

Aus der Klinik für Anästhesiologie
der Ludwig-Maximilians-Universität München
Direktor: Prof. Dr. med. B. Zwißler

**MiRNA Bioinformatik als Werkzeug medizinischer Forschung –
Etablierung eines Datenbank- und Zielvorhersagesystems zur
Bearbeitung von miRNA-bezogenen Fragestellungen.**

Dissertation
zum Erwerb des Doktorgrades der Medizin
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

Vorgelegt von
Ludwig Christian Giuseppe Hinske
aus Ahaus

2011

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter: Prof. Dr. B. Zwißler

Mitberichterstatter: Prof. Dr. A.-L. Boulesteix

Prof. Dr. D. Rujescu

Mitbetreuung durch den
promovierten Mitarbeiter: Dr. med. Dr. rer. nat. Simone Kreth

Dekan: Prof. Dr. med. Dr. h.c. M. Reiser, FACR, FRCR

Tag der mündlichen Prüfung: 03.03.2011

Inhaltsverzeichnis

1 EINLEITUNG	7
1.1 DAS GENOM	7
1.2 DAS TRANSKRIPTOM	8
1.3 MICRORNA ÜBERBLICK	8
1.4 MICRORNA BIOGENESE	10
1.5 MICRORNA WIRKMECHANISMEN	11
1.6 MICRORNA REGULATION	12
1.7 MICRORNA ZIEL-VORHERSAGE	13
1.7.1 TARGETSCAN	14
1.7.2 MIRANDA	15
1.7.3 RNA22	15
1.7.4 PITA	16
1.7.5 MIRTARGET2	16
1.7.6 PICTAR	17
1.7.7 TARBASE	17
1.8 BEDEUTUNG VON MIRNAS IN DER MEDIZIN	18
2 PROBLEMATIK UND ZIELSETZUNG	20
3 MATERIAL UND METHODEN	25
3.1 ÖFFENTLICH VERFÜGBARE DATENSÄTZE	25
3.2 DATENBANKEN UND MICRORNA KLASSIFIKATION	25
3.3 ZIELVORHERSAGE-METHODEN	26
3.4 DATENBANKDESIGN	26
3.5 ERMITTLUNG DES SIGNIFIKANZNIVEAUS VON MIRNAS, DIE IHR HOST-GEN KONTROLLIEREN	27
3.6 STATISTISCHE TESTS	28
4 ERGEBNISSE	29
4.1 ENTWICKLUNG EINES KONSENSUSVERFAHRENS ZUR MIRNA-ZIELVORHERSAGE	29
4.2 ANALYSE AN EINEM PRAKTISCHEN BEISPIEL: DIE ROLLE VON MIRNAS IN DER SEPSIS	33
4.3 CHARAKTERISTIKA INTRONISCHER MIRNAS	39
4.3.1 VERHÄLTNIS MIRNA – HOST-GEN	41
4.3.2 CHARAKTERISTIKA VON HOST- UND ZIEL-GENEN	43
5 DISKUSSION	46
5.1 ETABLIERUNG EINES KONSENS-VERFAHRENS FÜR 6 ZIEL-VORHERSAGE- METHODEN	47
5.2 SEPSIS UND MIRNA-BIOINFORMATIK	49
5.3 INTRONISCHE MIRNAS	53
5.4 LIMITATIONEN	55
5.5 ZUSAMMENFASSUNG	55
DANKSAGUNG	66

Abbildungsverzeichnis

ABBILDUNG 1. SCHEMATISCHE SKIZZE EINER BIOMEDIZINISCHEN miRNA-UNTERSUCHUNG	21
ABBILDUNG 2. AUC IN ABHÄNGIGKEIT VOM PREDICTION AGREEMENT	32
ABBILDUNG 3. VISUALISIERUNG DER INTERAKTIONEN PRO- UND ANTIAPOPTOTISCHER ZIEL- GENE.	51

Tabellenverzeichnis

TABELLE 1. VERGLEICH VERSCHIEDENER miRNA-ZIELVORHERSAGE-ALGORITHMEN AN PROTEIN-EXPRESSIONSÄNDERUNGEN	30
TABELLE 2. VERGLEICH VERSCHIEDENER miRNA-ZIELVORHERSAGE-ALGORITHMEN AN mRNA- EXPRESSIONSÄNDERUNGEN	31
TABELLE 3. LOGISTISCHES REGRESSIONSMODELL	33
TABELLE 4. KEGG-ANALYSE HOCHKONFIDENTER ZIELE VON HSA-miR-342 UND HSA-miR-150.	36
TABELLE 5. KEGG-ANALYSE HOCHKONFIDENTER ZIELE VON HSA-miR-486 UND HSA-miR-182.	38
TABELLE 6. VERTEILUNG INTERGENISCHER, INTRONISCHER UND EXONISCHER miRNAs IN VERSCHIEDENEN SPEZIES	40
TABELLE 7. TEST DER ROBUSTHEIT MITTELS WAHLVERFAHREN	42
TABELLE 8. KONSERVIERUNG INTRONISCHER miRNA - HOST-GEN PAARE.	43
TABELLE 9. CHARAKTERISTIKA DER HOST- UND ZIEL-GENE INTRONISCHER miRNAs	45

1 Einleitung

1.1 Das Genom

Mit der Entdeckung der Erbsubstanz durch den Chemiker Friedrich Miescher im Jahre 1871 wurde eine neue Ära in der Medizin eingeläutet. Seither folgten innerhalb weniger Jahrzehnte weitere Meilensteine in der Erforschung des menschlichen Erbgutes. 1953 beschrieben James Watson und Francis Crick erstmals die doppelhelikale Struktur des genetischen Trägermoleküls, der Desoxyribonukleinsäure (DNA, engl. Deoxyribose nucleic acid) [1].

Die Bausteine der DNA werden als Nukleotide bezeichnet, die aus je 3 Bestandteilen bestehen: Einer organischen Base, einem Zuckermolekül (Monosaccharid) und einer Phosphorsäure. Bei der Base kann es sich um Adenin (A), Guanin (G), Thymin (T) oder Cytosin (C) handeln. Aufgrund ihrer Base werden die Abkürzungen A, G, T und C auch für die entsprechenden Nukleotide verwendet.

Die biologischen Informationen sind als Gene, d.h. als Sequenzen dieser Nukleotide kodiert. Es handelt sich dabei um semantisch-funktionell zusammengehörige Bereiche, die strukturelle Gemeinsamkeiten aufweisen, wie beispielsweise regulatorische Sequenzen, kodierende (Exons) und nicht-kodierende (Introns) Bereiche. Dabei ist anzumerken, dass unter „nicht-kodierend“ die fehlende Umsetzung zum fertigen Protein zu verstehen ist und

nicht das Fehlen biologisch relevanter Information bedeutet. Die Summe aller Gene wird unter dem Begriff „Genom“ zusammengefasst. Seit April 2003 gilt die Sequenz des menschlichen Genoms offiziell als vollständig entschlüsselt [2].

1.2 Das Transkriptom

Die im Zellkern liegenden Genabschnitte werden von RNA-Polymerasen „abgelesen“, die aus der ursprünglichen DNA-Sequenz ein komplementäres Ribonukleinsäure (RNA) Molekül erzeugen. Dieser Vorgang, die „Transkription“, erzeugt das sogenannte „Transkriptom“, welches die Summe aller transkribierten RNA Moleküle in der Zelle ist.

Lange Zeit wurde postuliert, dass jedes Gen nach der Transkription in ein Protein translatiert wird, d.h. die Funktion der transkribierten RNA also einzig aus der Übersetzungsarbeit vom genetischen Bauplan des Erbguts (DNA) zum Eiweiss besteht. In den vergangenen zwei Jahrzehnten zeigte sich jedoch, dass es neben protein-kodierenden Genen auch solche gibt, die in nicht-kodierende, biologisch aktive RNA (engl.: non-coding ribonucleic acid, ncRNA) transkribiert werden. Nicht-kodierende RNA-Moleküle, wie z.B. microRNA (miRNA), transferRNA (tRNA) oder Ribozyme übernehmen dabei wichtige Funktionen in der Genregulation, Replikation, Translation und dem Abbau von Molekülen.

1.3 MicroRNA Überblick

1993 entdeckten Lee, Feinbaum und Ambros in den Zellen des Fadenwurms *Caenorhabditis elegans* ein ungewöhnlich kurzes, einstrangiges RNA-Molekül,

dessen Funktion in der Wissenschaft Aufsehen erregte und das Verständnis über die Regulation von Genen fundamental erweitern sollte. In dem Wurm wurde das Gen „*lin-4*“ nach der Transkription zwar in eine RNA, im weiteren Verlauf jedoch nicht in ein Eiweiß übersetzt; stattdessen griff es selbst regulierend in die Proteinsynthese ein. *Lin-4* lagerte sich an die messengerRNA (mRNA) des Gens *LIN-14* und blockierte so die weiteren Produktionsschritte, die normalerweise zum Eiweiß führen [3]. Mit dieser Erkenntnis wurde ein biomolekularer Mechanismus aufgedeckt, der das Verständnis über die Genregulation grundlegend veränderte: Diese erfolgt nicht, wie vorher angenommen, ausschliesslich über spezielle Proteine, sondern auch über biologisch aktive RNA (ncRNA, siehe oben).

Erst sieben Jahre später, im Jahr 2000, konnten Pasquinelli et al. ein weiteres nicht-kodierendes RNA-Molekül mit ähnlichen genregulierenden Eigenschaften („*let-7*“) in sämtlichen Spezies, erstmals auch im Menschen, nachweisen [4]. Dies deutete nun auf einen universellen Mechanismus hin und löste in der Wissenschaft eine intensive Suche nach weiteren microRNAs (miRNAs), wie sie aufgrund der geringen Molekülgröße von ca. zwanzig Basen nun genannt wurden, aus [5]. In den folgenden Jahren wurden viele weitere miRNA-Gene nachgewiesen [5-7]. Inzwischen sind in der zentralen Datenbank für miRNA „miRBase“ über 4000 Gensequenzen der unterschiedlichen Spezies registriert [8-10]. Derzeit wird postuliert, dass alleine im Menschen zwischen 500 und 1000 verschiedene miRNAs für die Genregulation von ca. 30 Prozent der Gene verantwortlich sind [11, 12].

Inzwischen weiß man, dass den miRNAs sehr wichtige Funktionen in grundlegenden zellulären Prozessen zukommen. Tierexperimente (u. a. am Fadenwurm, der Taufliege oder der Maus) zeigten, dass miRNAs beispielsweise an der Koordination von Entwicklungsvorgängen, der Zelldifferenzierung oder der Ausbildung des Immunsystems beteiligt sind. In der Medizin gewannen die genregulatorischen Fähigkeiten von miRNAs unter anderem durch deren Rolle in der Tumorgenese an Bedeutung (siehe Kapitel 1.8) [3, 13-16].

Nachdem die Eigenschaften der miRNA als zusätzliche Ebene der Genregulation erkannt und bereits zahlreiche miRNA-Gene entschlüsselt wurden, arbeitet die Wissenschaft derzeit daran, die Funktionsmechanismen der RNA-Moleküle genauer zu verstehen. Vieles über die Interaktion von miRNA zu ihren Zielgenen ist nach wie vor ungeklärt [17].

1.4 MicroRNA Biogenese

Die Gene der miRNAs können, im Gegensatz zum fertigen miRNA-Molekül, bis zu mehrere tausend Basenpaare (kb) lang sein. Nach der Transkription durch „Pol II-Polymerasen“ entsteht das primäre miRNA-Transkript (pri-miRNA) [18]. Da die beiden Enden des entstandenen RNA-Stranges mehr oder weniger komplementäre Basenabfolgen besitzen, lagern sich diese aneinander und bilden eine charakteristische haarnadelförmige Struktur (engl.: hairpin) [19].

Danach schneidet ein Enzym den Teil mit der Schleife ab. Es verbleibt ein Doppelstrang aus zwei komplementären RNA-Ketten, der ca. 70 Nukleotide lang ist und als Vorläufer-miRNA bezeichnet wird (engl.: precursor miRNA, pre-

miRNA). Die pre-miRNA enthält einen zwei Nukleotide langen Überhang am 5'-Ende, der zum Transport des Moleküls aus dem Zellkern (Nukleus) heraus dient [20-22]. Da miRNAs ihre Ziele durch komplementäre Basenpaarung erkennen, müssen sie als Einzelstrang vorliegen. Diese Trennung der beiden noch aneinander gelagerten RNA-Stränge erfolgt im Zytosol. Aus einem Strang entsteht die eigentliche miRNA, der andere wird abgebaut [23].

1.5 MicroRNA Wirkmechanismen

Ihre genregulierende Wirkung kann eine miRNA erst nach Integration in eine aus mehreren Proteinen bestehende Struktur, dem „RNA-induced silencing complex“ (Risc), entfalten. Der Komplex aus Risc und miRNA bindet an die gesuchte mRNA, die eine zur miRNA komplementäre Basensequenz in der 3'-UTR-Region besitzt [24-26]. Ist die Ziel-mRNA gefunden, wird ihre weitere Verarbeitung durch den miRNA-Risc-Komplex blockiert, oder sie wird sogar degradiert. Wann welcher Mechanismus greift, hängt vermutlich vom Grad der Basenpaarung zwischen der mRNA und der miRNA ab, ist aber noch nicht verstanden. Es scheint, dass eine gute Bindung durch Komplementarität besonders im Bereich der Nukleotide 2-8 (gezählt vom 5' Ende der miRNA) zur 3'-UTR-Sequenz der mRNA erforderlich und die Thermodynamik der RNA-RNA-Interaktion entsprechend günstig ist [27].

Wie genau miRNAs die Protein-Spiegel der Ziel-Gene regulieren, ist ebenfalls noch nicht eindeutig geklärt. Bislang gibt es Hinweise auf mehrere verschiedene Mechanismen, die von Kodegradation der entstehenden Proteinkette parallel zur

Translation [13] über verfrühte Polysomen-Dissoziation [28] bis zu kompetitiver Hemmung von translationsinitiierenden Faktoren reichen [29].

Die kurze Basenabfolge einer miRNA und ihrer Zielsequenz ist potentiell zu vielen mRNA-Sequenzen komplementär. Daher kann eine einzelne miRNA die Aktivität vieler Gene steuern. Umgekehrt kann die Translation einer mRNA auch gleichzeitig von verschiedenen miRNAs blockiert werden. Dies ergibt ein komplexes und differenziertes Regulationsnetzwerk, das zeit- und zellspezifisch arbeitet und dessen Strukturen und Interaktionen allenfalls im Ansatz verstanden sind.

1.6 MicroRNA Regulation

Nachdem zunächst die Suche nach neuen miRNAs sowie die Vorhersage und Validierung deren Ziel-mRNA im wissenschaftlichen Fokus stand, wird zunehmend die Frage nach regulatorischen Prinzipien, die die miRNA-Transkription an sich beeinflussen, wichtiger Bestandteil der Forschung.

Es können derzeit drei grundlegende Prinzipien unterschieden werden:

- (i) Regulation durch koordinierte Expression
- (ii) Regulation durch eigenständige cis-regulatorische Elemente
- (iii) Post-transkriptionelle Regulation.

MiRNAs können in intragen und intergen unterteilt werden. Intergenische miRNAs befinden sich außerhalb der Bereiche gegenwärtig annotierter Genkomplexe. Intragenische miRNA-Gene liegen in den Introns und, selten, in

den Exons sogenannter Wirtsgene. Es konnte gezeigt werden, dass das Wirtsgen inklusive miRNA transkribiert, und die miRNA noch vor dem Splicing-Prozess herausgeschnitten wird [30]. Daraus folgt eine koordinierte Expression von Wirtsgen-mRNA und miRNA. Ähnlich funktioniert das Prinzip der Polycistrone: Es sind miRNA-Gene beschrieben worden, die initial gemeinsam transkribiert werden, bevor sie anschließend als Einzeltranskripte den oben beschriebenen Reifungsprozessen unterliegen [31].

Einige miRNA-Gene scheinen allerdings, wie viele protein-kodierenden Gene auch, eigene cis-regulatorische Elemente zu besitzen. Das bedeutet, dass ihre Transkription zum Beispiel über Transkriptionsfaktoren reguliert werden können. Das Prinzip der post-transkriptionellen Regulation wurde unlängst postuliert. Durch Modifikation der Enzyme der miRNA-Biogenese könnte die miRNA-Expression beeinflusst werden, allerdings sind die Details dieses Systems unverstanden [32, 33].

1.7 MicroRNA Ziel-Vorhersage

Wie bereits beschrieben kann eine miRNA viele Ziel-Gene regulieren und ein Ziel-Gen von mehreren miRNAs kontrolliert werden [34]. Aufgrund der Vielzahl dieser Interaktionen ist ein klassischer, laborexperimenteller Ansatz zur Analyse der Funktion einer miRNA kaum durchführbar. Bis heute gibt es allerdings keine molekularen Hochdurchsatzverfahren, um die Ziel-Transkripte der miRNAs schnell und in großer Zahl zu bestimmen oder aufgestellte Hypothesen zu validieren. Aus diesem Grund sind verschiedene rechnergestützte Methoden entwickelt worden, um Interaktionen zwischen miRNA und mRNA vorherzusagen.

Sechs der bekanntesten Methoden, TargetScan, miRanda, PITA, RNA22, MirTarget2, PicTar werden im Folgenden vorgestellt. Biochemisch validierte miRNA-mRNA-Interaktionen werden in der Datenbank TarBase gesammelt.

1.7.1 TargetScan

Lewis et al. [35] entwickelten bereits im Jahr 2003 einen Algorithmus namens TargetScan. Zentrale Elemente dieses Algorithmus sind die Berechnung der sekundären RNA-Struktur sowie evolutionärer Konservierung bestimmter Sequenzbereiche von 3'-UTR Motiven. Potentielle Ziele werden durch Watson-Crick-Komplementarität der 5'- Kernsequenz der miRNA und der 3'-UTR Zielsequenz des jeweiligen Transkripts identifiziert. Im nächsten Schritt werden die umliegenden Basenpaare mit dem „RNAfold“-Algorithmus angeglichen und die freien Bindungsenergien berechnet [36]. Die Energiewerte sämtlicher Ziele werden z-transformiert und anschließend sortiert. Der TargetScan-Algorithmus nutzt weiterhin die Annahme, dass funktionell wichtige Bereiche im Laufe der Evolution „konserviert“ werden, also einer geringeren Mutationsrate als erwartet unterliegen. Dieses Phänomen wird auch als „cross-species conservation“ bezeichnet. Es ist also zu erwarten, dass für die Genexpression nicht wichtige Sequenzbereiche im 3'-UTR einer normalen Mutationsrate unterliegen, während Zielsequenzen für miRNA-Bindung aber größtenteils konserviert werden. Beispielsweise konnten die Autoren zeigen, dass das Signal-zu-Rausch-Verhältnis von 2:1 auf 4,6:1 steigt, wenn zusätzlich zur Sequenzkonservierung zwischen Maus und Mensch auch die Sequenzen von Ratte und Pufferfisch mit eingebunden werden. Allerdings geht dieser Gewinn an Spezifität mit einer Reduktion von Sensitivität einher. Die geschätzte Falsch-Positiv-Rate bewegt sich

zwischen 22% und 31%, abhängig von der Spezies und den Parametern. Interessanterweise ist TargetScan nach wie vor der Goldstandard in vielen Experimenten, obwohl es einer der ältesten bisher publizierten Algorithmen ist.

1.7.2 miRanda

Ein weiterer populärer Algorithmus, der im Jahr 2003 entwickelt wurde, ist miRanda. Ursprünglich entwickelt von Enright et al. [37], adaptierten John und Kollegen miRanda für die Anwendung am menschlichen Genom [38]. Der Algorithmus nutzt dieselben Prinzipien wie TargetScan, allerdings erfolgt die Punkteberechnung unterschiedlich. Die von den Autoren geschätzte Falsch-Positiv-Rate liegt zwischen 24% und 39%.

1.7.3 RNA22

RNA22 ist konzeptionell sehr unterschiedlich zu den bisher besprochenen Ansätzen. Miranda et al. [39] benutzen den „TEIRESIAS variable length motif finding“-Algorithmus [40], um aus reifen miRNA-Sequenzen Motive zu extrahieren. Die statistische Signifikanz eines jeden Motivs wird durch ein Second-order Markov Modell geprüft. Der Grundgedanke ist der, dass nach dem „guilty-by-association“-Prinzip [41] für jede putative Zielsequenz ein Mitgliedschaftsgrad errechnet werden kann. Die Autoren schätzen die Falsch-Positiv-Rate zwischen 19% and 26% und die Sensitivität auf zwischen 36% bis zu 95%, abhängig vom Trainingsdatensatz. Die große Stärke dieses Ansatzes ist, dass zuerst ein potentielles Ziel identifiziert wird und erst im Anschluss die dazu gehörige miRNA ermittelt wird. Theoretisch ist es also möglich, dass miRNA

Ziele entdeckt werden, deren zugehörige miRNA noch gar nicht beschrieben wurde.

1.7.4 PITA

PITA ist ein neuerer Algorithmus, der 2007 von Kertesz und dessen Mitarbeitern publiziert wurde [27]. Das zentrale Prinzip ist auch hier die sekundäre RNA-Struktur, allerdings im Rahmen eines „Zugänglichkeits-Punktesystems“. Zugänglichkeit bedeutet, dass zusätzlich zur Bindungsenergie einer miRNA-mRNA Interaktion auch die Energie einberechnet wird, die zur Trennung von komplementären mRNA Basenpaaren notwendig ist. Durch Reporter-Gen Assay Experimente konnten die Autoren weiterhin zeigen, dass die mRNA-Trennungsenergie 3 Nukleotide aufwärts und 15 Nukleotide abwärts der potentiellen miRNA Zielsequenz mit der translationalen Repression korreliert. Im Gegensatz zu TargetScan oder miRanda basiert dieser Ziel-Vorhersage-Algorithmus nicht auf Sequenz-Konservierung.

1.7.5 MirTarget2

Linsley et al. [42] haben systematisch Änderungen der mRNA Expression nach Überexpression verschiedener miRNAs gemessen. MirTarget2 [43, 44] nutzt ein Support-Vector-Machine (SVM) basiertes Modell, das auf dem Boden verschiedener Sequenz-Merkmale reduzierte mRNA-Expression vorhersagt. Die Area-Under-Curve (AUC) lag bei 0,79 (10-fach Cross-Validierung). Das prädiktive Potential schätzen die Autoren als vergleichbar mit TargetScan ein. Während die Stärke dieses Ansatzes die direkte Nutzung biologischer

Observation ist, darf nicht vergessen werden, dass eine Änderung von mRNA-Expression auch durch indirekte Effekte erzielt werden kann, wie beispielsweise durch Regulierung eines mit dem putativen Ziel assoziierten Transkriptionsfaktors.

1.7.6 PicTar

Krek et al. veröffentlichten 2005 PicTar [45]. Auch wenn bekannte Elemente wie freie Bindungsenergie und Sequenzkonservierung benutzt werden, wird die Spezifität erhöht durch die Annahme, dass koregulierte miRNAs ähnliche Zieltranskripte haben. Daher benutzen sie einen probabilistischen Algorithmus, der bereits erfolgreich zur Identifikation von Transkriptionsfaktor-bindenden Sequenzen (engl.: transcription factor binding sites, TFBS) eingesetzt worden ist. Die zwei verfügbaren Versionen von PicTar unterscheiden sich im Wesentlichen durch die Anzahl der Spezies zur Berechnung der Sequenzkonservierung.

1.7.7 TarBase

Anders als die bereits beschriebenen Vorhersage-Algorithmen ist TarBase eine Datenbank für experimentell validierte miRNA-mRNA-Interaktionen für verschiedene Organismen [46]. Jeder Eintrag enthält Informationen über die miRNA, die Ziel-mRNA, Information, ob translationale Repression oder Degradation beobachtet wurde, sowie die Referenz zur Original-Publikation und die Art des durchgeführten Experiments. Leider ist die Anzahl der Einträge vergleichsweise gering.

1.8 Bedeutung von miRNAs in der Medizin

Aufgrund ihrer genregulatorischen Eigenschaften und ihrer Funktion als Modulatoren in der Proteinsynthese verwundert es nicht, dass miRNAs eine zentrale Rolle bei vielfältigen biologischen Prozessen spielen. Schätzungen gehen davon aus, dass mehr als 30% des menschlichen Genoms von miRNAs reguliert werden könnten [11, 12, 34, 47]. Man erhofft sich durch zunehmende Erkenntnisse der miRNA-Wirkmechanismen neue diagnostische und therapeutische Ansätze für Krankheiten zu entwickeln, die bislang als schwer heilbar oder sogar als palliativ betrachtet werden müssen [48-50]. Ein zentraler Bereich der miRNA-Forschung ist die Tumorphathogenese. Mehrere Studien konnten miRNA-Gensignaturen extrahieren, die mit dem Vorliegen einer malignen Transformation des Gewebes, oder sogar mit der Prognose korrelieren [51, 52]. Schetter et al. beispielsweise zeigten, dass das Expressionsmuster von 37 miRNAs beim Colon-Karzinom nicht nur maligne veränderte Mukosa sicher identifiziert sondern auch mit der 5-Jahres-Überlebensrate korreliert. Maligne Veränderungen des Immunsystems konnten ebenfalls mehrfach mit verändertem miRNA-Expressionsmuster assoziiert werden [14, 47, 53, 54]. Ein weiterer wichtiger Bereich ist die Funktion viraler miRNAs, die die Genexpression des Wirts zu Vorteil des Virus zu beeinflussen scheinen, wie beim Herpes-, HI- und Epstein-Barr-Virus beschrieben [55]. Seit kurzem wird die Rolle von miRNAs bei Remodeling-Prozessen des Herzens und als Biomarker zur Myokardinfarkt Diagnostik diskutiert [56, 57].

Obwohl es Hinweise gibt, dass miRNAs für die korrekte Funktion und Entwicklung des Immunsystems eine tragende Rolle spielen könnten, ist ihre

Funktion in der Pathogenese von Erkrankungen, die mit einer Dysregulation des Immunsystems einhergehen, wie beispielsweise der Sepsis, weitestgehend unbekannt. Die Sepsis stellt eine lebensbedrohliche Erkrankung dar und führt trotz aufwendigster Behandlungsstrategien in einem hohen Prozentsatz zum Tode. Die grundlegende Problematik des Krankheitsbildes besteht in einer äußerst komplexen, durch die Invasion pathogener Keime ausgelösten systemischen Immunreaktion. Diese ist grundsätzlich zur Keimabwehr notwendig, jedoch führt bei der Sepsis eine unkontrollierte, überschießende Aktivierung von Immunzellen zu einer Schädigung von körpereigenen Geweben bis hin zum Organversagen [58]. Trotz intensiver Forschungsbemühungen ist die Pathophysiologie der Sepsis nur wenig verstanden; erkennbar ist, dass es sich um ein heterogenes dynamisches Syndrom handelt, das durch ein Ungleichgewicht im „inflammatorischen Netzwerk“ verursacht wird und an dem unterschiedliche biologische Systeme und Zelltypen beteiligt sind [59].

Eine erste Arbeit, die sich mit dem Thema miRNA-Regulation bei septischen Patienten beschäftigt, wurde im Juli 2009 von Vasilescu et al. publiziert. Laut dieser Arbeit gibt es mindestens vier differentiell exprimierte miRNAs in Leukozyten septischer Patienten (verglichen mit nicht-septischen Patienten). Die Autoren haben selektiv drei Ziele von hsa-miR-150 herausgegriffen und eine inverse Korrelation zwischen dem Plasmaspiegel dieser miRNAs und den drei Sepsis relevanten Markern Interleukin 18, Interleukin 10 und TNF-alpha beschrieben. Weitere Analysen, die die pathogenetische Bedeutung dieser Erkenntnis beleuchten könnten, wurden nicht durchgeführt [60].

2 Problematik und Zielsetzung

Seit ihrer Entdeckung hat die klinische Bedeutung von miRNAs stetig zugenommen. MiRNAs sind kleine, ca. 21 Nukleotide lange, nicht-kodierende Einzelstrang-RNA-Moleküle, die durch Watson-Crick-Komplementarität zu den 3'-UTR ihrer Ziel-mRNA zur Reduktion der Expression führen, sowohl durch translationale Repression als auch beschleunigte mRNA-Degradation. Aufgrund der wachsenden Erkenntnisse über die Rolle von miRNAs in pathogenetischen Prozessen wie beispielsweise der Tumorentwicklung [61], kardiovaskulären Erkrankungen [56] und Erkrankungen des Immunsystems [14, 62], hat die zielgerichtete Analyse von miRNAs entscheidende Bedeutung erlangt.

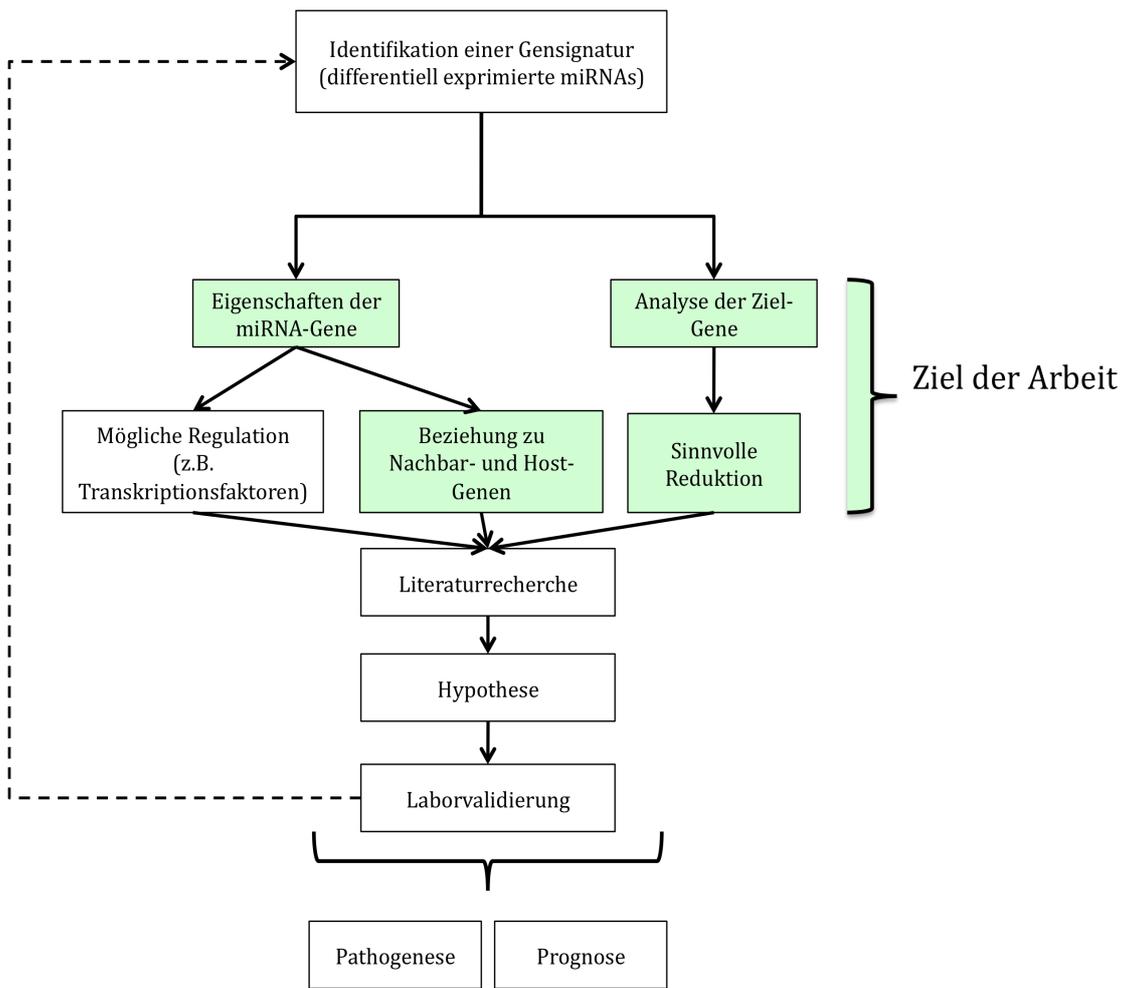


Abbildung 1. Schematische Skizze einer biomedizinischen miRNA-Untersuchung

Abbildung 1 stellt den Ablauf einer Untersuchung der Relevanz von Genregulation durch miRNAs schematisch dar. Diese beginnt in vielen Fällen damit, miRNAs zu identifizieren, deren Expression charakteristisch für das Zielgewebe ist (siehe beispielsweise [48, 52, 60, 63]). Diese miRNA-Genexpressionssignatur wird anschließend auf ihre diagnostische und prognostische Relevanz hin untersucht. Solche Signaturen sind prinzipiell geeignet, um Marker zu etablieren oder prognostische Aussagen zu treffen [52, 57]. Allerdings können sie auch als Grundlage pathogenetischer Forschung dienen. Es ist naheliegend, dass diejenigen miRNAs, deren Expressionsänderung

pathognomisch für einen Krankheitszustand sind, ebenfalls relevant für die Entstehung der Erkrankung sein könnten („Guilty-by-Association“-Prinzip [41]). Im nächsten Schritt müssen also funktionelle Aspekte der entsprechenden miRNAs durch Untersuchung der miRNA-Gene selbst und durch Untersuchung ihrer Ziele beleuchtet werden. Literaturrecherche der vorhergesagten Ziele und beobachteten miRNA-Gen-Eigenschaften können dann zur Generierung einer Hypothese führen oder eine bereits existierende Hypothese stützen oder schwächen. Eine letztendliche Laborvalidierung ist im besten Fall in der Lage, die beobachtete Gensignatur zu erklären.

Bei diesem Prozess gibt es zwei kritische Punkte:

- (i) Informationsverfügbarkeit
- (ii) Informationsvielfalt

(i) Informationsverfügbarkeit.

Viele Informationen über miRNA-Gene sind nicht direkt, nicht im Zusammenhang oder lediglich auf Einzelfallbasis verfügbar. Beispielsweise erlaubt mirBase [8, 64] als miRNA-Referenz-Datenbank zwar die Anzeige der Sequenz und Position des miRNA-Gens und, seit kurzem, auch die Einteilung in intragenisch und intergenisch, allerdings besteht nicht die Möglichkeit, Charakteristika einer größeren Gruppe von miRNAs oder assoziierten Genen zu untersuchen. Beispielsweise ist es über die Online-Präsenz nicht möglich, eine Liste aller intragenisch gelegenen miRNAs und deren Intron-/Exonpositionen und Hostgen-Größe zu erhalten. Untersuchungen auf Ebene des gesamten Genoms sind daher stark eingeschränkt.

(ii) Informationsvielfalt.

Während neue Technologien wie zum Beispiel Illumina/Solexa RNA-Sequencing [65-67] die Zahl an bekannten miRNAs vervielfachen, gibt es bislang noch keine Methoden zur Hochdurchsatzvalidierung von miRNA-mRNA-Interaktionen, die letztlich die biologische Relevanz einer miRNA determinieren. Das daraus resultierende Ungleichgewicht zwischen bekannten miRNAs und mangelnder Kenntnis um deren biologische Bedeutung stellt die Forschungsgemeinde vor komplexe Aufgaben. Verschiedene biologische Beobachtungen und Hypothesen um den miRNA-Zielerkennungs-Mechanismus haben zur Entwicklung sehr unterschiedlicher Zielvorhersage-Algorithmen geführt, die häufig sehr sensitiv sind, jedoch auf Kosten einer hohen Falsch-Positiv-Rate [68]. Beispielsweise listet TargetScan für die vier in der Sepsis potentiell relevanten miRNAs hsa-miR-150, hsa-miR-342, hsa-miR-182 und hsa-miR-486 insgesamt 13953 potentielle Ziel-Gene (ca. 50% der insgesamt bekannten protein-kodierenden Gene) [60]. Eine klassische, laborchemische Untersuchung der Gesamtheit dieser Vorhersagen ist ein langwieriger, kostspieliger und ebenso ineffizienter Prozess. Auch das Durchsuchen von Literaturdatenbanken, um eine entsprechende Vorauswahl zu treffen, ist für mehr als 1000 Gene sehr anspruchsvoll. Die Anzahl von Zielen muss daher sinnvoll eingegrenzt werden. Ein möglicher Ansatz wird von Sethupathy et al. beschrieben, die die Schnittmenge der Zielvorhersagen von PicTar und TargetScan als guten Kompromiss zwischen Sensitivität und Spezifität erachten [69]. Dieser Ansatz ist in der Datenbank miRGen verwirklicht [70, 71]. Allerdings ist so keine Aussage über die Konfidenz einer jeweiligen

Zielvorhersage und somit keine Skalierbarkeit möglich. Ein weiteres Problem entsteht dadurch, dass jede Ziel-Vorhersagemethode lediglich auf einigen wenigen biologischen Hypothesen beruht. Dies birgt die Gefahr der Einführung eines Bias. Ist beispielsweise ein grundlegendes Prinzip einer Methode die Konservierung der Zielsequenzen im 3'-UTR der potentiellen Zielgene, wird die Auswahl an vorhergesagten Zielen an evolutionär älteren Genen angereichert sein. Somit ist die Entscheidung für oder gegen eine Zielvorhersage-Methode eine das Ergebnis unmittelbar bestimmende Entscheidung.

Ziel der vorliegenden Untersuchung war es daher, eine Datenbank zu erstellen, die aktuelle Informationen über miRNAs, das menschliche Genom, beschriebene Gentranskripte und miRNA-mRNA-Interaktionen enthält. Basierend auf diesen Daten sollte ein Konsensus-Verfahren zur miRNA-mRNA-Interaktionsvorhersage entwickelt und an einem öffentlich verfügbaren Datensatz validiert werden. Dieses System sollte die Untersuchung von miRNAs im Kontext klinischer Fragestellung ermöglichen, indem eine sinnvolle Reduktion der vorhergesagten miRNA-mRNA-Interaktionen auf eine experimentell handhabbare Anzahl erreicht wird.

Die Möglichkeiten des Systems wurden exemplarisch an einem praktischen, für die Intensivmedizin relevanten Ansatz untersucht.

3 Material und Methoden

3.1 Öffentlich verfügbare Datensätze

Um Sensitivität, Spezifität und AUC der verschiedenen Ziel-Vorhersage-Methoden sowie der Anzahl übereinstimmender Methoden („Prediction Agreement“ = PA) abzuschätzen, wurde der öffentlich verfügbare Datensatz von Selbach et al. [72] benutzt (<http://psilac.mdc-berlin.de>). Die Autoren bestimmten die relative Änderung der Proteinexpression nach Überexpression von hsa-miR-1, hsa-miR-16, hsa-miR-30a, hsa-miR-155 und hsa-let-7b. Auch Änderungen der mRNA-Expression wurden erfasst.

3.2 Datenbanken und microRNA Klassifikation

Grundlage der genomischen Koordinaten und Sequenzinformation bildet der hg18-Build der University of California, Santa Cruz (<http://genome.ucsc.edu>) [73, 74]. Informationen über genetische Koordinaten und Konservierung von miRNA-Genen in unterschiedlichen Organismen wurden der miRNA-Datenbank mirBase Release 11 (April 2008) entnommen [8, 64]. Die Koordinaten des Transkriptionsstarts und -stopps, Exons und Introns der „RNA Reference Sequences“-Datenbank (<http://www.ncbi.nlm.nih.gov/RefSeq>; Release 31) wurden verwendet [75]. Homologe, protein-kodierender Gene wurden mittels der HomoloGene-Datenbank NCBI Release 61 ermittelt (<http://www.ncbi.nlm.nih.gov/homologene>). Intragenische miRNAs sind definiert als die Subgruppe der miRNAs, deren genomische Position mit der eines protein-kodierenden Transkripts überlappen; gibt es Überschneidungen mit als Exon klassifizierten Regionen, wurden die miRNAs als „exonisch“ klassifiziert,

andernfalls als „intronisch“. Die genomische Distanz zweier miRNAs wurde nur errechnet, wenn beide miRNA-Gene auf dem selben Chromosom und DNA-Strang liegen.

3.3 Zielvorhersage-Methoden

Zielvorhersagen der Methoden TargetScan Release 4.2 [35] (April 2008), PITA catalog version 6 [27] (August 2008), MirTarget2 (mirDB) Version 2.0 [43, 44] (Dezember 2007), miRanda [38] (September 2008), RNA22 [39] (November 2006) und PicTar 5-way [45] wurden verwendet. Validierte miRNA-Ziel-Interaktionen wurden aus TarBase Version 5.0c [46] (Juni 2008) integriert. Es wurden Datensätze mit dem Evidenztyp „True“ oder „Microarray“ importiert.

3.4 Datenbankdesign

Zur Konstruktion der Datenbank wurde das frei verfügbare Datenbanksystem „MySQL“ (<http://www.mysql.de>) benutzt. Für jeden Organismus wurde eine eigene Datenbank mit gleichem Aufbau erstellt. Die erste Tabelle enthält generelle Gen-Informationen, wie das offizielle Gensymbol, die zugehörige Entrez-ID-Nummer, ENSEMBL-ID-Nummer und Beschreibung. In einer zweiten Tabelle sind synonyme Gensymbole mit dem entsprechenden Eintrag des offiziellen Gensymbols verlinkt. Anschließend wurden allgemeine Informationen der RefSeq-Datenbank mit den offiziellen Gensymbolen verlinkt, sowie Exonpositionskoordinaten mit den RefSeq-Genen. Introns wurden identifiziert als nicht exon-annotierte Bereiche eines Gens und konsekutiv in 5'->3'-Richtung durchnummeriert (beginnend mit 1). Die Tabelle „mirBase“ enthält genomische

Informationen sämtlicher bekannter miRNA-Gene und deren offizielle Bezeichnung. Die genomische Position wurde verglichen mit Positionskordinaten der RefSeq-Tabellen und die entsprechende miRNA als intragenisch oder intergenisch klassifiziert. Intragenische miRNAs wurden subklassifiziert in intronisch und exonisch. Für die Host-Gen Analyse wurden lediglich die miRNAs berücksichtigt, deren Gen auf dem gleichen Strang wie das Host-Gen liegt. Die Zielvorhersagen der entsprechenden Methoden wurden in jeweils eigene Tabellen geladen. Jeder Eintrag ist verlinkt mit der miRNA-ID, dem offiziellen Gensymbol, dem Score und der Anzahl an vorhergesagten Bindungsstellen. Der Zugriff auf die Datenbank erfolgte durch die Programmiersprache Python 2.6.

3.5 Ermittlung des Signifikanzniveaus von miRNAs, die ihr Host-Gen kontrollieren

Die statistische Signifikanz der Anzahl an beobachteten miRNAs, die potentiell ihr Host-Gen regulieren, wurde folgendermaßen berechnet. Die Indikatorfunktion $I(C_i', C)$ wurde definiert als

$$I(C_i', C) \begin{cases} 1 & \text{wenn } C_i' \geq C \\ 0 & \text{andernfalls} \end{cases}, \text{ wobei } C \text{ die dem Modell entsprechend beobachtete}$$

Anzahl an miRNAs ist, die ihr eigenes Host-Gen kontrollieren und C_i' die jeweils beobachtete Anzahl nach zufälliger Permutation der Host-Gene durch potentielle Zielgene. Daher kann die Wahrscheinlichkeit, dass eine zufällig beobachtete Zuordnung von miRNAs zu Zielgenen eine gleich große oder höhere Zahl an

miRNAs gibt, die ihr Host-Gen regulieren könnten folgendermaßen geschätzt werden:

$$p(C' \geq C) = \frac{\sum_{i=1}^{1000} I(C_i', C)}{n}, \text{ wobei } n \text{ der Anzahl an Permutationen entspricht (in}$$

dieser Arbeit war $n=1000$).

3.6 Statistische Tests

Charakteristika der Host- ($n=246$) und putativen Zielgene ($n=326$) intronischer miRNAs wurden mittels Mann-Whitney-U Test untersucht. Die Subgruppen-Analyse für miRNAs, die ihr Host-Gen kontrollieren könnten, wurde mit dem Kruskal-Wallis Test und post-hoc Mann-Whitney-U Test mit Bonferroni-Korrektur durchgeführt. Die Gene der Kontrollgruppe wurden zufällig unter Erhaltung der Chromosomen-Verteilung ausgewählt ($n=2460$ für Host-Kontrollgruppen-Gene und $n=3260$ für Ziel-Kontrollgruppen-Gene). Zur Signifikanzberechnung der evolutionären Konservierung der Assoziation von miRNAs und Host-Gen wurde der „2-sample test for equality of proportions“ mit Kontinuitätskorrektur gewählt.

Der automatisierte stufenweise Vorwärtsselektionsalgorithmus wurde in R 2.7.1 implementiert. Er beginnt mit einem leeren Modell. Der Reihe nach werden mögliche Prädiktoren addiert, von denen derjenige mit der höchsten Signifikanz beibehalten wird ($p\text{-Wert} < 0.05$). Verliert durch Addition einer weiteren Variable ein Prädiktor seine Signifikanz ($p\text{-Wert} \geq 0.05$), wird er aus dem Modell entfernt.

Der Algorithmus stoppt, wenn dem Modell kein weiterer signifikanter Prädiktor hinzugefügt werden kann.

4 Ergebnisse

4.1 Entwicklung eines Konsensusverfahrens zur miRNA-Zielvorhersage

Wie bereits erwähnt, gibt es verschiedene Zielvorhersage-Algorithmen, die jeweils auf wenigen Hypothesen der miRNA-Ziel-Interaktion beruhen. In dieser Arbeit wurde ein Konsensusverfahren untersucht. Das „Prediction Agreement“ (PA) einer miRNA-mRNA-Interaktion wurde als die Anzahl an in dieser Vorhersage übereinstimmenden Methoden definiert. Dabei wird nicht berücksichtigt, welche Methoden in ihren Vorhersagen übereinstimmen, was einen möglichen Bias durch einzelne Methoden reduzieren sollte. Die Leistungsfähigkeit dieser Methode wurde an dem öffentlich verfügbaren Datensatz von Selbach und Kollegen gemessen und mit den einzelnen Methoden verglichen. Selbach et al. haben fünf miRNAs, hsa-miR-1, hsa-miR-16, hsa-miR-30a, hsa-miR-155 und hsa-let-7b in HeLa-Zellen überexprimiert und die resultierenden Veränderungen der Protein- und mRNA-Expression gemessen [72]. Die miRNA-mRNA-Interaktionsvorhersagen der sechs Algorithmen miRanda [38], TargetScan [35], PITA [27], MirTarget2 [43, 44], PicTar [45] und RNA22 [39], sowie die literaturbasierte Datenbank TarBase [46] wurden als Prädiktoren für verminderte Expression von mRNA und Proteinen evaluiert. Zur Bestimmung der Leistungsfähigkeit der einzelnen Vorhersagemethoden wurde die Präzision, die Falsch-Positiv-Rate, die Sensitivität, und die Fläche unter der Receiver-Operator-Characteristics (ROC)-Kurve bestimmt (AUC). Die AUC wurde hier als Maß

benutzt, wie gut die Vorhersagen mit den gemessenen Werten korrelieren, ohne abhängig von der Definition eines Schwellenwertes zu sein (im Gegensatz zur Berechnung bzw. Schätzung der Sensitivität, Präzision und Falsch-Positiv-Rate).

Basierend auf pSilac Protein Log2-fache Expressionsänderungen < -0.1 (Präzision und Sensitivität) und > 0.1 (Falsch-Positiv-Rate)								
miRNAs Ziel-Vorhersage-Methode	Anzahl an Vorhersagen	Geschätzte korrekte Vorhersagen	Präzision	Geschätzte falsche Vorhersagen	Geschätzte Falsch-Positiv-Rate	Zahl an Protein-Expressionsänderungen < -0.1	Geschätzte Sensitivität	AUC
PicTar	201	117	0.58	28	0.14	2032	0.06	0.7
TargetScan	2423	1072	0.44	462	0.19	2032	0.53	0.6
PITA	5429	1849	0.34	1381	0.25	2032	0.91	0.5
RNA22	856	289	0.34	205	0.24	2032	0.14	0.5
MirTarget2	364	248	0.68	39	0.11	2032	0.12	0.7
Miranda	1331	577	0.43	282	0.21	2032	0.28	0.6
Tarbase	26	24	0.92	1	0.04	2032	0.01	0.9
PA >= 2	2865	1192	0.42	607	0.21	2032	0.59	0.6

Tabelle 1. Vergleich verschiedener miRNA-Zielvorhersage-Algorithmen an Protein-Expressionsänderungen

Basierend auf pSilac mRNA Log2-fache Expressionsänderungen < -0.1 (Präzision und Sensitivität) und > 0.1 (Falsch-Positiv-Rate)

miRNAs Ziel-Vorhersage-Methode	Anzahl an Vorhersagen	Geschätzte korrekte Vorhersagen	Präzision	Geschätzte falsche Vorhersagen	Geschätzte Falsch-Positiv-Rate	Zahl an Protein-Expressionsänderungen < -0.1	Geschätzte Sensitivität	AUC
PicTar	1018	505	0.50	191	0.19	12311	0.04	0.63
TargetScan	14151	5566	0.39	3605	0.25	12311	0.45	0.57
PITA	30569	11223	0.37	9007	0.29	12311	0.91	0.55
RNA22	4985	1750	0.35	1469	0.29	12311	0.14	0.55
MirTarget2	2028	1084	0.53	363	0.18	12311	0.09	0.65
Miranda	7530	3259	0.43	1866	0.25	12311	0.26	0.55
Tarbase	109	94	0.86	8	0.07	12311	0.01	0.82
PA >= 2	16299	6518	0.40	4258	0.26	12311	0.53	0.55

Tabelle 2. Vergleich verschiedener miRNA-Zielvorhersage-Algorithmen an mRNA-Expressionsänderungen

Alle Methoden erzielen in den Proteinexpressionsänderungsmessungen (Tabelle 1) eine höhere Präzision und geringere Falsch-Positiv-Rate als in den mRNA-Expressionsänderungsmessungen (Tabelle 2). Auch die geschätzte Sensitivität ist bei den meisten Vorhersagemethoden geringer. Die höchste Sensitivität, allerdings auch niedrigste Präzision, wird von PITA erreicht, während Tarbase bzw. MirTarget2 die höchste Präzision erreichen, auf Kosten der Sensitivität.

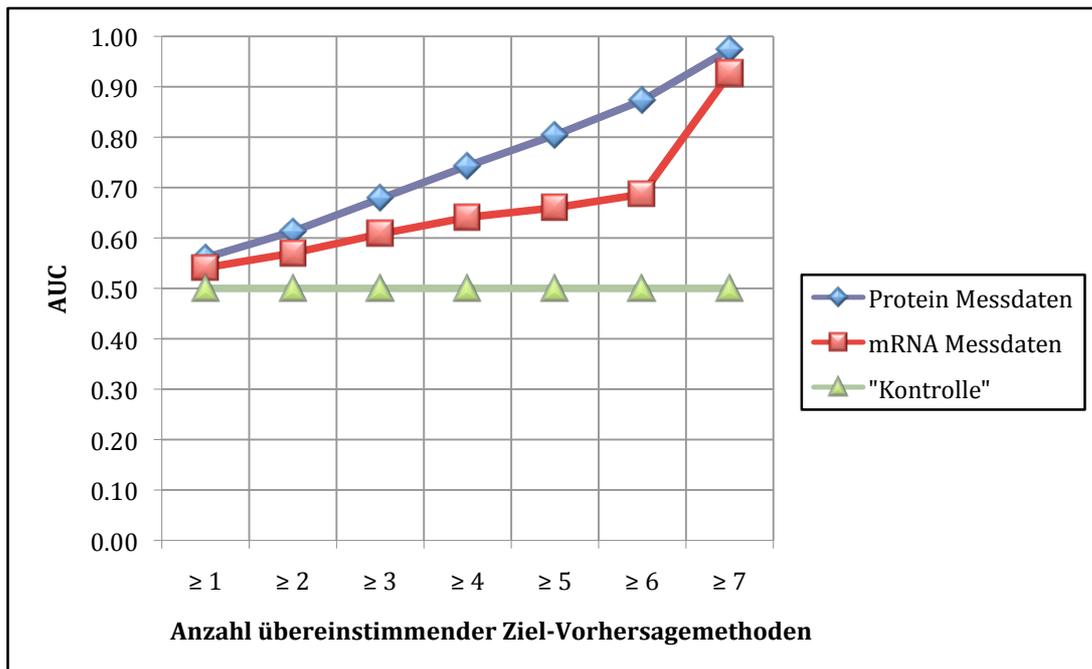


Abbildung 2. AUC in Abhängigkeit vom Prediction Agreement

Es besteht ein fast linearer Zusammenhang zwischen der AUC und dem Prediction Agreement in dem Protein-Expressionsänderungsdatensatz. In den mRNA-Expressionsänderungsmessungen ist ebenfalls für ein Prediction Agreement von bis zu ≥ 6 Methoden ein linearer Zusammenhang sichtbar, der allerdings eine wesentlich geringere Steigung aufweist, als die analoge Kurve in dem Protein-Expressionsänderungsdatensatz.

Im nächsten Schritt wurde ein Prädiktionsmodell für Änderung der log₂-fachen Proteinexpression von kleiner als -0,1 mittels eines automatisierten stufenweisen Vorwärtsselektionsalgorithmus entwickelt.

Variable	Koeffizient	Standard-Fehler	p-Wert
Achsenabschnitt	-1.296	0.024	2.00E-16
Prediction Agreement	0.209	0.029	6.71E-13
MirTarget 2	0.785	0.132	2.81E-09
TargetScan	0.398	0.075	1.17E-07
TarBase	2.421	0.745	1.20E-03

Tabelle 3. Logistisches Regressionsmodell.

Das beste vom automatisierten Algorithmus mit logistischer Regression gefundene Modell enthält vier signifikante Prädiktoren. Das Prediction Agreement ist die Methode mit der höchsten Signifikanz, aber dem niedrigsten Koeffizienten, gefolgt von MirTarget2, TargetScan und TarBase. Um Overfitting zu vermeiden und einen ungebiasteten Schätzwert eines automatisch entwickelten logistischen Regressionsmodells zu erhalten, wurde eine 5-fach Crossvalidierung mit zehn Wiederholungen durchgeführt. Die dabei durchschnittlich erzielte AUC lag bei 0,59, bei guter Kalibrierung (Test nach Hosmer-Lemeshow > 0.05 in 10 Wiederholungen).

4.2 Analyse an einem praktischen Beispiel: Die Rolle von miRNAs in der Sepsis

Kürzlich wurden vier miRNAs, hsa-miR-150, hsa-miR-182, hsa-miR-342 und hsa-miR-486, identifiziert, die in Leukozyten septischer Patienten differentiell exprimiert sind [60]. Allerdings gibt es bislang keine weiteren Untersuchungen, die die pathogenetische Rolle dieser miRNAs und damit eine mögliche Eignung

als Biomarker der Sepsis eingehend untersuchen. Wie Abbildung 1 darlegt, könnten diese miRNAs über die Funktion ihrer Ziel-mRNAs charakterisiert werden. Allerdings beläuft sich die Summe der von den sechs in der vorliegenden Arbeit verwendeten Zielvorhersage-Methoden insgesamt auf 13953 vorhergesagte miRNA-mRNA-Interaktionen. Ziel ist es, mithilfe des erstellten Datenbanksystems und Konsensusverfahrens eine mögliche Herangehensweise für eine solche Untersuchung zu entwickeln.

Die Expression von hsa-miR-342 und hsa-miR-150 ist in Leukozyten septischer Patienten signifikant reduziert, während die Expression von hsa-miR-182 und hsa-miR-486 signifikant erhöht ist [60]. Zwei von den vier differentiell exprimierten miRNAs, hsa-miR-342 und hsa-miR-486, sind zudem intragenische miRNAs; hsa-miR-342 befindet sich fast mittig im Intron 3 (Länge: 25879 bp) des Gens *EVL* auf Chromosom 14. Das Gen für hsa-miR-486 liegt im Intron 42 des *ANK1*-Gens in einem Abstand von 921 Basenpaare zum vorherigen Exon. Somit ergeben sich zwei verschiedene Aspekte, die analysiert werden sollten. Einerseits können die Ziel-Gene der jeweils hoch- und runterregulierten miRNAs untersucht werden. Andererseits ist die Interaktion zwischen intronischer miRNA und Host-Gen zu untersuchen, welche Hinweise auf die Funktion der miRNA oder die des Host-Gens geben kann.

Analyse der „Kyoto Encyclopedia of Genes and Genomes“ (KEGG)-Ontologie [76] für Ziele von hsa-miR-342 (Prediction Agreement ≥ 4 ; 146 Ziele) und von hsa-miR-150 (Prediction Agreement ≥ 4 ; 196 Ziele).

KEG G-ID	Biomolekularer Pfad	p-Wert	Potentielle Zielgene
4010	MAPK signaling pathway	0.003	FLNC; IL1A; PAK2; PDGFRA; NLK; ZAK; PPP3R1; BDNF; TGFB1; MAP3K12; MAPKAPK3; MAP3K14
4660	T cell receptor signaling pathway	0.007	NFAT5; PAK2; PPP3R1; VAV2; CBL; MAP3K14
4510	Focal adhesion	0.011	COL4A6; FLNC; ITGA3; ITGB3; SHC4; PAK2; PDGFRA; VAV2; CCND2
4310	Wnt signaling pathway	0.020	NFAT5; NLK; PPP3R1; SKP1; WNT10B; CAMK2G; CCND2
4012	ErbB signaling pathway	0.022	NRG1; SHC4; PAK2; CAMK2G; CBL
4350	TGF-beta signaling pathway	0.024	ID4; INHBA; SKP1; SP1; TGFB1
5214	Glioma	0.032	E2F3; SHC4; PDGFRA; CAMK2G
533	Keratan sulfate biosynthesis	0.035	B3GNT1; FUT8
4810	Regulation of actin cytoskeleton	0.041	ITGA3; ITGAX; ITGB3; PAK2; PDGFRA; PFN2; TIAM1; VAV2
4020	Calcium signaling pathway	0.045	ADORA2A; ADRB1; PDGFRA;

			PPP3R1; SLC8A1; TACR3; CAMK2G
5220	Chronic myeloid leukemia	0.052	E2F3; SHC4; TGFBR1; CBL
4210	Apoptosis	0.071	IL1A; IRAK2; PPP3R1; MAP3K14
910	Nitrogen metabolism	0.073	CTH; CA12
4640	Hematopoietic cell lineage	0.076	IL1A; IL7; ITGA3; ITGB3
4512	ECM-receptor interaction	0.078	COL4A6; ITGA3; ITGB3; SV2B
4360	Axon guidance	0.089	SEMA3A; NFAT5; PAK2; PPP3R1; SRGAP3
4120	Ubiquitin mediated proteolysis	0.094	UBE2J1; UBE2R2; SKP1; UBE2D2; CBL

Tabelle 4. KEGG-Analyse hochkonfidenter Ziele von hsa-miR-342 und hsa-miR-150.

Analyse der KEGG-Ontologie [76] für Ziele von hsa-miR-486 (Prediction Agreement ≥ 4 , 97 Ziele) und von hsa-miR-182 (Prediction Agreement ≥ 5 , 124 Ziele).

KEG G-ID	Biomolekularer Pfad	p-Wert	Potentielle Zielgene
5210	Colorectal cancer	<0.001	DCC; IGF1R; SMAD2; PDGFRA; PIK3R1; CYCS; RAC1; BCL2
5218	Melanoma	<0.001	FGF7; FGF9; IGF1R; MITF; PDGFRA; PIK3R1; PTEN
5215	Prostate cancer	<0.001	CREB1; FOXO1; IGF1R; PDGFRA; PIK3R1; PTEN; BCL2
4510	Focal adhesion	0.001	IGF1R; ARHGAP5; SHC4; PDGFRA; PIK3R1; PAK6; PTEN; RAC1; BCL2
5214	Glioma	0.001	IGF1R; SHC4; PDGFRA; PIK3R1; PTEN
5030	Amyotrophic lateral sclerosis (ALS)	0.002	RAC1; BCL2; SLC1A2
4810	Regulation of actin cytoskeleton	0.004	DIAPH2; FGF7; FGF9; PDGFRA; PIK3R1; PAK6; RAC1; ARHGEF7
4360	Axon guidance	0.005	DCC; EPHA3; PPP3R1; PAK6; RAC1; RASA1
4520	Adherens junction	0.014	IGF1R; SMAD2; RAC1; YES1

4070	Phosphatidylinositol signaling system	0.016	OCRL; SKIP; PIK3R1; PTEN
4210	Apoptosis	0.021	PIK3R1; CYCS; PPP3R1; BCL2
562	Inositol phosphate metabolism	0.023	OCRL; SKIP; PTEN
5222	Small cell lung cancer	0.023	PIK3R1; CYCS; PTEN; BCL2
4530	Tight junction	0.026	CTTN; PTEN; EXOC4; YES1; MAGI1
5213	Endometrial cancer	0.027	FOXO3; PIK3R1; PTEN
401	Novobiocin biosynthesis	0.037	TAT
4662	B cell receptor signaling pathway	0.046	PIK3R1; PPP3R1; RAC1
4010	MAPK signaling pathway	0.046	FGF7; FGF9; PDGFRA; PPP3R1; RAC1; RASA1; BDNF
5216	Thyroid cancer	0.051	TPM3; NCOA4
5211	Renal cell carcinoma	0.055	PIK3R1; PAK6; RAC1
5040	Huntington`s disease	0.061	RASA1; BDNF
950	Alkaloid biosynthesis I	0.061	TAT
5212	Pancreatic cancer	0.063	SMAD2; PIK3R1; RAC1
4370	VEGF signaling pathway	0.063	PIK3R1; PPP3R1; RAC1
4650	Natural killer cell mediated cytotoxicity	0.080	SHC4; PIK3R1; PPP3R1; RAC1
4012	ErbB signaling pathway	0.095	SHC4; PIK3R1; PAK6
4910	Insulin signaling pathway	0.096	FOXO1; SHC4; SKIP; PIK3R1

Tabelle 5. KEGG-Analyse hochkonfidenter Ziele von hsa-miR-486 und hsa-miR-182.

Beide miRNA Zielgen-Listen enthalten mehrere biomolekulare Pfade, die eng mit den Signalkaskaden der Immunreaktion assoziiert sind (TGF-beta Signalweg,

T- und B-Zell-Rezeptor-Signalweg, Natural killer cell mediated cytotoxicity, mTOR-Signalweg, fokale Adhäsion, Apoptose). Sowohl die über- als auch die unterexprimierten miRNAs enthalten Zielvorhersagen aus der Apoptose-Signalkaskade. Bis auf PPP3R1 sind die Zielgene allerdings unterschiedlich. Die zwei Tabellen zeigen exemplarisch, wie interessante Zielgene strukturiert identifiziert werden können.

4.3 Charakteristika intronischer miRNAs

Sowohl das Gen für hsa-miR-342 als auch für hsa-miR-486 liegen in intronischen Bereichen sogenannter Host-Gene. Allerdings ist die funktionelle Bedeutung dieser Lagebeziehung weitestgehend unklar und weder *EVL* noch *ANK1* sind funktionell ausreichend charakterisiert, um in der KEGG-Datenbank [76] gelistet zu sein. Li et al. spekulierten, dass es einen Zusammenhang zwischen der Funktion des Hosts und der Funktion der miRNA geben könnte, allerdings ohne die Hypothese weiter zu untersuchen [77]. Das in der vorliegenden Arbeit entwickelte System bietet die Möglichkeit, Gruppen von Genen in verschiedenen Spezies zu untersuchen, um sich diesem Gedanken zu nähern.

Organism	Intra- genische miRNAs		Inter- genische miRNAs	Intragenische miRNAs	
	Intronisch	Exonisch		miRNAs auf Host- Gen- Strang	miRNAs auf gegenüber liegendem Strang
<i>Homo sapiens</i>	296 (42.6 %)	37 (5.3 %)	362 (52.1 %)	282 (84.7 %)	51 (15.3 %)
<i>Mus musculus</i>	171 (35.4 %)	30 (6.2 %)	282 (58.4 %)	163 (78.2 %)	38 (21.8 %)
<i>Canis familiaris</i>	3 (1.5 %)	0 (0 %)	201 (98.5 %)	2 (66.7 %)	1 (33.3 %)
<i>Gallus gallus</i>	50 (10.7 %)	1 (0.2 %)	418 (89.1 %)	46 (90.2 %)	5 (9.8 %)
<i>Danio rerio</i>	48 (15.0 %)	1 (0.3 %)	271 (84.7 %)	39 (79.6 %)	10 (20.4 %)
<i>Drosophila melanogaster</i>	65 (42.8 %)	2 (1.3 %)	85 (55.9 %)	53 (79.1 %)	14 (20.9 %)
<i>Caenorhabditis elegans</i>	51 (33.1 %)	1 (0.6 %)	102 (66.2 %)	33 (63.6 %)	19 (36.5 %)

Tabelle 6. Verteilung intergenischer, intronischer und exonischer miRNAs in verschiedenen Spezies

Die Verteilung bekannter miRNA-Gene auf dem Genom verschiedener Spezies ist sehr unterschiedlich (Tabelle 6). Bei *H.sapiens* befinden sich 47,9% der

miRNA-Gene innerhalb bekannter protein-kodierender Gene. Bei *C.familiaris* hingegen sind es lediglich 1,5%.

4.3.1 Verhältnis miRNA – Host-Gen

Die Expression des Gens *ANK1* ist mit einem Prediction Agreement von 3 Methoden Ziel der eigenen intronischen miRNA hsa-miR-486, während *EVL* kein vorhergesagtes Ziel seiner intronischen miRNA hsa-miR-342 ist. Eine Abfrage des hier vorgestellten Datenbanksystems ergibt, dass insgesamt bis zu 20% aller bekannten intronischen miRNAs (insgesamt 56 in 49 Host-Genen) ihr eigenes Host-Gen regulieren könnten (Prediction Agreement ≥ 2 Methoden; geschätzter *p-Wert* < 0.001). Mittels Gruppierung der Vorhersage-Methoden und der Einführung eines Wählverfahrens kann geschätzt werden, dass zwischen 12 und 56 miRNAs ihren eigenen Host kontrollieren. Diese Zahl war in jedem Durchlauf signifikant höher, als erwartet.

Gruppe 1	Gruppe 2	Anzahl miRNAs, die ihr Host-Gen regulieren	p-Wert
PITA, miRanda, TargetScan	PicTar 5, RNA22, MirTarget2	55	< 0.001
PicTar 5, miRanda, TargetScan	PITA, RNA22, MirTarget2	12	0.023
PITA, RNA22, TargetScan	PicTar 5, miRanda, MirTarget2	33	0.012
PITA, miRanda, MirTarget2	PicTar 5, RNA22, TargetScan	41	< 0.001

Tabelle 7. Test der Robustheit mittels Wahlverfahren

MiRNA-Host-Paare sind auch häufiger konserviert (Tabelle 7), wenn das Host-Gen auch ein potentiell Ziel der miRNA ist. Signifikanz kann für die Konservierung von miRNA-Host-Gen-Paaren zwischen Mensch und Huhn gezeigt werden (Tabelle 8).

Organismen	miRNA-Host Paare (die das Host-Gen kontrollieren)		miRNA-Host Paare (die nicht das Host- Gen kontrollieren)		<i>p</i> -Wert
	Konserviert	Gesamt	Konserviert	Gesamt	
<i>Homo sapiens</i> –	18 (35.2 %)	51	41 (24.5 %)	167	0.18
<i>Mus musculus</i>					
<i>Homo sapiens</i> –	1 (2.12 %)	47	0	142	0.56
<i>Canis familiaris</i>					
<i>Homo sapiens</i> –	5 (12.5 %)	40	1 (0.71 %)	139	0.001
<i>Gallus gallus</i>					

Tabelle 8. Konservierung intronischer miRNA - Host-Gen Paare.

4.3.2 Charakteristika von Host- und Ziel-Genen

Sowohl *EVL* als auch *ANK1* sind mit 78722 bp beziehungsweise 144396 bp vergleichsweise lange Gene mit vielen Introns (13 beziehungsweise 43). Eigenschaften wie Länge, Anzahl an Introns, Länge der nichttranslatierten Regionen (3'- und 5'-UTR) und AU-reiche Elemente können allerdings Aufschluss über die Funktion von Genen geben. Es wurden alle Hostgene intragenischer miRNAs untersucht, deren miRNA auf dem gleichen DNA-Strang wie das Host-Gen liegen.

Eigenschaft	Gen-Set	Median[Range] Host/Ziel	Median[Range] Kontrolle	Verhältnis	p-Wert
Gesamtlänge (Basenpaare)	Host -	84871.0	29324.5	2.89	< 2.2e-
	Gene	[2792-2220381]	[599-2304633]		16
	Ziel- Gene	83747.5 [2366-2220381]	30232.5 [218-2220381]	2.77	< 2.2e- 16
Introns	Host -	13	8	1.62	4.3e-13
	Gene	[1-88]	[1-105]		
	Ziel- Gene	10.5 [0-78]	8 [0-311]	1.31	9.77e- 07
5'UTR- Länge (Basenpaare)	Host -	279.5	298.5	0.94	0.25
	Gene	[0-385608]	[0-1098107]		
	Ziel- Gene	439.5 [0-460277]	282.5 [0-1098107]	1.56	2.32e- 08
3'UTR- Länge (Basenpaare)	Host -	1218.5	872	1.4	4.71e-
	Gene	[0-535884]	[0-321862]		05
	Ziel- Gene	1764 [171-11799]	872 [0-72058]	2.2	< 2.2e- 16
ARE (absolut)	Host -	2.0	1.0	2.0	3.89e-
	Gene	[0-1794]	[0-592]		04
	Ziel- Gene	5.0 [0-47]	2.0 [0-187]	2.5	< 2.2e- 16
ARE (per kb)	Host -	1.9	1.49	1.26	0.012
	Gene	[0-2.74]	[0-0.045]		

	Ziel-	2.69	1.63	1.65	< 2.2e-
	Gene	[0-14.22]	[0-76.92]		16
5' UTR GC-	Host -	0.6	0.59	1.06	0.015
Anteil	Gene	[0.31-0.95]	[0-1]		
	Ziel-	0.59	0.58	1.01	0.71
	Gene	[0.26-1]	[0-1]		

Tabelle 9. Charakteristika der Host- und Ziel-Gene intronischer miRNAs.

Host-Gene intragenischer miRNAs sind insgesamt länger als Gene der Kontrollgruppe. Sie haben mehr und auch längere Introns (Tabelle 9). Das 3'-UTR-Sequenz ist signifikant länger und es kommt häufiger das Pentamer AUUUA vor. Die Gruppe hochkonfidenter Zielgene (Prediction Agreement ≥ 6) zeigt ähnliche Merkmale mit signifikant längeren Genen, mehr Introns, längeren 3'- und 5'-UTR-Sequenzen und mehr AUUUA-Pentameren. In der Subanalyse zeigt sich, dass sich lediglich die Host-Gene, die potentiell von ihren intragenischen miRNAs reguliert werden, durch signifikant längere 3'-UTR-Sequenzen auszeichnen. Die anhand von *EVL* und *ANK1* identifizierten Charakteristika sind offensichtlich ein Merkmal intronischer miRNAs generell, aber auch ihrer Zielgene. Das Host-Gen von hsa-miR-342, *EVL*, ist insgesamt 78722 bp lang, enthält 13 Introns mit einer 498 bp langen 3'-UTR-Sequenz. *ANK1* ist insgesamt 144396 bp lang, enthält 43 Introns und eine 8240 bp lange 3'-UTR-Sequenz. Lediglich die Anzahl der AU-reichen Elemente ist mit 11 für *ANK1* und 1 für *EVL* unter dem Erwartungswert der Analyse.

5 Diskussion

In der aktuellen molekularmedizinischen Forschung sind miRNAs ein zentraler Bestandteil mit direktem Einfluss auf Pathogenese [56, 78] und Diagnostik [52, 79] und, wahrscheinlich in näherer Zukunft, auch der Therapie vieler Erkrankungen. Sie gehören zur Klasse der nicht-kodierenden RNA-Moleküle, die eingebunden in einen Proteinkomplex an mRNA binden und die Translation dieser beeinflussen können [72]. Unser Wissen über molekularbiologische Prozesse, insbesondere miRNAs, vergrößert sich exponentiell und es wird immer anspruchsvoller, diese Informationen sinnvoll zu nutzen. Einerseits sind sie auf viele Quellen verteilt, die zusammengebracht werden müssen. Andererseits müssen sie je nach Fragestellung flexibel strukturiert werden können. Diese Arbeit behandelt die Entwicklung einer Datenbank, die das Generieren von biologischen Hypothesen zur Funktion von miRNAs unterstützen und das experimentelle Vorgehen leiten soll. Ein zentrales Element der Datenbank ist die Integration der Information aus mehreren miRNA-mRNA-Interaktionsvorhersage-Algorithmen. Die Möglichkeiten dieses Systems werden dann basierend auf den Ergebnissen einer kürzlich veröffentlichten Arbeit von Vasilescu et al. [60] demonstriert. Die Autoren identifizierten ein verändertes Expressionsmuster für vier miRNAs, hsa-miR-150, hsa-miR-342, hsa-miR-182 und hsa-mir-486 in Leukozyten septischer Patienten, ohne weitere Hypothesen experimentell zu verfolgen. Exemplarisch wurden daher aus den publizierten Informationen anhand der in dieser Arbeit entwickelten Datenbank biologische Hypothesen generiert und experimentelle Ansätze vorgestellt, mit denen diese überprüft werden könnten.

5.1 Etablierung eines Konsens-Verfahrens für 6 Ziel-Vorhersage-Methoden

Wie bereits beschrieben, gibt es bislang keine Möglichkeit der genomweiten Validierung von miRNA-mRNA-Interaktionen, die benötigt werden würden, um die potentiell große Anzahl von Interaktionen einer miRNA zu determinieren. Eine Vielzahl von miRNA-mRNA-Interaktionsvorhersage-Programmen wurde bereits entwickelt. Auf der Basis dieser Vorhersagen werden Experimente geplant und Hypothesen formuliert (beispielsweise in [80]). Die Übereinstimmung der Vorhersagen ist aufgrund der methodischen Verschiedenheit jedoch gering [81] und reicht in unserer Datenbank von 3,5% (MirTarget2 und RNA22) bis 21,2% (PITA und TargetScan). Hingegen ist die Anzahl der vorhergesagten Interaktionen hoch (je nach Methode durchschnittlich ca. 200 bis ca. 6000 Zielgene pro miRNA) und die Auswahl interessanter Ziele basierend auf Gen-Namen praktisch unmöglich.

Vasilescu et al. [60], die relevante miRNA-Expressionsänderungen in Leukozyten septischer Patienten beschrieben haben, reduzierten die möglichen Ziel-Gene von hsa-miR-150 durch Kombination von miRGen- [71] und RNA22-Zielvorhersagen [39]. miRGen bietet über ein Web-Interface die Möglichkeit, die Schnittmenge oder Vereinigungsmenge mehrerer bekannter Zielvorhersage-Programme zu nutzen (in diesem Fall miRanda, TargetScan und PicTar), nicht aber die Kombination als Konsensusverfahren. Wie bereits beschrieben (Kapitel 2) kann aber durch Benutzung eines Konsensusverfahrens ein potentiell eingebrachter Bias reduziert werden.

Das in der vorliegenden Arbeit vorgestellte Modell integriert die Vorhersagen von insgesamt sechs Zielvorhersage-Methoden und einer Datensammlung validierter miRNA-Ziel-Interaktionen in einer Datenbank. Der lineare Zusammenhang zwischen AUC und Prediction Agreement bedeutet, dass das Prediction Agreement (d.h. die Anzahl übereinstimmender Vorhersagen für eine Interaktion) als gutes Maß für die Konfidenz einer Vorhersage betrachtet werden kann. Dabei wird bislang noch nicht die Anzahl an [82] oder Distanz zwischen [83] Bindungsstellen einer einzelnen Methode berücksichtigt, oder ob die Vorhersagen verschiedener Methoden auch innerhalb der 3'-UTR-Sequenz denselben Bereich betreffen. Der Zusammenhang zwischen Prediction Agreement und AUC ist weniger ausgeprägt für mRNA-Expressionsänderungen. Dieser Unterschied ist erklärbar durch die Tatsache, dass sowohl translationale Repression als auch mRNA-Degradation zu reduzierter Proteinexpression führen, während nur letzterer Mechanismus zu einem messbaren Unterschied der mRNA-Expression führt. Die Ungenauigkeit der Messergebnisse von MicroArrays muss ebenfalls in Betracht gezogen werden, da die optimale Datenprozessierung nach wie vor kontrovers diskutiert wird und direkten Einfluss auf das Ergebnis haben kann [84-87].

Zusammengefasst erscheint das in dieser Arbeit entwickelte Konsensus-Verfahren sinnvoll, da ein linearer Zusammenhang zwischen AUC und der Anzahl der verwendeten Methoden gezeigt werden konnte. Aufgrund der Unterschiedlichkeit der verwendeten Zielvorhersage-Algorithmen wird die Gefahr der Einbringung eines Bias reduziert und gleichzeitig die Möglichkeit geschaffen, Sensitivität und Spezifität durch das Prediction Agreement zu kontrollieren.

5.2 Sepsis und miRNA-Bioinformatik

Sepsis ist nach wie vor ein großes Problem der Intensivstationen, das mit hoher Letalität verbunden ist [88]. Die Pathogenese ist komplex und weitestgehend unverstanden und die Scores, die zur Bewertung der Schwere des Krankheitsbilds dienen, orientieren sich primär an Symptomen [59, 89, 90]. Eine amerikanische Studie schätzte 2001, dass mindestens ebenso viele Patienten an der Sepsis versterben, wie am Herzinfarkt und dass die Inzidenz der Sepsis in kommenden Jahren weiter zunimmt [88]. Während die Rolle von miRNAs bereits mit Dysregulationen des Immunsystems im Rahmen von malignen Veränderungen beschrieben wurde [53, 54], ist die Datenlage zur Beteiligung von miRNAs an der Pathogenese der Sepsis spärlich. Unlängst wurde von Vasilescu et al. die differentielle Expression von vier miRNAs in Leukozyten septischer Patienten beschrieben, die mit dem klinischen Bild korrelieren könnte [60]. Allerdings wurde die pathogenetische Bedeutung dieser miRNA-Signatur nicht hinreichend untersucht. Aufbauend auf den Ergebnissen dieser Studie wurden daher in der vorliegenden Arbeit zwei Listen mit potentiellen Ziel-mRNAs erstellt. Die erste Liste enthält Ziele der in septischen Patienten vermehrt exprimierten miRNAs hsa-miR-486 und hsa-miR-182 (Prediction Agreement ≥ 4 bzw. ≥ 5), die zweite Liste Ziel-mRNAs der vermindert exprimierten hsa-miR-150 und hsa-miR-342 (Prediction Agreement ≥ 4). Anschließend wurden diese dann mittels der Ontologie „Kyoto Encyclopedia of Genes and Gene Products“ [76] biomolekularen Pfaden zugeordnet (Tabelle 4 und Tabelle 5). Beide Listen enthalten Ziele aus verschiedenen für die Sepsis relevanten Signalkaskaden, die näher untersucht werden könnten. Interessant ist in diesem Zusammenhang

beispielsweise die Rolle des Apoptose-Signalwegs. Apoptose, im Gegensatz zur Nekrose, ist der programmierte, von der Zelle selbst organisierte Zelltod [91]. Auf der Basis äußerer (extrinsischer) oder innerer (intrinsischer) Faktoren werden Kaspasen aktiviert, die ihrerseits entweder selbst oder via Rekrutierung von DNAsen zytotoxisch aktiv werden, ohne dabei das Nachbargewebe zu schädigen (Review siehe [92]). Das Überleben der Zelle wird durch die Balance pro- und antiapoptotischer Faktoren bestimmt. Proapoptotische Stimuli sind beispielsweise Fas, TNF-Alpha und Interleukin-1 [93, 94] und antiapoptotische Stimuli unter anderem Interleukin-2, Interleukin-3 und Interleukin-4 [92, 95]. In mehreren Studien wurde eine im Rahmen der Sepsis auftretende ausgeprägte und prognostisch relevante Apoptose lymphatischer Zellen gefunden [96-98]. Im Mausmodell konnte sogar durch Reduktion der Apoptose lymphatischer Zellen das Überleben verbessert werden [99]. Pathogenetische Mechanismen sind weitestgehend unklar und therapeutische Erfolge beim Menschen eher gering. Molekulargenetische Änderungen rücken immer mehr in den Fokus der Sepsis-Forschung. McDunn et al. beschreiben bereits 6h nach Sepsis-Induktion im Mausmodell sichtbare Veränderungen des Transkriptom zugunsten proapoptotischer Gene [100]. Beide Ziel-Gen-Listen (Tabelle 4 und Tabelle 5) enthalten hochkonfidente Ziel-Gene, die mit dieser Signalkaskade in Verbindung gebracht werden können. Ein Gen, PPP3R1, taucht in beiden Listen auf und ist somit nicht eindeutig funktionell einzugliedern. Ziel-Gene der unterexprimierten miRNAs (*IL1A*, *IRAK2* und *MAP3K14*) können der pro-apoptotischen Signalkaskade zugeordnet werden [101-103], während die Ziel-Gene *PIK3R1* und *BCL2* der überexprimierten miRNAs mit der durch Interleukin-3 initiierten antiapoptotischen Signalkaskade assoziiert sind [104-106]. Weitere Beispiele lassen

sich für potentielle Ziele mit geringerem Prediction Agreement finden; die Signalwege sind schematisch in Abbildung 3 dargestellt.

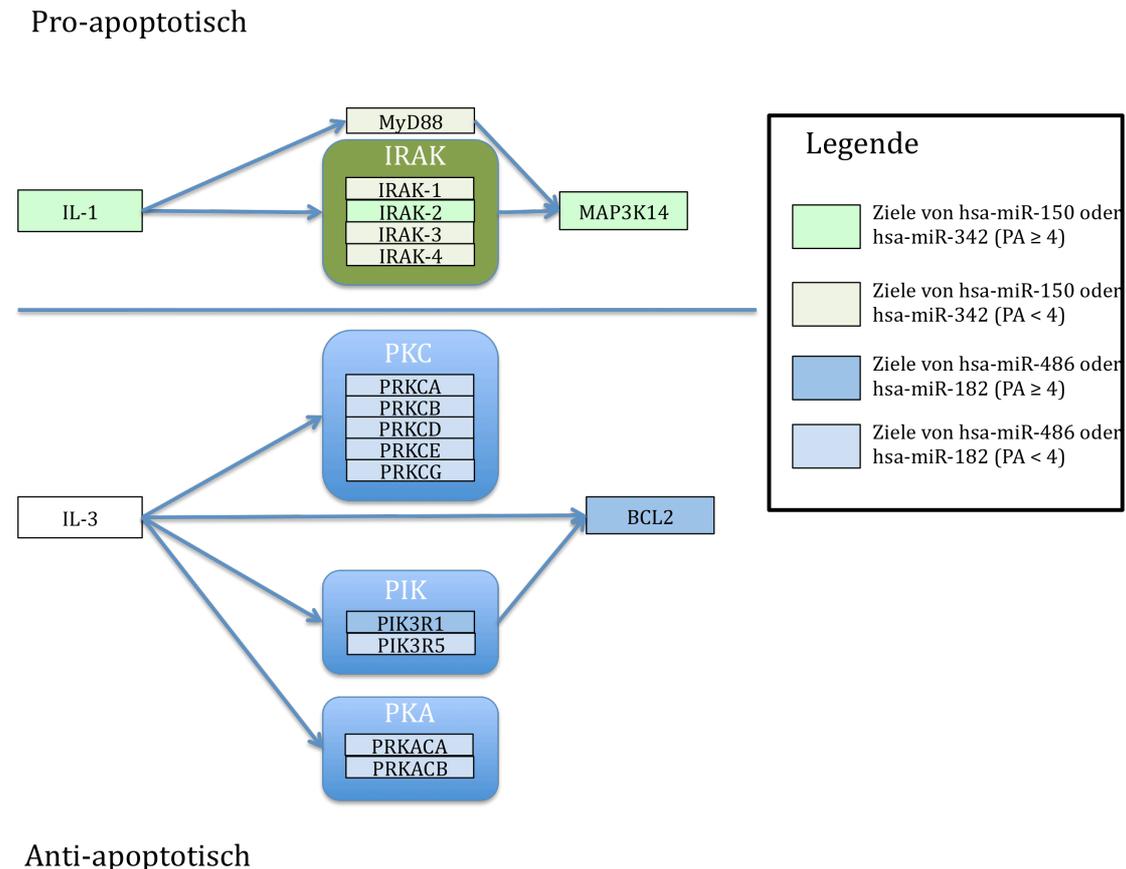


Abbildung 3. Visualisierung der Interaktionen pro- und antiapoptotischer Ziel-Gene.

Einige weniger konfidente Zielvorhersagen unterstützen das Modell der Balance pro- und antiapoptotischer Faktoren durch hsa-miR-150, hsa-miR-342, hsa-miR-182 und hsa-miR-486. IRAK1: hsa-miR-150 (PA 3), IRAK3: hsa-miR-342 (PA 2), IRAK4: hsa-miR-150, hsa-miR-342, (auch hsa-miR-182) (PA 2), MYD88: hsa-miR-150 (PA 2), FAS (CD95): hsa-miR-342 (PA 2); PPP3CA: hsa-miR-182 (PA 2), PPP3CB: hsa-miR-486 (PA 3), PRKACA: hsa-miR-182 (PA 2), PRKACB: hsa-miR-182 (PA 3), PIK3R5: hsa-miR-182 (PA 3), PRKCA: hsa-miR-182 (auch hsa-miR-150), PRKCB: hsa-miR-486, PRKCG: hsa-miR-182, PRKCD: hsa-miR-486 (PA 2 für alle), PRKCE: hsa-miR-182, hsa-miR-486 (PA 3 beide miRNAs). Die Interaktionen sind in vereinfachter Form der KEGG-Ontologie, sowie aktueller Literatur entnommen.

Diese Beobachtung ist gut in Einklang zu bringen mit der beschriebenen lymphozytären Apoptose. Die miRNAs hsa-miR-150 und hsa-miR-342 scheinen proapoptotische Faktoren zu kontrollieren. Durch die beobachtete Reduktion deren Expression könnte die vorhergesagte negative Kontrolle entfallen und das zelluläre Gleichgewicht in Richtung Apoptose verschoben werden. Dieser Effekt

würde durch gesteigerte Expression von miRNAs verstärkt werden, die anti-apoptotische Faktoren regulieren, wie vermutlich hsa-miR-182 und hsa-miR-486. Ein solches Modell vermag auch den Zusammenhang der Korrelation der beobachteten miRNA-Expressionsänderungen mit klinischen Bild der Sepsis erklären [60]. Wenn das Ausmaß der Apoptose lymphatischer Zellen mit dem klinischen Bild korreliert und die Expression bestimmter miRNAs mit dem Ausmaß an Apoptose, dann korreliert logischerweise ebenfalls die Expression dieser miRNAs mit dem klinischen Bild.

Um die Verbindung zwischen *in-silico* Analyse und *in-vivo* Beobachtung herzustellen, bedarf es molekularbiologischer Methoden. Ein sich anschließender experimenteller Ansatz könnte in erster Instanz versuchen, einige der vorhergesagten miRNA-mRNA-Interaktionen zu validieren. Dazu wird durch Einbringen von Vektoren oder von pre-miRNAs die gewünschte miRNA artifiziell überexprimiert oder durch Einbringen von perfekt sequenzkomplementären RNA-Molekülen inhibiert [72]. Ein sichtbarer Effekt sollte durch die gezielte Mutation der 3'-UTR-Zielsequenz umkehrbar sein. Anschließend könnte zellspezifisch (Makrophagen, B- und T-Zellen) *in-vitro* untersucht werden, ob die Apoptose-Signalkaskade durch diese Art der veränderten miRNA-Expressionsmuster beeinflusst wird. Ebenso interessant wäre die Untersuchung der Abhängigkeit der Expression der vier beschriebenen miRNAs hsa-miR-150, hsa-miR-342, hsa-miR-182 und hsa-miR-486 von bekannten pro- und antiapoptotischen Faktoren. Dazu könnten die entsprechenden Zellkulturen 6h, 12h und 24h mit Interleukin-1, TNF-Alpha, Fas-Ligand und Interleukin-3 inkubiert und die jeweiligen miRNA-Konzentrationen

in den Zellen bestimmt werden. Anders als bisherige therapeutische Versuche durch beispielsweise selektive Blockade des Interleukin-1-Rezeptors wäre es nämlich durch Veränderung der miRNA-Expression möglich, Einfluss auf mehrere Schaltstellen der Zell-Balance Einfluss zu nehmen und möglicherweise effektivere therapeutische Optionen bereit zu stellen.

5.3 Intronische miRNAs

Integration der Lokalisationsinformation der miRNA-Gene zeigte (Kapitel 4.3), dass es sich bei hsa-miR-342 und hsa-miR-486 um intronische miRNAs handelt. Es konnte mehrfach gezeigt werden, dass die Expression intronischer miRNAs mit der Expression ihrer Host-Gene in vielen Fällen hoch korreliert. Diesem Phänomen liegt die Kotranskription von miRNA- und Host-Gen zugrunde [31]. Der Zweck dieser Koregulation auf Ebene des Transkriptoms ist nach wie vor nicht genau verstanden, allerdings liegt die Vermutung nahe, dass ein funktioneller Zusammenhang zwischen Host-Gen und miRNA bestehen könnte. Das Datenbanksystem wurde benutzt, um charakteristische Eigenschaften für eine Klasse von Genen zu identifizieren. Die Analyse zeigt, dass sich Host- und Zielgene in vielerlei Hinsicht ähneln (Tabelle 9): beide sind signifikant länger, haben mehr Introns und besitzen längere 3'-UTR-Sequenzen, als die jeweilige Kontrollgruppe. In der Subgruppenanalyse zeigte sich allerdings, dass lediglich die Host-Gene längere 3'-UTR-Sequenzen besitzen, die laut Vorhersage (Prediction Agreement ≥ 2) von ihrer eigenen miRNA reguliert werden (wie im Beispiel von hsa-miR-486 und ANK1). Die Transkription ist für die Zelle ein energieaufwendiger Prozess. Gene, die in hoher Zahl transkribiert werden müssen, sind daher tendenziell kurz, haben dementsprechend wenig Introns und

kurze regulatorische Sequenzen [107, 108]. Die Merkmale der Host- und Zielgene sind daher eher typisch für komplexe Genprodukte, deren Expression einer starken regulatorischen Kontrolle unterliegt. Dieser Gedanke wird gestützt durch die Tatsache, dass die 3'-UTR-Sequenzen der Host-Gene signifikant mehr AU-reiche Elemente enthalten. AU-reiche Elemente spielen eine zentrale Bedeutung für die Stabilität und die Überlebenszeit des Transkripts [109-111]. Erst kürzlich wurde diese Beobachtung mit der regulatorischen Maschinerie der miRNAs in Verbindung gebracht [112]. Interessant ist allerdings, dass weder *ANK1* noch *EVL* mehr AUUUA-Sequenzblöcke enthalten (relativ zur UTR-Länge). Eine mögliche Erklärung könnte in der fehlenden Analyse weiterer regulativer Elemente liegen. Viele weitere Faktoren beeinflussen die mRNA-Stabilität, wie beispielsweise die unlängst beschriebenen GU-reichen Elemente, die in dieser Analyse nicht berücksichtigt wurden [113, 114].

Die Möglichkeit der genomweiten Untersuchung von Host- und Ziel-Genen erlaubt die Charakterisierung dieser Gen-Gruppen. Die Ähnlichkeit der identifizierten Merkmale erhärtet die Theorie eines funktionellen Zusammenhangs. In dem hier beschriebenen Fall wäre es also sinnvoll, die bislang wenig charakterisierten Gene *ANK1* (hsa-miR-486) und *EVL* (hsa-miR-342) im Zusammenhang mit anderen Ziel-Genen zu untersuchen. In Anknüpfung an Kapitel 5.2 könnte beispielsweise ein experimenteller Ansatz Änderungen der Apoptose-Kaskade nach Silencing dieser zwei Gen-Produkte untersuchen. Ebenfalls interessant wäre es, die Interaktion von hsa-miR-486 mit seinem Host-Gen *ANK1* zu validieren und die biologische Signifikanz durch Mutation der potentiellen Bindungsstelle zu untersuchen.

5.4 Limitationen

Das in der vorliegenden Arbeit beschriebene System beinhaltet für die moderne medizinische Forschung viele Vorteile, jedoch ist die Nutzung eines solchen Systems nicht trivial. Das zu benutzende Prediction Agreement muss je nach Fragestellung adaptiert werden, abhängig davon, ob Sensitivität oder Spezifität priorisiert wird. Hypothesen, die aus der Analyse der KEGG-Datenbank biomolekularer Signalkaskaden generiert werden, basieren auf der Interpretation des Benutzers, wie auch auf verfügbarem Wissen. Beispielsweise sind *EVL* und *ANK1*, die Host-Gene von hsa-miR-342 und hsa-miR-486, keiner Signalkaskade der KEGG-Datenbank zugeordnet.

Von technischer Seite ist darauf zu achten, dass alle genomischen Koordinaten konsistent zueinander sind. Jeder neue Release einer UCSC Genome-Assembly benutzt leicht unterschiedliche Koordinaten, daher muss bei der Integration von Geninformationen (zum Beispiel RefSeq) immer berücksichtigt werden, auf welche Genome-Assembly sich die Positionskordinaten beziehen. Dementsprechend sollte auch versucht werden, möglichst aktuelle Versionen zu benutzen, da sich die Annotationen bestimmter Bereiche als Intron oder Exon im Laufe kurzer Zeit ändern können.

5.5 Zusammenfassung

Mit Beginn der Ära der Hochdurchsatz-Sequenzierung und Transkriptionsmessungen ist das Angebot an genetischer Information in den letzten Jahren

exponentiell gestiegen. Die Entdeckung der miRNAs als regulative Elemente mit weitreichendem Einfluss hat dabei eine Schlüsselrolle in der Erforschung von diagnostischen Möglichkeiten, pathogenetischen Prozessen und therapeutischen Konzepten vieler Krankheiten eingenommen. Mit der stetig wachsenden Informationsvielfalt steigt allerdings auch die Komplexität der Informationsverarbeitung und -aufbereitung. In der vorliegenden Arbeit wurde daher eine miRNA Datenbank konzipiert und evaluiert, die verfügbare Informationen handhabbar macht und bei der Generierung von Hypothesen hilft. Um Aussagen über die biologische Bedeutung von miRNAs treffen zu können, werden miRNA-mRNA-Interaktions-Vorhersagealgorithmen benutzt und so mögliche Ziel-mRNAs identifiziert. Aufgrund der beschriebenen Limitationen (Kapitel 2) wurde in dieser Arbeit ein Konsensusverfahren zur Ziel-Vorhersage etabliert und validiert, das das Prediction Agreement als Maß der Konfidenz einer Interaktion nutzt. Exemplarisch wurde dieses Verfahren eingesetzt, um vier miRNAs im Kontext der Apoptose-Signalkaskade zu beleuchten. Die Gene von zwei dieser vier miRNAs befinden sich in Introns proteinkodierender Gene (Host-Gene). Mithilfe der erstellten Datenbank ließen sich Charakteristika von Host-Genen extrahieren, die denen der Ziel-Gene ähneln. Die Summe der Beobachtungen erlaubt die Spekulation, dass die bislang biologisch wenig charakterisierten Host-Gene potentiell in funktionellem Zusammenhang zu den Ziel-Genen der miRNAs stehen.

Am Beispiel von bei Sepsis differentiell exprimierten miRNAs konnte in der vorliegenden Arbeit gezeigt werden, wie durch die Entwicklung einer bioinformatischen Datenbank schwer handhabbare Datenmengen und –

strukturen genutzt werden können, um die Entwicklung klinisch relevanter Hypothesen zu leiten. Die Möglichkeiten eines solchen Systems sind allerdings nicht ausgeschöpft. Je nach Fragestellung können weitere Daten integriert (Informationen über Promotor-Bereiche, Sequenzen, Protein-Protein-Interaktionen, weitere miRNA-/mRNA-Expressionsmessungen) und direkt analysiert werden. Mit zunehmendem Fortschritt biologischer Forschung und Methodik wird auch die informationsverarbeitende Methodik einen immer größeren Stellenwert einnehmen und der Bedarf an Datenbanksystemen und Konzepten zur strukturierten Analyse und Eingrenzung der Informationsvielfalt wird stetig steigen.

1. Watson JD, Crick FH: **Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.** *Nature* 1953, **171**:737-738.
2. Consortium EP, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, et al: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.** *Nature* 2007, **447**:799-816.
3. Lee RC, Feinbaum RL, Ambros V: **The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*.** *Cell* 1993, **75**:843-854.
4. Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Hayward DC, Ball EE, Degan B, Müller P, et al: **Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA.** *Nature* 2000, **408**:86-89.
5. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T: **Identification of novel genes coding for small expressed RNAs.** *Science* 2001, **294**:853-858.
6. Lau NC, Lim LP, Weinstein EG, Bartel DP: **An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*.** *Science* 2001, **294**:858-862.
7. Lee RC, Ambros V: **An extensive class of small RNAs in *Caenorhabditis elegans*.** *Science* 2001, **294**:862-864.
8. Griffiths-Jones S, Grocock RJ, Van Dongen S, Bateman A, Enright AJ: **miRBase: microRNA sequences, targets and gene nomenclature.** *Nucleic Acids Research* 2006, **34**:D140-144.
9. Griffiths-Jones S: **miRBase: the microRNA sequence database.** *Methods Mol Biol* 2006, **342**:129-138.
10. Griffiths-Jones S, Saini HK, Van Dongen S, Enright AJ: **miRBase: tools for microRNA genomics.** *Nucleic Acids Research* 2008, **36**:D154-158.
11. Lewis B, Burge C, Bartel D: **Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets.** *Cell* 2005, **120**:15-20.
12. Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M: **Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals.** *Nature* 2005, **434**:338-345.
13. Seggerson K, Tang L, Moss EG: **Two genetic circuits repress the *Caenorhabditis elegans* heterochronic gene *lin-28* after translation initiation.** *Dev Biol* 2002, **243**:215-225.
14. Baltimore D, Boldin MP, O'Connell RM, Rao DS, Taganov KD: **MicroRNAs: new regulators of immune cell development and function.** *Nat Immunol* 2008, **9**:839-845.
15. Thermann R, Hentze MW: ***Drosophila* miR2 induces pseudo-polysomes and inhibits translation initiation.** *Nature* 2007, **447**:875-878.
16. Shi X-B, Tepper CG, deVere White RW: **Cancerous miRNAs and their regulation.** *Cell Cycle* 2008, **7**:1529-1538.

17. Eulalio A, Huntzinger E, Izaurralde E: **Getting to the root of miRNA-mediated gene silencing.** *Cell* 2008, **132**:9-14.
18. Lee Y, Kim M, Han J, Yeom K-H, Lee S, Baek SH, Kim VN: **MicroRNA genes are transcribed by RNA polymerase II.** *EMBO J* 2004, **23**:4051-4060.
19. Kim VN, Han J, Siomi MC: **Biogenesis of small RNAs in animals.** *NATURE REVIEWS MOLECULAR CELL BIOLOGY* 2009, **10**:126-139.
20. Bohnsack MT, Czaplinski K, Gorlich D: **Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs.** *RNA* 2004, **10**:185-191.
21. Yi R, Qin Y, Macara IG, Cullen BR: **Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs.** *Genes Dev* 2003, **17**:3011-3016.
22. Lund E, Güttinger S, Calado A, Dahlberg JE, Kutay U: **Nuclear export of microRNA precursors.** *Science* 2004, **303**:95-98.
23. Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD: **Asymmetry in the assembly of the RNAi enzyme complex.** *Cell* 2003, **115**:199-208.
24. Nielsen CB, Shomron N, Sandberg R, Hornstein E, Kitzman J, Burge CB: **Determinants of targeting by endogenous and exogenous microRNAs and siRNAs.** *RNA* 2007, **13**:1894-1910.
25. Hofacker IL: **How microRNAs choose their targets.** *Nat Genet* 2007, **39**:1191-1192.
26. Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R: **Human RISC couples microRNA biogenesis and posttranscriptional gene silencing.** *Cell* 2005, **123**:631-640.
27. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E: **The role of site accessibility in microRNA target recognition.** *Nat Genet* 2007, **39**:1278-1284.
28. Petersen CP, Bordeleau M-E, Pelletier J, Sharp PA: **Short RNAs repress translation after initiation in mammalian cells.** *Mol Cell* 2006, **21**:533-542.
29. Kiriakidou M, Tan GS, Lamprinaki S, De Planell-Saguer M, Nelson PT, Mourelatos Z: **An mRNA m7G cap binding-like motif within human Ago2 represses translation.** *Cell* 2007, **129**:1141-1151.
30. Kim Y-K, Kim VN: **Processing of intronic microRNAs.** *EMBO J* 2007, **26**:775-783.
31. Baskerville S, Bartel DP: **Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes.** *RNA* 2005, **11**:241-247.
32. Thomson JM, Newman M, Parker JS, Morin-Kensicki EM, Wright T, Hammond SM: **Extensive post-transcriptional regulation of microRNAs and its implications for cancer.** *Genes Dev* 2006, **20**:2202-2207.
33. Obernosterer G, Leuschner PJF, Alenius M, Martinez J: **Post-transcriptional regulation of microRNA expression.** *RNA* 2006, **12**:1161-1167.

34. Baek D, Villén J, Shin C, Camargo FD, Gygi SP, Bartel DP: **The impact of microRNAs on protein output.** *Nature* 2008, **455**:64-71.
35. Lewis BP, Shih I-h, Jones-Rhoades MW, Bartel DP, Burge CB: **Prediction of mammalian microRNA targets.** *Cell* 2003, **115**:787-798.
36. Hofacker I, Fontana W, Stadler P, Bonhoeffer L: **Fast folding and comparison of RNA secondary structures.** *Monatshefte für Chemie/Chemical Monthly* 1994.
37. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS: **MicroRNA targets in Drosophila.** *Genome Biol* 2003, **5**:R1.
38. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS: **Human MicroRNA targets.** *PLoS Biol* 2004, **2**:e363.
39. Miranda KC, Huynh T, Tay Y, Ang Y-S, Tam W-L, Thomson AM, Lim B, Rigoutsos I: **A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes.** *Cell* 2006, **126**:1203-1217.
40. Rigoutsos I, Floratos A: **Combinatorial pattern discovery in biological sequences: The TEIRESIAS algorithm.** *Bioinformatics* 1998, **14**:55-67.
41. Doolittle RF, Hunkapiller MW, Hood LE, Devare SG, Robbins KC, Aaronson SA, Antoniades HN: **Simian sarcoma virus onc gene, v-sis, is derived from the gene (or genes) encoding a platelet-derived growth factor.** *Science* 1983, **221**:275-277.
42. Linsley PS, Schelter J, Burchard J, Kibukawa M, Martin MM, Bartz SR, Johnson JM, Cummins JM, Raymond CK, Dai H, et al: **Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression.** *Mol Cell Biol* 2007, **27**:2240-2252.
43. Wang X, El Naqa IM: **Prediction of both conserved and nonconserved microRNA targets in animals.** *Bioinformatics* 2008, **24**:325-332.
44. Wang X: **miRDB: a microRNA target prediction and functional annotation database with a wiki interface.** *RNA* 2008, **14**:1012-1017.
45. Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N: **Combinatorial microRNA target predictions.** *Nat Genet* 2005, **37**:495-500.
46. Sethupathy P, Corda B, Hatzigeorgiou AG: **TarBase: A comprehensive database of experimentally supported animal microRNA targets.** *RNA* 2006, **12**:192-197.
47. Urbich C, Kuehbacher A, Dimmeler S: **Role of microRNAs in vascular diseases, inflammation, and angiogenesis.** *Cardiovasc Res* 2008, **79**:581-588.
48. Barbarotto E, Schmittgen TD, Calin GA: **MicroRNAs and cancer: profile, profile, profile.** *Int J Cancer* 2008, **122**:969-977.
49. Fiore R, Siegel G, Schratt G: **MicroRNA function in neuronal development, plasticity and disease.** *Biochim Biophys Acta* 2008, **1779**:471-478.
50. Hariharan M, Scaria V, Pillai B, Brahmachari SK: **Targets for human encoded microRNAs in HIV genes.** *Biochemical and Biophysical Research Communications* 2005, **337**:1214-1218.

51. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, et al: **MicroRNA expression profiles classify human cancers.** *Nature* 2005, **435**:834-838.
52. Schetter AJ, Leung SY, Sohn JJ, Zanetti KA, Bowman ED, Yanaihara N, Yuen ST, Chan TL, Kwong DLW, Au GKH, et al: **MicroRNA Expression Profiles Associated With Prognosis and Therapeutic Outcome in Colon Adenocarcinoma.** *JAMA: The Journal of the American Medical Association* 2008, **299**:425-436.
53. Agirre X, Jiménez-Velasco A, San José-Enériz E, Garate L, Bandrés E, Cordeu L, Aparicio O, Saez B, Navarro G, Vilas-Zornoza A, et al: **Down-regulation of hsa-miR-10a in chronic myeloid leukemia CD34+ cells increases USF2-mediated cell growth.** *Mol Cancer Res* 2008, **6**:1830-1840.
54. Le Y, Xu L, Lu J, Fang J, Nardi V, Chai L, Silberstein LE: **FAK silencing inhibits leukemogenesis in BCR/ABL-transformed hematopoietic cells.** *Am J Hematol* 2009, **84**:273-278.
55. Grassmann R, Jeang K-T: **The roles of microRNAs in mammalian virus infection.** *Biochim Biophys Acta* 2008, **1779**:706-711.
56. Divakaran V, Mann D: **The emerging role of microRNAs in cardiac remodeling and heart failure.** *Circ Res* 2008, **103**:1072.
57. D'Alessandra Y, Devanna P, Limana F, Straino S, Di Carlo A, Brambilla PG, Rubino M, Carena MC, Spazzafumo L, De Simone M, et al: **Circulating microRNAs are new and sensitive biomarkers of myocardial infarction.** *European heart journal* 2010.
58. Dellinger RP, Carlet JM, Masur H, Gerlach H, Calandra T, Cohen J, Gea-Banacloche J, Keh D, Marshall JC, Parker MM, et al: **Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock.** *Crit Care Med* 2004, **32**:858-873.
59. Stearns-Kurosawa DJ, Osuchowski MF, Valentine C, Kurosawa S, Remick DG: **The Pathogenesis of Sepsis.** *Annual review of pathology* 2010.
60. Vasilescu C, Rossi S, Shimizu M, Tudor S, Veronese A, Ferracin M, Nicoloso M, Barbarotto E, Popa M, Stanciu O: **MicroRNA Fingerprints Identify miR-150 as a Plasma Prognostic Marker in Patients with Sepsis.** 2009.
61. Gaur A, Jewell DA, Liang Y, Ridzon D, Moore JH, Chen C, Ambros VR, Israel MA: **Characterization of microRNA expression levels and their biological correlates in human cancer cell lines.** *Cancer Research* 2007, **67**:2456-2468.
62. Tili E, Michaille J-J, Costinean S, Croce CM: **MicroRNAs, the immune system and rheumatic disease.** *Nature clinical practice Rheumatology* 2008, **4**:534-541.
63. Iorio MV, Ferracin M, Liu C-G, Veronese A, Spizzo R, Sabbioni S, Magri E, Pedriali M, Fabbri M, Campiglio M, et al: **MicroRNA gene expression deregulation in human breast cancer.** *Cancer Research* 2005, **65**:7065-7070.
64. Griffiths-Jones S: **The microRNA Registry.** *Nucleic Acids Research* 2004, **32**:D109-111.
65. Fox S, Filichkin S, Mockler TC: **Applications of ultra-high-throughput sequencing.** *Methods Mol Biol* 2009, **553**:79-108.

66. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, et al: **Accurate whole human genome sequencing using reversible terminator chemistry.** *Nature* 2008, **456**:53-59.
67. Morozova O, Marra MA: **Applications of next-generation sequencing technologies in functional genomics.** *Genomics* 2008, **92**:255-264.
68. Didiano D, Hobert O: **Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions.** *Nat Struct Mol Biol* 2006, **13**:849-851.
69. Sethupathy P, Megraw M, Hatzigeorgiou AG: **A guide through present computational approaches for the identification of mammalian microRNA targets.** *Nat Meth* 2006, **3**:881-886.
70. Alexiou P, Vergoulis T, Gleditsch M, Prekas G, Dalamagas T, Megraw M, Grosse I, Sellis T, Hatzigeorgiou AG: **miRGen 2.0: a database of microRNA genomic information and regulation.** *Nucleic Acids Research* 2010, **38**:D137-141.
71. Megraw M, Sethupathy P, Corda B, Hatzigeorgiou AG: **miRGen: a database for the study of animal microRNA genomic organization and function.** *Nucleic Acids Research* 2007, **35**:D149-155.
72. Selbach M, Schwanhäusser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N: **Widespread changes in protein synthesis induced by microRNAs.** *Nature* 2008, **455**:58-63.
73. Karolchik D, Hinrichs AS, Kent WJ: **The UCSC Genome Browser.** *Curr Protoc Bioinformatics* 2009, **Chapter 1**:Unit1.4.
74. Mangan ME, Williams JM, Kuhn RM, Lathe WC: **The UCSC genome browser: what every molecular biologist should know.** *Curr Protoc Mol Biol* 2009, **Chapter 19**:Unit19.19.
75. Pruitt KD, Tatusova T, Maglott DR: **NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.** *Nucleic Acids Research* 2005, **33**:D501-504.
76. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Research* 2000, **28**:27-30.
77. Li S-C, Tang P, Lin W-C: **Intronic microRNA: discovery and biological implications.** *DNA and Cell Biology* 2007, **26**:195-207.
78. Hammond SM: **MicroRNAs as oncogenes.** *Curr Opin Genet Dev* 2006, **16**:4-9.
79. Blenkinson C, Goldstein LD, Thorne NP, Spiteri I, Chin S-F, Dunning MJ, Barbosa-Morais NL, Teschendorff AE, Green AR, Ellis IO, et al: **MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype.** *Genome Biol* 2007, **8**:R214.
80. Zhao Y, Samal E, Srivastava D: **Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis.** *Nature* 2005, **436**:214-220.
81. Rajewsky N: **microRNA target predictions in animals.** *Nat Genet* 2006, **38 Suppl**:S8-13.
82. Doench JG, Petersen CP, Sharp PA: **siRNAs can function as miRNAs.** *Genes Dev* 2003, **17**:438-442.

83. Saetrom P, Heale BSE, Snøve O, Aagaard L, Alluin J, Rossi JJ: **Distance constraints between microRNA target sites dictate efficacy and cooperativity.** *Nucleic Acids Research* 2007, **35**:2333-2342.
84. Huber W, von Heydebreck A, Sülthmann H, Poustka A, Vingron M: **Variance stabilization applied to microarray data calibration and to the quantification of differential expression.** *Bioinformatics* 2002, **18 Suppl 1**:S96-104.
85. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: **Exploration, normalization, and summaries of high density oligonucleotide array probe level data.** *Biostatistics (Oxford, England)* 2003, **4**:249-264.
86. Rao Y, Lee Y, Jarjoura D, Ruppert AS, Liu C-G, Hsu JC, Hagan JP: **A comparison of normalization techniques for microRNA microarray data.** *Statistical applications in genetics and molecular biology* 2008, **7**:Article22.
87. Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, Smyth GK: **A comparison of background correction methods for two-colour microarrays.** *Bioinformatics* 2007, **23**:2700-2707.
88. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR: **Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care.** *Crit Care Med* 2001, **29**:1303-1310.
89. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, Reinhart CK, Suter PM, Thijs LG: **The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine.** *Intensive Care Med* 1996, **22**:707-710.
90. Vincent JL, De Mendonça A, Cantraine F, Moreno R, Takala J, Suter PM, Sprung CL, Colardyn F, Blecher S: **Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European Society of Intensive Care Medicine.** *Crit Care Med* 1998, **26**:1793-1800.
91. Monti D, Troiano L, Tropea F, Grassilli E, Cossarizza A, Barozzi D, Pelloni MC, Tamassia MG, Bellomo G, Franceschi C: **Apoptosis--programmed cell death: a role in the aging process?** *Am J Clin Nutr* 1992, **55**:1208S-1214S.
92. Wesche DE, Lomas-Neira JL, Perl M, Chung C-S, Ayala A: **Leukocyte apoptosis and its significance in sepsis and shock.** *J Leukoc Biol* 2005, **78**:325-337.
93. De Freitas I, Fernández-Somoza M, Essenfled-Sekler E, Cardier JE: **Serum levels of the apoptosis-associated molecules, tumor necrosis factor-alpha/tumor necrosis factor type-I receptor and Fas/FasL, in sepsis.** *Chest* 2004, **125**:2238-2246.

94. Hogquist KA, Nett MA, Unanue ER, Chaplin DD: **Interleukin 1 is processed and released during apoptosis.** *Proc Natl Acad Sci USA* 1991, **88**:8485-8489.
95. Xu YX, Wichmann MW, Ayala A, Cioffi WG, Chaudry IH: **Trauma-hemorrhage induces increased thymic apoptosis while decreasing IL-3 release and increasing GM-CSF.** *J Surg Res* 1997, **68**:24-30.
96. Kasten K, Tschöp J, Adediran S, Hildeman D, Caldwell C: **T Cells are potent early mediators of the host response to sepsis.** *Shock* 2010, **34**:327.
97. Hotchkiss RS, Swanson PE, Freeman BD, Tinsley KW, Cobb JP, Matuschak GM, Buchman TG, Karl IE: **Apoptotic cell death in patients with sepsis, shock, and multiple organ dysfunction.** *Crit Care Med* 1999, **27**:1230-1251.
98. Rajan G, Sleigh JW: **Lymphocyte counts and the development of nosocomial sepsis.** *Intensive Care Med* 1997, **23**:1187.
99. Hotchkiss RS, Swanson PE, Knudson CM, Chang KC, Cobb JP, Osborne DF, Zollner KM, Buchman TG, Korsmeyer SJ, Karl IE: **Overexpression of Bcl-2 in transgenic mice decreases apoptosis and improves survival in sepsis.** *J Immunol* 1999, **162**:4148-4156.
100. McDunn J, Turnbull I, Polpitiya A, Tong A, MacMillan S, Osborne D, Hotchkiss R, Colonna M, Cobb J: **Splenic CD4+ T cells have a distinct transcriptional response six hours after the onset of sepsis.** *Journal of the American College of Surgeons* 2006, **203**:375.
101. Flannery S, Bowie AG: **The interleukin-1 receptor-associated kinases: Critical regulators of innate immune signalling.** *Biochemical pharmacology* 2010.
102. Nuñez C, Cansino JR, Bethencourt F, Pérez-Utrilla M, Fraile B, Martínez-Onsurbe P, Olmedilla G, Paniagua R, Royuela M: **TNF/IL-1/NIK/NF-kappa B transduction pathway: a comparative study in normal and pathological human prostate (benign hyperplasia and carcinoma).** *Histopathology* 2008, **53**:166-176.
103. O'Neill LA, Greene C: **Signal transduction pathways activated by the IL-1 receptor family: ancient signaling machinery in mammals, insects, and plants.** *J Leukoc Biol* 1998, **63**:650-657.
104. Ito T, Deng X, Carr B, May WS: **Bcl-2 phosphorylation required for anti-apoptosis function.** *J Biol Chem* 1997, **272**:11671-11673.
105. Creson TK, Yuan P, Manji HK, Chen G: **Evidence for involvement of ERK, PI3K, and RSK in induction of Bcl-2 by valproate.** *J Mol Neurosci* 2009, **37**:123-134.
106. Craddock BL, Welham MJ: **Interleukin-3 induces association of the protein-tyrosine phosphatase SHP2 and phosphatidylinositol 3-kinase with a 100-kDa tyrosine-phosphorylated protein in hemopoietic cells.** *J Biol Chem* 1997, **272**:29281-29289.
107. Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA: **Selection for short introns in highly expressed genes.** *Nat Genet* 2002, **31**:415-418.

108. Kertesz S, Kerényi Z, Merai Z, Bartos I, Palfy T: **Both introns and long 3'-UTRs operate as cis-acting elements to trigger nonsense-mediated decay in** *Nucleic Acids Research* 2006.
109. Grzybowska E, Wilczynska A, Siedlecki J: **Regulatory Functions of 3' UTRs.** *Biochemical and Biophysical Research Communications* 2001.
110. Mazumder B, Seshadri V, Fox PL: **Translational control by the 3'-UTR: the ends specify the means.** *Trends in Biochemical Sciences* 2003, **28**:91-98.
111. Mignone F, Gissi C, Liuni S, Pesole G: **Untranslated regions of mRNAs.** *Genome Biol* 2002.
112. Jing Q, Huang S, Guth S, Zarubin T, Motoyama A, Chen J, Di Padova F, Lin S-C, Gram H, Han J: **Involvement of microRNA in AU-rich element-mediated mRNA instability.** *Cell* 2005, **120**:623-634.
113. Rattenbacher B, Beisang D, Wiesner DL, Jeschke JC, von Hohenberg M, St Louis-Vlasova IA, Bohjanen PR: **Analysis of CUGBP1 targets identifies GU-repeat sequences that mediate rapid mRNA decay.** *Mol Cell Biol* 2010, **30**:3970-3980.
114. Vlasova IA, Tahoe NM, Fan D, Larsson O, Rattenbacher B, Sternjohn JR, Vasdewani J, Karypis G, Reilly CS, Bitterman PB, Bohjanen PR: **Conserved GU-rich elements mediate mRNA decay by binding to CUG-binding protein 1.** *Mol Cell* 2008, **29**:263-270.

Danksagung

Viele Menschen haben mich auf dem Weg geprägt, der in dieser Arbeit mündete. Ganz besonders möchte ich mich bei meinem Doktorvater Prof. Dr. B. Zwißler für das in mich gesetzte Vertrauen und die stete Unterstützung und Förderung bedanken. Großer Dank gilt ebenfalls Frau Dr. Dr. S. Kreth für ihre unermüdliche Hilfsbereitschaft und Unterstützung bei der Durchführung und Gestaltung der Arbeit sowie Privatdozent Dr. M. Jacob für die vielen hilfreichen Diskussionen. Prof. Ohno-Machado und Dr. Pedro Galante gilt besonderer Dank dafür, dass sie mich für das Feld der miRNA-Bioinformatik begeisterten und mir grundlegende Konzepte bioinformatischer Algorithmen näher brachten. Der größte Dank aber gilt meiner Familie, meiner Frau Patricia und meinem Sohn Ben Luca, die mir eine liebende Umgebung gegeben und gleichermaßen genug Freiraum gelassen haben, um mich dieser Arbeit widmen zu können.