The GC content of the respective seed sequences ranged from 16.7% in the seed region of miR-129-5p to 66.7% in the seed regions of miR-129-1-3p and miR-34a-5p and was not related to the correlation between binding and RLU reduction. Notably, for reporter plasmids containing more than one miRNA binding site, we did not observe a significant correlation between the number of miRNA binding sites and the reduction in RLU (Supplementary Fig. 4).

Finally, we summarized the MTI results from the UTR level to the gene level based on the above criteria. We identified a total of 214 potential target genes for at least one of the tested miRNAs (miR-129-5p, miR-129-1-3p, miR-133b, and miR-873-5p; Supplementary Table 8). Considering the MTIs at the gene level as edges in a bipartite graph with 214 genes potentially targeted by at least one miRNA of the four miRNAs results in $214 \times 4 = 856$ edges (Fig. 3g). Overall, we provide computational evidence for 372 of the 856 edges (43.5%). Based on our experimental data, we considered a gene to be repressed if one of its UTR-based MTIs showed a significant reduction to at most 90%. By analyzing the experimental results of the 372 edges with positive computational evidence, we reached a confirmation rate of 76.9% (i.e., we validated 286 target genes). Overall, 286 of the 856 edges (33.4%) in the bipartite graph had computational and experimental target evidence. Three genes (CXCL12 (C-X-C motif chemokine ligand 12), RAB3B (RAB3B, a member of the RAS oncogene family), and SYT1 (synaptotagmin 1)) were targets of all four miRNAs. There were 23 additional genes that were targets of 3 of the four miRNAs and 52 genes that were targets of two of the four miRNAs.

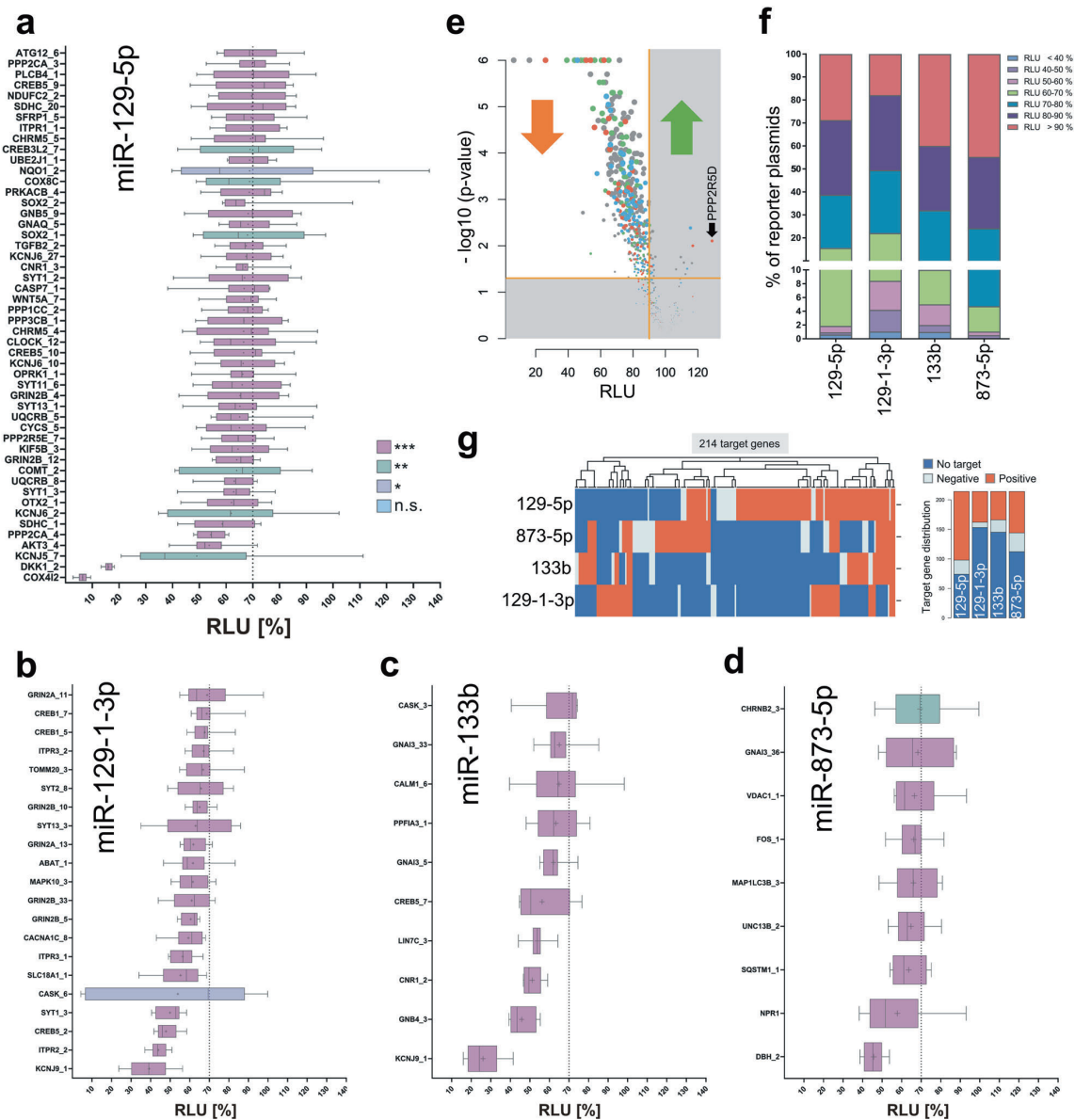


Fig. 2 Analyzing the target efficacy of the PD-specific library. **a** Reporter assay results for miR-129-5p. The extracted reporter plasmids corresponding to the predicted target genes of the four PD-associated miRNAs were analyzed for miRNA targeting ability via semiautomated HiTmIR assays. First, 293T cells were cotransfected with the respective miRNA expression plasmid and different reporter plasmids. After 48 h, the cells were lysed, and luciferase activity was determined. The RLU of the respective reporter plasmid was normalized to that of the corresponding cotransfected miRNA with the empty vector pMIR-RNL-TK. The reporter plasmids with a detected RLU <70% are shown. LiMTaR was performed four times in technical duplicates. The colors of the single bars represent the corresponding *p* values. **b** Reporter assay results for miR-129-1-3p. **c** Reporter assay results for miR-133b. **d** Reporter assay results for miR-873-5p. **e** RLUs of all MTIs for all miRNAs measured by the reporter assay versus the negative decade logarithm of the *p* value. The colors represent the four miRNAs (blue: miR-133b; red: miR-873-5p; green: miR-129-5p; gray: miR-129-1-3p). All points in the upper left corner met our criterion of a reduction in gene activity at a significant level. **f** Overall distribution of miRNA target regulation. The analyzed reporter plasmids were categorized by the detected RLU. **g** Heatmap showing the targeting results as an adjacency matrix. Dark blue cells are sets of genes and miRNAs for which no computational evidence exists and for which information has not been obtained. Light blue represents negative validation results (false-positive predictions), and red represents verified interactions.

In total, 101 genes were targeted by exactly one miRNA, and no experimental evidence could be found for 35 of the original 214 target genes.

To put the results in the context of currently available targetomes, we compared the number of MTIs to the number of

formerly validated MTIs deposited in miRTarBase. Compared to previously published MTIs entered in miRTarBase, we extended the number of MTIs validated by a reporter assay for each miRNA. Including the results of our previous study, we increased the number of validated target gene interactions from 46 to 115 for

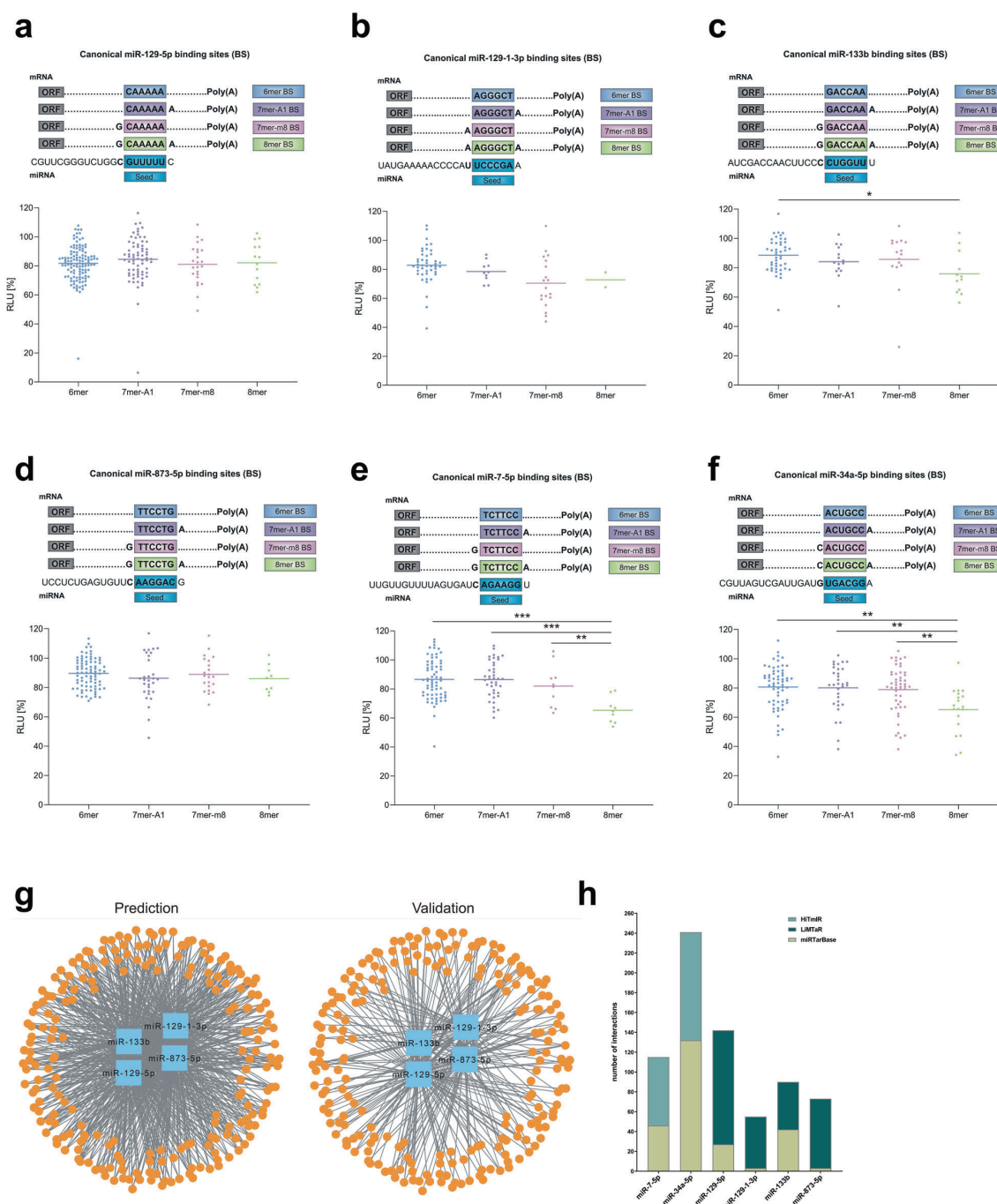


Fig. 3 Impact of the miRNA binding sites within a reporter plasmid on miRNA-dependent regulation. We categorized the results by the type of corresponding miRNA binding site within the 3'UTR sequence that was correlated with the respective RLU. The sequences of the different miRNA binding sites for each miRNA are depicted at the top of each subpanel. **a** The 3'UTR sequence of miR-129-5p. **b** The 3'UTR sequences of miR-129-1-3p. **c** The 3'UTR sequences of miR-133b. **d** The 3'UTR sequence of miR-873-5p. **e** The 3'UTR sequence of miR-7-5p. **f** The 3'UTR sequence of miR-34a-5p. **g** Graph-based validation. Left panel: Complete target network. The four miRNAs are presented in the middle as blue nodes, and the target genes are presented as smaller organ nodes. Right panel: validation of the target gene network. All edges without experimental evidence from the previous graph were removed. **h** Comparison of miRNA–target gene interactions in miRTarBase versus positive miRNA–target gene interactions in miRTarBase for miR-7-5p, miR-34a-5p, miR-129-5p, miR-129-1-3p, miR-133b and miR-873-5p. HiTmIR= high-throughput miRNA interaction reporter assay, LIMTaR= library-based miRNA target gene reporter assay (this study), and miRTarBase= microRNA–target interaction database.

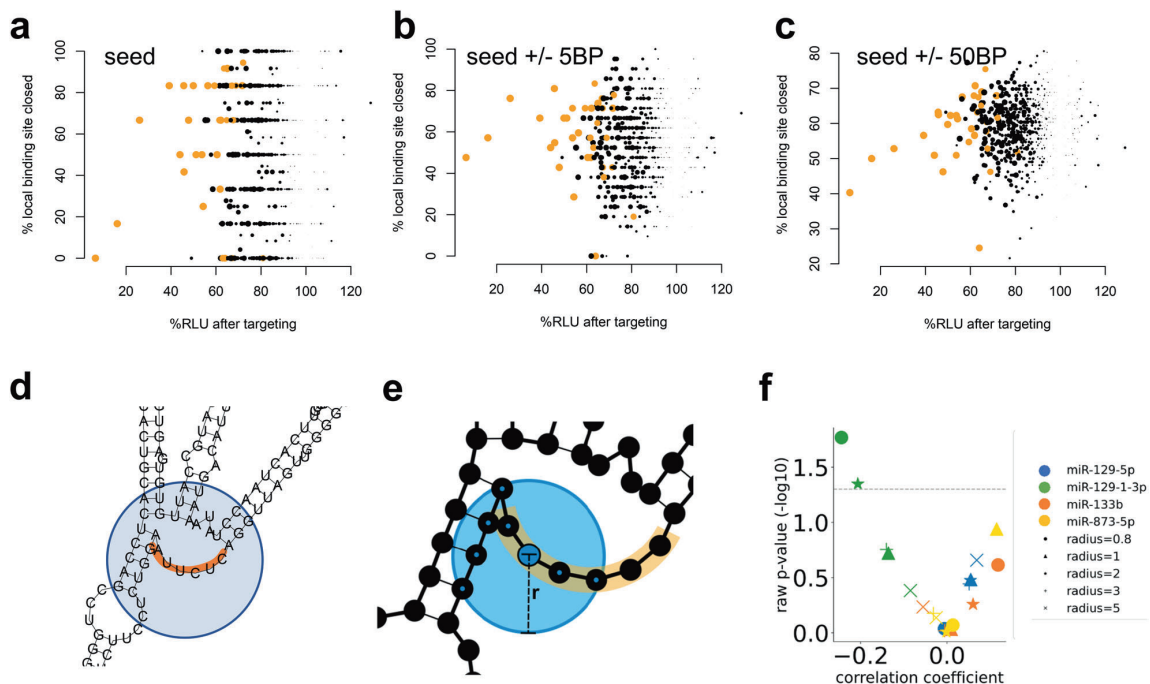


Fig. 4 **Correlation of the local secondary structure with the repression of target genes.** **a** The scatter plot shows the percentage of closed bases that are bound (y-axis) against the reduction in the RLU in %. The expected pattern is a positive correlation: the more closed a binding site is, the lower the reduction in repression should be. The orange dots represent significantly (p value $< 10^{-5}$) reduced target sites. **b** The results when the binding site was extended by a window of 5 bases in the 3' and 5' directions. **c** The results when the binding site was extended by a window of 50 bases in the 3' and 5' directions. **d** Schematic drawing of an example seed region (highlighted in orange) in a secondary structure. The circle in blue shows the area around the binding site. **e** For each base in the binding site, a circle with variable radius (r) was used to determine the number of nucleotides in proximity (bases in the circle are black circles with blue dots). We tested values of 0.8–5 length units for the radius to compute a coverage score for the binding site. **f** Scatter plot describing the correlation between the coverage score and the RLU (x-axis) and the respective significance value (y-axis). The analysis was performed for each miRNA (colors) with different radii (shapes). For miR-129-1-3p, we observed a pattern corresponding to the original hypothesis, however, the other miRNAs did not display this pattern.

miR-7-5p ($\Delta 2.5$ -fold), from 132 to 244 for miR-34a-5p ($\Delta 1.8$ -fold), from 27 to 142 for miR-129-5p ($\Delta 10.1$ -fold), from 3 to 55 for miR-129-1-3p ($\Delta 18.3$ -fold), from 42 to 90 for miR-133b ($\Delta 2.1$ -fold), and from 3 to 73 for miR-873-5p ($\Delta 24.3$ -fold) (Fig. 3h). Notably, the opposite comparison demonstrated that none of the MTIs being negative in our assay has been found positive in the miRTarBase.

Computational analysis highlights the limited influence of the predicted local secondary structure

One of the original goals was to develop a resource that facilitates improved target prediction by providing standardized reporter assay results. In the last section, we described the effects of the target sites on the miRNA targeting efficacy. One piece of information that is not included in most target predictors is the local secondary structure of the 3'UTR, not to mention more complex information such as the tertiary structure. However, RNA structures are known to generate natural cooperation between single-stranded RNA-binding proteins and 3'UTRs³⁸. Not surprisingly, the secondary structure seems to affect the binding of specific genes, as demonstrated for miR-159 in plants³⁹. Our standardized dataset allowed us to test this hypothesis at a larger scale. While we acknowledge that the cloned 3'UTRs do not necessarily reflect the physiological situation (this remains one of the primary challenges of reporter assays), the cloned sequences and the predictions made from the same sequences are consistent.

Thus, sites that are less accessible in the 3'UTR due to folding onto other sites in the same 3'UTR might impair the targeting

process. We thus compared the fraction of the seed binding region within the 3'UTR that was paired with the reduction in the RLU. We did not detect a significant correlation for any of the miRNAs (Fig. 4a), indicating that the local folding structure has a limited influence on the reporter assays. By extending the consideration to 5- (Fig. 4b) and 50-base windows (Fig. 4c) to the 3' and 5' end of the 3'UTR, we identified a tendency toward a significant reduction in the number of MTIs ($p < 10^{-5}$). The latter aspect indicates that while the local secondary structure of the 3'UTR seems to have a limited influence, the overall secondary, and even the tertiary structure, may have an influence.

This investigation requires a more precise model of the region surrounding each binding site. We implemented a score that represents the region surrounding the binding site (Fig. 4d). More precisely, we defined a circle with a variable diameter surrounding each nucleotide in the seed region of the secondary structure and counted the bases within this circle (Fig. 4e). The score was calculated by adding up the counts for each seed base and normalized by the number of bases in the seed. A high score reflected a rather closed secondary structure, while a low score reflected an open structure. By testing the radii between 0.8 and 6 length units, we identified patterns that differed significantly between miRNAs. Depending on the miRNA, we observed weak positive correlations between the coverage score and the RLU (which reflects the original assumption of better binding when the binding site is open) but also significant negative correlations (Fig. 4f). For miR-129-1-3p, we identified significant negative correlations between the coverage score and the RLU. These

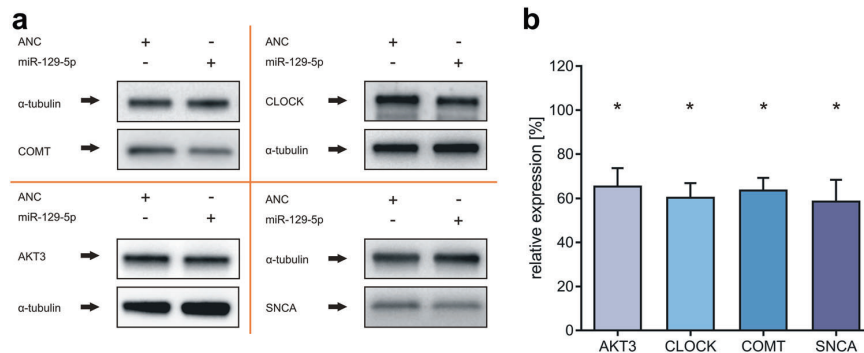


Fig. 5 **Western blot analysis of miR-129-5p.** SH-SY5Y cells were transfected with either ANC or the miRNA mimic. After 48 h, the changes in the protein expression of AKT3, CLOCK, COMT, and SNCA induced by miR-129-5p were determined using specific antibodies against these proteins. Western blot analysis for each protein was performed for three independent replicates. **a** Results of the analysis of the regulation of SNCA, CLOCK, AKT3, and COMT protein expression by miR-129-5p. **b** Quantitative analysis of the changes in the protein expression of AKT3, CLOCK, COMT, and SNCA induced by miR-129-5p are shown in the bar chart (mean $\pm$ standard deviation). Asterisks indicate a significant reduction.

examples highlight that developing a general model for improved targeting of miRNAs via secondary structure is challenging and that factors beyond the 3'UTR secondary structure also have to be considered. A tool to visualize the binding sites of one or multiple miRNAs within a UTR is freely available (<https://ccb-web.cs.uni-saarland.de/utr-seremato>).

Effect of miRNA overexpression on endogenous protein expression

To complement the expression analysis of miRNAs determined by the dual luciferase assays with both exogenous reporter plasmids and exogenous miRNA expression plasmids, we analyzed the effect of exogenous miRNA on endogenous proteins by Western blotting. For validation by Western blotting, we selected miR-129-5p, which has been previously associated with PD²³. We selected proteins that have also been previously associated with PD and were identified as potential mRNA targets of miR-129-5p by our reporter assays. The neuronal cell line SH-SY5Y was transfected with the respective miRNA mimics, and RT-qPCR was used to evaluate the ectopic expression of miR-129-5p (Supplementary Fig. 2). Ectopic expression of miR-129-5p reduced the expression of AKT3 (AKT serine/threonine kinase 3) to 65.8%, CLOCK (clock circadian regulator) to 60.7%, COMT (catechol-O-methyltransferase) to 64%, and SNCA (α -synuclein) to 59% (Fig. 5a, b). In summary, Western blot analysis confirmed the MTIs observed by dual luciferase assays. This finding agrees with our previous findings, which confirmed the results of dual luciferase assays and Western blot analyses^{17,23,40–42}.

DISCUSSION

Understanding the interactions between miRNAs and target genes is central to the characterization of molecular signaling cascades, including signaling cascades, in a disease context. As mentioned above, there are more than 44,571,700 potential miRNA target interactions (MTIs). Computational approaches are certainly the first choice for tackling the task of elucidating complex networks. Ultimately, however, these approaches require experimental approaches to verify the predicted MTIs. We employed a systematic approach for MTI determination under standardized conditions by using a 3'UTR library enriched for genes associated with Parkinson's disease. We analyzed this library by more than 13,000 single transfections for a set of four miRNAs that have been previously associated with PD. Mature miR-129-5p and miR-129-1-3p, which are both processed from the same precursor, premiR-129, were significantly upregulated after the induction of a PD-like

phenotype in LUHMES cells, as shown in our previous study²³. A study by Kim et al. showed that the precursor premiR-129-2, which can also be processed into the mature miR-129-5p, was enriched in the midbrain of PD patients³³. The same study showed a significant reduction in the expression of the miR-133b precursor. A significantly reduced miR-133b level was also demonstrated in the plasma of PD patients and in different models mimicking a PD-like phenotype^{33,43–45}. miR-873-5p has been previously described to be neuroprotective in a neuroinflammatory model of PD³⁴.

Depending on the tested miRNA, differences were detected in the overall RLU value distribution, the number of significant target gene 3'UTR reporter plasmids and the positive rate. We found that the effectiveness of miRNAs strongly varies between distinct signaling pathways. Especially for miR-873-5p, most reporter plasmids showed only a slight reduction in the RLU, a small number of significant target genes and a low positive rate compared to those of the other tested miRNAs. Although miRNAs that regulate only small numbers of target genes in the dopamine and PD-associated pathways might be less relevant for these signaling pathways, these miRNAs might nevertheless impact the pathogenesis of the disease via other signaling pathways. The miR-873-5p, which has been primarily described as neuroprotective in the context of neuroinflammation, may regulate target genes directly associated with the PD-associated immune response³⁴. Although our study focused on the generation of miRNA–target gene networks in affected dopaminergic neurons, identifying target genes of miR-873-5p that are associated with the immune system could further help to elucidate the role of this miRNA in PD pathogenesis, especially in the context of neuroinflammation.

While previous high-throughput studies on the impact of miRNA binding often did not exclude indirect effects^{46,47}, our high-throughput analysis focused on direct interactions between a miRNA binding site and the respective miRNA. The results of the analysis of the binding sites of miR-34a-5p, miR-7-5p, miR-129-1-3p and miR-133b were comparable to those of other miRNAs^{46,47}. The higher the number of potential binding nucleotides within a 3'UTR was, the greater the effect of miRNA-induced regulation was. This, however, did not apply to reporter plasmids containing a 7-mer-m8 binding site that showed a low or no reduction in the RLU. In contrast, no significant effect between the different types of binding sites was detected for the miRNAs miR-129-5p and miR-873-5p. Several additional factors may contribute to MTIs, including the surrounding sequence in the 3'UTR of the target gene and the secondary structure of the mRNA^{47–50}. The latter hypothesis is supported by in silico predictions of the local

secondary structure of mRNAs around the miRNA binding site, indicating reduced accessibility of the miRNA binding site for genes with high RLUs.

Independent of the mRNA secondary structure, the mere flanking sequences, for example, 3' complementarity, may also influence the MTI as well as the nucleotide composition of the miRNA seed region⁵¹. Of particular interest in this context are the differences in the GC content ranging from 16.7% in the seed region of miR-129-5p to 66.7% in the seed regions of miRNAs miR-34a-5p and miR-129-1-3p. Notably, miRNAs with a low GC content in the seed region were associated with noncanonical targeting⁵¹. In addition to the analysis of specific MTIs, MTI networks that are largely predicted only by in silico tools await further identification and validation via experimental approaches. As recently shown, target sites within mRNAs can act cooperatively, resulting in greater repression of the target mRNA than that caused by independent action at each site⁵². Similarly, the number of miRNAs, the number of targets, and the availability of free RISCs are important, as each plays a role in the development and status of MTI networks.

The functional assays that are required to contribute to MTI network analyses have severe limitations and biases. To analyze the interactive effects of two or more miRNAs, transfection assays can be used, but they do not allow us to define the stoichiometric relationships between the miRNAs and their target MTIs. The interpretability of cotransfection assays is further impacted by the variable and/or frequently low transfection efficacy. Furthermore, the cellular background can also affect the functionality of MTIs⁵³. Different cell types express varying amounts of different RNA-binding proteins (RBPs)⁵⁴, which can affect the ability of a miRNA to regulate target genes. RBPs can modulate the secondary structure of target 3'UTRs, enhancing the binding capability of miRNAs⁵⁵. On the other hand, RBPs binding to target 3'UTRs can prevent miRNAs from binding by masking their binding motifs. This can occur when RBPs bind to the same site as the miRNA or when they bind to a nearby site that overlaps with the miRNA binding site^{56,57}.

Independent of these challenges, the presented systematic 3' UTR library-based approach for targetome determination provides a basis for subsequent MTI network analyses. To acknowledge the complexity of this scenario, we employed a standardized library-based functional approach with 13,536 single transfections and identified a very dense MTI network with 447 MTIs for PD. Our approach to studying MTIs in a disease context can serve as a foundation to experimentally validate millions of potential MTIs, especially in specific disease phenotypes. From our study, we derive the following concrete scenarios and next steps: (1) Standardization of reporter assay experiments is essential for obtaining sufficient high-quality data and training artificial intelligence approaches to improve the prediction of targets. (2) It is mandatory to perform standardized experiments at the same scale for cooperative binding scenarios. Only with the respective datasets paired with improved computational models can the effects and side effects of miRNA-mediated therapies be modeled, facilitating miRNA therapeutic applications.

REFERENCES

- Lai, E. C. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.* **30**, 363–364 (2002).
- Lytle, J. R., Yario, T. A. & Steitz, J. A. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proc. Natl Acad. Sci. USA* **104**, 9667–9672 (2007).
- Moretti, F., Thermann, R. & Hentze, M. W. Mechanism of translational regulation by miR-2 from sites in the 5' untranslated region or the open reading frame. *RNA* **16**, 2493–2502 (2010).
- Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
- Giraldez, A. J. et al. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**, 75–79 (2006).
- Mathonnet, G. et al. MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. *Science* **317**, 1764–1767 (2007).
- Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
- Simeone, I. et al. Prediction and pan-cancer analysis of mammalian transcripts involved in target directed miRNA degradation. *Nucleic Acids Res.* **50**, 2019–2035 (2022).
- Lirussi, L. et al. A regulatory network comprising let-7 miRNA and SMUG1 is associated with good prognosis in ER+ breast tumours. *Nucleic Acids Res.* **50**, 10449–10468 (2022).
- Imperatore, J. A. et al. Highly conserved s2m element of SARS-CoV-2 dimerizes via a kissing complex and interacts with host miRNA-1307-3p. *Nucleic Acids Res.* **50**, 1017–1032 (2022).
- Bryant, C. J. et al. Discovery of novel microRNA regulators of ribosome biogenesis. *bioRxiv* <https://doi.org/10.1101/2023.02.17.526327> (2023).
- Condrat, C. E. et al. miRNAs as biomarkers in disease: latest findings regarding their role in diagnosis and prognosis. *Cells* <https://doi.org/10.3390/cells9020276> (2020).
- Diener, C., Keller, A. & Meese, E. Emerging concepts of miRNA therapeutics: from cells to clinic. *Trends Genet.* **38**, 613–626 (2022).
- Fehlmann, T. et al. Evaluating the use of circulating microRNA profiles for lung cancer detection in symptomatic patients. *JAMA Oncol.* **6**, 714–723, (2020).
- Fehlmann, T. et al. Common diseases alter the physiological age-related blood microRNA profile. *Nat. Commun.* **11**, 5958 (2020).
- Hong, D. S. et al. Phase 1 study of MRX34, a liposomal miR-34a mimic, in patients with advanced solid tumours. *Br. J. Cancer* <https://doi.org/10.1038/s41416-020-0802-1> (2020).
- Hart, M. et al. Wrinkle in the plan: miR-34a-5p impacts chemokine signaling by modulating CXCL10/CXCL11/CXCR3-axis in CD4(+), CD8(+) T cells, and M1 macrophages. *J. Immunother. Cancer* <https://doi.org/10.1136/jitc-2020-001617> (2020).
- Chu, Y. et al. Argonaute binding within 3'-untranslated regions poorly predicts gene repression. *Nucleic Acids Res.* **48**, 7439–7453 (2020).
- Alles, J. et al. An estimate of the total number of true human miRNAs. *Nucleic Acids Res.* **47**, 3353–3364 (2019).
- Kern, F. et al. What's the target: understanding two decades of in silico microRNA-target prediction. *Brief. Bioinform.* **21**, 1999–2010 (2020).
- Kuhn, D. E. et al. Experimental validation of miRNA targets. *Methods* **44**, 47–54 (2008).
- Kilikevicius, A., Meister, G. & Corey, D. R. Reexamining assumptions about miRNA-guided gene silencing. *Nucleic Acids Res.* **50**, 617–634 (2022).
- Kern, F. et al. Validation of human microRNA target pathways enables evaluation of target prediction tools. *Nucleic Acids Res.* **49**, 127–144 (2021).
- Balestrino, R. & Schapira, A. H. V. Parkinson disease. *Eur. J. Neurol.* **27**, 27–42 (2020).
- Leggio, L. et al. microRNAs in Parkinson's disease: from pathogenesis to novel diagnostic and therapeutic approaches. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms18122698> (2017).
- Craig, D. W. et al. RNA sequencing of whole blood reveals early alterations in immune cells and gene expression in Parkinson's disease. *Nat. Aging* **1**, 734–747 (2021).
- Kern, F. et al. Deep sequencing of sncRNAs reveals hallmarks and regulatory modules of the transcriptome during Parkinson's disease progression. *Nat. Aging* **1**, 309–322 (2021).
- Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
- Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
- Slenter, D. N. et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.* **46**, D661–D667 (2018).
- Croft, D. et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39**, D691–D697 (2011).
- Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubock, R. & Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Res.* **36**, W70–W74 (2008).
- Kim, J. et al. A MicroRNA feedback circuit in midbrain dopamine neurons. *Science* **317**, 1220–1224 (2007).
- Wu, J. et al. The inhibition of miR-873 provides therapeutic benefit in a lipopolysaccharide-induced neuroinflammatory model of Parkinson's disease. *Oxid. Med. Cell Longev.* **2020**, 8735249 (2020).
- Liu, X. et al. Interleukin-17 (IL-17)-induced microRNA 873 (miR-873) contributes to the pathogenesis of experimental autoimmune encephalomyelitis by targeting A20 ubiquitin-editing enzyme. *J. Biol. Chem.* **289**, 28971–28986 (2014).
- Erdmann, K., Kaulke, K., Rieger, C., Wirth, M. P. & Fuessel, S. Induction of alpha-methylacyl-CoA racemase by miR-138 via up-regulation of beta-catenin in prostate cancer cells. *J. Cancer Res. Clin. Oncol.* **143**, 2201–2210 (2017).

37. Matsui, M. et al. Promoter RNA links transcriptional regulation of inflammatory pathway genes. *Nucleic Acids Res.* **41**, 10086–10109 (2013).
38. Lin, Y. H. & Bundschuh, R. RNA structure generates natural cooperativity between single-stranded RNA binding proteins targeting 5' and 3'UTRs. *Nucleic Acids Res.* **43**, 1160–1169 (2015).
39. Zheng, Z. et al. Target RNA secondary structure is a major determinant of miR159 efficacy. *Plant Physiol.* **174**, 1764–1778 (2017).
40. Hart, M. et al. miR-34a: a new player in the regulation of T cell function by modulation of NF-kappaB signaling. *Cell Death Dis.* **10**, 46 (2019).
41. Hart, M. et al. Identification of miR-34a-target interactions by a combined network based and experimental approach. *Oncotarget* **7**, 34288–34299 (2016).
42. Hart, M. et al. miR-34a as hub of T cell regulation networks. *J. Immunother. Cancer* **7**, 187 (2019).
43. Zhao, N., Jin, L., Fei, G., Zheng, Z. & Zhong, C. Serum microRNA-133b is associated with low ceruloplasmin levels in Parkinson's disease. *Parkinsonism Relat. Disord.* **20**, 1177–1180 (2014).
44. Zhang, X. et al. Reduced circulating levels of miR-433 and miR-133b are potential biomarkers for Parkinson's disease. *Front. Cell. Neurosci.* **11**, 170 (2017).
45. Dong, L. G. et al. MiR-133b inhibits MPP⁺-induced apoptosis in Parkinson's disease model by inhibiting the ERK1/2 signaling pathway. *Eur. Rev. Med. Pharmacol. Sci.* **24**, 11192–11198 (2020).
46. Nielsen, C. B. et al. Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA* **13**, 1894–1910 (2007).
47. Grimson, A. et al. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* **27**, 91–105 (2007).
48. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
49. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nat. Genet.* **39**, 1278–1284 (2007).
50. Long, D. et al. Potent effect of target structure on microRNA function. *Nat. Struct. Mol. Biol.* **14**, 287–294 (2007).
51. Wang, X. Composition of seed sequence is a major determinant of microRNA targeting patterns. *Bioinformatics* **30**, 1377–1383 (2014).
52. Briskin, D., Wang, P. Y. & Bartel, D. P. The biochemical basis for the cooperative action of microRNAs. *Proc. Natl Acad. Sci. USA* **117**, 17764–17774 (2020).
53. Diener, C., Keller, A. & Meese, E. The miRNA-target interactions: an underestimated intricacy. *Nucleic Acids Res.* **52**, 1544–1557 (2023).
54. Kakumani, P. K. AGO-RBP crosstalk on target mRNAs: Implications in miRNA-guided gene silencing and cancer. *Transl. Oncol.* **21**, 101434 (2022).
55. Kim, S. et al. The regulatory impact of RNA-binding proteins on microRNA targeting. *Nat. Commun.* **12**, 5057 (2021).
56. Li, Y., Estep, J. A. & Karginov, F. V. Transcriptome-wide Identification and Validation of Interactions between the miRNA Machinery and HuR on mRNA Targets. *J. Mol. Biol.* **430**, 285–296 (2018).
57. Cottrell, K. A., Chaudhari, H. G., Cohen, B. A. & Djuranovic, S. PTRE-seq reveals mechanism and interactions of RNA binding proteins and miRNAs. *Nat. Commun.* **9**, 301 (2018).

M. Hart et al.

945

ACKNOWLEDGEMENTS

This work was supported by the European Union's Seventh Framework Program for Research, Technological Development and Demonstration [grant number: 600841] and by the Michael J. Fox Foundation [grant number: 14446]. L.K. was supported by a grant from the Stefan-Morsch Foundation.

FUNDING

Open Access funding enabled and organized by Projekt DEAL.

COMPETING INTERESTS

A.K. was a scientific advisor of Firalis, a company that develops miRNA biomarkers for neurological disorders.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s12276-024-01202-5>.

Correspondence and requests for materials should be addressed to Martin Hart.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

3.7 CHARACTERIZING EXPRESSION CHANGES IN NONCODING RNAs DURING AGING AND HETEROCHRONIC PARABIOSIS ACROSS MOUSE TISSUES

nature biotechnology



Article

<https://doi.org/10.1038/s41587-023-01751-6>

Characterizing expression changes in noncoding RNAs during aging and heterochronic parabiosis across mouse tissues

Received: 16 August 2022

Accepted: 15 March 2023

Published online: 27 April 2023

Check for updates

Viktoria Wagner¹, Fabian Kern^{1,2,3}, Oliver Hahn², Nicholas Schaum², Nicole Ludwig⁴, Tobias Fehlmann¹, Annika Engel¹, Dominic Henn⁵, Shusruto Rishik¹, Alina Isakova⁶, Michelle Tan⁶, Rene Sit⁶, Norma Neff⁶, Martin Hart⁴, Eckart Meese⁴, Steve Quake⁶, Tony Wyss-Coray^{6,7,8} & Andreas Keller^{1,2,3,8} ✉

Molecular mechanisms of organismal and cell aging remain incompletely understood. We, therefore, generated a body-wide map of noncoding RNA (ncRNA) expression in aging (16 organs at ten timepoints from 1 to 27 months) and rejuvenated mice. We found molecular aging trajectories are largely tissue-specific except for eight broadly deregulated microRNAs (miRNAs). Their individual abundance mirrors their presence in circulating plasma and extracellular vesicles (EVs) whereas tissue-specific ncRNAs were less present. For miR-29c-3p, we observe the largest correlation with aging in solid organs, plasma and EVs. In mice rejuvenated by heterochronic parabiosis, miR-29c-3p was the most prominent miRNA restored to similar levels found in young liver. miR-29c-3p targets the extracellular matrix and secretion pathways, known to be implicated in aging. We provide a map of organism-wide expression of ncRNAs with aging and rejuvenation and identify a set of broadly deregulated miRNAs, which may function as systemic regulators of aging via plasma and EVs.

One primary risk factor for cancer, diabetes, cardiovascular disorders and neurodegenerative diseases is aging¹. Therefore, understanding the underlying mechanisms of this complex process is essential to improve quality of life by developing new therapies. Finding the most promising therapy target is challenging, as it is not possible for a single level of omics data to explain whether the changes discovered are causative to or the result of aging². Epigenetic markers like DNA

methyations have been identified as promising aging biomarkers^{3,4}. Current research efforts, including transcriptomic studies of major organs in aged mice^{5,6}, largely lack information covering the whole RNA diversity, for example, the diverse classes of noncoding RNAs (ncRNAs). Attempting to better differentiate cause and effect, we herein present the corresponding ncRNA dataset to the Tabula Muris Senis (TMS) cohort⁵. Such RNAs are part of epigenetic reprogramming and altered

¹Clinical Bioinformatics, Saarland University, Saarbrücken, Germany. ²Department of Neurology and Neurological Sciences, Stanford University, Stanford, CA, USA. ³Helmholtz Institute for Pharmaceutical Research Saarland (HIPS)–Helmholtz Centre for Infection Research (HZI), Saarland University Campus, Saarbrücken, Germany. ⁴Department of Human Genetics, Saarland University, Saarland, Germany. ⁵Department of Plastic Surgery, University of Texas Southwestern Medical Center, Dallas, TX, USA. ⁶Department of Bioengineering, Stanford University, Stanford, CA, USA. ⁷The Phil and Penny Knight Initiative for Brain Resilience, Stanford University, Stanford, CA, USA. ⁸These authors jointly supervised this work: Tony Wyss-Coray, Andreas Keller.

✉e-mail: twc@stanford.edu; andreas.keller@ccb.uni-saarland.de

intercellular communication, which have been described as hallmarks of aging^{1,7}. Further, they can have a role in intercellular communication via extracellular vesicles (EV)⁸. MicroRNAs (miRNA), a class of ncRNAs, target messenger RNA (mRNA) through base-pair binding and thereby regulate gene expression via post-transcriptional gene silencing^{7,9}. Furthermore, miRNAs act as age-specific disease biomarkers¹⁰ and have been identified as regulators in aging-associated phenotypes¹¹.

We analyzed eight classes of ncRNAs in TMS separately and together with the existing single-cell and bulk mRNA datasets⁵. The previously observed tissue-driven shifts in gene expression with aging that correlate with corresponding protein levels in plasma could be caused by epigenetic regulation mechanisms mediated by ncRNA. Furthermore, these may not only be implicated in aging but also have a role in the regenerative effects observed in aging interventions such as heterochronic parabiosis. Regenerative activities within young blood with translational implications for aged liver, muscle and brain have been observed before¹². Therefore, we performed ncRNA sequencing of tissue samples following heterochronic parabiosis experiments, in which a young (3–4 months) and an aged (19 months) mouse share a common blood circulation. Our two datasets describe age- and rejuvenation-related ncRNA expression changes to reveal the potential of ncRNAs as targets for new pharmaceutical approaches.

Results

Mapping of ncRNA expression across mouse organs

We sequenced 771 tissue samples of the TMS cohort to map molecular shifts across the whole organism during healthy aging (Fig. 1a). The protocol enriches for small ncRNA, especially mature miRNAs. Even though full-length reads cover only small ncRNAs (miRNAs or piwi-interacting RNAs (piRNAs)) completely, the protocol generates measurable fragments of longer ncRNAs. This sequencing strategy extends the existing mRNA TMS dataset⁵ with miRNA, piRNA, long ncRNA (lncRNA), small nucleolar RNA (snoRNA), small nuclear RNA (snRNA), transfer RNA (tRNA), ribosomal RNA (rRNA) and small Cajal body-specific RNA (scaRNA). The tissue sample collection includes 16 solid tissues of C57BL/6JN mice (bone (femurs and tibiae), brain (hemibrain), brown adipose tissue (BAT, interscapular depot), gonadal adipose tissue (GAT, inguinal depot), heart, kidney, limb muscle (tibialis anterior), liver, lung, bone marrow, mesenteric adipose tissue (MAT), pancreas, skin, small intestine (duodenum), spleen and subcutaneous adipose tissue (SCAT, posterior depot)). The selected time course covers the mouse lifespan from a developmental age of 1 month up to 27 months (males: aged 1, 3, 6, 9, 12, 15, 18, 21, 24 and 27 months; females: aged 1, 3, 6, 9, 12, 15, 18 and 21 months). With up to six mice per timepoint, the study covers a maximum of 960 samples (16 organs × 10 timepoints × 6 replicates). As not all mice survived to the later timepoints and we further excluded 26 low-quality RNA samples, we finally included 771 high-quality samples in the study (Supplementary Table 1).

We mapped resulting sequencing reads against 87,590 ncRNA sequences (Fig. 1a, left column) derived from established reference databases (miRNAs: miRBase22, tRNAs: GtRNAdb18.1, piRNA: RNACentral15, all other ncRNAs: Ensembl 100). Altogether, we detected reads mapping to 58,422 different ncRNAs (Fig. 1a, middle column), with miRNAs being the most abundant class. An average of 36.2% of reads across tissues mapped to miRNAs (Extended Data Fig. 1a). The distribution of reads to RNA classes, however, varied substantially between tissues ($P < 0.05$, Kruskal–Wallis test; Extended Data Fig. 1b). We thus asked whether the variation in read distribution is related to the length of representatives. We generated aligned sequence profiles to quantify the length of sequences covered with our reads. First, we explored the percentage of sequence length covered versus the sequence reference length. Even for very long sequences exceeding 10,000 bases, we partially recovered large fractions or even the complete sequence (Extended Data Fig. 1c). We also computed the maximal assembly for each RNA, that is, the longest contiguous read mapping,

and compared it with the sequence reference length (Extended Data Fig. 1d). Even though the fraction of the sequence covered by the maximal assembly decreased for larger RNAs (Spearman's $\rho = -0.43$), we verified that throughout all RNA classes, we were still able to reproduce for a subset up to 100% of the full-length reference (Supplementary Table 2). Therefore, we decided to include the full dataset as a reference for future studies while implementing restrictive filtering steps to increase the reliability of our data. Especially the somatic piRNAs exceeded the expected counts, likely driven by artifacts in piRNA annotation¹³ and calling for additional quality control filters. We first retained piRNAs encoded in prepachytene piRNA genomic clusters^{14,15} to minimize the number of false positive hits. Next, we removed low abundant features across all ncRNA classes, keeping those with at least 1 read mapped per million (rpmm) in at least one sample, resulting in the abundant dataset (Fig. 1a, right column). Applying this stringent filtering, the number of piRNAs in our dataset decreased from 43,799 detected down to 43 abundant, likely removing most falsely annotated features¹³.

Clustering by ncRNA expression using t-distributed stochastic neighbor embedding (t-SNE), samples split into tissue-specific groups (Fig. 1b). One cluster contained skin, GAT and SCAT samples, which likely can be explained by their biological and functional close relationship of containing similar cell types. To check whether relevant biological factors outweigh technical ones, we performed a principal variance component analysis. The highest proportion of variance in the data was explained by tissue identity (Extended Data Fig. 2a). Annotating the t-SNE plot by animal sex revealed a uniform spread, excluding it as a major driver of the observed variance (Extended Data Fig. 2b).

Following our main objective to identify organ-specific aging trajectories, we added a tissue-specific, that is, local filtering to check whether ncRNA expression changed not only between tissues but also with age (cf. Methods). On the locally-filtered data, we calculated read count percentages for all RNA classes. As for the detected reads, we observed tissue-specific distributions (Fig. 1c). Analyzing those over time, we identified two clusters of tissues (Extended Data Fig. 2c). One exhibited a stable count distribution (mean variance < 4.5%) and the other showed high variance within the count distribution (mean variance > 4.5%). Specifically, 3 of 16 tissues showed high variance (brain, BAT and limb muscle), while most tissues (13–16) were characterized by a stable read distribution (including marrow and liver) (Fig. 1d and Extended Data Fig. 3). In the brain, the share of snRNA reads decreased from 77.9% to 10.0%, while the share of miRNAs increased from 9.1% to 28.5%. In BAT, the miRNA share grew steadily from 4.1% to 26.4% and the rRNA share dropped from 62.7% to 27.8%.

The observed variations of the RNA classes prompted us to assess the expression changes during aging for the individual ncRNAs. Therefore, we determined the Spearman rank correlation of age with the expression of every ncRNA in each tissue separately. We identified 31 tRNA fragments that were substantially differentially expressed between 3 and 21 months (two-sided t -test, P adjust < 0.05). Eight tRNA fragments showed increased expression (in brain and lung) and 23 showed decreased expression with age (in bone, limb muscle, skin and GAT). tRNA-related metabolism, transcription, modification and derivatives have vital roles in aging and longevity of organisms, as tRNA expression decreases with age¹⁶. We further observed that miRNAs displayed the strongest correlations with age over all tissues (exceeding the interval of -0.5 to 0.5 ; Fig. 1e). Given that miRNAs were captured in full-length by our sequencing platform, their high abundance across tissues and the fact that they exhibited the largest effect size, we further focused on miRNAs for downstream analysis.

MiRNA lifespan trajectories are largely tissue specific

For the intersection of miRNAs expressed in all tissues, we observed more markers being correlated positively than negatively with age (Fig. 2a and Supplementary Table 3). In contrast, large sets of miRNAs

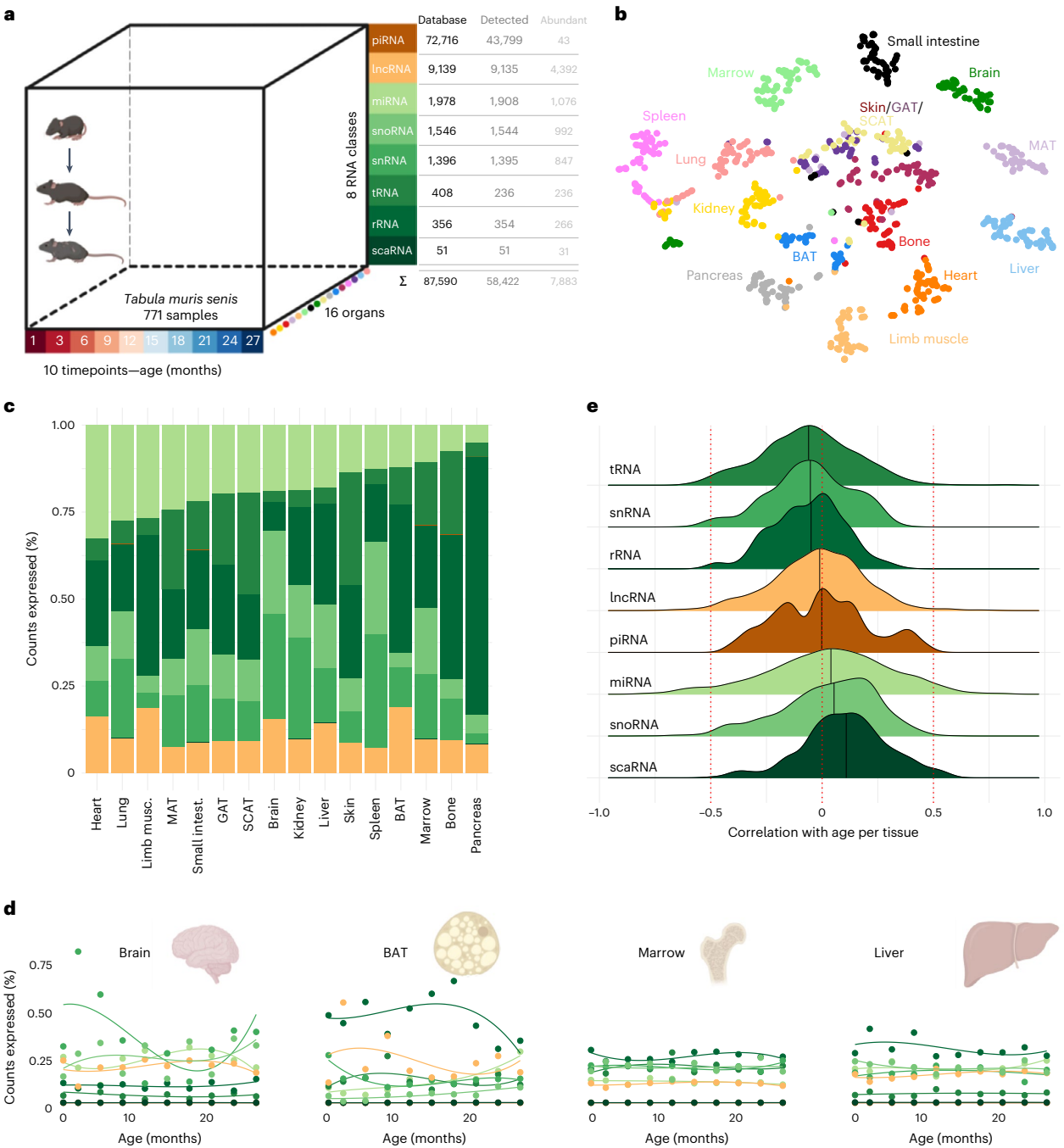


Fig. 1 | Atlas of noncoding RNA expression along the mouse lifespan. a, Study overview—data of the aging (TMS) cohort, consisting of mouse samples collected from 16 different tissues at ten different timepoints throughout the lifespan with maximal six replicates per timepoint varying due to sample and sequencing quality. A total of 771 samples were sequenced, and the reads were annotated to the 87,590 different RNA reference sequences from eight RNA classes. Of the RNAs in the databases listed on the left, 58,422 different ncRNAs were annotated in the raw reads and we found 7,883 noncoding features as abundant expressed in our TMS aging cohort. Created with BioRender. **b**, t-SNE visualization of all

samples of the TMS cohort over all detected noncoding RNAs, colored by tissue of origin. **c**, Percentage of counts per RNA class, calculated on total counts per tissue after local filtering for all tissues in the TMS cohort, color coded by RNA-class color legend as indicated in **a**. **d**, Variation of mean count distribution per RNA class over the lifespan of the mouse in the brain, BAT, marrow and liver; calculated count percentages per sample after local filtering. Created with BioRender. **e**, Density plot of Spearman rank correlation of all expressed noncoding RNA with age in each individual tissue grouped by RNA classes, density scaled individually for every RNA class.

Article

<https://doi.org/10.1038/s41587-023-01751-6>

were correlated with age in a specific tissue. For example, six miRNAs were negatively correlated exclusively in limb muscle and 37 were positively correlated only in BAT. One of these miRNAs, miR-107, regulates insulin sensitivity and is postulated as a target for the treatment of type 2 diabetes and obesity⁴⁷. Its increase in aging could be connected to the fact that age is a risk factor for diabetes⁴.

Certain miRNAs were linearly correlated with age in more than one tissue (Fig. 2b). MiR-29a-3p was positively correlated in eight tissues, and miR-300-3p, miR-487b-3p and miR-541-5p were negatively correlated in five tissues each. Based on these observations, we separated the miRNAs into the following three classes: nonaging-related, local aging and global aging miRNAs. Local aging miRNAs were defined as correlated with age in at least one tissue exceeding the interval of -0.5 to 0.5 . We accordingly defined miRNAs correlated with age in more than five different tissues as globally aging. Following these definitions, we identified the three mentioned negatively correlated miRNAs together with five positively correlated miRNAs (miR-29a-3p, miR-29c-3p, miR-155-5p, miR-184-3p and miR-1895) as globally aging.

We then examined whether nonlinear age-related expression changes occur as well. Using the 3m timepoint as a baseline, we calculated foldchanges (FC) for all later timepoints and respective *P* values. Based on this analysis, we determined the number of deregulated miRNAs (Fig. 2c). Most were deregulated in BAT, driven by the large fraction of positively correlated local aging miRNAs. Investigating the brain, we found a peak at the ages 12 and 18 months with a count of 412 and 427 deregulated miRNAs, respectively. Most of all substantially deregulated miRNAs (77.6%) in brain showed the strongest effect at 12 and 18 months (Fig. 2d, Extended Data Fig. 4a,b and Supplementary Table 4). The higher count of deregulated miRNA at certain timepoints matched our expectation, as we hypothesized that miRNAs were responsible for the regulation of the previously reported transcriptome changes¹⁸. We further confirmed that those effects were not driven by lowly expressed features—we projected the mean expression against the FC for all tissues and all ncRNAs per timepoint (Extended Data Fig. 5). In line with our assumption, substantial FCs could be observed across all expression scales.

To identify common patterns within the nonlinear changes over time, we calculated *z* scores for all miRNAs being expressed in every tissue. Each miRNA in every single tissue was displayed as an aging trajectory and clustered across all organs. Ten of the 20 clusters obtained were composed mainly of one tissue; thus, we propose the existence of organ-specific miRNA time course signatures (Fig. 2e). Half of the miRNAs in cluster 2, with a peak at 3 months and a late increase again at 24 months, originated from the skin. Cluster 9, which showed a peak at 12 and 18 months, was composed of 61.2% brain miRNAs. The expression of miRNAs in cluster 13 increased continuously from the age of 6 months on, and this trajectory was specific for BAT (70.6%). In summary, we determined 10 of the 20 clusters to be tissue-specific, with at least 30% of miRNA originating from a single tissue (Extended Data Fig. 6).

The global aging miRNAs marked an exception to this tissue-specific clustering. Trajectories from more than five different

organs for seven global aging miRNAs clustered together. For instance, we found the trajectories of miR-29a-3p and miR-29c-3p from ten and eight different tissues in cluster 20, respectively (Fig. 2f). The expression of miRNAs within this cluster increased continuously with age. This consistent signature could be indicative of the regulation of key pathways across all organs upon aging. Thus, we investigated the relationship between miRNA and mRNA expression closer.

Transcriptome changes mirrored by global aging miRNAs

The previous analyses suggested five miRNAs as cross-organ aging markers increasing with age (Fig. 2b). Following the biological mechanism, we expected repression of target genes with aging. We chose to identify potential new targets in an unbiased manner by correlating miRNA with mRNA expression levels from the TMS dataset⁵. In the first step, we defined targets by exhibiting a significant inverse correlation ($r < -0.4$, $P < 0.05$). To support the validity of our approach, we checked the share of predicted miRNA–mRNA interactions with conserved binding sites for the miRNAs. As a control, we compared this number against the share of conserved binding sites in the miRNA–mRNA interactions predicted via positive correlation. For 7.3% (9 of 122) of the miRNA–mRNA interactions identified via inverse correlation, we found at least one conserved miRNA binding site, as compared to the 2.1% (120 of 54,992) miRNA–mRNA interactions in the control set (Fisher's exact test, $P = 0.0018$). Because a gene can contain multiple binding sites across multiple 3' UTRs and different site types exhibit different strengths, we repeated the analysis for each type of binding site. The amount of conserved 8mer binding sites is 6.3 times higher as compared to the control (4.91% inv. correlation, 0.78% control; $P = 0.0006$), for conserved 7mer-8m binding sites 3.8 times higher (4.92% inv. correlation, 1.30% control; $P = 0.0062$) and for conserved 7mer-1a binding sites 9.0 times higher (2.45% inv. correlation, 0.27% control; $P = 0.0064$).

The filtered target gene sets showed distinct overlaps (cf. Methods; Fig. 3a and Supplementary Table 5). Three of the six targets are shared among all miRNAs, *Eln*, *Col1a1* and *Col3a1*, which have a role in protein digestion and absorption and encode extracellular matrix (ECM) proteins. These are already validated targets for miR-29b-1/miR-29a (ref. 19). Overall, enriched processes for all targets were dominated by ECM-associated processes, such as ECM organization, collagen fibril organization and ECM-receptor interaction (Fig. 3b)²⁰. Senescent cells are known to exhibit altered expression and organization of ECM and the 'senescence-associated secretory phenotype'^{1,21}. Our data suggest that these effects could be regulated by global aging miRNAs. Another part of the network composed of mainly Y-chromosome-coded proteins contained proteins related to 'ubiquitin-proteasome dependent proteolysis' (*Usp9y*), histone modification introducing proteins (*Kdm5d*) and probable transcriptional activators (*Zfy1*, *Zfy2*). Hence, other layers of regulation mechanisms are targeted. The 'AGE-RAGE signaling pathway in diabetic complications' and 'dysregulated miRNA targeting in insulin/PI3K-AKT signaling' were enriched, supporting our suggestion of the importance of miRNA regulation in nutrient sensing.

Consistent with the enriched pathways for the targets of cross-organ aging miRNAs were the enriched pathways for the targets of the local aging miRNAs. The 'PI3K-AKT signaling pathway', 'protein

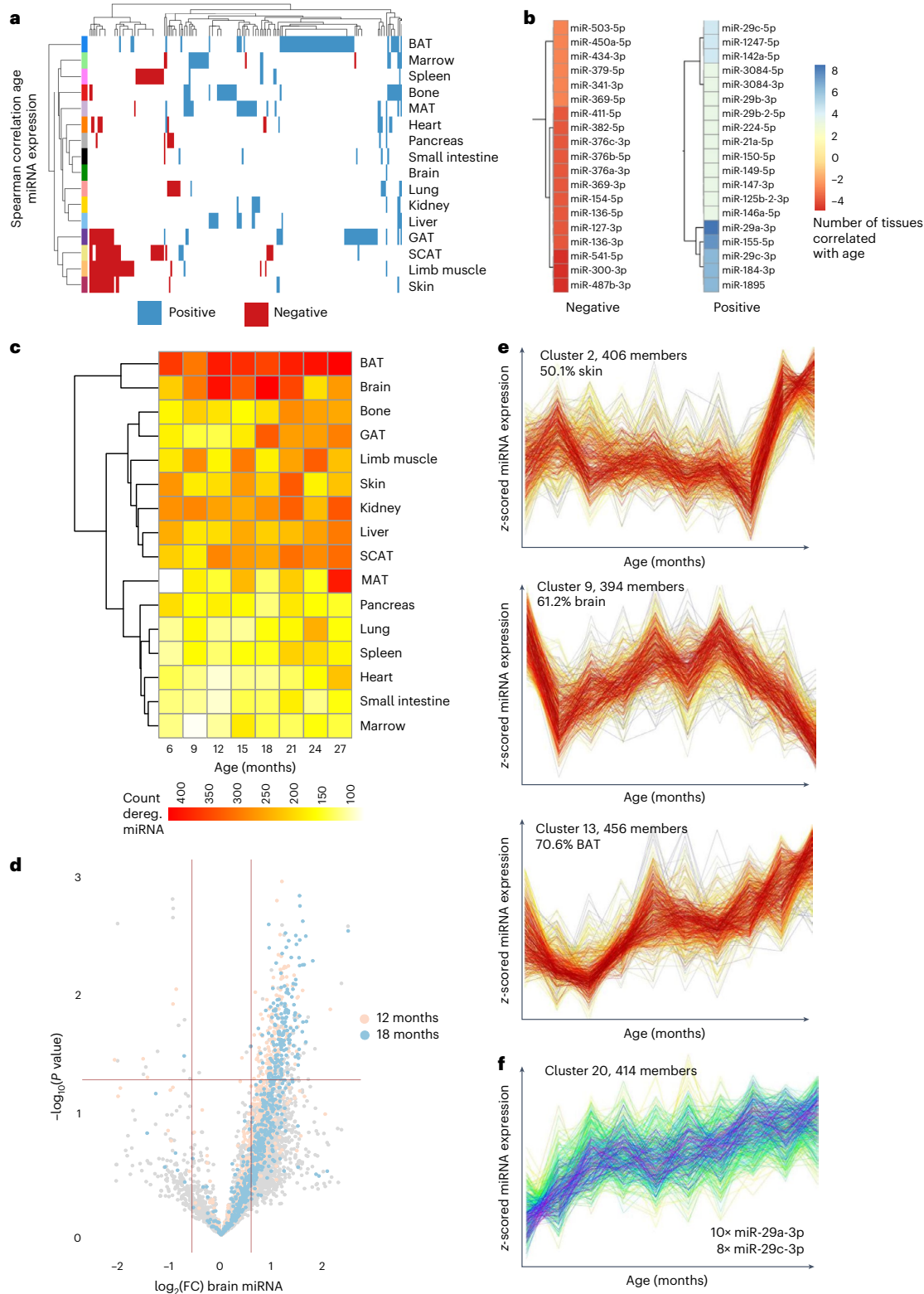
Fig. 2 | Global and tissue-specific miRNA expression patterns with aging.

a, Heatmap of Spearman rank correlation values of the intersection of miRNAs expressed in all tissues, color coded for positively correlated in blue ($r > 0.5$), negatively correlated in red ($r < -0.5$) and not correlated in white ($-0.5 < r < 0.5$). **b**, Heatmap of miRNAs (anti-) correlated with age in at least two tissues, colored by number of tissues (anti-) correlated and divided into miRNA positively and negatively correlated with age. **c**, Heatmap for the count of deregulated miRNAs in each tissue at each subsequent timepoint. Deregulated miRNAs are determined by calculating the foldchange of all later timepoints versus 3 months of age and miRNA with foldchanges $< 2/3$ or $> 3/2$ are considered deregulated. **d**, Volcano plot of all miRNAs expressed in brain, $\log_2(\text{FC})$ versus $-\log_{10}(P\text{values})$

(two-sided *t*-test) calculated between mice aged 3 months and all later timepoints with comparisons for 12 months (light red) and 18 months (light blue) highlighted. **e**, Whole organism miRNA trajectory clustering—*z*-scored trajectories of each expressed miRNA in each tissue over the entire lifespan of the mice were calculated. These trajectories were grouped into 20 clusters. Three clusters are displayed as examples, showing tissue-specific miRNA signatures. Cluster 2 is composed of 50.1% miRNAs originating from skin, cluster 9 of 61.2% miRNAs from brain and cluster 13 of 70.6% miRNAs from BAT. **f**, Cluster 20 of the whole organism miRNA trajectory clustering—the cluster contains two global aging miRNA, miR-29a-3p from ten different tissues and miR-29c-3p from eight different tissues.

digestion and adsorption', 'metabolic pathways', 'adipocytokine signaling pathway' and 'insulin resistance' were found among the top 20 locally enriched pathways in targeted mRNAs (Extended Data Fig. 6b).

Through a reduction of miRNA expression during aging, the repression of gene expression is potentially reduced or even lost. Gene targets for global aging miRNAs reducing repression with age were



identified via correlation (Fig. 2b) ($r < -0.4$, $P < 0.05$). The expression of three miRNAs, miR-300-3p, miR-487b-3p and miR-541-3p, decreased during the lifespan in five tissues. The overlap of their potential targets was high, with 138 of 327 predicted interactions (Fig. 3c and Supplementary Table 6). The identified targets exhibited a functional enrichment for pathways related to immune system processes, such as ‘cytokine–cytokine receptor interaction’, ‘Th1 and Th2 cell differentiation’, ‘Th17 cell differentiation’, ‘chemokine signaling pathway’ and ‘NF- κ B signaling pathway’. The network was particularly dense in its center, with targets related to ‘adaptive immunity’, ‘immunoglobulin’, ‘hematopoietic lineage’, ‘immune receptor activity’ and ‘cytokine activity’ (Fig. 3d). We also determined the locally enriched pathways for all miRNAs in every individual tissue whose expression decreased upon aging via inverse correlation with mRNA targets. These were similarly dominated by immune-related processes (Extended Data Fig. 6c). As immune senescence and inflammation are hallmarks of aging⁴, it is crucial to further investigate these potentially age-sensitive regulation mechanisms.

We chose the global aging miR-29c-3p as an example for further investigation. In liver and kidney, expression increased monotonically over the lifespan as well as in BAT but at a lower baseline expression (Fig. 3e). In the lung, the steep increase during early adulthood ends at approximately 12 months of age. A general trend of miR-29c-3p expression increase was present in all tissues, but expression levels and the course of increase showed tissue-specific patterns (Extended Data Fig. 7).

miR-29c-3p exhibits an organ-specific rejuvenation response

Expansive beneficial effects on cognition, muscle strength and bone repair have been observed for heterochronic parabiosis via a shared common circulation, or systemic infusions of young blood²². We sequenced tissue samples from a parabiosis intervention cohort to determine whether the young blood in aged individuals influences small ncRNA expression. The cohort was composed of 176 samples from six different organs of isochronic young (IY) and aged (IA), and heterochronic young (HY) and aged mice (HA) (Supplementary Table 1). Rejuvenation, the reversion of aging aspects, is the desired outcome of the intervention. However, it is accompanied by accelerated aging, the negative effect of the young sharing their blood with the old. In our study, the rejuvenation effect was measured by comparing the expression levels in IA mice with those detected in HA mice. In turn, the accelerated aging effect was defined by the difference between IY and HY mice (Extended Data Fig. 8a). Healthy aging was defined as the comparison of mice from the TMS cohort aged 3 and 21 months (AGE), closely matching the age distribution of the parabiosis cohort at takedown. Clustering the samples using t-SNE revealed tissue identity as major driver of variance across the experimental groups (Extended Data Fig. 8b,c).

We assigned deregulated miRNAs to the following groups: either (1) uniquely deregulated in rejuvenation (REJ unique) or in accelerated aging (ACC unique), or (2) deregulated in physiological aging as well as rejuvenation (REJ up and AGE down, REJ down and AGE up) or accelerated aging (AGE and ACC up/down). We found 233 uniquely deregulated miRNAs in rejuvenation and 43 in accelerated aging

(Fig. 4a). Intriguingly, 17 age-related miRNAs were deregulated in the opposite direction in REJ. No miRNAs were deregulated in AGE and in the same direction in ACC, but the uniquely rejuvenated miRNAs were enriched in certain pathways in MAT (‘insulin resistance’, ‘adipocytokine pathway’, ‘type 2 diabetes mellitus’), which again have a role in nutrient sensing.

For three global aging miRNAs, we discovered that changes in expression observed during healthy aging can be partially reversed in response to parabiosis. For miR-29c-3p, we measured a strong rejuvenation effect in the liver, four times higher than the effect of accelerated aging (Fig. 4b). The other two global aging miRNAs miR-184-3p in the liver and miR-300-5p in GAT showed similar trends of reversed expression but with a lower magnitude (Extended Data Fig. 9a,b). Considering the pronounced globally aging versus local rejuvenation profile of miR-29c-3p, we chose to explore systemic effectors and mediators of these signals.

Expression of circulating mir-29 family increases with aging

MiRNAs can circulate in the plasma and EVs between organs. We thus assessed the abundance of miR-29c-3p in both plasma and the vesicle-bound fraction using an independent cohort²³. Analyzing the expression at five timepoints across the lifespan from 2 to 18 months allowed us to correlate and compare the abundance of the miRNA in plasma and EVs. We observed an increase of miR-29c-3p expression correlated with age for both fractions ($r = 0.56$ (plasma) and 0.65 (EVs)). The share of positive global aging miRNAs detected as circulating was higher than the share of local aging miRNAs (38.3%) (Fig. 4c). This supports our hypothesis that miRNAs traveling via the shared circulatory system could have a role in the positive effect of parabiosis.

MiR-29c-3p could regulate gene expression in pathways resulting in health improvements by entering the tissue via the blood in vesicles. Recently, miRNAs with certain sequences were shown to be more likely secreted in small EVs, and their capability to inhibit target genes in recipient cells is enhanced⁸. One so-called EXOmotif is CNGGNC, which is very similar to a sequence found in the mature mmu-miR-29c-3p CUGGUG. We performed luciferase assay experiments to validate our predictions for mir-29-family members on target genes related to aging. *Lox* and *Adamts17* were validated as high confidence targets, and *Vash1* was validated as a low confidence target (Extended Data Fig. 9c,d). Previously known targets from the literature (*Eln*, *Col1a1*, *Col1a2*, *Col3a1* and *Adam12*), as well as *Lox* and *Adamts17*, are components in ECM processes (Fig. 4d), supporting our hypothesis that mir-29-family members have a crucial role in organismal aging due to their repressive regulatory function on these targets.

Discussion

We extended the TMS and parabiosis transcriptome datasets by bulk ncRNA sequencing and combined the data to highlight interactions of biomolecules and their functions to reveal potential regulatory mechanisms of aging. We report organ-specific trajectories during aging for miRNAs using organism-wide clustering. We thereby observed enrichment of miRNAs in pathways related to insulin resistance, especially for adipose tissue organ-specific miRNA trajectories. These results relate the miRNA expression changes to deregulated nutrient sensing.

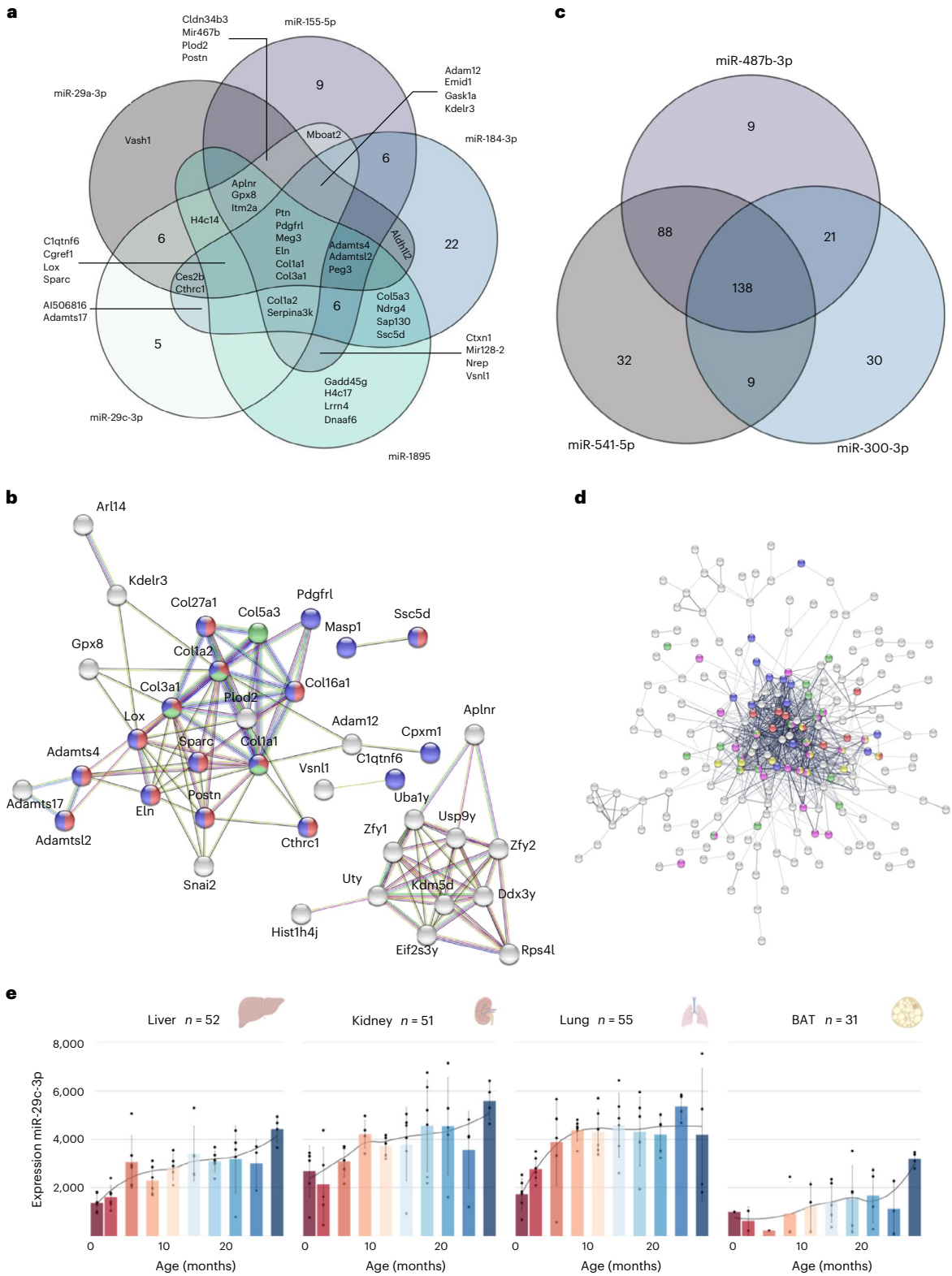
Fig. 3 | mRNA target correlation analysis for global aging miRNAs. a, Venn diagram of predicted target transcripts of the five global aging miRNAs positively correlated with age in most tissues. Targets are identified via inverse correlation of expression values (Spearman’s $\rho < -0.4$, Spearman’s statistics $P < 0.05$, two-sided); only miRNA–mRNA target predictions were selected that are correlated in at least two tissues for one of the five miRNAs. **b**, STRING network for all connected proteins encoded by target transcripts of the global aging miRNAs positively correlated with age (a); nodes are color coded for pathways in red for ‘ECM’, purple for ‘secreted’ and green for ‘dysregulated miRNA targeting in insulin PI3K–AKT signaling’. **c**, Venn diagram of predicted target

transcripts of the three global aging miRNAs negatively correlated with age in most tissues. **d**, STRING network for all connected proteins encoded by target transcripts of the negatively with age correlated global aging miRNA negatively correlated with age (c); nodes are color coded for pathways in red for ‘immune receptor activity’, in purple for ‘cytokine activity’, in yellow for ‘hematopoietic cell lineage’, in pink for ‘adaptive immunity’ and in green for ‘immunoglobulin’. **e**, Expression of miR-29c-3p in reads per mapped million in the liver ($r = 0.69$), kidney ($r = 0.56$), lung ($r = 0.51$) and BAT ($r = 0.48$) over the mouse lifespan (mean per timepoint $\pm$ s.d.), created with BioRender.

Article

<https://doi.org/10.1038/s41587-023-01751-6>

Moreover, we identified global aging miRNAs negatively and positively correlated with age. The increased expression levels of miR-29c-3p in age are partially reversible through heterochronic parabiosis. miR-29c is known as a negative regulator of RAG1 in B cells in mice and humans. Overexpression of miR-29c thereby reduces V(D)J recombination²⁴, which is a major process shaping the immune



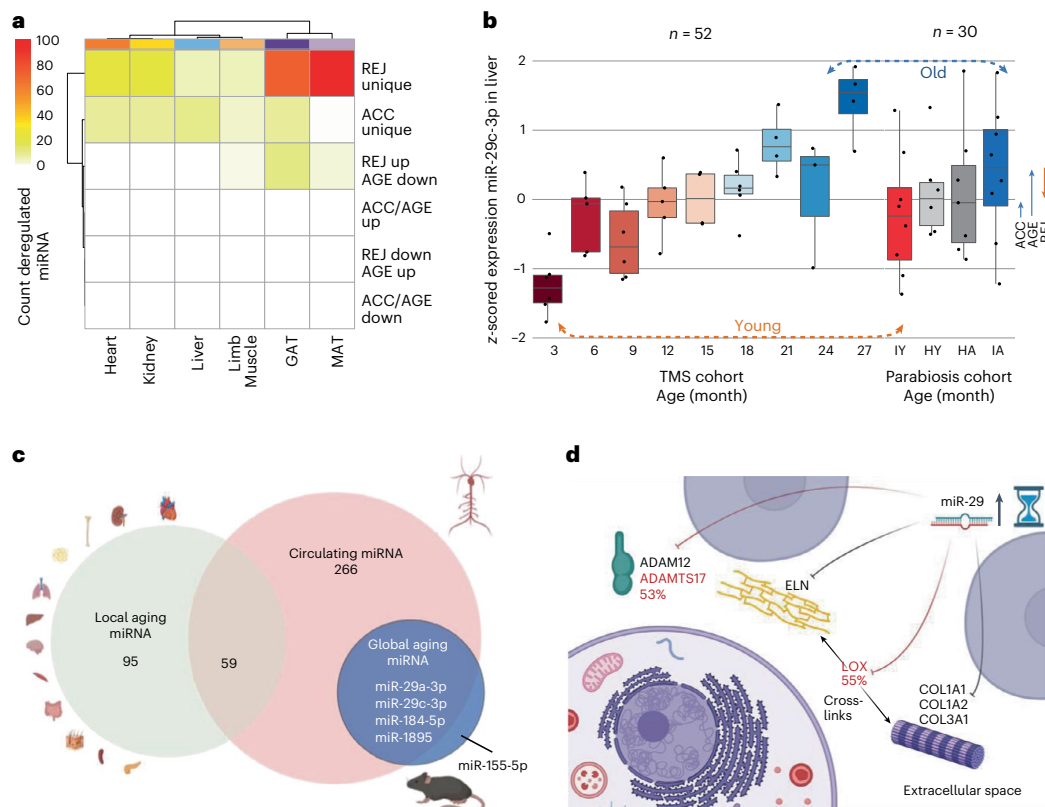


Fig. 4 | Compiled systemic mechanisms of global aging miRNAs. a, Heatmap of the absolute numbers of deregulated miRNAs in paradosis for the six comparisons in all analyzed tissues—uniquely deregulated in REJ, uniquely deregulated in ACC, in REJ up-regulated and healthy aging (AGE) downregulated, and vice versa, in AGE and ACC up-regulated or downregulated, all values >100 set to 100. **b**, Boxplot of expression z-scores for miR-29c-3p per timepoint in the liver for healthy aging from the TMS cohort and z-scores of miR-29c-3p in the liver from the parabiosis cohort for the following four different groups: IY, HY, HA and IA; arrows on the right indicate the effect size of aging in the parabiosis cohort, rejuvenation and accelerated aging. Boxes span the first to the third quartile

with the median value represented as line inside the box. The whiskers show the maximum and minimum values or values up to 1.5 times the interquartile range above and below the first or third quartile if outliers are present. **c**, Venn diagram of local aging miRNAs (exclusively positively correlated with age in one tissue; green), circulating miRNAs (all miRNAs expressed in EVs and serum, red) and global aging miRNAs (positively correlated with age in more than five tissues, blue). **d**, Summary of validated targets of miR-29 family members, shown in their cellular location. These targets are key proteins in processes related to the ECM; newly validated targets are shown in red. Created with BioRender.

system repertoire, to support clearance of infectious agents, infected cells and cells on the verge of malignant transformation¹. The global increase in this miRNA in several tissues and the already-known regulation of the immune system suggest that the immune senescence aggravating the aging phenotype could be caused by this development. Another age-related pathology is the process of cellular senescence, which is regulated by the TGF- β /Smad pathway. TGF- β signaling involves miR-29-induced loss of H4K20me3 to promote senescence²⁵. In a brain-specific miR-29 knockdown mouse, sex-specific effects on lifespan and reproduction were observed²⁶. To prove that miR-29 has a causal role in processes responsible for cellular aging and rejuvenation, detailed knockdown or knockout experiments are needed.

Future studies should also focus on gathering single-cell miRNA data to explain which cell types are responsible for the expression of miRNA aging markers. High-throughput single-cell sequencing and vesicle sequencing could help us to distinguish between cellular miRNA expression and vesicles. MiRNAs can be transported via EVs and thereby mediate the regulation of aging-related processes²⁷. Also, miRNA-mediated gene silencing, which we based our targeting analysis on and used for validation, is only one mode of action of gene expression regulation. Other modes of action worth mentioning

are miRNA-mediated translational activation, miRNA-mediated transcriptional and post-transcriptional gene regulation within the nucleus²⁸. A more detailed view of cell-type-specific and vesicular expression might explain why we found distinct miRNA trajectories of aging in adipose tissue while the strongest rejuvenation effects for global aging miRNAs, especially miR-29c-3p, occurred in the liver. The miRNA is known to be expressed highest in T and B cells²⁴, but is also expressed in liver hepatocytes (Extended Data Fig. 10) and reported as a potential tumor suppressor in human^{29,30}. Hence, revealing the responsible cell type could help illuminate which mechanisms modulate miRNA expression levels. It is also necessary to discern which changes impact the transcriptome and proteome in different tissues and cellular compositions, as miRNA targetomes can differ across cell types³¹. Currently, only a few different protocols for single-cell miRNA sequencing exist and no high-throughput gold standard is available³².

Another limitation of the study is the known issues of small RNA library production as adapter ligation bias, adapter dimer contamination, polymerase chain reaction (PCR) amplification bias, barcode bias and the influence of RNA degradation on ncRNA profiles^{33–35}. The challenge of sequencing mainly fragments for six of the eight RNA classes is

Article

<https://doi.org/10.1038/s41587-023-01751-6>

related to these. Potentially, a major part of piRNA reads in the somatic tissues could have been derived from piRNA-sized fragments of other ncRNAs. These fragments are annotated in piRNA databases even though their biogenesis is perhaps independent of the PIWI pathway¹³. However, these piRNA-like small RNA are known to have important roles outside of the germline³⁶. tRNA-derived small RNAs, which have a biological role by inhibiting translation or regulating gene expression, are studied likewise in aging and age-related diseases³⁷. We decided not to exclude these data, so our study can be used as a reference for future studies aiming to analyze for instance tRNA-derived fragments or piRNA-like small RNAs in more detail. Of note, the fragments of longer ncRNAs are not necessarily surrogates of the full-length mature transcripts but can occur due to degradation processes. The biological function of respective mapping results remains to be explored.

In summary, our study provides a rich resource for biologists across many disciplines, as ncRNAs for all major organs across the entire lifespan of the mouse were sequenced. Reference data for healthy aging are important, because miRNAs are promising candidates for age-specific disease biomarkers¹⁰, and patterns of physiological aging must be defined not only in blood but also in every solid organ to promote the development of successful RNA-based therapies.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41587-023-01751-6>.

References

- Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
- Valdes, A. M., Glass, D. & Spector, T. D. Omics technologies and the study of human ageing. *Nat. Rev. Genet.* **14**, 601–607 (2013).
- Levine, M. E. et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging* **10**, 573–591 (2018).
- Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* **19**, 371–384 (2018).
- Schaum, N. et al. Ageing hallmarks exhibit organ-specific temporal signatures. *Nature* **583**, 596–602 (2020).
- Tabula Muris, C. et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
- Bartel, D. P. Metazoan microRNAs. *Cell* **173**, 20–51 (2018).
- Garcia-Martin, R. et al. MicroRNA sequence codes for small extracellular vesicle release and cellular retention. *Nature* **601**, 446–451 (2022).
- Tkach, M. & Thery, C. Communication by extracellular vesicles: where we are and where we need to go. *Cell* **164**, 1226–1232 (2016).
- Fehlmann, T. et al. Common diseases alter the physiological age-related blood microRNA profile. *Nat. Commun.* **11**, 5958 (2020).
- Singh, J. et al. Aging-associated changes in microRNA expression profile of internal anal sphincter smooth muscle: role of microRNA-133a. *Am. J. Physiol. Gastrointest. Liver Physiol.* **311**, G964–G973 (2016).
- Castellano, J. M., Kirby, E. D. & Wyss-Coray, T. Blood-borne revitalization of the aged brain. *JAMA Neurol.* **72**, 1191–1194 (2015).
- Tosar, J. P., Rovira, C. & Cayota, A. Non-coding RNA fragments account for the majority of annotated piRNAs expressed in somatic non-gonadal tissues. *Commun. Biol.* **1**, 2 (2018).
- Li, X. Z. et al. An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. *Mol. Cell* **50**, 67–81 (2013).
- Ding, D. et al. TDRD5 binds piRNA precursors and selectively enhances pachytene piRNA processing in mice. *Nat. Commun.* **9**, 127 (2018).
- Zhou, Z., Sun, B., Yu, D. & Bian, M. Roles of tRNA metabolism in aging and lifespan. *Cell Death Dis.* **12**, 548 (2021).
- Trajkovski, M. et al. MicroRNAs 103 and 107 regulate insulin sensitivity. *Nature* **474**, 649–653 (2011).
- Ximerakis, M. et al. Single-cell transcriptomic profiling of the aging mouse brain. *Nat. Neurosci.* **22**, 1696–1708 (2019).
- van Rooij, E. et al. Dysregulation of microRNAs after myocardial infarction reveals a role of miR-29 in cardiac fibrosis. *Proc. Natl Acad. Sci. USA* **105**, 13027–13032 (2008).
- Szklarczyk, D. et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).
- Mavrogonatou, E., Pratsinis, H., Papadopoulou, A., Karamanos, N. K. & Kletsas, D. Extracellular matrix alterations in senescent cells and their significance in tissue homeostasis. *Matrix Biol.* **75–76**, 27–42 (2019).
- Palovics, R. et al. Molecular hallmarks of heterochronic parabiosis at single-cell resolution. *Nature* **603**, 309–314 (2022).
- Kern, F. et al. Vesicle-bound regulatory RNAs are associated with tissue aging. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.05.07.443093> (2021).
- Kumari, R. et al. MicroRNA miR-29c regulates RAG1 expression and modulates V(D)J recombination during B cell development. *Cell Rep.* **36**, 109390 (2021).
- Lyu, G. et al. Addendum: TGF-beta signaling alters H4K20me3 status via miR-29 and contributes to cellular senescence and cardiac aging. *Nat. Commun.* **9**, 4134 (2018).
- Takeda, T. & Tanabe, H. Lifespan and reproduction in brain-specific miR-29-knockdown mouse. *Biochem. Biophys. Res. Commun.* **471**, 454–458 (2016).
- Zhang, Y. et al. Hypothalamic stem cells control ageing speed partly through exosomal miRNAs. *Nature* **548**, 52–57 (2017).
- O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. Overview of MicroRNA biogenesis, mechanisms of actions, and circulation. *Front. Endocrinol.* **9**, 402 (2018).
- Wu, H. et al. miR-29c-3p regulates DNMT3B and LATS1 methylation to inhibit tumor progression in hepatocellular carcinoma. *Cell Death Dis.* **10**, 48 (2019).
- Patil, A. H., Baran, A., Brehm, Z. P., McCall, M. N. & Halushka, M. K. A curated human cellular microRNAome based on 196 primary cell types. *Gigascience* **11**, giac083 (2022).
- Clark, P. M. et al. Argonaute CLIP-seq reveals miRNA targetome diversity across tissue types. *Sci. Rep.* **4**, 5947 (2014).
- Hucker, S. M. et al. Single-cell microRNA sequencing method comparison and application to cell lines and circulating lung tumor cells. *Nat. Commun.* **12**, 4316 (2021).
- Raabe, C. A., Tang, T. H., Brosius, J. & Rozhdestvensky, T. S. Biases in small RNA deep sequencing data. *Nucleic Acids Res.* **42**, 1414–1426 (2014).
- Buschmann, D. et al. Toward reliable biomarker signatures in the age of liquid biopsies—how to standardize the small RNA-seq workflow. *Nucleic Acids Res.* **44**, 5995–6018 (2016).
- Ludwig, N. et al. Bias in recent miRBase annotations potentially associated with RNA quality issues. *Sci. Rep.* **7**, 5162 (2017).
- Yan, Z. et al. Widespread expression of piRNA-like molecules in somatic tissues. *Nucleic Acids Res.* **39**, 6596–6607 (2011).
- Yuan, Y. et al. tRNA-derived fragments as new hallmarks of aging and age-related diseases. *Aging Dis.* **12**, 1304–1322 (2021).

Article<https://doi.org/10.1038/s41587-023-01751-6>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this

article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Article

<https://doi.org/10.1038/s41587-023-01751-6>**Methods****Samples**

Mouse tissues of the aging cohort were obtained, and RNA was isolated as previously described³. From the National Institute on Ageing colony (Charles River) male and virgin female C57BL/6JN mice were shipped to the Veterinary Medical unit at the VA Palo Alto. The mice were housed on a 12-h light/dark cycle at 20–24 °C with food and water provided *ad libitum*. Humidity was monitored daily and between 23% and 55%. Mice from both cohorts were anesthetized with 2.5% vol/vol avertin, and mice were weighed and shaved. Blood was drawn via cardiac puncture before transcatheterial perfusion with 20 ml PBS. Dissection of organs was performed in the following order and then instantly frozen on dry ice: pancreas, spleen, brain, heart, lung, kidney, mesenteric adipose tissue, intestine (duodenum), gonadal adipose tissue, muscle (tibialis anterior), skin (dorsal), subcutaneous adipose tissue (inguinal pad), brown adipose tissue (interscapular pad), bone and bone marrow (femurs and tibiae). Bulk RNA samples of the heterochronic parabiosis cohort consisting of male C57BL/6JN, C57BL/6J and C57BL/6-Tg(UBC-GFP)30Scha/J mice were collected as previously described²² (Supplementary Table 1). The 3- to 4.5-month-old and 19-month-old mice were housed under the same conditions as the aging cohort mice. Suturing together the peritoneum of adjacent flanks of two mice, forming a continuous peritoneal cavity, accomplished the aging intervention parabiosis via the peritoneal method. To enable coordinated movement after surgery, adjacent knee joints and elbow joints were joined with nylon monofilament sutures, as well as skin with surgical autoclips. These procedures were conducted with aseptic conditions on heating pads, with mice under continuous isoflurane anesthesia. Mice were injected with Baytril (5 µg g⁻¹), buprenorphine and 0.9% (wt/vol) sodium chloride to avoid infection, limit pain and promote hydration, as previously described in ref. 22. For 5 weeks, the pairs remained together, and organs were collected. First, heart, liver, kidney, then MAT and GAT, and finally limb muscle were collected in this order, all within 30–40 min. All animal care and procedures were carried out in accordance with institutional guidelines approved by the VA Palo Alto Committee on Animal Research (Protocol, LU01736). RNA was isolated according to the manufacturer's protocol with the miRNeasy Kit (Qiagen, 217084). All RNA samples were shipped to the Institute of Human Genetics. Samples of the TMS cohort were additionally precipitated due to salt contamination. In brief, 150 ng of RNA was mixed with 3 M NAAC (pH 7.0) and 100% EtOH and incubated overnight at -20 °C. This was centrifuged at 20,817g at 4 °C for 60 min. Supernatant was discarded and the pellet was washed with 80% EtOH, followed by another centrifugation for 30 min (20,817g, 4 °C). Supernatant was again discarded, the pellet was dried on ice and resuspended in 50 µl 1x TE buffer. Quality control of concentration was performed with the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific), and the RNA integrity was determined using the Agilent RNA 6000 Nano Kit (Agilent Technologies, 5067-1512) for randomized samples of the cohorts.

Sample size, randomization and blinding

No sample size choice was performed before the study. During mouse dissection, order and preparation of 96-well plates for cDNA creation randomization was performed. No blinding was performed; the authors were aware of all data and metadata-related variables during the entire course of the study.

Library preparation

Small RNA library preparation was performed using the MGIEasy Small RNA Library Prep Kit (Item 940-000196-00) and the high-throughput MGI SP-960 sample prep system according to the manufacturer's protocols. In brief, 3'- and 5'-adapters were ligated to the RNA, and reverse transcription (RT) was performed using an RT primer, in which sample-specific barcodes were incorporated. The resulting cDNA was

amplified in a PCR with 21 cycles. The amplification product was size selected and purified using AMPure Beads XP (Beckman Coulter). The size of the purified PCR products was checked using an Agilent DNA 1000 Kit (Agilent Technologies), and the concentration was determined using Qubit 1X dsDNA High Sensitivity (Thermo Fisher Scientific). For each library, 16 samples, barcoded with barcodes 1–4, 13–16 and 25–32, were pooled in an equimolar fashion at a concentration of 4.56 ng µl⁻¹. Pooled libraries were circularized and sent for sequencing. A total of 65 libraries consisting of 947 samples were analyzed in the project.

Sequencing and data analysis

Samples were sequenced single-ended on the BGISEQ500RS using the High-throughput Sequencing Set (SE50) (Small RNA) as a service provided by BGI. The sequencing data were processed with miRMaster 2.0 using standard settings³⁸ and mapped read percentages were generated. Data analysis was performed using RStudio Software v4.0.3 with the following packages: viper v1.26.0, data.table v1.14.2, ggrepel v0.9.1, ggvenn v0.1.9, M3C v1.14.0, ggridges v0.5.3, forcats v0.5.1, purrr v0.3.4, tidyr v1.2.0, tibble v3.1.6, ggplot2 v3.3.5, tidyverse v1.3.1, viridisLite v0.4.0, ColorBrewer v1.1-2, reshape2 v1.4.4, pheatmap v1.0.12, Mfuzz v2.52.0, DynDoc v1.70.0, widgetTools v1.70.0, e1071 v1.7-9, stringr v1.4.0, dplyr v1.0.8, readr v2.1.2 and Biobase v2.52.0.

Samples were excluded if fewer than 2 million aligned reads were detected while allowing one mismatch per read. Using Bowtie (v1.2.3.), reads were mapped against the RNA sequence derived from the respective databases (miRNAs: miRBase 22, tRNAs: GtRNAdb 18.1, piRNA: RNACentral 15, all other ncRNAs: Ensembl 100). Only the first paralog was retained for analysis, additional paralogs are listed in Supplementary Table 7. As the lengths of the mature ncRNAs matched with our sequencing read length exclusively for miRNAs and piRNAs³⁹, we calculated detailed covered sequence length statistics. These analyses verify that not only random fragments were sequenced for the other ncRNA classes. Such fragments can occur as a result of a physiological process like tRNA-derived fragments and have a regulatory role in aging³⁷ or can be products of postmortem RNA degradation. The amount and distribution of degradation fragments are highly influenced by the RNA quality³⁵. Covered read length, reference read length, longest covered region, covered percentage reference length, longest mapping read, total reads mapping and average covered read length are listed in Supplementary Table 2 for all detected ncRNA. All mature ncRNAs are represented by their highest counting precursor.

Percentages of aligned reads per sample derived from the miRMaster analysis were used to calculate mean percentages within each tissue and each timepoint. As a first filtering step, piRNAs were filtered for those encoded in prepachytene genomic piRNA clusters as an established method to identify true somatic piRNAs^{14,15}.

For global analyses (analyses independent of the organ), samples were filtered for 1 rpmm in at least one sample in the cohort. As a global analysis, we performed a t-SNE and a principal variance component analysis (PVCA). All samples were clustered in an unweighted t-SNE with a seed set to 40 using the M3C package. A t-SNE is an optimized dimensionality reduction method used for the visualization of high-dimensional data⁴⁰. A PVCA was used to estimate the variability of biological and technical parameters. Data dimensionality is reduced through a principal component analysis.

For local analyses, tissue-specific patterns were considered, further requiring that ncRNAs were expressed with 1 rpmm in at least 10% of the samples for each tissue. Percentages of counts per RNA class and tissue were calculated with the total number of counts within a tissue and the respective mean number of counts of the RNA classes. Percentages of counts per tissue and timepoint were calculated with the percentage of counts per sample after local filtering with the total RNA class counts. Timepoint percentages were calculated as means of the corresponding samples at each timepoint and in each tissue. Spearman rank correlations with age of each ncRNA expressed in each

Article

<https://doi.org/10.1038/s41587-023-01751-6>

tissue were calculated and illustrated in a density plot grouped by RNA class. Spearman rank correlation between miRNA expression and age was categorized into positively ($r > 0.5$) and negatively ($r < -0.5$) correlated and annotated with P values (Supplementary Table 8). Based on this categorization, the number of tissues in which a miRNA was (anti-) correlated with age was determined. MiRNA FCs in each tissue were computed with the mean expression of each later timepoint, always comparing against 3 months. Features with mean expression of 0 for 3 months of age were excluded from this analysis. FCs lower or higher than 2/3 and 3/2, respectively, were considered deregulated. P values were only calculated with t -tests for comparisons with at least three samples per timepoint and adjusted for each tissue and timepoint separately with the Benjamini and Hochberg method. For the volcano plots, $\log_2(\text{FC})$ was calculated and all FC equal to 0 were discarded. Volcano plots for each tissue were generated with the $-\log_{10}(P\text{ values})$ versus the $\log_2(\text{foldchange})$ and colored by timepoint. Organism-wide miRNA trajectory clustering was performed using the Mfuzz package, which clustered based on fuzzy c-means algorithms, and the number of clusters c between 2 and 20 was individually determined for each clustering using the minimum centroid distance measure. For the organism-wide clustering of the z-scored miRNAs over all tissues, 20 was determined as optimal. A cluster was deemed tissue-specific if at least 30% of the miRNAs in a cluster were tissue-specific.

The coding transcriptome data for the same samples were obtained from the previous study⁵. mRNA targets of miRNAs were identified via negative correlation ($P < 0.05$, $r < -0.4$). For the local miRNA–mRNA interaction analysis, miRNAs exceeding the age-correlation interval between -0.5 and 0.5 were considered. The more stringent filtering approach for aging miRNAs was chosen to discover a small set of strong candidates from the millions of possible miRNA–mRNA interactions. For the global analysis, we considered targets inversely correlated with either the positive global aging miRNAs (miR-29a-3p, miR-29c-3p, miR-155-5p, miR-184-3p and miR-1895) or the negative global aging miRNAs (miR-300-3p, miR-487b-3p and miR-541-5p) in at least two tissues, to obtain the filtered target gene set. Using STRING, the protein–protein association network database²⁰, we illustrated known connections between proteins encoded by the predicted mRNA targets of the global miRNAs.

For pathway enrichment analysis, an overrepresentation analysis (ORA) was performed with the target mRNAs of global miRNAs using GeneTrail 3.2 (ref. 41). An ORA was performed to identify the pathways negatively and positively regulated locally in all tissues through the local aging miRNAs. The standard parameters were used, with FDR adjustment and 0.001 as significance level. Heatmaps for positive and negative regulation of miRNA on target mRNA expression were generated with the top 20 and 25 nondisease-related pathways, respectively, (lowest P values) regulated in most tissues.

For the parabiosis cohort, sequenced samples were analyzed with the same filtering criteria as TMS samples. We detected 50,776 ncRNAs in the raw reads of this cohort and we filtered 5,248 abundant ncRNAs for the global analysis (t-SNE). To quantify the effects of parabiosis in each tissue, FCs were calculated between IY and HY mice for the effect termed accelerated aging (ACC) and between IA and HA mice for the effect termed REJ. The effects of physiological aging (AGE) were defined as the FC between 3- and 21-month-old mice from the TMS cohort, corresponding to the ages of young and aged mice in the parabiosis experiment. As previously, FCs exceeding the interval of 2/3 and 3/2 and a significant P value ($P < 0.05$) were considered as deregulated.

Expression data from the EVs study were obtained as previously reported in ref. 23. miRNAs were considered as expressed if they were detected with at least 1 rpmm for more than 10% of the samples of one group. The intersection between miRNAs expressed in circulating plasma and EVs and either global or local aging miRNAs was determined and visualized as a Venn diagram.

Cell lines

The HEK 293T (ACC 635) was purchased from the German collection of microorganisms and cell cultures (Deutsche Sammlung von Mikroorganismen und Zellkulturen, DSMZ). STR fingerprinting by DSMZ confirmed the authenticity of the cell line. The cells were cultivated with DMEM (Life Technologies) supplemented with Penicillin (100 U ml⁻¹), Streptomycin (100 µg ml⁻¹) and 10% (vol/vol) FCS and passaged two times a week for not longer than 3 months.

miRNA expression plasmid and 3'UTR reporter plasmids

The cloning of the pSG5-miR-29a expression plasmid was described previously⁴². Targets for reporter plasmids were selected based on the predicted target genes for miR-29 from Fig. 3a, as miR-29a-3p and miR-29c-3p have the same seed sequence. Only target genes with at least a 7mer binding site and the lowest possible hamming distance between human and mouse 3'UTR and binding sites were selected. The 3'UTR reporter constructs were synthesized and cloned into reporter plasmid pMIR-RNL-TK using SpeI and SacI restriction sites by GeneArt (Life Technologies GmbH). The reporter plasmid pMIR-COL1A2, which was identified by ref. 43 as direct target gene of miR-29a-3p, served as positive control. The results of the control experiments are given in Extended Data Fig. 7d. The complete list of all tested 3'UTR sequences, including the respective NM accession number, is given in Supplementary Table 9.

High-throughput miRNA interaction reporter assay

High-throughput analysis of reporter constructs was conducted by a liquid handling system and described previously in ref. 44. In brief, HEK 293T cells were seeded at 3.2×10^4 cells per well in a 96-well plate using a liquid handling system epMotion 5,075 (Eppendorf). Twenty-four hours after seeding, cells were transfected with 50 ng per well of either reporter plasmid pMIR-RNL-TK, with or without insert, and 200 ng per well of miRNA expression plasmid pSG5-miR-29a or the empty expression vector pSG5. Forty-eight hours post-transfection, HEK 293T cells were lysed and the lysates were measured using a GloMax Navigator microplate luminometer (Promega) using Luciferase substrates of the Dual-Luciferase Reporter Assay System (Promega). High-throughput miRNA interaction reporter assay was conducted four times in technical duplicates.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data generated in this study are freely accessible from the Gene Expression Omnibus (GSE217458, GSE222857). Databases used in this study are miRBase 22 (<https://www.mirbase.org/>), GtRNSdb 18.1 (<http://gtrnadb.ucsc.edu/>), RNCentral 15 (<https://rnacentral.org/>) and Ensembl 100 (<https://useast.ensembl.org/index.html>).

Code availability

All analyses have been carried out using freely available software packages. Custom code used to analyze the RNA-seq data and datasets generated and/or processed in the current study is available from the corresponding authors upon request.

References

- Fehlmann, T. et al. miRMaster 2.0: multi-species non-coding RNA sequencing analyses at scale. *Nucleic Acids Res.* **49**, W397–W408 (2021).
- Fehlmann, T. et al. A high-resolution map of the human small non-coding transcriptome. *Bioinformatics* **34**, 1621–1628 (2018).
- van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

Article

<https://doi.org/10.1038/s41587-023-01751-6>

41. Gerstner, N. et al. GeneTrail 3: advanced high-throughput enrichment analysis. *Nucleic Acids Res.* **48**, W515–W520 (2020).
42. Szczyrba, J. et al. Analysis of argonaute complex bound mRNAs in DU145 prostate carcinoma cells reveals new miRNA target genes. *Prostate Cancer* **2017**, 4893921 (2017).
43. Pan, H. et al. LncRNA LIFR-AS1 promotes proliferation and invasion of gastric cancer cell via miR-29a-3p/COL1A2 axis. *Cancer Cell Int.* **21**, 7 (2021).
44. Kern, F. et al. Validation of human microRNA target pathways enables evaluation of target prediction tools. *Nucleic Acids Res.* **49**, 127–144 (2021).

Acknowledgements

We thank all members of the Wyss-Coray and Meese lab, as well as all members of the Keller lab for feedback and support. We would like to thank the C. Zuckerberg Biohub. This study is funded by the M.J. Fox Foundation (MJFF-021418; A. K. and T.W.-C.), the Schaller-Nikolich Foundation (A.K.) and Saarland University. Computational resources used within this study were financed through the DFG project 466168626 (A.K.).

Author contributions

V.W. performed all sequencing experiments, analyzed the data and wrote the manuscript. O.H., N.S., A.I., M.T., R.S., N.N., S.Q. and T.W.-C. collected the samples. N.L. helped perform sequencing experiments. T.F., F.K., A. E. and S.R. helped with data analysis and visualization. M.H. performed luciferase assay experiments. A.K., T.W.-C. and N.S. designed the research. O.H., A.K., T.W.-C.,

F.K., D.H. and N.S. reviewed the manuscript. The project was supervised by E.M., A.K. and T.W.-C., who were also responsible for the funding acquisition.

Funding

Open access funding provided by Universität des Saarlandes.

Competing interests

A.K. is a member of the scientific advisory board of Firalis. The remaining authors declare no competing interests.

Additional information

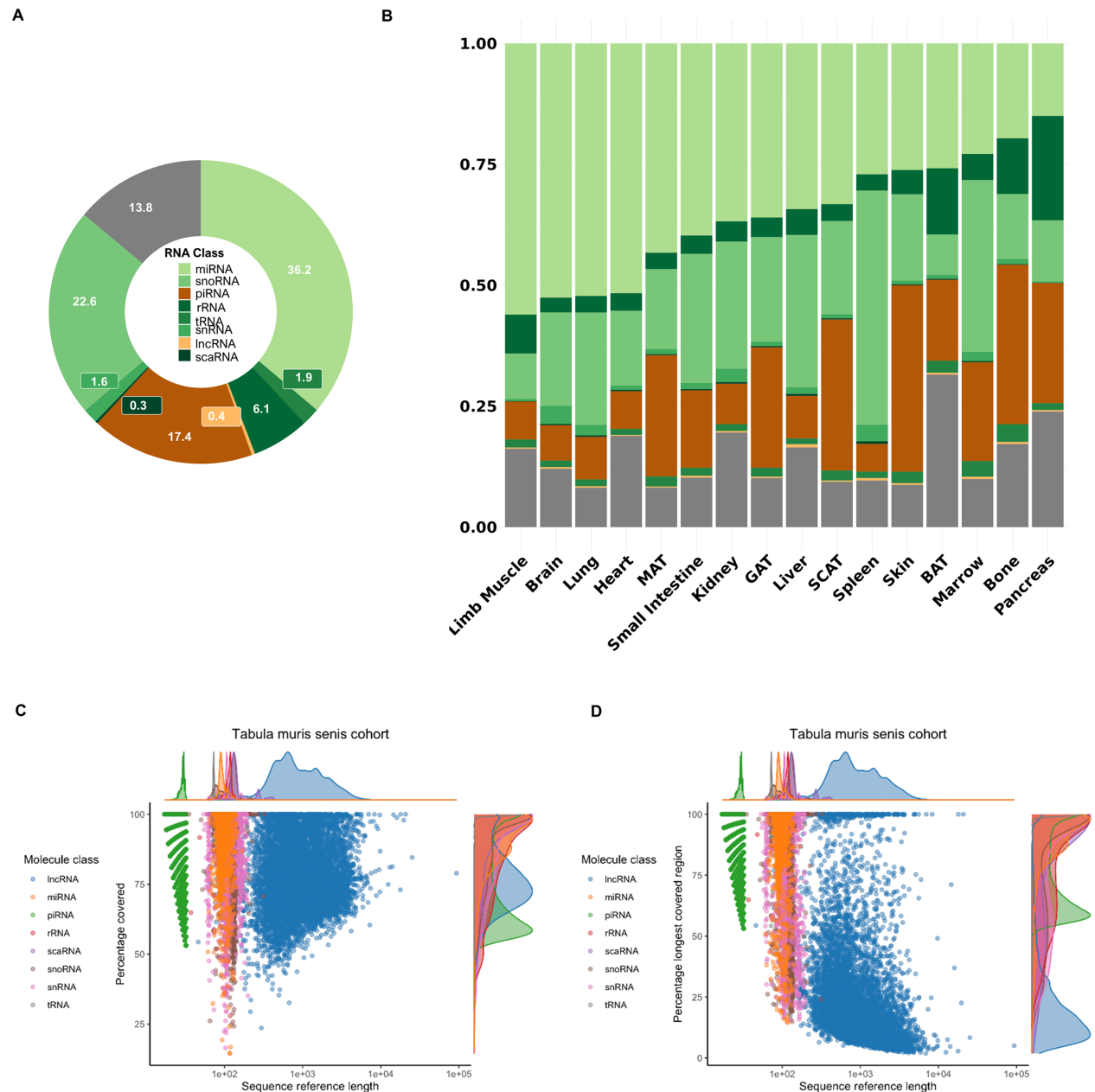
Extended data is available for this paper at <https://doi.org/10.1038/s41587-023-01751-6>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41587-023-01751-6>.

Correspondence and requests for materials should be addressed to Tony Wyss-Coray or Andreas Keller.

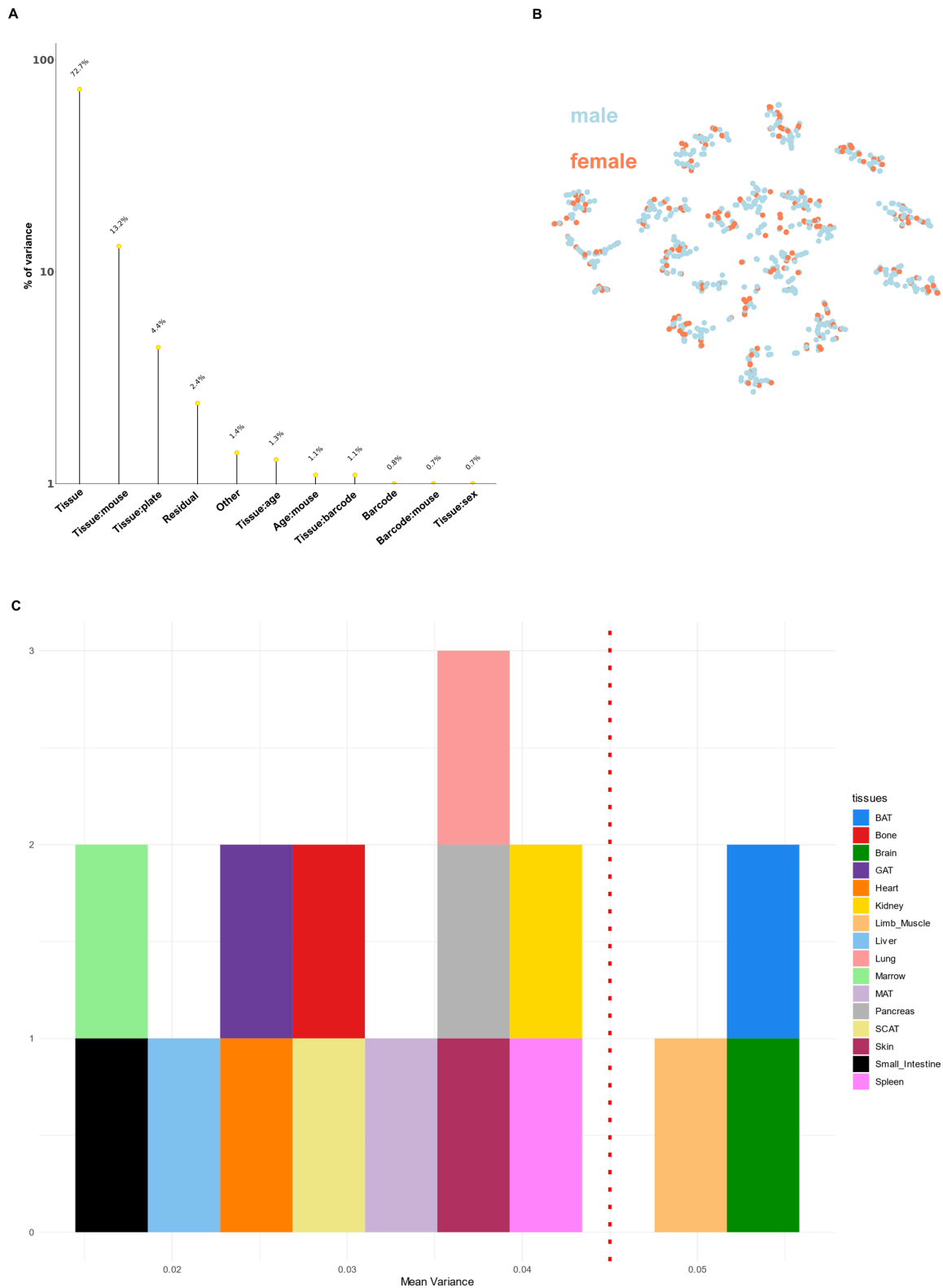
Peer review information *Nature Biotechnology* thanks Frank Slack and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

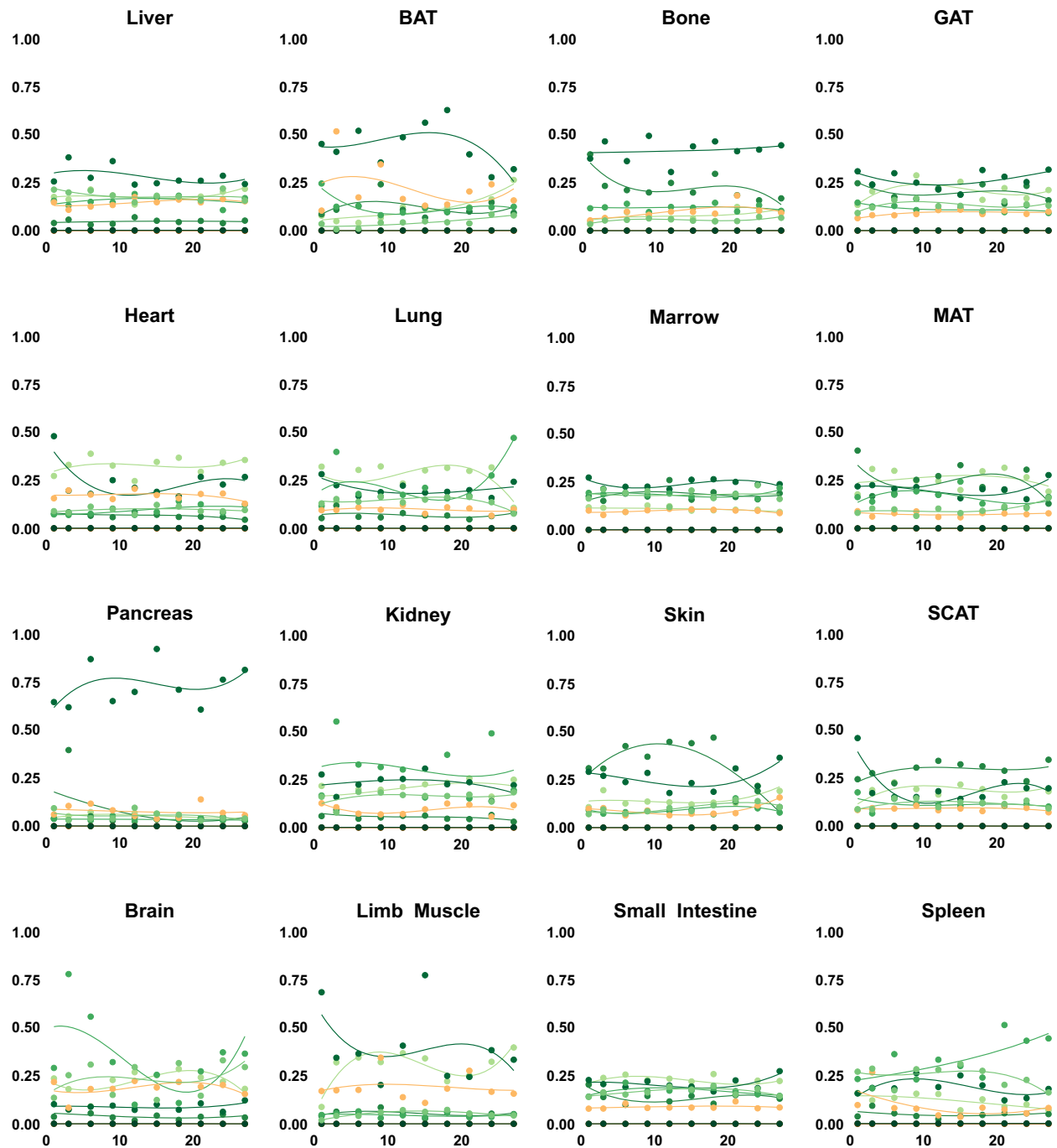


Extended Data Fig. 1 | Non-coding RNA read distribution. (A) Averaged percentage of mapped reads for all RNA classes over all TMS cohort samples. (B) Mapped RNA class read distribution in percent for each tissue, sorted descending by miRNA share. (C) Covered length analysis: Percentage covered of reference sequence with mapped reads versus sequence reference length,

colored by RNA class for all ncRNAs detected in the raw reads for the TMS cohort (D) Percentage of longest covered region calculated with the maximal connected read length versus sequence reference length for all ncRNAs detected in the raw reads for the TMS cohort.

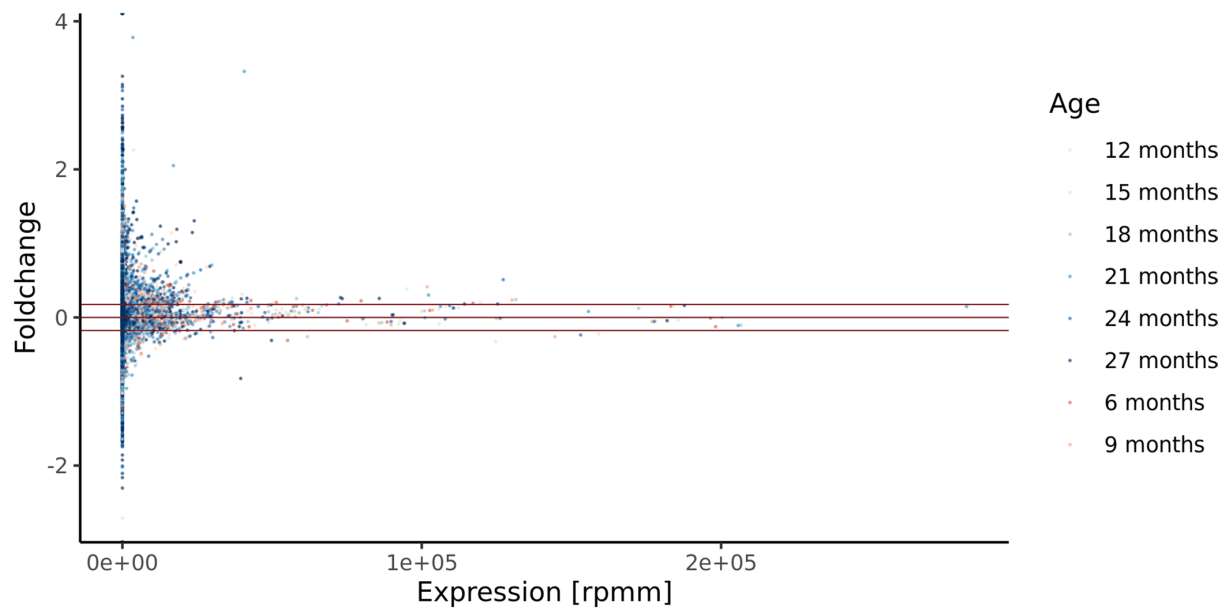


Extended Data Fig. 2 | Variance analysis for ncRNA expression levels. (A) Principal variance component analysis of biological components (sex, age, tissue, mouse) and technical components (barcode, plate); factors connected with ':' were linearly combined. (B) Visualization of all samples of the TMS cohort over all abundant non-coding RNAs as a t-SNE, colored by sex. (C) Histogram of mean variance calculated in abundant count percentages of all RNA classes. Calculations for each tissue after local filtering over the time course. Threshold determined at 4.5% to separate highly variable and stable tissues.

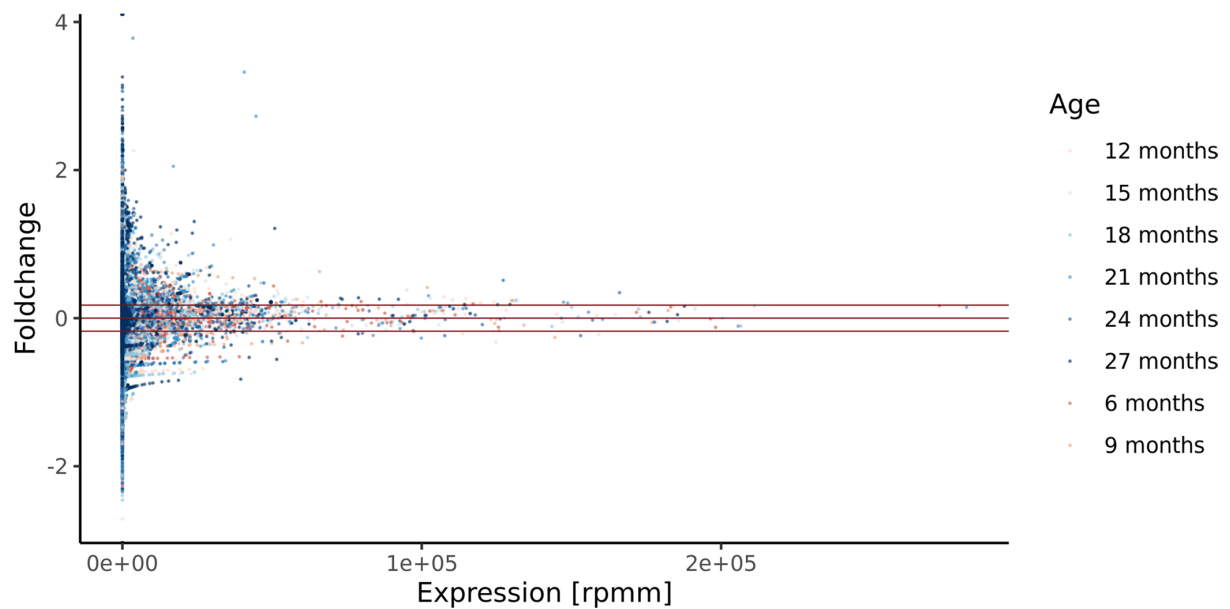


Extended Data Fig. 3 | Mean count percentages per RNA class over the lifespan in all organs. Mean count percentages per RNA class and over all timepoints, for each tissue in the TMS dataset. Calculations based on count shares per sample after local filtering.

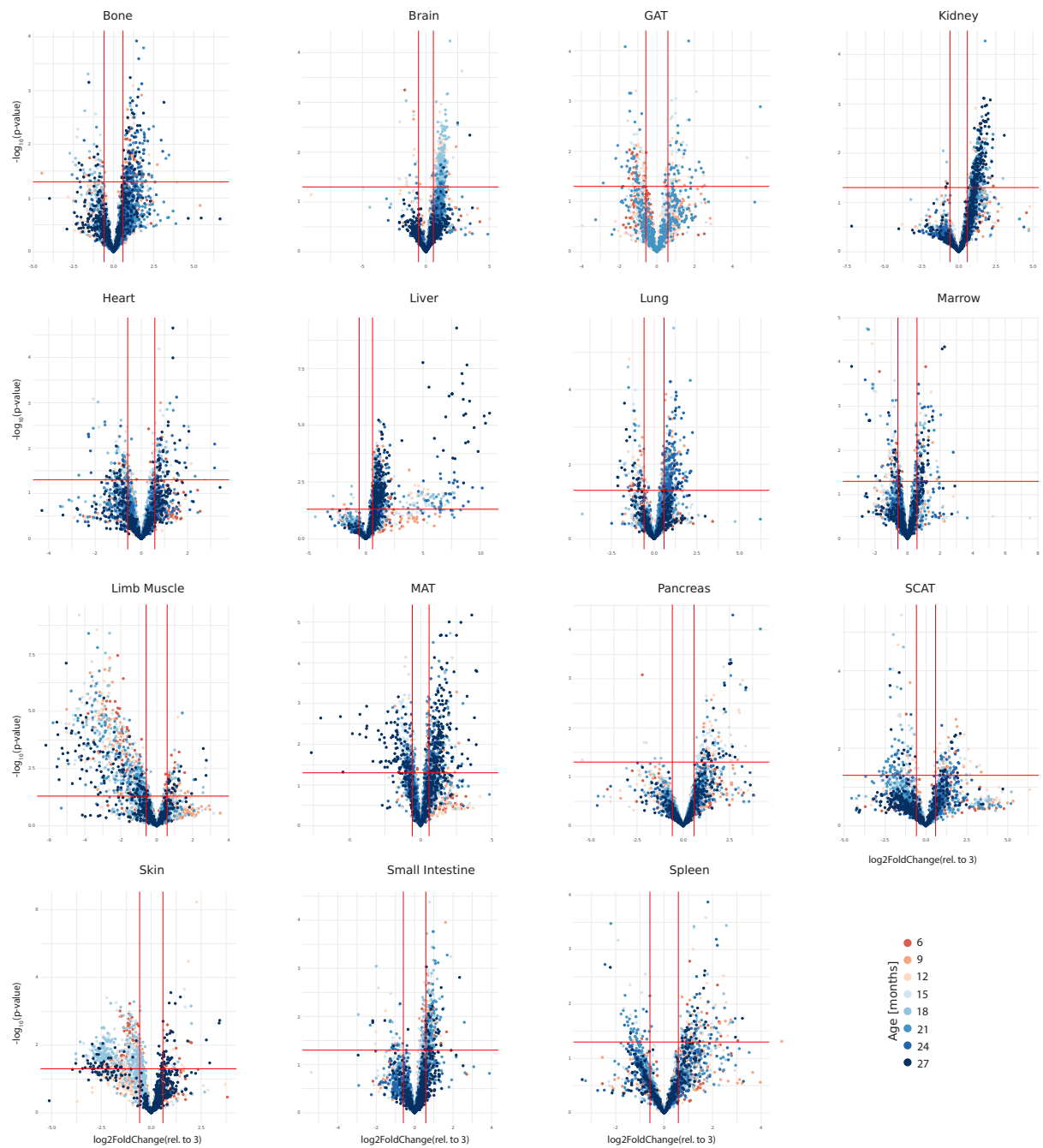
A



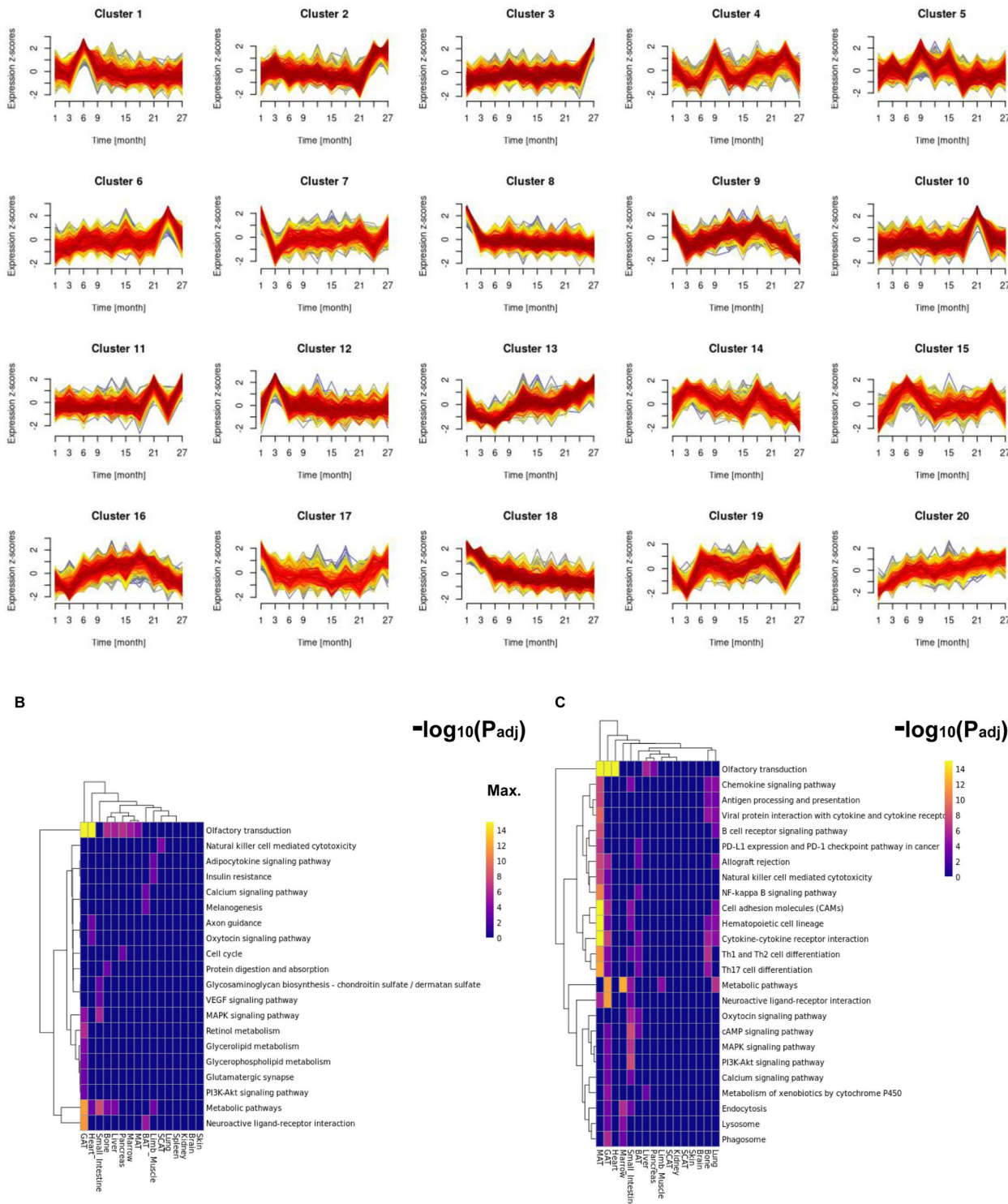
B



Extended Data Fig. 4 | Mean expression vs. log2 FC in all tissues for all ncRNAs. Log2 foldchanges (each timepoint vs 3 months) compared to mean expression values in reads per million mapped of all non-coding RNA (A) and only miRNA (B), red lines as reference for 2/3 and 3/2, colored by timepoint (as indicated in Fig. 1a).

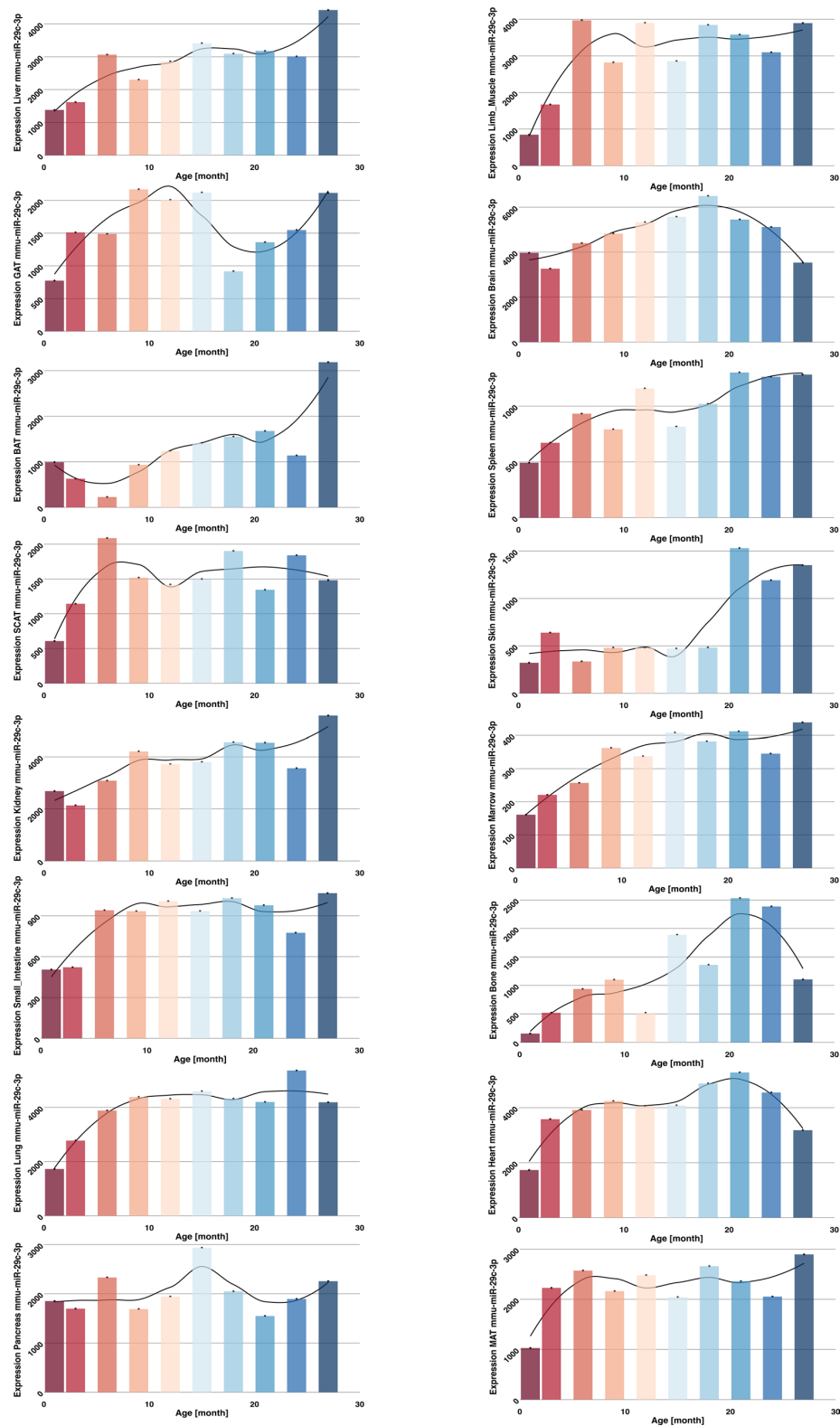


Extended Data Fig. 5 | Volcano plots for comparison 3 m to all later timepoints per organ. Volcano plots for each tissue, colored by timepoints (color code as indicated in Fig. 1a). Log2(FC) are plotted versus $-\log_{10}(\text{p-values})$ calculated in a two-sided t-test.

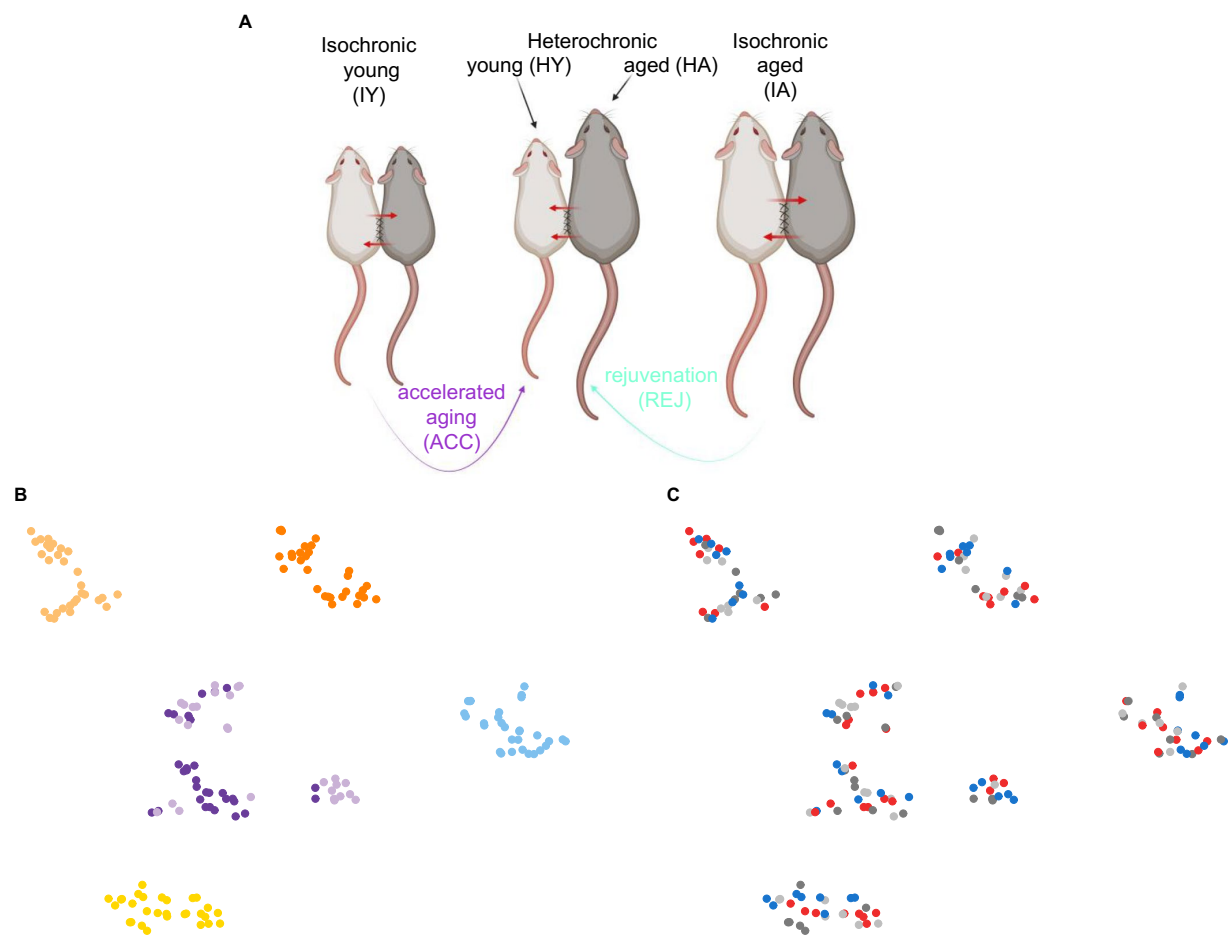


Extended Data Fig. 6 | Whole organism miRNA aging trajectory clustering and enriched pathways of mRNAs targeted by local aging miRNAs. (A) Fuzzy c-means organism-wide z-scored trajectory clustering for all miRNAs in all tissues into 20 clusters. (B) Top 20 non-disease-related locally enriched significant pathways overlapping between tissues and mRNA targets identified via negative correlation with the aging miRNA set ($r > 0.5$, with age) in each individual tissue.

(C) Top 25 non-disease-related locally enriched significant pathways overlapping between tissues. mRNA targets identified via negative correlation with miRNA correlated negatively with aging ($r < -0.5$) in each individual tissue (identified via overrepresentation analysis, hypergeometric test, two-sided, adjustment FDR, t-test < 0.001).

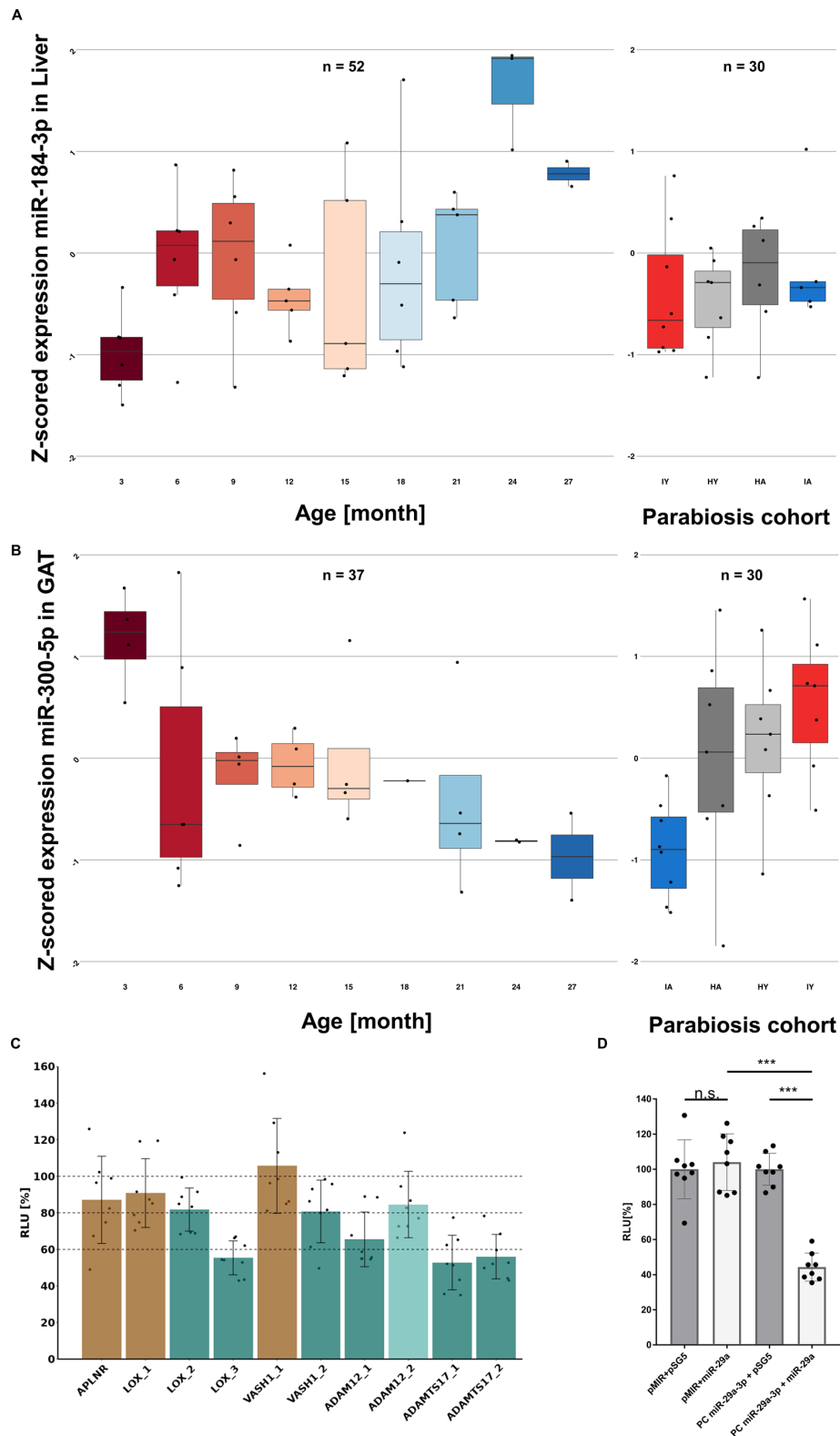


Extended Data Fig. 7 | miR-29c-3p expression per tissue. Expression of global aging miRNA miR-29c-3p as reads per mapped million for all tissues at all measured timepoints.



Extended Data Fig. 8 | Experimental design for the parabiosis study and variance analysis of samples. (A) Schematic plot of the experimental aging intervention and heterochronic parabiosis. Visualization of all samples of the

parabiosis cohort over all detected non-coding features in a t-SNE, colored by tissue (B, color code as indicated in Fig. 1a) and treatment group (C, color code as indicated in Fig. 4b).



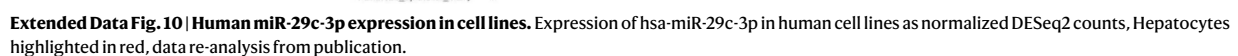
Extended Data Fig. 9 | See next page for caption.

Article

<https://doi.org/10.1038/s41587-023-01751-6>

Extended Data Fig. 9 | Observed rejuvenation effects of global aging miRNAs and validation experiments for predicted targets of miR-29c-3p. z-scored expression in the healthy aging cohort (TMS) for every timepoint and z-scored expression in the parabiosis cohort for the four different groups (IY, HY, HA and IA) for miR-184-3p in the liver (A) and miR-300-5p in GAT (B). The first to the third quartile are covered within the boxes with the median value shown as line inside the box. Maximum and minimum values are shown as whiskers or values up to 1.5-times the interquartile range above and below the first or third quartile if outliers are present. (C) Measured RLU (relative light unit) in percent for all target constructs for miR-29; reduction to under 80% determines a high confidence target, reduction between 80-90% low confidence target (n = 4 biologically independent experiments for each target, each as technical duplicates). Color coded in dark green for significant reduction with a p-value ≤ 0.01 (ADAMTS17_1:

0,0000004, ADAMTS17_2: 0,00000007, ADAM12_1: 0,00001, LOX_2: 0,0006, LOX_3: 0,000000001, VASH1_2: 0,0067), light green for significant reduction with a p-value ≤ 0.05 (AD-AM12_2: 0,0306) and in light brown for non-significant reduction with a p-value ≥ 0.05 (n.s.; APLNR: 0,1485, LOX_1: 0,1892, VASH1_1: 0,5457). P-values were calculated by an unpaired t-test. Data are shown as mean $\pm$ SD. (D) Control experiment results luciferase assay, negative control: vector alone and vector with miRNA and positive control with vector (pMIR+ pSG5; pMIR+miR-29a, PC miR-29a-3p + pSG5) and positive control (PC miR-29a-3p + miR-29a) (n = 4 biologically independent experiments, each as technical duplicates; *** = p-value ≤ 0.001 , n.s. = p-value ≥ 0.05). P-values were calculated by a Welch's t-test (pMIR+pSG5 vs. pMIR+miR-29a = 0,6314; PC miR-29a-3p+miR-29a vs. PC miR-29a-3p + pSG5 p-value $< 0,0001$; PC miR-29a-3p+miR-29a vs. pMIR+miR-29a p-value $< 0,0001$.) Data are shown as mean $\pm$ SEM.



DISCUSSION

This work presents the development and application of a flexible computational framework tailored to the analysis of microRNA (miRNA) datasets. The versatility of this framework is evident from its use in a range of applied studies. It successfully processed large, high-dimensional bulk datasets, revealing condition-specific miRNA expression patterns related to variables such as disease status, treatment, tissue type, age, and sex. Furthermore, the framework was adapted to accommodate single-cell miRNA datasets and therefore expand its applicability to cutting-edge research domains.

The field of tissue-specific bulk studies is particularly interesting, as it allows for efficient investigation of heterogeneity within tissues. Recent research, including this study, has also highlighted the importance of investigating sex differences in miRNA expression, particularly in aging-related contexts [159–162]. While bulk datasets provide valuable insights, they inherently mask the complex, cell-specific expression patterns that single-cell approaches can uncover.

Single-cell resolution studies represent a promising direction for the future of miRNA research [167, 349–353]. Recent advancements have yielded the first insights into miRNA dynamics at the cellular level. Currently, the scalability and high costs of single-cell high-throughput miRNA sequencing protocols present significant challenges for conducting large-scale, cell-type-resolved studies. Hence, a substantial effort is required to establish accurate, efficient and reproducible workflows in the laboratory [4, 182, 191, 195]. Despite these challenges, it is likely only a matter of time before these methods are refined and commercialized, much like the trajectory observed for bulk and single-cell messenger-RNA (mRNA) sequencing technologies, where costs have significantly decreased over time [178, 354–357].

Nonetheless, challenges persist, particularly for single-cell miRNA studies. The inherently low abundance of miRNAs per cell poses a substantial issue and requires a balance between maximizing sequencing depth to detect weak signals and minimizing technical noise [358, 359]. Additionally, single-cell mRNA investigations have been shown to be affected by batch effects [257, 360–363]. To detect these effects the Principal Variance Component Analysis (PVCA) is included in our framework which also indicates the proportion of variances for experimental parameters [253, 271]. More generally, different library preparation protocols or profiling techniques introduce a bias in the obtained expression profile [364–366].

The expected popularity of fine-grained bulk up to high-throughput single-cell miRNA studies or even the use of novel sequencing techniques [367, 368] leads to large, multi-faceted datasets. Once a dataset is processed to an expression table, it can be handled by our framework, regardless whether one dimension of the table corresponds to real samples, cells or even localized spots. Any gained granularity only leads to additional attributes which can be captured in a metadata table, for example the cell to sample matching in single-cell resolved datasets. Scaling our computational framework to larger datasets compared to the ones encountered within this thesis is straightforward due to the use of workflow management systems, which parallelize independent workloads. Furthermore, exchanging implemented methods with novel, highly-efficient ones leads to a direct performance improvement [369]. In general, adding new applications or profiling specific methods is encouraged by the modular architecture of the framework and the workflow management system.

OUTLOOK

Our web-based tool *SingmiR* [5] provides quality control and initial downstream analyses. However, future efforts should focus on streamlining multiple analysis tools regarding standardized interfaces and developing modular pipelines capable of processing diverse types of expression data. Further, interfacing with existing internal and external tools via a seamless data transfer facilitates user accessibil-

ity and integration into broader research workflows. For example, linking the tools *miRMaster* to the enrichment analysis platforms *miEAA* [328] enabled more comprehensive analyses [184]. Following this approach, the results of *SingmiR* could be linked to multiple web services, among those *miEAA*, *miRPathDB*, *miRTargetLink*, *miRNATissueAtlas* [78, 319, 328, 329, 348, 370, 371]. Additionally, integrating our computational tools into universal workflow management platforms, such as Galaxy [304], would significantly enhance their usability, accessibility and visibility.

To accommodate for emerging high-throughput single-cell miRNA technologies, extending the alignment abilities of *SingmiR* to multiplexed single-cell datasets represents a promising direction for further development [372].

Tissue-resolved miRNA studies currently offer an efficient means to address cellular heterogeneity. In this work, the sex and age-resolved tissue-specific datasets [7] and (Planned I) exemplify the potential of this approach. Notably, for these miRNA datasets sample-matched mRNA datasets are present [150, 151]. Therefore, tighter integration of the analyses between the two ribonucleic acid (RNA) types and their expression profiles promises a deeper understanding of regulatory networks. Building upon this approach, future studies could expand this scheme into multi-omics research by incorporating, for example, miRNA, mRNA and proteomic datasets. These extensions would provide a more general view of biological systems and are particularly relevant for clinical applications, such as exploring the molecular mechanisms underlying neurodegenerative diseases [373–377].

CONCLUSION

In summary, the approaches presented in this dissertation constitute a powerful framework for miRNA analysis. Through the integration of single-cell sequencing, computational pipeline development, and tissue-resolved analyses, this work paves the way to uncover intricate miRNA dynamics in any application scenario, from time-series to case-control setups.

Personally, on the one hand, approaching the research field coming from a mathematics background, a theoretical discipline, enabled treating the research questions with minor biologically driven expectations. While this approach occasionally introduced challenges due to the limited reliance on biological assumptions, the adoption of a data-driven approach allowed for an unbiased exploration of the datasets. On the other hand, this work encompassed adapting established tools to the specific requirements of miRNA studies, demonstrating the utility of tailoring existing methodologies to novel applications. Bridging the insights gained from mathematical foundations and molecular biological application studies, harbors strong potential for the development of novel, integrated methodologies in future research.

BIBLIOGRAPHY

- [1] Fabian Kern, Thomas Kuhn, Nicole Ludwig, Martin Simon, Laura Gröger, et al. "Ageing-associated small RNA cargo of extracellular vesicles." In: *RNA biology* 20 (2023), pp. 482–494. URL: <https://doi.org/10.1080/15476286.2023.2234713>.
- [2] Annika Engel, Nicole Ludwig, Friederike Grandke, Viktoria Wagner, Fabian Kern, et al. "Skin treatment with non-thermal plasma modulates the immune system through miR-223-3p and its target genes." In: *RNA Biology* 21.1 (2024), pp. 31–44. URL: <https://doi.org/10.1080/15476286.2024.2361571>.
- [3] Caroline Diener, Konstantin Thüre, Annika Engel, Martin Hart, Andreas Keller, et al. "Paving the way to a neural fate - RNA signatures in naive and trans-differentiating Mesenchymal Stem Cells." In: *European Journal of Cell Biology* 103 (2024), p. 151458. URL: <https://doi.org/10.1016/j.ejcb.2024.151458>.
- [4] Stefan Scheuermann, Sarah Hücker, Annika Engel, Nicole Ludwig, Philipp Lebhardt, et al. "A novel approach to generate enzyme-free single cell suspensions from archived tissues for miRNA sequencing." In: *SLAS Technology* 29.3 (2024), p. 100133. URL: <https://doi.org/10.1016/j.slas.2024.100133>.
- [5] Annika Engel, Shusruto Rishik, Pascal Hirsch, Verena Keller, Tobias Fehlmann, et al. "SingmiR: a single-cell miRNA alignment and analysis tool." In: *Nucleic Acids Research* 52.W1 (2024), W374–W380. URL: <https://doi.org/10.1093/nar/gkae225>.
- [6] Martin Hart, Fabian Kern, Claudia Fecher-Trost, Lena Krammes, Ernesto Aparicio, et al. "Experimental capture of miRNA targetomes: disease-specific 3'UTR library-based miRNA targetomics for Parkinson's disease."

- In: *Experimental & molecular medicine* 56 (2024). URL: <https://doi.org/10.1038/s12276-024-01202-5>.
- [7] Viktoria Wagner, Fabian Kern, Oliver Hahn, Nick Schaum, Nicole Ludwig, et al. "Characterizing expression changes in noncoding RNAs during aging and heterochronic parabiosis across mouse tissues." In: *Nature Biotechnology* (2023), pp. 1–10. URL: <https://doi.org/10.1038/s41587-023-01751-6>.
- [8] Olaf Schmidt. *Genetik und Molekularbiologie*. Vol. 1. Springer, 2017. ISBN: 978-3-662-50274-7. URL: <https://doi.org/10.1007/978-3-662-50274-7>.
- [9] Sarah A. Signor, Sergey V. Nuzhdin. "The evolution of gene expression in cis and trans." In: *Trends in Genetics* 34.7 (2018), pp. 532–544. URL: <https://doi.org/10.1016/j.tig.2018.03.007>.
- [10] Peter W. Harrison, Alison E. Wright, Judith E. Mank. "The evolution of gene expression and the transcriptome–phenotype relationship." In: *Seminars in cell & developmental biology*. Vol. 23. 2. Elsevier. 2012, pp. 222–229. URL: <https://doi.org/10.1016/j.semcdb.2011.12.004>.
- [11] Takashi Ikeda, Takafusa Hikichi, Hisashi Miura, Hirofumi Shibata, Kanae Mitsunaga, et al. "Srf destabilizes cellular identity by suppressing cell-type-specific gene expression programs." In: *Nature communications* 9.1 (2018), p. 1387. URL: <https://doi.org/10.1038/s41467-018-03748-1>.
- [12] Wenjia Shi, Wenzhe Ma, Liyang Xiong, Mingyue Zhang, Chao Tang. "Adaptation with transcriptional regulation." In: *Scientific reports* 7.1 (2017), p. 42648. URL: <https://doi.org/10.1038/srep42648>.
- [13] Roadmap Epigenomics Consortium, Anshul Kundaje, Wouter Meuleman, Jason Ernst, Misha Bilenky, et al. "Integrative analysis of 111 reference human epigenomes." In: *Nature* 518.7539 (2015), pp. 317–330. URL: <https://doi.org/10.1038/nature14248>.
- [14] Rudolf Jaenisch, Adrian Bird. "Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals." In: *Nature genetics* 33.3 (2003), pp. 245–254. URL: <https://doi.org/10.1038/ng1089>.

- [15] Lindsay M. Reynolds, Jackson R. Taylor, Jingzhong Ding, Kurt Lohman, Craig Johnson, et al. "Age-related variations in the methylome associated with gene expression in human monocytes and T cells." In: *Nature communications* 5.1 (2014), p. 5366. URL: <https://doi.org/10.1038/ncomms6366>.
- [16] Andrew J. Bannister, Tony Kouzarides. "Regulation of chromatin by histone modifications." In: *Cell research* 21.3 (2011), pp. 381–395. URL: <https://doi.org/10.1038/cr.2011.22>.
- [17] Rebecca F. Lowdon, Bo Zhang, Misha Bilenky, Thea Mauro, Daofeng Li, et al. "Regulatory network decoded from epigenomes of surface ectoderm-derived cell types." In: *Nature communications* 5.1 (2014), p. 5442. URL: <https://doi.org/10.1038/ncomms6442>.
- [18] Elizabeta Gjoneska, Andreas R. Pfenning, Hansruedi Mathys, Gerald Quon, Anshul Kundaje, et al. "Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease." In: *Nature* 518.7539 (2015), pp. 365–369. URL: <https://doi.org/10.1038/nature14252>.
- [19] Donata Vercelli. "Genetics, epigenetics, and the environment: switching, buffering, releasing." In: *Journal of Allergy and Clinical Immunology* 113.3 (2004), pp. 381–386. URL: <https://doi.org/10.1016/j.jaci.2004.01.752>.
- [20] Danny Leung, Inkyung Jung, Nisha Rajagopal, Anthony Schmitt, Siddarth Selvaraj, et al. "Integrative analysis of haplotype-resolved epigenomes across human tissues." In: *Nature* 518.7539 (2015), pp. 350–354. URL: <https://doi.org/10.1038/nature14217>.
- [21] Suk-Youl Park, Jeong-Sun Kim. "A short guide to histone deacetylases including recent progress on class II enzymes." In: *Experimental & molecular medicine* 52.2 (2020), pp. 204–212. URL: <https://doi.org/10.1038/s12276-020-0382-4>.
- [22] S. Eden, T. Hashimshony, I. Keshet, H. Cedar, A. W. Thorne. "DNA methylation models histone acetylation." In: *Nature* 394.6696 (1998), pp. 842–842. URL: <https://doi.org/10.1038/29680>.

- [23] Johannes Bohacek, Isabelle M. Mansuy. "Epigenetic inheritance of disease and disease risk." In: *Neuropsychopharmacology* 38.1 (2013), pp. 220–236. URL: <https://doi.org/10.1038/npp.2012.110>.
- [24] Andrew J. Fritz, Mohammed El Dika, Rabail H. Toor, Princess D. Rodriguez, Stephen J. Foley, et al. "Epigenetic-Mediated Regulation of Gene Expression for Biological Control and Cancer: Cell and Tissue Structure, Function, and Phenotype." In: *Nuclear, Chromosomal, and Genomic Architecture in Biology and Medicine* (2022), pp. 339–373. URL: https://doi.org/10.1007/978-3-031-06573-6_12.
- [25] Jing Yang, Jin Xu, Wei Wang, Bo Zhang, Xianjun Yu, et al. "Epigenetic regulation in the tumor microenvironment: molecular mechanisms and therapeutic targets." In: *Signal transduction and targeted therapy* 8.1 (2023), p. 210. URL: <https://doi.org/10.1038/s41392-023-01480-x>.
- [26] Priyanka Dey Talukdar, Urmi Chatterji. "Transcriptional co-activators: emerging roles in signaling pathways and potential therapeutic targets for diseases." In: *Signal Transduction and Targeted Therapy* 8.1 (2023), p. 427. URL: <https://doi.org/10.1038/s41392-023-01651-w>.
- [27] Yuan Cheng, Cai He, Manni Wang, Xuelei Ma, Fei Mo, et al. "Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials." In: *Signal transduction and targeted therapy* 4.1 (2019), p. 62. URL: <https://doi.org/10.1038/s41392-019-0095-0>.
- [28] Seung Hyuk Choi, Seongjae Kim, Katherine A. Jones. "Gene expression regulation by CDK12: a versatile kinase in cancer with functions beyond CTD phosphorylation." In: *Experimental & molecular medicine* 52.5 (2020), pp. 762–771. URL: <https://doi.org/10.1038/s12276-020-0442-9>.
- [29] Yang Dong, Rongfu Tu, Hudan Liu, Guoliang Qing. "Regulation of cancer cell metabolism: oncogenic MYC in the driver's seat." In: *Signal transduction and targeted therapy* 5.1 (2020), p. 124. URL: <https://doi.org/10.1038/s41392-020-00235-2>.

- [30] Ping Song, Fan Yang, Hongchuan Jin, Xian Wang. "The regulation of protein translation and its implications for cancer." In: *Signal transduction and targeted therapy* 6.1 (2021), p. 68. URL: <https://doi.org/10.1038/s41392-020-00444-9>.
- [31] Paula Dobosz, Przemysław A. Stempor, Miguel Ramírez Moreno, Natalia A. Bulgakova. "Transcriptional and post-transcriptional regulation of checkpoint genes on the tumour side of the immunological synapse." In: *Heredity* 129.1 (2022), pp. 64–74. URL: <https://doi.org/10.1038/s41437-022-00533-1>.
- [32] Nan Zhang, Xinjia Wang, Yu Li, Yiwei Lu, Chengcheng Sheng, et al. "Mechanisms and therapeutic implications of gene expression regulation by circRNA-protein interactions in cancer." In: *Communications Biology* 8.1 (2025), p. 77. URL: <https://doi.org/10.1038/s42003-024-07383-z>.
- [33] Anthony S. Zannas, Tobias Wiechmann, Nils C. Gassen, Elisabeth B. Binder. "Gene-stress-epigenetic regulation of FKBP5: clinical and translational implications." In: *Neuropsychopharmacology* 41.1 (2016), pp. 261–274. URL: <https://doi.org/10.1038/npp.2015.235>.
- [34] Rama P. Bhatia, Hande A. Kirit, Alexander V. Predeus, Jonathan P. Bollback. "Transcriptomic profiling of Escherichia coli K-12 in response to a compendium of stressors." In: *Scientific Reports* 12.1 (2022), p. 8788. URL: <https://doi.org/10.1038/s41598-022-12463-3>.
- [35] Samuel J. Virolainen, Andrew VonHandorf, Kenyatta C. M. F. Viel, Matthew T. Weirauch, Leah C. Kottyan. "Gene-environment interactions and their impact on human health." In: *Genes & Immunity* 24.1 (2023), pp. 1–11. URL: <https://doi.org/10.1038/s41435-022-00192-6>.
- [36] Kemal Keseroglu, Oriana QH Zinani, Sevdnur Keskin, Hannah Seawall, Eslim E. Alpay, et al. "Stochastic gene expression and environmental stressors trigger variable somite segmentation phenotypes." In: *Nature Communications* 14.1 (2023), p. 6497. URL: <https://doi.org/10.1038/s41467-023-42220-7>.

- [37] Cheng Bei, Junhao Zhu, Peter H. Culviner, Mingyu Gan, Eric J. Rubin, et al. "Genetically encoded transcriptional plasticity underlies stress adaptation in *Mycobacterium tuberculosis*." In: *Nature Communications* 15.1 (2024), p. 3088. URL: <https://doi.org/10.1101/2023.08.20.553992>.
- [38] Adam Frankish, Sílvia Carbonell-Sala, Mark Diekhans, Irwin Jungreis, Jane E. Loveland, et al. "GENCODE: reference annotation for the human and mouse genomes in 2023." In: *Nucleic Acids Research* 51.D1 (2022), pp. D942–D949. URL: <https://doi.org/10.1093/nar/gkac1071>.
- [39] Alexander F. Palazzo, Eliza S. Lee. "Non-coding RNA: what is functional and what is junk?" In: *Frontiers in genetics* 6 (2015), p. 2. URL: <https://doi.org/10.3389/fgene.2015.00002>.
- [40] John S. Mattick, Igor V. Makunin. "Non-coding RNA." In: *Human Molecular Genetics* 15.suppl-1 (2006), R17–R29. URL: <https://doi.org/10.1093/hmg/ddlo46>.
- [41] Ye Su, Haijiang Wu, Alexander Pavlosky, Ling-Lin Zou, Xinna Deng, et al. "Regulatory non-coding RNA: new instruments in the orchestration of cell death." In: *Cell death & disease* 7.8 (2016), e2333–e2333. URL: <https://doi.org/10.1038/cddis.2016.210>.
- [42] Luisa Statello, Chun-Jie Guo, Ling-Ling Chen, Maite Huarte. "Gene regulation by long non-coding RNAs and its biological functions." In: *Nature reviews Molecular cell biology* 22.2 (2021), pp. 96–118. URL: <https://doi.org/10.1038/s41580-020-00315-9>.
- [43] John S. Mattick, Paulo P. Amaral, Piero Carninci, Susan Carpenter, Howard Y. Chang, et al. "Long non-coding RNAs: definitions, functions, challenges and recommendations." In: *Nature reviews Molecular cell biology* 24.6 (2023), pp. 430–447. URL: <https://doi.org/10.1038/s41580-022-00566-8>.
- [44] Sumin Yang, Sung-Hyun Kim, Eunjeong Yang, Mingon Kang, Jae-Yeol Joo. "Molecular insights into regulatory RNAs in the cellular machinery." In: *Experimental & Molecular Medicine* (2024), pp. 1–15. URL: <https://doi.org/10.1038/s12276-024-01239-6>.

- [45] Laura Poliseno, Martina Lanza, Pier Paolo Pandolfi. "Coding, or non-coding, that is the question." In: *Cell Research* 34.9 (2024), pp. 609–629. URL: <https://doi.org/10.1038/s41422-024-00975-8>.
- [46] Je-Hyun Yoon, Kotb Abdelmohsen, Myriam Gorospe. "Posttranscriptional gene regulation by long noncoding RNA." In: *Journal of molecular biology* 425.19 (2013), pp. 3723–3730. URL: <https://doi.org/10.1016/j.jmb.2012.11.024>.
- [47] Hui Ling, Muller Fabbri, George A. Calin. "MicroRNAs and other non-coding RNAs as targets for anticancer drug development." In: *Nature reviews Drug discovery* 12.11 (2013), pp. 847–865. URL: <https://doi.org/10.1038/nrd4140>.
- [48] Mahlon B. Hoagland, Mary Louise Stephenson, Jesse F. Scott, Liselotte I. Hecht, Paul C. Zamecnik. "A soluble ribonucleic acid intermediate in protein synthesis." In: *J. biol. Chem* 231.1 (1958), pp. 241–257. URL: [https://doi.org/10.1016/S0021-9258\(19\)77302-5](https://doi.org/10.1016/S0021-9258(19)77302-5).
- [49] Suki Albers, Elizabeth C. Allen, Nikhil Bharti, Marcos Davyt, Disha Joshi, et al. "Engineered tRNAs suppress nonsense mutations in cells and in vivo." In: *Nature* 618.7966 (2023), pp. 842–848. URL: <https://doi.org/10.1038/s41586-023-06133-1>.
- [50] Harry F. Noller, Rachel Green, Gabriele Heilek, Vernita Hoffarth, Alexander Hüttenhofer, et al. "Structure and function of ribosomal RNA." In: *Biochemistry and cell biology* 73.11-12 (1995), pp. 997–1009. URL: <https://doi.org/10.1139/095-107>.
- [51] Shaheen Sulthana, Georgeta N. Basturea, Murray P. Deutscher. "Elucidation of pathways of ribosomal RNA degradation: an essential role for RNase E." In: *Rna* 22.8 (2016), pp. 1163–1171. URL: <https://doi.org/10.1261/rna.056275.116>.
- [52] Jian Kang, Natalie Brajanovski, Keefe T. Chan, Jiachen Xuan, Richard B. Pearson, et al. "Ribosomal proteins and human diseases: molecular mechanisms and targeted therapy." In: *Signal transduction and targeted*

therapy 6.1 (2021), p. 323. URL: <https://doi.org/10.1038/s41392-021-00728-8>.

- [53] Luca F. R. Gebert, Ian J. MacRae. "Regulation of microRNA function in animals." In: *Nature reviews Molecular cell biology* 20.1 (2019), pp. 21–37. URL: <https://doi.org/10.1038/s41580-018-0045-7>.
- [54] Antti P. Aalto, Amy E. Pasquinelli. "Small non-coding RNAs mount a silent revolution in gene expression." In: *Current opinion in cell biology* 24.3 (2012), pp. 333–340. URL: <https://doi.org/10.1016/j.ceb.2012.03.006>.
- [55] Irfan A. Qureshi, Mark F. Mehler. "Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease." In: *Nature Reviews Neuroscience* 13.8 (2012), pp. 528–541. URL: <https://doi.org/10.1038/nrn3234>.
- [56] Zhenyao Chen, Xin Chen, Ping Chen, Shanxun Yu, Fengqi Nie, et al. "Long non-coding RNA SNHG20 promotes non-small cell lung cancer cell proliferation and migration by epigenetically silencing of P21 expression." In: *Cell death & disease* 8.10 (2017), e3092–e3092. URL: <https://doi.org/10.1038/cddis.2017.484>.
- [57] Chen Li, Yu-Qing Ni, Hui Xu, Qun-Yan Xiang, Yan Zhao, et al. "Roles and mechanisms of exosomal non-coding RNAs in human health and diseases." In: *Signal transduction and targeted therapy* 6.1 (2021), p. 383. URL: <https://doi.org/10.1038/s41392-021-00779-x>.
- [58] Yuliang Zhang, Hongliang Liu, Min Niu, Ying Wang, Rong Xu, et al. "Roles of long noncoding RNAs in human inflammatory diseases." In: *Cell Death Discovery* 10.1 (2024), p. 235. URL: <https://doi.org/10.1038/s41420-024-02002-6>.
- [59] Eleni Anastasiadou, Leni S. Jacob, Frank J. Slack. "Non-coding RNA networks in cancer." In: *Nature Reviews Cancer* 18.1 (2018), pp. 5–18. URL: <https://doi.org/10.1038/nrc.2017.99>.
- [60] Zhao Zhang, Jian Zhang, Lixia Diao, Leng Han. "Small non-coding RNAs in human cancer: function, clinical utility, and characterization." In: *Oncology*

- gene* 40.9 (2021), pp. 1570–1577. URL: <https://doi.org/10.1038/s41388-020-01630-3>.
- [61] S. John Liu, Ha X. Dang, Daniel A. Lim, Felix Y. Feng, Christopher A. Maher. “Long noncoding RNAs in cancer metastasis.” In: *Nature Reviews Cancer* 21.7 (2021), pp. 446–460. URL: <https://doi.org/10.1038/s41568-021-00353-1>.
- [62] Michela Coan, Simon Haefliger, Samir Ounzain, Rory Johnson. “Targeting and engineering long non-coding RNAs for cancer therapy.” In: *Nature Reviews Genetics* (2024), pp. 1–18. URL: <https://doi.org/10.1038/s41576-024-00693-2>.
- [63] BaoQing Chen, Mihnea P. Dragomir, Chen Yang, Qiaoqiao Li, David Horst, et al. “Targeting non-coding RNAs to overcome cancer therapy resistance.” In: *Signal transduction and targeted therapy* 7.1 (2022), p. 121. URL: <https://doi.org/10.1038/s41392-022-00975-3>.
- [64] Yi-Ying Wu, Hung-Chih Kuo. “Functional roles and networks of non-coding RNAs in the pathogenesis of neurodegenerative diseases.” In: *Journal of Biomedical Science* 27.1 (2020), p. 49. URL: <https://doi.org/10.1186/s12929-020-00636-z>.
- [65] Evgenia Salta, Bart De Strooper. “Noncoding RNAs in neurodegeneration.” In: *Nature Reviews Neuroscience* 18.10 (2017), pp. 627–640. URL: <https://doi.org/10.1038/nrn.2017.90>.
- [66] Nour Shaheen, Ahmed Shaheen, Mahmoud Osama, Abdulqadir J. Nashwan, Vishal Bharmauria, et al. “MicroRNAs regulation in Parkinson’s disease, and their potential role as diagnostic and therapeutic targets.” In: *npj Parkinson’s Disease* 10.1 (2024), p. 186. URL: <https://doi.org/10.1038/s41531-024-00791-2>.
- [67] Marharyta Smal, Domenico Memoli, Elena Alexandrova, Domenico Di Rosa, Ylenia D’Agostino, et al. “Small non-coding RNA transcriptomic profiling in adult and fetal human brain.” In: *Scientific Data* 11.1 (2024), p. 767. URL: <https://doi.org/10.1038/s41597-024-03604-6>.

- [68] Rosalind C. Lee, Rhonda L. Feinbaum, Victor Ambros. "The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*." In: *cell* 75.5 (1993), pp. 843–854. URL: [https://doi.org/10.1016/0092-8674\(93\)90529-y](https://doi.org/10.1016/0092-8674(93)90529-y).
- [69] Ana Kozomara, Maria Birgaoanu, Sam Griffiths-Jones. "miRBase: from microRNA sequences to function." In: *Nucleic acids research* 47.D1 (2019), pp. D155–D162. URL: <https://doi.org/10.1093/nar/gky1141>.
- [70] Xuezhong Cai, Curt H. Hagedorn, Bryan R. Cullen. "Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs." In: *Rna* 10.12 (2004), pp. 1957–1966. URL: <https://doi.org/10.1261/rna.7135204>.
- [71] Yoontae Lee, Kipyoun Jeon, Jun-Tae Lee, Sunyoung Kim, V. Narry Kim. "MicroRNA maturation: stepwise processing and subcellular localization." In: *The EMBO Journal* 21.17 (2002), pp. 4663–4670. URL: <https://doi.org/10.1093/emboj/cdf476>.
- [72] Yoontae Lee, Chiyoun Ahn, Jinju Han, Hyounjeong Choi, Jaekwang Kim, et al. "The nuclear RNase III Drosha initiates microRNA processing." In: *Nature* 425.6956 (2003), pp. 415–419. URL: <https://doi.org/10.1038/nature01957>.
- [73] Markus T. Bohnsack, Kevin Czaplinski, Dirk Görlich. "Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs." In: *Rna* 10.2 (2004), pp. 185–191. URL: <https://doi.org/10.1261/rna.5167604>.
- [74] Elsebet Lund, Stephan Guttinger, Angelo Calado, James E. Dahlberg, Ulrike Kutay. "Nuclear export of microRNA precursors." In: *science* 303.5654 (2004), pp. 95–98. URL: <https://doi.org/10.1126/science.1090599>.
- [75] David P. Bartel. "Metazoan micrornas." In: *Cell* 173.1 (2018), pp. 20–51. URL: <https://doi.org/10.1016/j.cell.2018.03.006>.
- [76] Erno Wienholds, Marco J. Koudijs, Freek J. M. van Eeden, Edwin Cuppen, Ronald H. A. Plasterk. "The microRNA-producing enzyme Dicer1 is essen-

- tial for zebrafish development." In: *Nature genetics* 35.3 (2003), pp. 217–218. URL: <https://doi.org/10.1038/ng1251>.
- [77] Tomoko Kawamata, Yukihide Tomari. "Making risc." In: *Trends in biochemical sciences* 35.7 (2010), pp. 368–376. URL: <https://doi.org/10.1016/j.tibs.2010.03.009>.
- [78] Fabian Kern, Jeremy Amand, Ilya Senatorov, Alina Isakova, Christina Backes, et al. "miRSwitch: detecting microRNA arm shift and switch events." In: *Nucleic Acids Research* 48.W1 (2020), W268–W274. URL: <https://doi.org/10.1093/nar/gkaa323>.
- [79] Mengfan Pu, Jing Chen, Zhouteng Tao, Lingling Miao, Xinming Qi, et al. "Regulatory network of miRNA on its target: coordination between transcriptional and post-transcriptional regulation of gene expression." In: *Cellular and Molecular Life Sciences* 76 (2019), pp. 441–451. URL: <https://doi.org/10.1007/s00018-018-2940-7>.
- [80] Muriel Girard, Emmanuel Jacquemin, Arnold Munnich, Stanislas Lyonnet, Alexandra Henrion-Caude. "miR-122, a paradigm for the role of microRNAs in the liver." In: *Journal of hepatology* 48.4 (2008), pp. 648–656. URL: <https://doi.org/10.1016/j.jhep.2008.01.019>.
- [81] Benjamin P. Lewis, Christopher B. Burge, David P. Bartel. "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets." In: *cell* 120.1 (2005), pp. 15–20. URL: <https://doi.org/10.1016/j.cell.2004.12.035>.
- [82] Bart Krist, Urszula Florczyk, Katarzyna Pietraszek-Gremplewicz, Alicja Józkowicz, Jozef Dulak. "The role of miR-378a in metabolism, angiogenesis, and muscle biology." In: *International journal of endocrinology* 2015.1 (2015), p. 281756. URL: <https://doi.org/10.1155/2015/281756>.
- [83] Jennifer A. Broderick, William E. Salomon, Sean P. Ryder, Neil Aronin, Phillip D. Zamore. "Argonaute protein identity and pairing geometry determine cooperativity in mammalian RNA silencing." In: *Rna* 17.10 (2011), pp. 1858–1869. URL: <https://doi.org/10.1261/rna.2778911>.

- [84] Soken Tsuchiya, Yasushi Okuno, Gozoh Tsujimoto. "MicroRNA: biogenetic and functional mechanisms and involvements in cell differentiation and cancer." In: *Journal of pharmacological sciences* 101.4 (2006), pp. 267–270. URL: <https://doi.org/10.1254/jphs.CPJ06013X>.
- [85] Daniel W. Thomson, Cameron P. Bracken, Gregory J. Goodall. "Experimental strategies for microRNA target identification." In: *Nucleic acids research* 39.16 (2011), pp. 6845–6853. URL: <https://doi.org/10.1093/nar/gkr330>.
- [86] Xin Lai, Olaf Wolkenhauer, Julio Vera. "Understanding microRNA-mediated gene regulatory networks through mathematical modelling." In: *Nucleic acids research* 44.13 (2016), pp. 6019–6035. URL: <https://doi.org/10.1093/nar/gkw550>.
- [87] Peter Arner, Agn  Kulyt . "MicroRNA regulatory networks in human adipose tissue and obesity." In: *Nature Reviews Endocrinology* 11.5 (2015), pp. 276–288. URL: <https://doi.org/10.1038/nrendo.2015.25>.
- [88] Ning Liu, Eric N. Olson. "MicroRNA regulatory networks in cardiovascular development." In: *Developmental cell* 18.4 (2010), pp. 510–525. URL: <https://doi.org/10.1016/j.devcel.2010.03.010>.
- [89] Masaki Takasugi. "Emerging roles of extracellular vesicles in cellular senescence and aging." In: *Aging cell* 17.2 (2018), e12734. URL: <https://doi.org/10.1111/accel.12734>.
- [90] Yujia Yin, Huihui Chen, Yizhi Wang, Ludi Zhang, Xipeng Wang. "Roles of extracellular vesicles in the aging microenvironment and age-related diseases." In: *Journal of extracellular vesicles* 10.12 (2021), e12154. URL: <https://doi.org/10.1002/jev2.12154>.
- [91] Ionara Rodrigues Siqueira, Andressa De Souza Rodrigues, Marina Siqueira Flores, Eduarda Let cia Vieira Cunha, Madeleine Goldberg, et al. "Circulating Extracellular Vesicles and Particles Derived From Adipocytes: The Potential Role in Spreading MicroRNAs Associated With Cellular Senescence." In: *Frontiers in Aging* 3 (2022), p. 867100. URL: <https://doi.org/10.3389/fragi.2022.867100>.

- [92] Monica Correa Alfonzo, Ahmed Al Saedi, Sadanand Fulzele, Mark W. Hamrick. "Extracellular vesicles as communicators of senescence in musculoskeletal aging." In: *Journal of Bone and Mineral Research Plus* 6.11 (2022), e10686. URL: <https://doi.org/10.1002/jbm4.10686>.
- [93] John R. Chevillet, Qing Kang, Ingrid K. Ruf, Hilary A. Briggs, Lucia N. Vojtech, et al. "Quantitative and stoichiometric analysis of the microRNA content of exosomes." In: *Proceedings of the National Academy of Sciences* 111.41 (2014), pp. 14888–14893. URL: <https://doi.org/10.1073/pnas.1408301111>.
- [94] Heedoo Lee, Michael Groot, Mayra Pinilla-Vera, Laura E. Fredenburgh, Yang Jin. "Identification of miRNA-rich vesicles in bronchoalveolar lavage fluid: Insights into the function and heterogeneity of extracellular vesicles." In: *Journal of Controlled Release* 294 (2019), pp. 43–52. URL: <https://doi.org/10.1016/j.jconrel.2018.12.008>.
- [95] Rekha Nair, Livia Santos, Siddhant Awasthi, Thomas von Erlach, Lesley W. Chow, et al. "Extracellular vesicles derived from preosteoblasts influence embryonic stem cell differentiation." In: *Stem cells and development* 23.14 (2014), pp. 1625–1635. URL: <https://doi.org/10.1089/scd.2013.0633>.
- [96] Luciana De Luca, Stefania Trino, Iaria Laurenzana, Vittorio Simeon, Giovanni Calice, et al. "MiRNAs and piRNAs from bone marrow mesenchymal stem cell extracellular vesicles induce cell survival and inhibit cell differentiation of cord blood hematopoietic stem cells: a new insight in transplantation." In: *Oncotarget* 7.6 (2015), p. 6676. URL: <https://doi.org/10.18632/oncotarget.6791>.
- [97] Peter J. Quesenberry, Jason Aliotta, Maria Chiara Deregibus, Giovanni Camussi. "Role of extracellular RNA-carrying vesicles in cell differentiation and reprogramming." In: *Stem cell research & therapy* 6 (2015), pp. 1–10. URL: <https://doi.org/10.1186/s13287-015-0150-x>.
- [98] Paul D. Robbins, Adrian E. Morelli. "Regulation of immune responses by extracellular vesicles." In: *Nature Reviews Immunology* 14.3 (2014), pp. 195–208. URL: <https://doi.org/10.1038/nri3622>.

- [99] Samarjit Das, Marc K. Halushka. "Extracellular vesicle microRNA transfer in cardiovascular disease." In: *Cardiovascular pathology* 24.4 (2015), pp. 199–206. URL: <https://doi.org/10.1016/j.carpath.2015.04.007>.
- [100] Jie Chen, Chengping Hu, Pinhua Pan. "Extracellular vesicle microRNA transfer in lung diseases." In: *Frontiers in physiology* 8 (2017), p. 1028. URL: <https://doi.org/10.3389/fphys.2017.01028>.
- [101] Edit I. Buzas, Bence György, György Nagy, András Falus, Steffen Gay. "Emerging role of extracellular vesicles in inflammatory diseases." In: *Nature Reviews Rheumatology* 10.6 (2014), pp. 356–364. URL: <https://doi.org/10.1038/nrrheum.2014.19>.
- [102] Joana Cabral, Aideen E. Ryan, Matthew D. Griffin, Thomas Ritter. "Extracellular vesicles as modulators of wound healing." In: *Advanced Drug Delivery Reviews* 129 (2018), pp. 394–406. URL: <https://doi.org/10.1016/j.addr.2018.01.018>.
- [103] Florian Erhard, Jürgen Haas, Diana Lieber, Georg Malterer, Lukasz Jaskiewicz, et al. "Widespread context dependency of microRNA-mediated regulation." In: *Genome research* 24.6 (2014), pp. 906–919. URL: <https://doi.org/10.1101/gr.166702.113>.
- [104] Li-Fan Lu, Georg Gasteiger, I-Shing Yu, Ashutosh Chaudhry, Jing-Ping Hsin, et al. "A single miRNA-mRNA interaction affects the immune response in a context-and cell-type-specific manner." In: *Immunity* 43.1 (2015), pp. 52–64. URL: <https://doi.org/10.1016/j.immuni.2015.04.022>.
- [105] Jing-Ping Hsin, Yuheng Lu, Gabriel B. Loeb, Christina S. Leslie, Alexander Y. Rudensky. "The effect of cellular context on miR-155-mediated gene regulation in four major immune cell types." In: *Nature immunology* 19.10 (2018), pp. 1137–1145. URL: <https://doi.org/10.1038/s41590-018-0208-x>.
- [106] Terry S. Elton, Jack C. Yalowich. "Experimental procedures to identify and validate specific mRNA targets of miRNAs." In: *EXCLI journal* 14 (2015), p. 758. URL: <https://doi.org/10.17179/excli2015-319>.

- [107] Ander Muniategui, Rubén Nogales-Cadenas, Miguél Vázquez, Xabier L. Aranguren, Xabier Agirre, et al. "Quantification of miRNA-mRNA interactions." In: *PloS one* 7.2 (2012), e30766. URL: <https://doi.org/10.1371/journal.pone.0030766>.
- [108] Archana Shenoy, Robert H. Blelloch. "Regulation of microRNA function in somatic stem cell proliferation and differentiation." In: *Nature reviews Molecular cell biology* 15.9 (2014), pp. 565–576. URL: <https://doi.org/10.1038/nrm3854>.
- [109] Ling Guo, Robert C. H. Zhao, Yaojiong Wu. "The role of microRNAs in self-renewal and differentiation of mesenchymal stem cells." In: *Experimental hematology* 39.6 (2011), pp. 608–616. URL: <https://doi.org/10.1016/j.exphem.2011.01.011>.
- [110] Marina C. Costa, Szczepan Kurc, Anna Drożdż, Nuno Cortez-Dias, Francisco J. Enguita. "The circulating non-coding RNA landscape for biomarker research: lessons and prospects from cardiovascular diseases." In: *Acta Pharmacologica Sinica* 39.7 (2018), pp. 1085–1099. URL: <https://doi.org/10.1038/aps.2018.35>.
- [111] Leon Tribolet, Emily Kerr, Christopher Cowled, Andrew G. D. Bean, Cameron R. Stewart, et al. "MicroRNA biomarkers for infectious diseases: from basic research to biosensing." In: *Frontiers in Microbiology* 11 (2020), p. 1197. URL: <https://doi.org/10.3389/fmicb.2020.01197>.
- [112] Fabian Kern, Tobias Fehlmann, Ivo Violich, Eric Alsop, Elizabeth Hutchins, et al. "Deep sequencing of sncRNAs reveals hallmarks and regulatory modules of the transcriptome during Parkinson's disease progression." In: *Nature Aging* 1.3 (2021), pp. 309–322. URL: <https://doi.org/10.1038/s43587-021-00042-6>.
- [113] Nicole Ludwig, Tobias Fehlmann, Fabian Kern, Manfred Gogol, Walter Maetzler, et al. "Machine learning to detect Alzheimer's disease from circulating non-coding RNAs." In: *Genomics, Proteomics and Bioinformatics* 17.4 (2019), pp. 430–440. URL: <https://doi.org/10.1016/j.gpb.2019.09.004>.

- [114] Christina Backes, Eckart Meese, Andreas Keller. "Specific miRNA disease biomarkers in blood, serum and plasma: challenges and prospects." In: *Molecular diagnosis & therapy* 20 (2016), pp. 509–518. URL: <https://doi.org/10.1007/s40291-016-0221-4>.
- [115] Jin Wang, Jinyun Chen, Subrata Sen. "MicroRNA as biomarkers and diagnostics." In: *Journal of cellular physiology* 231.1 (2016), pp. 25–30. URL: <https://doi.org/10.1002/jcp.25056>.
- [116] Weili Huang. "MicroRNAs: biomarkers, diagnostics, and therapeutics." In: *Bioinformatics in MicroRNA research* (2017), pp. 57–67. URL: https://doi.org/10.1007/978-1-4939-7046-9_4.
- [117] Ursula S. Sandau, Jack T. Wiedrick, Trevor J. McFarland, Douglas R. Galasko, Zoe Fanning, et al. "Analysis of the longitudinal stability of human plasma miRNAs and implications for disease biomarkers." In: *Scientific Reports* 14.1 (2024), p. 2148. URL: <https://doi.org/10.1038/s41598-024-52681-5>.
- [118] Benzhe Su, Weiwei Wang, Xiaohui Lin, Shenglan Liu, Xin Huang. "Identifying the potential miRNA biomarkers based on multi-view networks and reinforcement learning for diseases." In: *Briefings in Bioinformatics* 25.1 (2024), bbad427. URL: <https://doi.org/10.1093/bib/bbad427>.
- [119] Hannah Walgrave, Lujia Zhou, Bart De Strooper, Evgenia Salta. "The promise of microRNA-based therapies in Alzheimer's disease: challenges and perspectives." In: *Molecular neurodegeneration* 16 (2021), pp. 1–16. URL: <https://doi.org/10.1186/s13024-021-00496-7>.
- [120] Caroline Diener, Andreas Keller, Eckart Meese. "Emerging concepts of miRNA therapeutics: from cells to clinic." In: *Trends in Genetics* 38.6 (2022), pp. 613–626. URL: <https://doi.org/10.1016/j.tig.2022.02.006>.
- [121] Chiranjib Chakraborty, Ashish Ranjan Sharma, Garima Sharma, Sang-Soo Lee. "Therapeutic advances of miRNAs: A preclinical and clinical update." In: *Journal of advanced research* 28 (2021), pp. 127–138. URL: <https://doi.org/10.1016/j.jare.2020.08.012>.

- [122] Taewan Kim, Carlo M. Croce. "MicroRNA: trends in clinical trials of cancer diagnosis and therapy strategies." In: *Experimental & molecular medicine* 55.7 (2023), pp. 1314–1321. URL: <https://doi.org/10.1038/s12276-023-01050-9>.
- [123] David De Gonzalo-Calvo, Jennifer Pérez-Boza, Joao Curado, Yvan Devaux, EU-CardioRNA COST Action CA17129. "Challenges of microRNA-based biomarkers in clinical application for cardiovascular diseases." In: *Clinical and Translational Medicine* 12.2 (2022). URL: <https://doi.org/10.1002/ctm2.585>.
- [124] Gareth Morris, Denis O'Brien, David C. Henshall. "Opportunities and challenges for microRNA-targeting therapeutics for epilepsy." In: *Trends in Pharmacological Sciences* 42.7 (2021), pp. 605–616. URL: <https://doi.org/10.1016/j.tips.2021.04.007>.
- [125] Lucio Barile, Giuseppe Vassalli. "Exosomes: Therapy delivery tools and biomarkers of diseases." In: *Pharmacology & therapeutics* 174 (2017), pp. 63–78. URL: <https://doi.org/10.1016/j.pharmthera.2017.02.020>.
- [126] Yun Wu, Melissa Crawford, Yicheng Mao, Robert J. Lee, Ian C. Davis, et al. "Therapeutic delivery of microRNA-29b by cationic lipoplexes for lung cancer." In: *Molecular Therapy-Nucleic Acids* 2 (2013). URL: <https://doi.org/10.1038/mtna.2013.14>.
- [127] Maria Angelica Cortez, David Valdecanas, Xiaochun Zhang, Yanai Zhan, Vikas Bhardwaj, et al. "Therapeutic delivery of miR-200c enhances radiosensitivity in lung cancer." In: *Molecular Therapy* 22.8 (2014), pp. 1494–1503. URL: <https://doi.org/10.1038/mt.2014.79>.
- [128] David S. Hong, Yoon-Koo Kang, Mitesh Borad, Jasgit Sachdev, Samuel Ejadi, et al. "Phase 1 study of MRX34, a liposomal miR-34a mimic, in patients with advanced solid tumours." In: *British journal of cancer* 122.11 (2020), pp. 1630–1637. URL: <https://doi.org/10.1038/s41416-020-0802-1>.
- [129] Xin Le Yap, Jun-An Chen. "Elucidation of how the Mir-23-27-24 cluster regulates development and aging." In: *Experimental & Molecular Medicine* (2024), pp. 1–9. URL: <https://doi.org/10.1038/s12276-024-01266-3>.

- [130] Li-Hua Sun, Dan Tian, Ze-Cheng Yang, Jin-Long Li. "Exosomal miR-21 promotes proliferation, invasion and therapy resistance of colon adenocarcinoma cells through its target PDCD4." In: *Scientific reports* 10.1 (2020), p. 8271. URL: <https://doi.org/10.1038/s41598-020-65207-6>.
- [131] Mustafa Kahraman, Anne Röske, Thomas Laufer, Tobias Fehlmann, Christina Backes, et al. "MicroRNA in diagnosis and therapy monitoring of early-stage triple-negative breast cancer." In: *Scientific reports* 8.1 (2018), p. 11584. URL: <https://doi.org/10.1038/s41598-018-29917-2>.
- [132] Nahid Arghiani, Khalid Shah. "Modulating microRNAs in cancer: Next-generation therapies." In: *Cancer Biology & Medicine* 19.3 (2021), p. 289. URL: <https://doi.org/10.20892/j.issn.2095-3941.2021.0294>.
- [133] Olivia Larne, Elena Martens-Uzunova, Zandra Hagman, Anders Edsjö, Giuseppe Lippolis, et al. "miQ—a novel microRNA based diagnostic and prognostic tool for prostate cancer." In: *International journal of cancer* 132.12 (2013), pp. 2867–2875. URL: <https://doi.org/10.1002/ijc.27973>.
- [134] Carlos López-Otín, Maria A. Blasco, Linda Partridge, Manuel Serrano, Guido Kroemer. "The hallmarks of aging." In: *Cell* 153.6 (2013), pp. 1194–1217. URL: <https://doi.org/10.1016/j.cell.2013.05.039>.
- [135] Carlos López-Otín, Maria A. Blasco, Linda Partridge, Manuel Serrano, Guido Kroemer. "Hallmarks of aging: An expanding universe." In: *Cell* 186.2 (2023), pp. 243–278. URL: <https://doi.org/10.1016/j.cell.2022.11.001>.
- [136] Catharina Bertram, Ralf Hass. "Cellular responses to reactive oxygen species-induced DNA damage and aging." In: *Biological Chemistry* 389.3 (2008), pp. 211–220. URL: <https://doi.org/10.1515/BC.2008.031>.
- [137] Carlotta Giorgi, Saverio Marchi, Ines C. M. Simoes, Ziyu Ren, Giampaolo Morciano, et al. "Mitochondria and reactive oxygen species in aging and age-related diseases." In: *International review of cell and molecular biology* 340 (2018), pp. 209–344. URL: <https://doi.org/10.1016/bs.ircmb.2018.05.006>.
- [138] Bennett G. Childs, Matej Durik, Darren J. Baker, Jan M. Van Deursen. "Cellular senescence in aging and age-related disease: from mechanisms

- to therapy." In: *Nature medicine* 21.12 (2015), pp. 1424–1435. URL: <https://doi.org/10.1038/nm.4000>.
- [139] Elizabeth H. Blackburn, Elissa S. Epel, Jue Lin. "Human telomere biology: a contributory and interactive factor in aging, disease risks, and protection." In: *Science* 350.6265 (2015), pp. 1193–1198. URL: <https://doi.org/10.1126/science.aab3389>.
- [140] Linda Partridge, Nicholas H. Barton. "Optimally, mutation and the evolution of ageing." In: *Nature* 362.6418 (1993), pp. 305–311. URL: <https://doi.org/10.1038/362305ao>.
- [141] Mariela Jaskelioff, Florian L. Muller, Ji-Hye Paik, Emily Thomas, Shan Jiang, et al. "Telomerase reactivation reverses tissue degeneration in aged telomerase-deficient mice." In: *Nature* 469.7328 (2011), pp. 102–106. URL: <https://doi.org/10.1038/nature09603>.
- [142] Shuling Song, Eric W.-F. Lam, Tamara Tchkonina, James L. Kirkland, Yu Sun. "Senescent cells: emerging targets for human aging and age-related diseases." In: *Trends in biochemical sciences* 45.7 (2020), pp. 578–592. URL: <https://doi.org/10.1016/j.tibs.2020.03.008>.
- [143] Steve Horvath. "DNA methylation age of human tissues and cell types." In: *Genome biology* 14 (2013), pp. 1–20. URL: <https://doi.org/10.1186/gb-2013-14-10-r115>.
- [144] Gregory Hannum, Justin Guinney, Ling Zhao, LI Zhang, Guy Hughes, et al. "Genome-wide methylation profiles reveal quantitative views of human aging rates." In: *Molecular cell* 49.2 (2013), pp. 359–367. URL: <https://doi.org/10.1016/j.molcel.2012.10.016>.
- [145] Reinier A. Boon, Kazuma Iekushi, Stefanie Lechner, Timon Seeger, Ariane Fischer, et al. "MicroRNA-34a regulates cardiac ageing and function." In: *Nature* 495.7439 (2013), pp. 107–110. URL: <https://doi.org/10.1038/nature11919>.
- [146] Takashi Ito, Shusuke Yagi, Munekazu Yamakuchi. "MicroRNA-34a regulation of endothelial senescence." In: *Biochemical and biophysical research*

communications 398.4 (2010), pp. 735–740. URL: <https://doi.org/10.1016/j.bbrc.2010.07.012>.

- [147] Ryan M. O’Connell, Dinesh S. Rao, Aadel A. Chaudhuri, David Baltimore. “Physiological and pathological roles for microRNAs in the immune system.” In: *Nature Reviews Immunology* 10.2 (2010), pp. 111–122. URL: <https://doi.org/10.1038/nri2708>.
- [148] Tobias Fehlmann, Benoit Lehallier, Nicholas Schaum, Oliver Hahn, Mustafa Kahraman, et al. “Common diseases alter the physiological age-related blood microRNA profile.” In: *Nature communications* 11.1 (2020), p. 5958. URL: <https://doi.org/10.1038/s41467-020-19665-1>.
- [149] Mehmet Somel, Song Guo, Ning Fu, Zheng Yan, Hai Yang Hu, et al. “MicroRNA, mRNA, and protein expression link development and aging in human and macaque brain.” In: *Genome research* 20.9 (2010), pp. 1207–1218. URL: <https://doi.org/10.1101/gr.106849.110>.
- [150] Oliver Hahn, Aulden G. Foltz, Micaiah Atkins, Blen Kedir, Patricia Moran-Losada, et al. “Atlas of the aging mouse brain reveals white matter as vulnerable foci.” In: *Cell* 186.19 (2023), pp. 4117–4133. URL: <https://doi.org/10.1016/j.cell.2023.07.027>.
- [151] Nicholas Schaum, Benoit Lehallier, Oliver Hahn, Róbert Pálovics, Shayan Hosseinzadeh, et al. “Ageing hallmarks exhibit organ-specific temporal signatures.” In: *Nature* 583.7817 (2020), pp. 596–602. URL: <https://doi.org/10.1038/s41586-020-2499-y>.
- [152] Tianxiao Huan, George Chen, Chunyu Liu, Anindya Bhattacharya, Jian Rong, et al. “Age-associated micro RNA expression in human peripheral blood is associated with all-cause mortality and age-related traits.” In: *Aging Cell* 17.1 (2018), e12687. URL: <https://doi.org/10.1111/accel.12687>.
- [153] Ulaş Işıldak, Mehmet Somel, Janet M. Thornton, Handan Melike Dönertaş. “Temporal changes in the gene expression heterogeneity during brain development and aging.” In: *Scientific reports* 10.1 (2020), p. 4080. URL: <https://doi.org/10.1038/s41598-020-60998-0>.

- [154] Daniel Glass, Ana Viñuela, Matthew N. Davies, Adaikalavan Ramasamy, Leopold Parts, et al. "Gene expression changes with age in skin, adipose tissue, blood and brain." In: *Genome biology* 14 (2013), pp. 1–12. URL: <https://doi.org/10.1186/gb-2013-14-7-r75>.
- [155] Sarah E. Harris, Valentina Riggio, Louise Evenden, Tamara Gilchrist, Sarah McCafferty, et al. "Age-related gene expression changes, and transcriptome wide association study of physical and cognitive aging traits, in the Lothian Birth Cohort 1936." In: *Aging (Albany NY)* 9.12 (2017), p. 2489. URL: <https://doi.org/10.18632/aging.101333>.
- [156] Stephen Frenk, Jonathan Houseley. "Gene expression hallmarks of cellular ageing." In: *Biogerontology* 19.6 (2018), pp. 547–566. URL: <https://doi.org/10.1007/s10522-018-9750-z>.
- [157] Kailiang Sun, Eric C. Lai. "Adult-specific functions of animal microRNAs." In: *Nature reviews genetics* 14.8 (2013), pp. 535–548. URL: <https://doi.org/10.1038/nrg3471>.
- [158] Donghong Cai, Jing-Dong J. Han. "Aging-associated lncRNAs are evolutionarily conserved and participate in NFκB signaling." In: *Nature Aging* 1.5 (2021), pp. 438–453. URL: <https://doi.org/10.1038/s43587-021-00056-0>.
- [159] Kira Sheinerman, Vladimir Tsivinsky, Aabhas Mathur, Debra Kessler, Beth Shaz, et al. "Age-and sex-dependent changes in levels of circulating brain-enriched microRNAs during normal aging." In: *Aging (Albany NY)* 10.10 (2018), p. 3017. URL: <https://doi.org/10.18632/aging.101613>.
- [160] Dantong Zhu, Matt Arnold, Brady A. Samuelson, Judy Z. Wu, Amber Mueller, et al. "Sex dimorphism and tissue specificity of gene expression changes in aging mice." In: *Biology of sex Differences* 15.1 (2024), p. 89. URL: <https://doi.org/10.1186/s13293-024-00666-4>.
- [161] Dongguang Guo, Yaqiong Ye, Junjie Qi, Xiaotong Tan, Yuan Zhang, et al. "Age and sex differences in microRNAs expression during the process of thymus aging." In: *Acta biochimica et biophysica Sinica* 49.5 (2017), pp. 409–419. URL: <https://doi.org/10.1093/abbs/gmx029>.

- [162] Joshua C. Kwekel, Vikrant Vijay, Varsha G. Desai, Carrie L. Moland, James C. Fuscoe. "Age and sex differences in kidney microRNA expression during the life span of F344 rats." In: *Biology of sex Differences* 6 (2015), pp. 1–16. URL: <https://doi.org/10.1186/s13293-014-0019-1>.
- [163] Monika S. Kowalczyk, Itay Tirosh, Dirk Heckl, Tata Nageswara Rao, Atray Dixit, et al. "Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells." In: *Genome research* 25.12 (2015), pp. 1860–1872. URL: <https://doi.org/10.1101/gr.192237.115>.
- [164] Jacob C. Kimmel, Lolita Penland, Nimrod D. Rubinstein, David G. Hendrickson, David R. Kelley, et al. "Murine single-cell RNA-seq reveals cell-identity-and tissue-specific trajectories of aging." In: *Genome research* 29.12 (2019), pp. 2088–2103. URL: <https://doi.org/10.1101/gr.253880.119>.
- [165] Methodios Ximerakis, Scott L. Lipnick, Brendan T. Innes, Sean K. Simmons, Xian Adiconis, et al. "Single-cell transcriptomic profiling of the aging mouse brain." In: *Nature neuroscience* 22.10 (2019), pp. 1696–1708. URL: <https://doi.org/10.1038/s41593-019-0491-3>.
- [166] The Tabula Muris Consortium. "A single-cell transcriptomic atlas characterizes ageing tissues in the mouse." In: *Nature* 583.7817 (2020), pp. 590–595. URL: <https://doi.org/10.1038/s41586-020-2496-1>.
- [167] Alex K. Shalek, Rahul Satija, Joe Shuga, John J. Trombetta, Dave Gennert, et al. "Single-cell RNA-seq reveals dynamic paracrine control of cellular variation." In: *Nature* 510.7505 (2014), pp. 363–369. URL: <https://doi.org/10.1038/nature13437>.
- [168] Michelle Boehm, Frank Slack. "A developmental timing microRNA and its target regulate life span in *C. elegans*." In: *Science* 310.5756 (2005), pp. 1954–1957. URL: <https://doi.org/10.1126/science.1115596>.
- [169] Eric M. Small, Eric N. Olson. "Pervasive roles of microRNAs in cardiovascular biology." In: *Nature* 469.7330 (2011), pp. 336–342. URL: <https://doi.org/10.1038/nature09783>.

- [170] Jun Guo, Xiuqing Huang, Lin Dou, Mingjing Yan, Tao Shen, et al. "Aging and aging-related diseases: from molecular mechanisms to interventions and treatments." In: *Signal Transduction and Targeted Therapy* 7.1 (2022), p. 391. URL: <https://doi.org/10.1038/s41392-022-01251-0>.
- [171] Chang-Gong Liu, George Adrian Calin, Stefano Volinia, Carlo M. Croce. "MicroRNA expression profiling using microarrays." In: *Nature protocols* 3.4 (2008), pp. 563–578. URL: <https://doi.org/10.1038/nprot.2008.14>.
- [172] Chang-Gong Liu, George Adrian Calin, Brian Meloon, Nir Gamliel, Cinzia Sevignani, et al. "An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues." In: *Proceedings of the National Academy of Sciences* 101.26 (2004), pp. 9740–9744. URL: <https://doi.org/10.1073/pnas.0403293101>.
- [173] J Michael Thomson, Joel S Parker, Scott M Hammond. "Microarray analysis of miRNA gene expression." In: *Methods in enzymology* 427 (2007), pp. 107–122. URL: [https://doi.org/10.1016/S0076-6879\(07\)27006-5](https://doi.org/10.1016/S0076-6879(07)27006-5).
- [174] Sara Goodwin, John D. McPherson, W. Richard McCombie. "Coming of age: ten years of next-generation sequencing technologies." In: *Nature reviews genetics* 17.6 (2016), pp. 333–351. URL: <https://doi.org/10.1038/nrg.2016.49>.
- [175] Barton E. Slatko, Andrew F. Gardner, Frederick M. Ausubel. "Overview of next-generation sequencing technologies." In: *Current protocols in molecular biology* 122.1 (2018), e59. URL: <https://doi.org/10.1002/cpmb.59>.
- [176] Alessandra Modi, Stefania Vai, David Caramelli, Martina Lari. "The Illumina Sequencing Protocol and the NovaSeq 6000 System." In: *Bacterial Pangenomics: Methods and Protocols*. Ed. by Alessio Mengoni, Giovanni Bacci, Marco Fondi. Springer US, 2021, pp. 15–42. ISBN: 978-1-0716-1099-2. URL: https://doi.org/10.1007/978-1-0716-1099-2_2.
- [177] H. P. J. Buermans, J. T. Den Dunnen. "Next generation sequencing technology: advances and applications." In: *Biochimica et Biophysica Acta (BBA)-*

- Molecular Basis of Disease* 1842.10 (2014), pp. 1932–1941. URL: <https://doi.org/10.1016/j.bbadis.2014.06.015>.
- [178] K.A. Wetterstrand. *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)*. last visited on 2025-01-28. URL: <https://www.genome.gov/sequencingcostsdata>.
- [179] Marcel Martin. “Cutadapt removes adapter sequences from high-throughput sequencing reads.” In: *EMBnet. journal* 17.1 (2011), pp. 10–12. URL: <https://doi.org/10.14806/ej.17.1.200>.
- [180] Shirley Tam, Ming-Sound Tsao, John D. McPherson. “Optimization of miRNA-seq data preprocessing.” In: *Briefings in Bioinformatics* 16.6 (2015), pp. 950–963. URL: <https://doi.org/10.1093/bib/bbv019>.
- [181] Ben Langmead, Steven L. Salzberg. “Fast gapped-read alignment with Bowtie 2.” In: *Nature methods* 9.4 (2012), pp. 357–359. URL: <https://doi.org/10.1038/nmeth.1923>.
- [182] Ashraful Haque, Jessica Engel, Sarah A. Teichmann, Tapio Lönnberg. “A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications.” In: *Genome medicine* 9 (2017), pp. 1–12. URL: <https://doi.org/10.1186/s13073-017-0467-4>.
- [183] Alexander Dobin, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, et al. “STAR: ultrafast universal RNA-seq aligner.” In: *Bioinformatics* 29.1 (2012), pp. 15–21. URL: <https://doi.org/10.1093/bioinformatics/bts635>.
- [184] Tobias Fehlmann, Fabian Kern, Omar Laham, Christina Backes, Jeffrey Solomon, et al. “miRMaster 2.0: multi-species non-coding RNA sequencing analyses at scale.” In: *Nucleic Acids Research* 49.W1 (2021), W397–W408. URL: <https://doi.org/10.1093/nar/gkab268>.
- [185] Marc R. Friedländer, Wei Chen, Catherine Adamidi, Jonas Maaskola, Ralf Einspanier, et al. “Discovering microRNAs from deep sequencing data using miRDeep.” In: *Nature biotechnology* 26.4 (2008), pp. 407–415. URL: <https://doi.org/10.1038/nbt1394>.

- [186] Christian Wake, Adam Labadorf, Alexandra Dumitriu, Andrew G. Hoss, Joli Bregu, et al. "Novel microRNA discovery using small RNA sequencing in post-mortem human brain." In: *BMC genomics* 17 (2016), pp. 1–9. URL: <https://doi.org/10.1186/s12864-016-3114-3>.
- [187] Fuchou Tang, Catalin Barbacioru, Yangzhou Wang, Ellen Nordman, Clarence Lee, et al. "mRNA-Seq whole-transcriptome analysis of a single cell." In: *Nature methods* 6.5 (2009), pp. 377–382. URL: <https://doi.org/10.1038/nmeth.1315>.
- [188] Qiaolin Deng, Daniel Ramsköld, Björn Reinius, Rickard Sandberg. "Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells." In: *Science* 343.6167 (2014), pp. 193–196. URL: <https://doi.org/10.1126/science.1245316>.
- [189] Diego Adhemar Jaitin, Ephraim Kenigsberg, Hadas Keren-Shaul, Naama Elefant, Franziska Paul, et al. "Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types." In: *Science* 343.6172 (2014), pp. 776–779. URL: <https://doi.org/10.1126/science.1247651>.
- [190] Alina Isakova, Norma Neff, Stephen R. Quake. "Single-cell quantification of a broad RNA spectrum reveals unique noncoding patterns associated with cell types and states." In: *Proceedings of the National Academy of Sciences* 118.51 (2021), e2113568118. URL: <https://doi.org/10.1073/pnas.2113568118>.
- [191] Lukas Heumos, Anna C. Schaar, Christopher Lance, Anastasia Litinet-skaya, Felix Drost, et al. "Best practices for single-cell analysis across modalities." In: *Nature Reviews Genetics* 24.8 (2023), pp. 550–572. URL: <https://doi.org/10.1038/s41576-023-00586-w>.
- [192] Nayi Wang, Ji Zheng, Zhuo Chen, Yang Liu, Burak Dura, et al. "Single-cell microRNA-mRNA co-sequencing reveals non-genetic heterogeneity and mechanisms of microRNA regulation." In: *Nature communications* 10.1 (2019), p. 95. URL: <https://doi.org/10.1038/s41467-018-07981-6>.

- [193] Xiliang Wang, Yao He, Qiming Zhang, Xianwen Ren, Zemin Zhang. "Direct comparative analyses of 10X genomics chromium and smart-seq2." In: *Genomics, Proteomics and Bioinformatics* 19.2 (2021), pp. 253–266. URL: <https://doi.org/10.1016/j.gpb.2020.02.005>.
- [194] Caixia Gao, Mingnan Zhang, Lei Chen. "The comparison of two single-cell sequencing platforms: BD rhapsody and 10x genomics chromium." In: *Current genomics* 21.8 (2020), pp. 602–609. URL: <https://doi.org/10.2174/1389202921999200625220812>.
- [195] Sarah M. Hücker, Tobias Fehlmann, Christian Werno, Kathrin Weidele, Florian Lüke, et al. "Single-cell microRNA sequencing method comparison and application to cell lines and circulating lung tumor cells." In: *Nature communications* 12.1 (2021), p. 4316. URL: <https://doi.org/10.1038/s41467-021-24611-w>.
- [196] Peipei Li, Yongjun Piao, Ho Sun Shon, Keun Ho Ryu. "Comparing the normalization methods for the differential analysis of Illumina high-throughput RNA-Seq data." In: *BMC bioinformatics* 16 (2015), pp. 1–9. URL: <https://doi.org/10.1186/s12859-015-0778-7>.
- [197] Tallulah S. Andrews, Vladimir Yu Kiselev, Davis McCarthy, Martin Hemberg. "Tutorial: guidelines for the computational analysis of single-cell RNA sequencing data." In: *Nature protocols* 16.1 (2021), pp. 1–9. URL: <https://doi.org/10.1038/s41596-020-00409-w>.
- [198] Sylvain Pradervand, Johann Weber, Jérôme Thomas, Manuel Bueno, Pratyaksha Wirapati, et al. "Impact of normalization on miRNA microarray expression profiling." In: *Rna* 15.3 (2009), pp. 493–501. URL: <http://www.rnajournal.org/cgi/doi/10.1261/rna.1295509>.
- [199] Pedro López-Romero, Manuel A. González, Sergio Callejas, Ana Dopazo, Rafael A. Irizarry. "Processing of Agilent microRNA array data." In: *BMC Research notes* 3 (2010), pp. 1–6. URL: <https://doi.org/10.1186/1756-0500-3-18>.

- [200] Ying Sha, John H. Phan, May D. Wang. "Effect of low-expression gene filtering on detection of differentially expressed genes in RNA-seq data." In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2015, pp. 6461–6464. URL: <https://doi.org/10.1109/EMBC.2015.7319872>.
- [201] Nadia M. Davidson, Alicia Oshlack. "Corset: enabling differential gene expression analysis for de novo assembled transcriptomes." In: *Genome biology* 15 (2014), pp. 1–14. URL: <https://doi.org/10.1186/s13059-014-0410-6>.
- [202] Winston Koh, Chen Tian Sheng, Betty Tan, Qian Yi Lee, Vladimir Kuznetsov, et al. "Analysis of deep sequencing microRNA expression profile from human embryonic stem cells derived mesenchymal stem cells reveals possible role of let-7 microRNA family in downstream targeting of hepatic nuclear factor 4 alpha." In: *BMC genomics* 11 (2010), pp. 1–15. URL: <https://doi.org/10.1186/1471-2164-11-S1-S6>.
- [203] Alicia Oshlack, Matthew J. Wakefield. "Transcript length bias in RNA-seq data confounds systems biology." In: *Biology direct* 4 (2009), pp. 1–10. URL: <https://doi.org/10.1186/1745-6150-4-14>.
- [204] Ali Mortazavi, Brian A. Williams, Kenneth McCue, Lorian Schaeffer, Barbara Wold. "Mapping and quantifying mammalian transcriptomes by RNA-Seq." In: *Nature methods* 5.7 (2008), pp. 621–628. URL: <https://doi.org/10.1038/nmeth.1226>.
- [205] Marc Sultan, Marcel H. Schulz, Hugues Richard, Alon Magen, Andreas Klingenhoff, et al. "A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome." In: *Science* 321.5891 (2008), pp. 956–960. URL: <https://doi.org/10.1126/science.1160342>.
- [206] Michael I. Love, Wolfgang Huber, Simon Anders. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." In: *Genome biology* 15 (2014), pp. 1–21. URL: <https://doi.org/10.1186/s13059-014-0550-8>.

- [207] Klaus Backhaus, Bernd Erichson, Sonja Gensler, Rolf Weiber, Thomas Weiber. "Clusteranalyse." In: *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Springer Fachmedien Wiesbaden, 2023, pp. 485–571. ISBN: 978-3-658-40465-9. URL: https://doi.org/10.1007/978-3-658-40465-9_8.
- [208] Samuel S. Shapiro, Martin B. Wilk. "An analysis of variance test for normality (complete samples)." In: *Biometrika* 52.3-4 (1965), pp. 591–611. URL: <https://doi.org/10.1093/biomet/52.3-4.591>.
- [209] Myles Hollander, Douglas A. Wolfe, Eric Chicken. *Nonparametric statistical methods*. Vol. 3. John Wiley & Sons, 2014. ISBN: 9781119196037. URL: <https://doi.org/10.1002/9781119196037>.
- [210] R. John Simes. "An improved Bonferroni procedure for multiple tests of significance." In: *Biometrika* 73.3 (1986), pp. 751–754. URL: <https://doi.org/10.1093/biomet/73.3.751>.
- [211] Yoav Benjamini, Yosef Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." In: *Journal of the Royal statistical society: series B (Methodological)* 57.1 (1995), pp. 289–300. URL: <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- [212] Jacob Cohen, Patricia Cohen, Stephen G. West, Leona S. Aiken. *Applied multiple regression/correlation analysis for the behavioral sciences*. Vol. 3. Routledge, 2002. ISBN: 9780203774441. URL: <https://doi.org/10.4324/9780203774441>.
- [213] B. R. Martin. "Chapter 11 - Hypothesis Testing II: Other Tests." In: *Statistics for Physical Science*. Ed. by B. R. Martin. Academic Press, 2012, pp. 221–242. ISBN: 978-0-12-387760-4. URL: <https://doi.org/10.1016/B978-0-12-387760-4.00011-1>.
- [214] Charles Spearman. "The proof and measurement of association between two things." In: (1961). URL: <https://doi.org/10.1093/ije/dyq191>.
- [215] In-Kwon Yeo, Richard A. Johnson. "A new family of power transformations to improve normality or symmetry." In: *Biometrika* 87.4 (2000), pp. 954–959. URL: <https://doi.org/10.1093/biomet/87.4.954>.

- [216] Maurice S. Bartlett. "The use of transformations." In: *Biometrics* 3.1 (1947), pp. 39–52. URL: <https://doi.org/10.2307/3001536>.
- [217] Jun Lu, Gad Getz, Eric A. Miska, Ezequiel Alvarez-Saavedra, Justin Lamb, et al. "MicroRNA expression profiles classify human cancers." In: *nature* 435.7043 (2005), pp. 834–838. URL: <https://doi.org/10.1038/nature03702>.
- [218] Li Guo, Tingming Liang, Jiafeng Yu, Quan Zou. "A comprehensive analysis of miRNA/isomiR expression with gender difference." In: *PloS one* 11.5 (2016), e0154955. URL: <https://doi.org/10.1371/journal.pone.0154955>.
- [219] SFM Häusler, A. Keller, P. A. Chandran, K. Ziegler, K. Zipp, et al. "Whole blood-derived miRNA profiles as potential new tools for ovarian cancer screening." In: *British journal of cancer* 103.5 (2010), pp. 693–700. URL: <https://doi.org/10.1038/sj.bjc.6605833>.
- [220] Ilknur Suer, Esra Guzel, Omer F. Karatas, Chad J. Creighton, Michael Ittmann, et al. "MicroRNAs as prognostic markers in prostate cancer." In: *The Prostate* 79.3 (2019), pp. 265–271. URL: <https://doi.org/10.1002/pros.23731>.
- [221] Julia Wilflingseder, Heinz Regele, Paul Perco, Alexander Kainz, Afschin Soleiman, et al. "miRNA profiling discriminates types of rejection and injury in human renal allografts." In: *Transplantation* 95.6 (2013), pp. 835–841. URL: <https://doi.org/10.1097/TP.ob013e318280b385>.
- [222] Patrick Wahl, Udo F. Wehmeier, Felix J. Jansen, Yvonne Kilian, Wilhelm Bloch, et al. "Acute effects of different exercise protocols on the circulating vascular microRNAs-16,-21, and-126 in trained subjects." In: *Frontiers in physiology* 7 (2016), p. 643. URL: <https://doi.org/10.3389/fphys.2016.00643>.
- [223] Jingchuan Wu, Junchi He, Xiaocui Tian, Yuetao Luo, Jianjun Zhong, et al. "microRNA-9-5p alleviates blood–brain barrier damage and neuroinflammation after traumatic brain injury." In: *Journal of Neurochemistry* 153.6 (2020), pp. 710–726. URL: <https://doi.org/10.1111/jnc.14963>.
- [224] Roy Navon, Hui Wang, Israel Steinfeld, Anya Tsalenko, Amir Ben-Dor, et al. "Novel rank-based statistical methods reveal microRNAs with differential

- expression in multiple cancer types." In: *PloS one* 4.11 (2009), e8003. URL: <https://doi.org/10.1371/journal.pone.0008003>.
- [225] Thalia Erbes, Marc Hirschfeld, Gerta Rücker, Markus Jaeger, Jasmin Boas, et al. "Feasibility of urinary microRNA detection in breast cancer patients and its potential as an innovative non-invasive biomarker." In: *BMC cancer* 15 (2015), pp. 1–9. URL: <https://doi.org/10.1186/s12885-015-1190-4>.
- [226] Donald R. Jensen. "Multivariate Statistical Distributions." In: *International Encyclopedia of Statistical Science*. Ed. by Miodrag Lovric. Springer Berlin Heidelberg, 2011, pp. 920–928. ISBN: 978-3-642-04898-2. URL: https://doi.org/10.1007/978-3-642-04898-2_400.
- [227] Xiangqin Cui, Gary A. Churchill. "Statistical tests for differential expression in cDNA microarray experiments." In: *Genome biology* 4 (2003), pp. 1–10. URL: <https://doi.org/10.1186/gb-2003-4-4-210>.
- [228] Mark D. Robinson, Davis J. McCarthy, Gordon K. Smyth. "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." In: *bioinformatics* 26.1 (2010), pp. 139–140. URL: <https://doi.org/10.1093/bioinformatics/btp616>.
- [229] Charlotte Soneson, Mauro Delorenzi. "A comparison of methods for differential expression analysis of RNA-seq data." In: *BMC bioinformatics* 14 (2013), pp. 1–18. URL: <https://doi.org/10.1186/1471-2105-14-91>.
- [230] Nicole Noren Hooten, Kotb Abdelmohsen, Myriam Gorospe, Ngozi Ejiogu, Alan B. Zonderman, et al. "microRNA expression patterns reveal differential expression of target genes with age." In: *PloS one* 5.5 (2010), e10724. URL: <https://doi.org/10.1371/journal.pone.0010724>.
- [231] Maha Abdellatif. "Differential expression of microRNAs in different disease states." In: *Circulation research* 110.4 (2012), pp. 638–650. URL: <https://doi.org/10.1161/CIRCRESAHA.111.247437>.
- [232] Pedro López-Romero. "Pre-processing and differential expression analysis of Agilent microRNA arrays using the AgiMicroRna Bioconductor library."

- In: *BMC genomics* 12 (2011), pp. 1–8. URL: <https://doi.org/10.1186/1471-2164-12-64>.
- [233] Zhifu Sun, Jared Evans, Aditya Bhagwate, Sumit Middha, Matthew Bockol, et al. “CAP-miRSeq: a comprehensive analysis pipeline for microRNA sequencing data.” In: *BMC genomics* 15 (2014), pp. 1–10. URL: <https://doi.org/10.1186/1471-2164-15-423>.
- [234] Tomasz Stokowy, Markus Eszlinger, Michał Świerniak, Krzysztof Fulariewicz, Barbara Jarzab, et al. “Analysis options for high-throughput sequencing in miRNA expression profiling.” In: *BMC research notes* 7 (2014), pp. 1–12. URL: <https://doi.org/10.1186/1756-0500-7-144>.
- [235] Eduardo Andrés-León, Rocío Núñez-Torres, Ana M. Rojas. “miARma-Seq: a comprehensive tool for miRNA, mRNA and circRNA analysis.” In: *Scientific reports* 6.1 (2016), p. 25749. URL: <https://doi.org/10.1038/srep25749>.
- [236] Yunshun Chen, Aaron T. L. Lun, Gordon K. Smyth. “From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline.” In: *F1000Research* 5 (2016). URL: <https://doi.org/10.12688/f1000research.8987.2>.
- [237] Jacob Cohen. *Statistical power analysis for the behavioral sciences*. Vol. 2. routledge, 1988. ISBN: 0-8058-0283-5. URL: <https://doi.org/10.4324/9780203771587>.
- [238] Xingsong Li, Xiaokang Yu, Yuting He, Yuhuan Meng, Jinsheng Liang, et al. “Integrated analysis of MicroRNA (miRNA) and mRNA profiles reveals reduced correlation between MicroRNA and target gene in cancer.” In: *BioMed research international* 2018.1 (2018), p. 1972606. URL: <https://doi.org/10.1155/2018/1972606>.
- [239] Fengjun Wang, Zhiguo Zheng, Jiangfeng Guo, Xianfeng Ding. “Correlation and quantitation of microRNA aberrant expression in tissues and sera from patients with breast tumor.” In: *Gynecologic oncology* 119.3 (2010), pp. 586–593. URL: <https://doi.org/10.1016/j.ygyno.2010.07.021>.

- [240] Ke-Hung Chien, Shih-Jen Chen, Jorn-Hon Liu, Lin-Chung Woung, Jiann-Torng Chen, et al. "Correlation of microRNA-145 levels and clinical severity of pterygia." In: *The Ocular Surface* 11.2 (2013), pp. 133–138. URL: <https://doi.org/10.1016/j.jtos.2012.12.001>.
- [241] Mark Z. Kos, Sobha Puppala, Dianne Cruz, Jennifer L. Neary, Ashish Kumar, et al. "Blood-based miRNA biomarkers as correlates of brain-based miRNA expression." In: *Frontiers in Molecular Neuroscience* 15 (2022), p. 817290. URL: <https://doi.org/10.3389/fnmol.2022.817290>.
- [242] Marcus Gry, Rebecca Rimini, Sara Strömberg, Anna Asplund, Fredrik Pontén, et al. "Correlations between RNA and protein expression profiles in 23 human cell lines." In: *BMC genomics* 10 (2009), pp. 1–14. URL: <https://doi.org/10.1186/1471-2164-10-365>.
- [243] Fredrik Edfors, Frida Danielsson, Björn M. Hallström, Lukas Käll, Emma Lundberg, et al. "Gene-specific correlation of RNA and protein levels in human cells and tissues." In: *Molecular systems biology* 12.10 (2016), p. 883. URL: <https://doi.org/10.15252/msb.20167144>.
- [244] Max Kotlyar, Stefanie Fuhrman, Alan Ableson, Roland Somogyi. "Spearman correlation identifies statistically significant gene expression clusters in spinal cord development and injury." In: *Neurochemical Research* 27 (2002), pp. 1133–1140. URL: <https://doi.org/10.1023/A:1020969208033>.
- [245] Laura E. Pascal, Lawrence D. True, David S. Campbell, Eric W. Deutsch, Michael Risk, et al. "Correlation of mRNA and protein levels: cell type-specific gene expression of cluster designation antigens in the prostate." In: *BMC genomics* 9 (2008), pp. 1–13. URL: <https://doi.org/10.1186/1471-2164-9-246>.
- [246] Yanyan Jiang, Jing Chen, Yunchuang Sun, Fan Li, Luhua Wei, et al. "Profiling of differentially expressed microRNAs in saliva of Parkinson's disease patients." In: *Frontiers in Neurology* 12 (2021), p. 738530. URL: <https://doi.org/10.3389/fneur.2021.738530>.

- [247] Y-h. Taguchi, Yoshiki Murakami. "Principal component analysis based feature extraction approach to identify circulating microRNA biomarkers." In: *PloS one* 8.6 (2013), e66714. URL: <https://doi.org/10.1371/journal.pone.0066714>.
- [248] Stacy L. Sell, Steven G. Widen, Donald S. Prough, Helen L. Hellmich. "Principal component analysis of blood microRNA datasets facilitates diagnosis of diverse diseases." In: *PloS one* 15.6 (2020), e0234185. URL: <https://doi.org/10.1371/journal.pone.0234185>.
- [249] Y-h. Taguchi, Mitsuo Iwadate, Hideaki Umeyama. "Principal component analysis-based unsupervised feature extraction applied to in silico drug discovery for posttraumatic stress disorder-mediated heart disease." In: *BMC bioinformatics* 16 (2015), pp. 1–26. URL: <https://doi.org/10.1186/s12859-015-0574-4>.
- [250] Andrew Feber, Liqiang Xi, James D. Luketich, Arjun Pennathur, Rodney J. Landreneau, et al. "MicroRNA expression profiles of esophageal cancer." In: *The Journal of thoracic and cardiovascular surgery* 135.2 (2008), pp. 255–260. URL: <https://doi.org/10.1016/j.jtcvs.2007.08.055>.
- [251] Nikolai Hecker, Carsten Stephan, Hans-Joachim Mollenkopf, Klaus Jung, Robert Preissner, et al. "A new algorithm for integrated analysis of miRNA-mRNA interactions based on individual classification reveals insights into bladder cancer." In: *PLoS One* 8.5 (2013), e64543. URL: <https://doi.org/10.1371/journal.pone.0064543>.
- [252] Gregory D. Miles, Michael Seiler, Lorna Rodriguez, Gunaretnam Rajagopal, Gyan Bhanot. "Identifying microRNA/mRNA dysregulations in ovarian cancer." In: *BMC research notes* 5 (2012), pp. 1–10. URL: <https://doi.org/10.1186/1756-0500-5-164>.
- [253] Koki Tsuyuzaki, Hiroyuki Sato, Kenta Sato, Itoshi Nikaido. "Benchmarking principal component analysis for large-scale single-cell RNA-sequencing." In: *Genome biology* 21.1 (2020), p. 9. URL: <https://doi.org/10.1186/s13059-019-1900-3>.

- [254] Federico Marini, Harald Binder. “pcaExplorer: an R/Bioconductor package for interacting with RNA-seq principal components.” In: *BMC bioinformatics* 20 (2019), pp. 1–8. URL: <https://doi.org/10.1186/s12859-019-2879-1>.
- [255] Kitty C. M. Verhoeckx, Sabina Bijlsma, Els M. de Groene, Renger F. Witkamp, Jan van der Greef, et al. “A combination of proteomics, principal component analysis and transcriptomics is a powerful tool for the identification of biomarkers for macrophage maturation in the U937 cell line.” In: *Proteomics* 4.4 (2004), pp. 1014–1028. URL: <https://doi.org/10.1002/pmic.200300669>.
- [256] Abubakar Abid, Martin J. Zhang, Vivek K. Bagaria, James Zou. “Exploring patterns enriched in a dataset with contrastive principal component analysis.” In: *Nature communications* 9.1 (2018), p. 2134. URL: <https://doi.org/10.1038/s41467-018-04608-8>.
- [257] Hoa Thi Nhu Tran, Kok Siong Ang, Marion Chevrier, Xiaomeng Zhang, Nicole Yee Shin Lee, et al. “A benchmark of batch-effect correction methods for single-cell RNA sequencing data.” In: *Genome biology* 21 (2020), pp. 1–32. URL: <https://doi.org/10.1186/s13059-019-1850-9>.
- [258] Sai Ma, Bing Zhang, Lindsay M. LaFave, Andrew S. Earl, Zachary Chiang, et al. “Chromatin potential identified by shared single-cell profiling of RNA and chromatin.” In: *Cell* 183.4 (2020), pp. 1103–1116. URL: <https://doi.org/10.1016/j.cell.2020.09.056>.
- [259] Van Hoan Do, Stefan Canzar. “A generalization of t-SNE and UMAP to single-cell multimodal omics.” In: *Genome Biology* 22.1 (2021), p. 130. URL: <https://doi.org/10.1186/s13059-021-02356-5>.
- [260] Lisa Sikkema, Ciro Ramírez-Suástegui, Daniel C. Strobl, Tessa E. Gillett, Luke Zappia, et al. “An integrated cell atlas of the lung in health and disease.” In: *Nature medicine* 29.6 (2023), pp. 1563–1577. URL: <https://doi.org/10.1038/s41591-023-02327-2>.
- [261] Eirik Høye, Bastian Fromm, Paul H. M. Böttger, Diana Domanska, Annette Torgunrud, et al. “A comprehensive framework for analysis of microRNA

- sequencing data in metastatic colorectal cancer." In: *NAR cancer* 4.1 (2022), zcab051. URL: <https://doi.org/10.1093/narcan/zcab051>.
- [262] Ruizhi Xiang, Wencan Wang, Lei Yang, Shiyuan Wang, Chaohan Xu, et al. "A comparison for dimensionality reduction methods of single-cell RNA-seq data." In: *Frontiers in genetics* 12 (2021), p. 646936. URL: <https://doi.org/10.3389/fgene.2021.646936>.
- [263] Valéry L. Payen, Arnaud Lavergne, Niki Alevra Sarika, Megan Colonval, Latifa Karim, et al. "Single-cell RNA sequencing of human liver reveals hepatic stellate cell heterogeneity." In: *Jhep Reports* 3.3 (2021), p. 100278. URL: <https://doi.org/10.1016/j.jhepr.2021.100278>.
- [264] Ian T. Jolliffe. *Principal Component Analysis*. Vol. 2. Springer New York, 2002. ISBN: 978-0-387-22440-4. URL: <https://doi.org/10.1007/b98835>.
- [265] Leland McInnes, John Healy, James Melville. "Umap: Uniform manifold approximation and projection for dimension reduction." In: *arXiv preprint arXiv:1802.03426* (2018). URL: <https://doi.org/10.48550/arXiv.1802.03426>.
- [266] Nicole Ludwig, Anne Hecksteden, Mustafa Kahraman, Tobias Fehlmann, Thomas Laufer, et al. "Spring is in the air: seasonal profiles indicate vernal change of miRNA activity." In: *Rna Biology* 16.8 (2019), pp. 1034–1043. URL: <https://doi.org/10.1080/15476286.2019.1612217>.
- [267] Mehmet Somel, Xiling Liu, Lin Tang, Zheng Yan, Haiyang Hu, et al. "MicroRNA-driven developmental remodeling in the brain distinguishes humans from other primates." In: *PLoS biology* 9.12 (2011), e1001214. URL: <https://doi.org/10.1371/journal.pbio.1001214>.
- [268] Pierre R. Bushel, Richard S. Paules, Scott S. Auerbach. "A comparison of the TempO-Seq S1500+ platform to RNA-Seq and microarray using rat liver mode of action samples." In: *Frontiers in genetics* 9 (2018), p. 485. URL: <https://doi.org/10.3389/fgene.2018.00485>.
- [269] Eva C. Schwarz, Christina Backes, Arne Knörck, Nicole Ludwig, Petra Leidinger, et al. "Deep characterization of blood cell miRNomes by NGS."

- In: *Cellular and Molecular Life Sciences* 73 (2016), pp. 3169–3181. URL: <https://doi.org/10.1007/s00018-016-2154-9>.
- [270] Anne Hecksteden, Petra Leidinger, Christina Backes, Stefanie Rheinheimer, Mark Pfeiffer, et al. “miRNAs and sports: tracking training status and potentially confounding diagnoses.” In: *Journal of Translational Medicine* 14 (2016), pp. 1–13. URL: <https://doi.org/10.1186/s12967-016-0974-x>.
- [271] Jianying Li, Pierre Bushel, Tzu-Ming Chu, Russ Wolfinger. “Principal Variance Components Analysis: Estimating Batch Effects in Microarray Gene Expression Data.” In: *Batch Effects and Noise in Microarray Experiments: Sources and Solutions*. Ed. by Andreas Scherer. John Wiley & Sons, Ltd, 2009, pp. 141–154. ISBN: 9780470685983. URL: <https://doi.org/10.1002/9780470685983.ch12>.
- [272] Vladimir Yu Kiselev, Tallulah S. Andrews, Martin Hemberg. “Challenges in unsupervised clustering of single-cell RNA-seq data.” In: *Nature Reviews Genetics* 20.5 (2019), pp. 273–282. URL: <https://doi.org/10.1038/s41576-018-0088-9>.
- [273] Eric A. Miska, Ezequiel Alvarez-Saavedra, Matthew Townsend, Akira Yoshii, Nenad Šestan, et al. “Microarray analysis of microRNA expression in the developing mammalian brain.” In: *Genome biology* 5 (2004), pp. 1–13. URL: <https://doi.org/10.1186/gb-2004-5-9-r68>.
- [274] Yaqing Si, Peng Liu, Pinghua Li, Thomas P. Brutnell. “Model-based clustering for RNA-seq data.” In: *Bioinformatics* 30.2 (2014), pp. 197–205. URL: <https://doi.org/10.1093/bioinformatics/btt632>.
- [275] Pablo D. Reeb, Sergio J. Bramardi, Juan P. Steibel. “Assessing dissimilarity measures for sample-based hierarchical clustering of RNA sequencing data using plasmode datasets.” In: *PLoS One* 10.7 (2015), e0132310. URL: <https://doi.org/10.1371/journal.pone.0132310>.
- [276] Ren Qi, Anjun Ma, Qin Ma, Quan Zou. “Clustering and classification methods for single-cell RNA-sequencing data.” In: *Briefings in bioinformatics* 21.4 (2020), pp. 1196–1208. URL: <https://doi.org/10.1093/bib/bbzo62>.

- [277] Yang Yang, Ning Huang, Luning Hao, Wei Kong. "A clustering-based approach for efficient identification of microRNA combinatorial biomarkers." In: *BMC genomics* 18 (2017), pp. 1–14. URL: <https://doi.org/10.1186/s12864-017-3498-8>.
- [278] Luis Tari, Chitta Baral, Seungchan Kim. "Fuzzy c-means clustering with prior biological knowledge." In: *Journal of biomedical informatics* 42.1 (2009), pp. 74–81. URL: <https://doi.org/10.1016/j.jbi.2008.05.009>.
- [279] Matthias E. Futschik, Nikola K. Kasabov. "Fuzzy clustering of gene expression data." In: *2002 IEEE World Congress on Computational Intelligence. 2002 IEEE International Conference on Fuzzy Systems. FUZZ-IEEE'02. Proceedings (Cat. No. 02CH37291)*. Vol. 1. IEEE. 2002, pp. 414–419. URL: <https://doi.org/10.1109/FUZZ.2002.1005026>.
- [280] Zhiwen Yu, Hantao Chen, Jane You, Jiming Liu, Hau-San Wong, et al. "Adaptive fuzzy consensus clustering framework for clustering analysis of cancer data." In: *IEEE/ACM transactions on computational biology and bioinformatics* 12.4 (2014), pp. 887–901. URL: <https://doi.org/10.1109/TCBB.2014.2359433>.
- [281] Liang Chen, Weinan Wang, Yuyao Zhai, Minghua Deng. "Deep soft K-means clustering with self-training for single-cell RNA sequence data." In: *NAR genomics and bioinformatics* 2.2 (2020), lqaa039. URL: <https://doi.org/10.1093/nargab/lqaa039>.
- [282] Lihong Peng, Xiongfei Tian, Geng Tian, Junlin Xu, Xin Huang, et al. "Single-cell RNA-seq clustering: datasets, models, and algorithms." In: *RNA biology* 17.6 (2020), pp. 765–783. URL: <https://doi.org/10.1080/15476286.2020.1728961>.
- [283] Teny Handhayani, Lely Hiryanto. "Intelligent kernel k-means for clustering gene expression." In: *Procedia Computer Science* 59 (2015), pp. 171–177. URL: <https://doi.org/10.1016/j.procs.2015.07.544>.
- [284] Stephanie C. Hicks, Ruoxi Liu, Yuwei Ni, Elizabeth Purdom, Davide Risso. "mbkmeans: Fast clustering for single cell data using mini-batch

- k-means." In: *PLoS computational biology* 17.1 (2021), e1008625. URL: <https://doi.org/10.1371/journal.pcbi.1008625>.
- [285] Daniel Müllner. "Modern hierarchical, agglomerative clustering algorithms." In: *arXiv preprint arXiv:1109.2378* (2011). URL: <https://doi.org/10.48550/arXiv.1109.2378>.
- [286] Sławomir T. Wierzchoń, Mieczysław A. Kłopotek. "Algorithms of Combinatorial Cluster Analysis." In: *Modern Algorithms of Cluster Analysis*. Ed. by Janusz Kacprzyk. Springer International Publishing, 2018, pp. 67–161. ISBN: 978-3-319-69308-8. URL: https://doi.org/10.1007/978-3-319-69308-8_3.
- [287] N. R. Pal, J. C. Bezdek. "On cluster validity for the fuzzy c-means model." In: *IEEE Transactions on Fuzzy Systems* 3.3 (1995), pp. 370–379. URL: <https://doi.org/10.1109/91.413225>.
- [288] Peter J. A. Cock, Tiago Antao, Jeffrey T. Chang, Brad A. Chapman, Cymon J. Cox, et al. "Biopython: freely available Python tools for computational molecular biology and bioinformatics." In: *Bioinformatics* 25.11 (2009), pp. 1422–1423. URL: <https://doi.org/10.1093/bioinformatics/btp163>.
- [289] Charles R. Harris, K. Jarrod Millman, Stéfan J. Van Der Walt, Ralf Gommers, Pauli Virtanen, et al. "Array programming with NumPy." In: *Nature* 585.7825 (2020), pp. 357–362. URL: <https://doi.org/10.1038/s41586-020-2649-2>.
- [290] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python." In: *Nature methods* 17.3 (2020), pp. 261–272. URL: <https://doi.org/10.1038/s41592-019-0686-2>.
- [291] Tyson Barrett, Matt Dowle, Arun Srinivasan, Jan Gorecki, Michael Chirico, et al. *data.table: Extension of 'data.frame'*. CRAN package version 1.16.4. 2024. URL: <https://doi.org/10.32614/CRAN.package.data.table>.
- [292] Pierre Bushel. *Principal Variance Component Analysis (PVCA)*. R package version 1.46.0. 2024. URL: <https://doi.org/doi:10.18129/B9.bioc.pvca>.

- [293] Matthias E. Futschik, Bronwyn Carlisle. “Noise-robust soft clustering of gene expression time-course data.” In: *Journal of bioinformatics and computational biology* 3.04 (2005), pp. 965–988. URL: <https://doi.org/10.1142/S0219720005001375>.
- [294] Lokesh Kumar, Matthias E. Futschik. *Mfuzz: A software package for soft clustering of microarray data*. last visited on 2025-01-31. 2007. URL: <https://doi.org/10.6026/97320630002005>.
- [295] Wolfgang Huber, Vincent J. Carey, Robert Gentleman, Simon Anders, Marc Carlson, et al. “Orchestrating high-throughput genomic analysis with Bioconductor.” In: *Nature methods* 12.2 (2015), pp. 115–121. URL: <https://doi.org/10.1038/nmeth.3252>.
- [296] Björn Grüning, Ryan Dale, Andreas Sjödin, Brad A. Chapman, Jillian Rowe, et al. “Bioconda: sustainable and comprehensive software distribution for the life sciences.” In: *Nature methods* 15.7 (2018), pp. 475–476. URL: <https://doi.org/10.1038/s41592-018-0046-7>.
- [297] M. Baker. “1,500 scientists lift.” In: *Nature* 533 (2016), pp. 452–454. URL: <https://doi.org/10.1038/533452a>.
- [298] Ron Edgar, Michael Domrachev, Alex E. Lash. “Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.” In: *Nucleic Acids Research* 30.1 (2002), pp. 207–210. URL: <https://doi.org/10.1093/nar/30.1.207>.
- [299] Tanya Barrett, Stephen E. Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F. Kim, et al. “NCBI GEO: archive for functional genomics data sets—update.” In: *Nucleic Acids Research* 41.D1 (2012), pp. D991–D995. URL: <https://doi.org/10.1093/nar/gks1193>.
- [300] Carl Boettiger. “An introduction to Docker for reproducible research.” In: *SIGOPS Oper. Syst. Rev.* 49.1 (2015), pp. 71–79. URL: <https://doi.org/10.1145/2723872.2723882>.

- [301] Gregory M. Kurtzer, Vanessa Sochat, Michael W. Bauer. "Singularity: Scientific containers for mobility of compute." In: *PLOS ONE* 12.5 (2017), pp. 1–20. URL: <https://doi.org/10.1371/journal.pone.0177459>.
- [302] Johannes Köster, Sven Rahmann. "Snakemake—a scalable bioinformatics workflow engine." In: *Bioinformatics* 28.19 (2012), pp. 2520–2522. URL: <https://doi.org/10.1093/bioinformatics/bts480>.
- [303] Paolo Di Tommaso, Maria Chatzou, Evan W. Floden, Pablo Prieto Barja, Emilio Palumbo, et al. "Nextflow enables reproducible computational workflows." In: *Nature biotechnology* 35.4 (2017), pp. 316–319. URL: <https://doi.org/10.1038/nbt.3820>.
- [304] Enis Afgan, Dannon Baker, Bérénice Batut, Marius Van Den Beek, Dave Bouvier, et al. "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update." In: *Nucleic Acids Research* 46.W1 (2018), W537–W544. URL: <https://doi.org/10.1093/nar/gky379>.
- [305] Pascal Hirsch, Leidy-Alejandra G. Molano, Annika Engel, Jens Zentgraf, Sven Rahmann, et al. "Mibianto: ultra-efficient online microbiome analysis through k-mer based metagenomics." In: *Nucleic Acids Research* 52.W1 (2024), W407–W414. URL: <https://doi.org/10.1093/nar/gkae364>.
- [306] Yannan Fan, Jianguo Xia. "miRNet—functional analysis and visual exploration of miRNA–target interactions in a network context." In: *Computational cell biology: Methods and protocols* (2018), pp. 215–233. URL: https://doi.org/10.1007/978-1-4939-8618-7_10.
- [307] Valerio Licursi, Federica Conte, Giulia Fiscon, Paola Paci. "MIENTURNET: an interactive web tool for microRNA-target enrichment and network-based analysis." In: *BMC bioinformatics* 20 (2019), pp. 1–10. URL: <https://doi.org/10.1186/s12859-019-3105-x>.
- [308] Justina Žurauskienė, Christopher Yau. "pcaReduce: hierarchical clustering of single cell transcriptional profiles." In: *BMC bioinformatics* 17 (2016), pp. 1–11. URL: <https://doi.org/10.1186/s12859-016-0984-y>.

- [309] Quan Zou, Yaozong Mao, Lingling Hu, Yunfeng Wu, Zhiliang Ji. "miR-Classify: an advanced web server for miRNA family classification and annotation." In: *Computers in biology and medicine* 45 (2014), pp. 157–160. URL: <https://doi.org/10.1016/j.compbimed.2013.12.007>.
- [310] Eric Bonnet, Ying He, Kenny Billiau, Yves Van De Peer. "TAPIR, a web server for the prediction of plant microRNA targets, including target mimics." In: *Bioinformatics* 26.12 (2010), pp. 1566–1568. URL: <https://doi.org/10.1093/bioinformatics/btq233>.
- [311] William Rennie, Chaochun Liu, C. Steven Carmack, Adam Wolenc, Shaveta Kanoria, et al. "STarMir: a web server for prediction of microRNA binding sites." In: *Nucleic acids research* 42.W1 (2014), W114–W118. URL: <https://doi.org/10.1093/nar/gku376>.
- [312] Li Guo, Jiafeng Yu, Tingming Liang, Quan Zou. "miR-isomiRExp: a web-server for the analysis of expression of miRNA at the miRNA/isomiR levels." In: *Scientific reports* 6.1 (2016), p. 23700. URL: <https://doi.org/10.1038/srep23700>.
- [313] Christophe Tav, Sébastien Tempel, Laurent Poligny, Fariza Tahi. "miR-NAFold: a web server for fast miRNA precursor prediction in genomes." In: *Nucleic acids research* 44.W1 (2016), W181–W184. URL: <https://doi.org/10.1093/nar/gkw459>.
- [314] Jin-Wu Nam, Jinhan Kim, Sung-Kyu Kim, Byoung-Tak Zhang. "ProMiR II: a web server for the probabilistic prediction of clustered, nonclustered, conserved and nonconserved microRNAs." In: *Nucleic acids research* 34.suppl_2 (2006), W455–W458. URL: <https://doi.org/10.1093/nar/gkl321>.
- [315] Cedric Laczny, Petra Leidinger, Jan Haas, Nicole Ludwig, Christina Backes, et al. "miRTrail-a comprehensive webserver for analyzing gene and miRNA patterns to enhance the understanding of regulatory mechanisms in diseases." In: *BMC bioinformatics* 13 (2012), pp. 1–13. URL: <https://doi.org/10.1186/1471-2105-13-36>.

- [316] Fidel Ramírez, Devon P. Ryan, Björn Grüning, Vivek Bhardwaj, Fabian Kilpert, et al. “deepTools2: a next generation web server for deep-sequencing data analysis.” In: *Nucleic acids research* 44.Web Server issue (2016), W160. URL: <https://doi.org/10.1093/nar/gkw257>.
- [317] Daniel J. Rigden, Xosé M. Fernández. “The 2022 Nucleic Acids Research database issue and the online molecular biology database collection.” In: *Nucleic acids research* 50.D1 (2022), pp. D1–D10. URL: <https://doi.org/10.1093/nar/gkad1173>.
- [318] William Rennie, Shaveta Kanoria, Chaochun Liu, Bibekanand Mallick, Dang Long, et al. “STarMirDB: A database of microRNA binding sites.” In: *RNA biology* 13.6 (2016), pp. 554–560. URL: <https://doi.org/10.1080/15476286.2016.1182279>.
- [319] Tim Kehl, Fabian Kern, Christina Backes, Tobias Fehlmann, Daniel Stöckel, et al. “miRPathDB 2.0: a novel release of the miRNA Pathway Dictionary Database.” In: *Nucleic acids research* 48.D1 (2020), pp. D142–D147. URL: <https://doi.org/10.1093/nar/gkz1022>.
- [320] M. Maragkakis, M. Reczko, V. A. Simossis, P. Alexiou, G. L. Papadopoulos, et al. “DIANA-microT web server: elucidating microRNA functions through target prediction.” In: *Nucleic Acids Research* 37.2 (2009), W273–W276. URL: <https://doi.org/10.1093/nar/gkp292>.
- [321] Jüri Reimand, Meelis Kull, Hedi Peterson, Jaanus Hansen, Jaak Vilo. “g: Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments.” In: *Nucleic acids research* 35.2 (2007), W193–W200. URL: <https://doi.org/10.1093/nar/gkm226>.
- [322] Henrik Nielsen, Jacob Engelbrecht, Søren Brunak, Gunnar Von Heijne. “Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.” In: *Protein engineering* 10.1 (1997), pp. 1–6. URL: <https://doi.org/10.1093/protein/10.1.1>.
- [323] Christina Backes, Qurratulain T. Khaleeq, Eckart Meese, Andreas Keller. “miEAA: microRNA enrichment analysis and annotation.” In: *Nucleic acids*

- research* 44.W1 (2016), W110–W116. URL: <https://doi.org/10.1093/nar/gkw345>.
- [324] Maarten Hamberg, Christina Backes, Tobias Fehlmann, Martin Hart, Benjamin Meder, et al. “MiRTargetLink—miRNAs, genes and interaction networks.” In: *International journal of molecular sciences* 17.4 (2016), p. 564. URL: <https://doi.org/10.3390/ijms17040564>.
- [325] Spyros Tastsoglou, Athanasios Alexiou, Dimitra Karagkouni, Giorgos Skoufos, Elissavet Zacharopoulou, et al. “DIANA-microT 2023: including predicted targets of virally encoded miRNAs.” In: *Nucleic acids research* 51.W1 (2023), W148–W153. URL: <https://doi.org/10.1093/nar/gkad283>.
- [326] Liis Kolberg, Uku Raudvere, Ivan Kuzmin, Priit Adler, Jaak Vilo, et al. “g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update).” In: *Nucleic acids research* 51.W1 (2023), W207–W212. URL: <https://doi.org/10.1093/nar/gkad347>.
- [327] Henrik Nielsen, Felix Teufel, Søren Brunak, Gunnar von Heijne. “SignalP: the evolution of a web server.” In: *Protein Bioinformatics*. Ed. by Frédérique Lisacek. Springer, 2024, pp. 331–367. ISBN: 978-1-0716-4007-4. URL: https://doi.org/10.1007/978-1-0716-4007-4_17.
- [328] Ernesto Aparicio-Puerta, Pascal Hirsch, Georges P. Schartz, Fabian Kern, Tobias Fehlmann, et al. “miEAA 2023: updates, new functional microRNA sets and improved enrichment visualizations.” In: *Nucleic acids research* 51.W1 (2023), W319–W325. URL: <https://doi.org/10.1093/nar/gkad392>.
- [329] Fabian Kern, Ernesto Aparicio-Puerta, Yongping Li, Tobias Fehlmann, Tim Kehl, et al. “miRTargetLink 2.0—interactive miRNA target gene and target pathway networks.” In: *Nucleic Acids Research* 49.W1 (2021), W409–W416. URL: <https://doi.org/10.1093/nar/gkab297>.
- [330] Dominik Seelow. “The 22nd annual Nucleic Acids Research Web Server Issue 2024.” In: *Nucleic Acids Research* 52.W1 (2024), W1–W6. URL: <https://doi.org/10.1093/nar/gkae492>.

- [331] Fabian Kern, Tobias Fehlmann, Andreas Keller. "On the lifetime of bioinformatics web services." In: *Nucleic Acids Research* 48.22 (2020), pp. 12523–12533. URL: <https://doi.org/10.1093/nar/gkaa1125>.
- [332] Tobias Fehlmann, Fabian Kern, Pascal Hirsch, Robin Steinhaus, Dominik Seelow, et al. "Aviator: a web service for monitoring the availability of web services." In: *Nucleic Acids Research* 49.W1 (2021), W46–W51. URL: <https://doi.org/10.1093/nar/gkab396>.
- [333] Fatemeh Shaker, Abbas Nikraves, Roghaye Arezumand, Seyed Hamid Aghaee-Bakhtiari. "Web-based tools for miRNA studies analysis." In: *Computers in biology and medicine* 127 (2020), p. 104060. URL: <https://doi.org/10.1016/j.compbio.2020.104060>.
- [334] Jiung-Wen Chen, Lisa Shrestha, George Green, André Leier, Tatiana T. Marquez-Lago. "The hitchhikers' guide to RNA sequencing and functional analysis." In: *Briefings in Bioinformatics* 24.1 (2023), bbac529. URL: <https://doi.org/10.1093/bib/bbac529>.
- [335] Fabian Kern, Christina Backes, Pascal Hirsch, Tobias Fehlmann, Martin Hart, et al. "What's the target: understanding two decades of in silico microRNA-target prediction." In: *Briefings in bioinformatics* 21.6 (2020), pp. 1999–2010. URL: <https://doi.org/10.1093/bib/bbz111>.
- [336] Tomas Tokar, Chiara Pastrello, Andrea E. M. Rossos, Mark Abovsky, Anne-Christin Hauschild, et al. "mirDIP 4.1—integrative database of human microRNA target predictions." In: *Nucleic acids research* 46.D1 (2018), pp. D360–D370. URL: <https://doi.org/10.1093/nar/gkx1144>.
- [337] Benjamin P. Lewis, I-hung Shih, Matthew W. Jones-Rhoades, David P. Bartel, Christopher B. Burge. "Prediction of mammalian microRNA targets." In: *Cell* 115.7 (2003), pp. 787–798. URL: [https://doi.org/10.1016/S0092-8674\(03\)01018-3](https://doi.org/10.1016/S0092-8674(03)01018-3).
- [338] Vikram Agarwal, George W. Bell, Jin-Wu Nam, David P. Bartel. "Predicting effective microRNA target sites in mammalian mRNAs." In: *elife* 4 (2015), e05005. URL: <https://doi.org/10.7554/elife.05005>.

- [339] Marc R. Friedländer, Sebastian D. Mackowiak, Na Li, Wei Chen, Nikolaus Rajewsky. “miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades.” In: *Nucleic acids research* 40.1 (2012), pp. 37–52. URL: <https://doi.org/10.1093/nar/gkr688>.
- [340] Ernesto Aparicio-Puerta, Cristina Gómez-Martín, Stavros Giannoukakos, José María Medina, Chantal Scheepbouwer, et al. “sRNAbench and sRNA-toolbox 2022 update: accurate miRNA and sncRNA profiling for model and non-model organisms.” In: *Nucleic acids research* 50.W1 (2022), W710–W717. URL: <https://doi.org/10.1093/nar/gkac363>.
- [341] Van-Quy Do, Chuc Hoang-Thi, Thanh-Truong Pham, Nhat-Le Bui, Dinh-Thai Kim, et al. “Computational tools supporting known miRNA identification.” In: *Progress in Molecular Biology and Translational Science* 203 (2024), pp. 225–242. URL: <https://doi.org/10.1016/bs.pmbts.2023.12.018>.
- [342] Anna Lukasik, Maciej Wójcikowski, Piotr Zielenkiewicz. “Tools4miRs—one place to gather all the tools for miRNA analysis.” In: *Bioinformatics* 32.17 (2016), pp. 2722–2724. URL: <https://doi.org/10.1093/bioinformatics/btw189>.
- [343] Ernesto Aparicio-Puerta, Ricardo Lebrón, Antonio Rueda, Cristina Gómez-Martín, Stavros Giannoukakos, et al. “sRNAbench and sRNAtoolbox 2019: intuitive fast small RNA profiling and differential expression.” In: *Nucleic acids research* 47.W1 (2019), W530–W535. URL: <https://doi.org/10.1093/nar/gkz415>.
- [344] Yuanwei Zhang, Bo Xu, Yifan Yang, Rongjun Ban, Huan Zhang, et al. “CPSS: a computational platform for the analysis of small RNA deep sequencing data.” In: *Bioinformatics* 28.14 (2012), pp. 1925–1927. URL: <https://doi.org/10.1093/bioinformatics/bts282>.
- [345] Jinyu Wu, Qi Liu, Xin Wang, Jiayong Zheng, Tao Wang, et al. “mirTools 2.0 for non-coding RNA discovery, profiling, and functional annotation based on high-throughput sequencing.” In: *RNA biology* 10.7 (2013), pp. 1087–1092. URL: <https://doi.org/10.4161/rna.25193>.

- [346] Masood Abu-Halima, Ulrike Fischer, Mohammad A. Al Smadi, Nicole Ludwig, Anissa Acheli, et al. "Single Sperm RNA signatures reveal MicroRNA biomarkers for male subfertility." In: *Journal of Assisted Reproduction and Genetics* 41.11 (2024), pp. 3119–3132. URL: <https://doi.org/10.1007/s10815-024-03264-w>.
- [347] Masood Abu-Halima, Andreas Keller, Lea Simone Becker, Ulrike Fischer, Annika Engel, et al. "Dynamic and static circulating cancer microRNA biomarkers—a validation study." In: *RNA biology* 20.1 (2023), pp. 1–9. URL: <https://doi.org/10.1080/15476286.2022.2154470>.
- [348] Ernesto Aparicio-Puerta, Pascal Hirsch, Georges P. Schmartz, Tobias Fehlmann, Verena Keller, et al. "isomiRdb: microRNA expression at isoform resolution." In: *Nucleic Acids Research* 51.D1 (2022), pp. D179–D185. URL: <https://doi.org/10.1093/nar/gkac884>.
- [349] Efthymia Papalexi, Rahul Satija. "Single-cell RNA sequencing to explore immune cell heterogeneity." In: *Nature Reviews Immunology* 18.1 (2018), pp. 35–45. URL: <https://doi.org/10.1038/nri.2017.76>.
- [350] Wenwen Shao, Zhiheng Lin, Zhikai Xiahou, Fu Zhao, Jue Xu, et al. "Single-cell RNA sequencing reveals that MYBL2 in malignant epithelial cells is involved in the development and progression of ovarian cancer." In: *Frontiers in Immunology* 15 (2024), p. 1438198. URL: <https://doi.org/10.3389/fimmu.2024.1438198>.
- [351] Kai Zheng, Lingmin Lin, Wei Jiang, Lin Chen, Xiyue Zhang, et al. "Single-cell RNA-seq reveals the transcriptional landscape in ischemic stroke." In: *Journal of Cerebral Blood Flow & Metabolism* 42.1 (2022), pp. 56–73. URL: <https://doi.org/10.1177/0271678X211026770>.
- [352] Semra Smajić, Cesar A. Prada-Medina, Zied Landoulsi, Jenny Ghelfi, Sylvie Delcambre, et al. "Single-cell sequencing of human midbrain reveals glial activation and a Parkinson-specific neuronal state." In: *Brain* 145.3 (2022), pp. 964–978. URL: <https://doi.org/10.1093/brain/awab446>.

- [353] Kang Guo, Jianing Luo, Dayun Feng, Lin Wu, Xin Wang, et al. "Single-cell RNA sequencing with combined use of bulk RNA sequencing to reveal cell heterogeneity and molecular changes at acute stage of ischemic stroke in mouse cortex penumbra area." In: *Frontiers in cell and developmental biology* 9 (2021), p. 624711. URL: <https://doi.org/10.3389/fcell.2021.624711>.
- [354] Christoph Ziegenhain, Beate Vieth, Swati Parekh, Björn Reinius, Amy Guillaumet-Adkins, et al. "Comparative analysis of single-cell RNA sequencing methods." In: *Molecular cell* 65.4 (2017), pp. 631–643. URL: <https://doi.org/10.1016/j.molcel.2017.01.023>.
- [355] Simone Picelli. "Single-cell RNA-sequencing: the future of genome biology is now." In: *RNA biology* 14.5 (2017), pp. 637–650. URL: <https://doi.org/10.1080/15476286.2016.1201618>.
- [356] Dragomirka Jovic, Xue Liang, Hua Zeng, Lin Lin, Fengping Xu, et al. "Single-cell RNA sequencing technologies and applications: A brief overview." In: *Clinical and translational medicine* 12.3 (2022), e694. URL: <https://doi.org/10.1002/ctm2.694>.
- [357] Valentine Svensson, Roser Vento-Tormo, Sarah A. Teichmann. "Exponential scaling of single-cell RNA-seq in the past decade." In: *Nature protocols* 13.4 (2018), pp. 599–604. URL: <https://doi.org/10.1038/nprot.2017.149>.
- [358] Jiahui Ji, Maryam Anwar, Enrico Petretto, Costanza Emanuelli, Prashant Kumar Srivastava. "PPMS: A framework to Profile Primary MicroRNAs from Single-cell RNA-sequencing datasets." In: *Briefings in Bioinformatics* 23.6 (2022), bbac419. URL: <https://doi.org/10.1093/bib/bbac419>.
- [359] Alina Isakova, Norma Neff, Stephen R. Quake. "Single cell profiling of total RNA using Smart-seq-total." In: *BioRxiv* (2020), pp. 2020–06. URL: <https://doi.org/10.1101/2020.06.02.131060>.
- [360] Po-Yuan Tung, John D. Blischak, Chiaowen Joyce Hsiao, David A. Knowles, Jonathan E. Burnett, et al. "Batch effects and the effective design of single-cell gene expression studies." In: *Scientific reports* 7.1 (2017), p. 39921. URL: <https://doi.org/10.1038/srep39921>.

- [361] Maren Büttner, Zhichao Miao, F. Alexander Wolf, Sarah A. Teichmann, Fabian J. Theis. "A test metric for assessing single-cell RNA-seq batch correction." In: *Nature methods* 16.1 (2019), pp. 43–49. URL: <https://doi.org/10.1038/s41592-018-0254-1>.
- [362] Geng Chen, Baitang Ning, Tielu Shi. "Single-cell RNA-seq technologies and related computational data analysis." In: *Frontiers in genetics* 10 (2019), p. 317. URL: <https://doi.org/10.3389/fgene.2019.00317>.
- [363] David Lähnemann, Johannes Köster, Ewa Szczurek, Davis J. McCarthy, Stephanie C. Hicks, et al. "Eleven grand challenges in single-cell data science." In: *Genome biology* 21 (2020), pp. 1–35. URL: <https://doi.org/10.1186/s13059-020-1926-6>.
- [364] Carsten A. Raabe, Thean-Hock Tang, Juergen Brosius, Timofey S. Rozhdestvensky. "Biases in small RNA deep sequencing data." In: *Nucleic acids research* 42.3 (2014), pp. 1414–1426. URL: <https://doi.org/10.1093/nar/gkt1021>.
- [365] Huajuan Shi, Ying Zhou, Erteng Jia, Min Pan, Yunfei Bai, et al. "Bias in RNA-seq library preparation: current challenges and solutions." In: *BioMed research international* 2021.1 (2021), p. 6647597. URL: <https://doi.org/10.1155/2021/6647597>.
- [366] Morgane Meistertzheim, Tobias Fehlmann, Franziska Drews, Marcello Pirritano, Gilles Gasparoni, et al. "Comparative analysis of biochemical biases by ligation-and template-switch-Based small RNA library preparation protocols." In: *Clinical Chemistry* 65.12 (2019), pp. 1581–1591. URL: <https://doi.org/10.1373/clinchem.2019.305045>.
- [367] Taishan Hu, Nilesh Chitnis, Dimitri Monos, Anh Dinh. "Next-generation sequencing technologies: An overview." In: *Human Immunology* 82.11 (2021), pp. 801–811. URL: <https://doi.org/10.1016/j.humimm.2021.02.012>.
- [368] Simon Ardui, Adam Ameer, Joris R. Vermeesch, Matthew S. Hestand. "Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics." In: *Nucleic acids research* 46.5 (2018), pp. 2159–2168. URL: <https://doi.org/10.1093/nar/gky066>.

- [369] Corey J. Nolet, Victor Lafargue, Edward Raff, Thejaswi Nanditale, Tim Oates, et al. “Bringing UMAP closer to the speed of light with GPU acceleration.” In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 1. 2021, pp. 418–426. URL: <https://doi.org/10.1609/aaai.v35i1.16118>.
- [370] Shusruto Rishik, Pascal Hirsch, Friederike Grandke, Tobias Fehlmann, Andreas Keller. “miRNATissueAtlas 2025: an update to the uniformly processed and annotated human and mouse non-coding RNA tissue atlas.” In: *Nucleic Acids Research* 53.D1 (2024), pp. D129–D137. URL: <https://doi.org/10.1093/nar/gkae1036>.
- [371] Valeria Palmieri, Christina Backes, Nicole Ludwig, Tobias Fehlmann, Fabian Kern, et al. “IMOTA: an interactive multi-omics tissue atlas for the analysis of human miRNA-target interactions.” In: *Nucleic Acids Research* 46.D1 (2017), pp. D770–D775. URL: <https://doi.org/10.1093/nar/gkx701>.
- [372] Yingwen Chen, Xuanqun Wang, Xing Na, Yingkun Zhang, Zan Li, et al. “Highly Multiplexed, Efficient, and Automated Single-Cell MicroRNA Sequencing with Digital Microfluidics.” In: *Small Methods* 8.3 (2024), p. 2301250. URL: <https://doi.org/10.1002/smt.202301250>.
- [373] Hansruedi Mathys, Jose Davila-Velderrain, Zhuyu Peng, Fan Gao, Shahin Mohammadi, et al. “Single-cell transcriptomic analysis of Alzheimer’s disease.” In: *Nature* 570.7761 (2019), pp. 332–337. URL: <https://doi.org/10.1038/s41586-019-1195-2>.
- [374] Devika Agarwal, Cynthia Sandor, Viola Volpato, Tara M. Caffrey, Jimena Monzón-Sandoval, et al. “A single-cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological disorders.” In: *Nature communications* 11.1 (2020), p. 4183. URL: <https://doi.org/10.1038/s41467-020-17876-0>.
- [375] Elliott Swanson, Cara Lord, Julian Reading, Alexander T. Heubeck, Palak C. Genge, et al. “Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq.” In: *Elife* 10 (2021), e63632. URL: <https://doi.org/10.7554/eLife.63632>.

- [376] Marlon Stoeckius, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K. Chattopadhyay, et al. "Large-scale simultaneous measurement of epitopes and transcriptomes in single cells." In: *Nature methods* 14.9 (2017), p. 865. URL: <https://doi.org/10.1038/nmeth.4380>.
- [377] Song Chen, Blue B. Lake, Kun Zhang. "High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell." In: *Nature biotechnology* 37.12 (2019), pp. 1452–1457. URL: <https://doi.org/10.1038/s41587-019-0290-0>.

ACKNOWLEDGMENTS

Zuallererst möchte ich meinem Betreuer Andreas Keller danken, der an mich geglaubt und mir eine Chance gegeben hat, obwohl ich aus einem ganz anderen Fachgebiet komme. Ich danke ihm für seine Unterstützung, seine Anleitung sowie für die vielen wertvollen Kommentare und sein konstruktives Feedback.

Mein Dank gilt außerdem allen Kolleginnen und Kollegen am Lehrstuhl und am Zentrum für Bioinformatik in Saarbrücken sowie allen Kollaborationspartnern, mit denen ich während meiner Zeit in Saarbrücken zusammenarbeiten durfte, insbesondere den Kolleginnen und Kollegen aus Homburg, Regensburg und Stanford.

Ein besonderer Dank geht an Sabine Lessel und Anna Priester, die stets erreichbar waren und unermüdlich bei allen bürokratischen Herausforderungen geholfen haben.

Im Speziellen möchte ich Tobias Fehlmann für seine große Unterstützung in der Anfangszeit danken, er stand stets mit Rat und Tat an meiner Seite und ist noch heute für Fragen immer offen. Georges Schmartz schätze ich nicht nur für seinen Humor und seine Direktheit, sondern auch für seine unermüdliche Hilfsbereitschaft. Ernesto Aparicio-Puerta und Jérémy Amand waren wertvolle Gesprächspartner, von deren enormem Wissen und Ideen ich profitieren durfte und die zu guten Freunden geworden sind. Mit Viktoria Wagner und Shusru-to Rishik habe ich viel Zeit an wichtigen Projekten meiner Doktorandenzeit verbracht. Danke für die bereichernden Diskussionen und die großartige Zusammenarbeit. Ich danke Matthias Flotho und Leidy Alejandra Gonzalez Molano für die vielen Gespräche beim Mittagessen.

Ein besonderer Begleiter während dieser Zeit war Fabian Kern, den ich mit unzähligen Fragen löchern konnte und der mir immer mit Rat und Tat zur Seite stand. Danke für die vielen Gespräche und die großartige gemeinsame Zeit.

Ich danke auch Friederike Grandke und Pascal Hirsch. Ihr seid nicht nur Kollegen, sondern weit mehr! Stets wart ihr für mich da, im selben oder im Nachbarbüro oder nur eine Chatnachricht entfernt. Ohne euch hätte ich es nicht geschafft.

Meinen Freunden aus Schul- und Studienzeiten danke ich für ihre Unterstützung und ihre beständige Begleitung. Ein besonderer Dank gilt meinen Freunden in Saarbrücken, die mich daran erinnern haben, dass es neben der Arbeit auch ein Leben gibt, und mir viele entspannte Abende und Wochenenden beschert haben.

Zuletzt möchte ich meiner Familie danken, insbesondere meinen Eltern, meinen Großeltern und Suthep ♥. Ihr habt all die Jahre an mich geglaubt und mich bei jedem meiner Schritte unterstützt. Danke, dass es euch gibt.

Aus datenschutzrechtlichen Gründen wird der Lebenslauf in der elektronischen Fassung der Dissertation nicht veröffentlicht.

