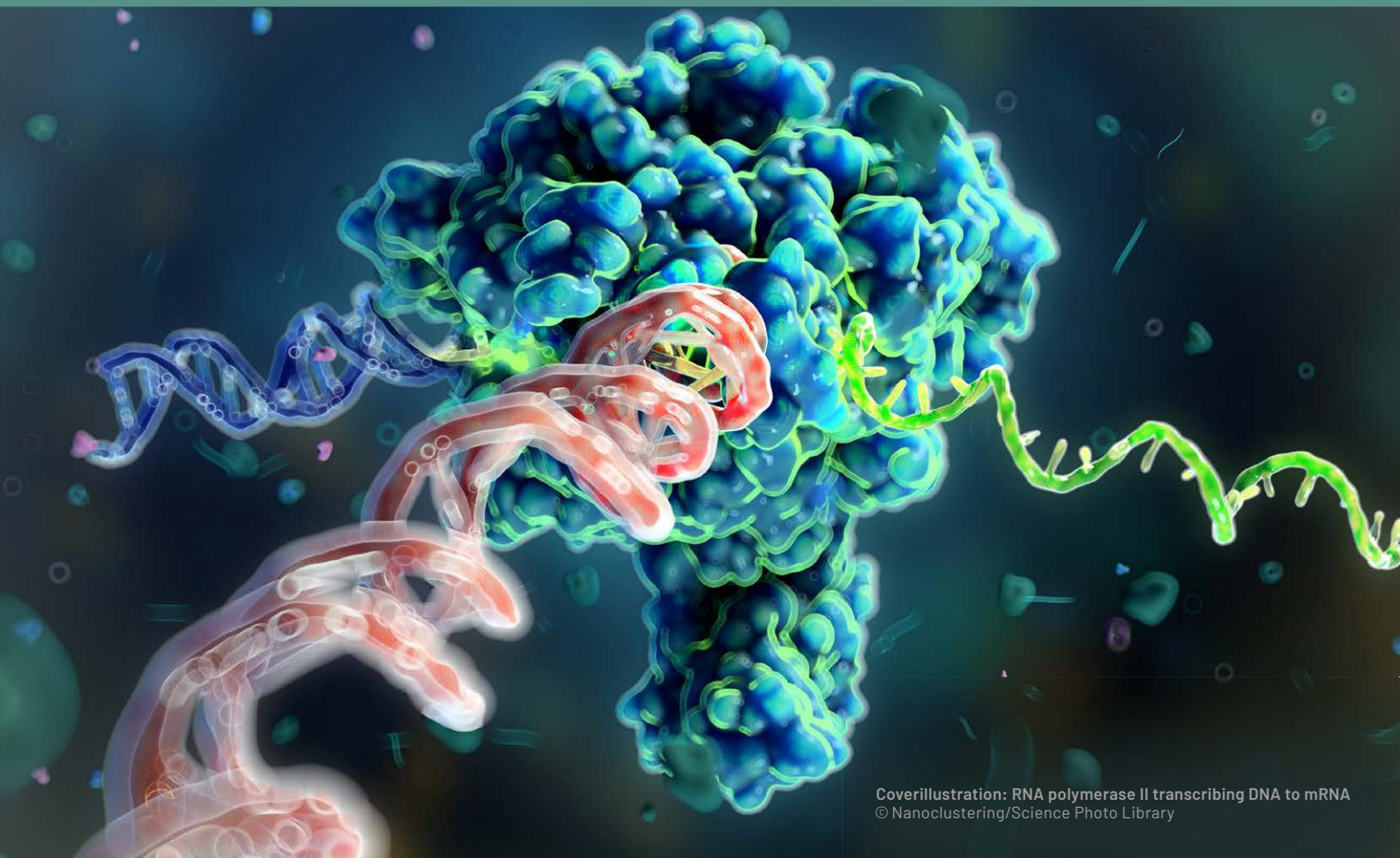


Im Fokus: RNA.
Eine aktuelle Bestandsaufnahme der
Arbeitsgruppe *Gentechnologiebericht*

In Focus: RNA.
A current stocktaking of the Working
Group *Gene Technology Report*

Boris Fehse, Jörn Walter, AG *Gentechnologiebericht* (Hrsg./Ed.)



Im Fokus: RNA. Eine aktuelle Bestandsaufnahme der Arbeitsgruppe *Gene Technologiebericht*

In Focus: RNA.
A current stocktaking of the Working
Group *Gene Technology Report*

Boris Fehse, Jörn Walter, *AG Gene Technologiebericht* (Hrsg./Ed.)

Berliner Institut für Gesundheitsforschung/Berlin Institute of Health at Charité (BIH)

Anna-Louisa-Karsch-Str 2 | 10178 Berlin, Germany

Redaktion/Editing: Hannah Schickl, Head of Office *Gene Technology Report*,
Research Fellow

Layout/Graphic: Polygraph Design

Druck/Printing: Druckerei Rüss

Übersetzung/Translation: Baker & Company

Auflage/Circulation: 500

Diese Veröffentlichung – ausgenommen anderweitig gekennzeichnete
Teile – ist unter der CC-Lizenz CC-BY 4.0 lizenziert.

This publication – except where otherwise noted – is licensed under
the Creative Commons License CC-BY 4.0.

<https://creativecommons.org/licenses/by/4.0>



ISBN: 978-3-00-074152-4

DOI: <http://dx.doi.org/10.17169/refubium-36831>

Berlin, Dezember/December 2022

Inhaltsverzeichnis

Vorwort	6
<i>Daniela Remus</i>	
1. Einleitung: Die Coronapandemie als Forschungsbeschleuniger	8
1.1 Der mRNA-Impfstoff gegen das Coronavirus	8
1.2 Voraussetzungen der Forschung	8
1.3 Die „berühmte“ mRNA	9
1.4 Viele verschiedene Arten von RNA	10
1.5 RNA-Technologie und klinische Forschung	11
<i>Mathias Munschauer und Jörg Vogel</i>	
2. Bekannte Typen von RNA und ihre biologischen Funktionen	13
2.1 Hintergrund	13
2.2 Messenger RNA (mRNA)	15
2.3 Ribosomale RNA (rRNA)	17
2.4 Transfer RNA (tRNA)	18
2.5 Small nucleolar RNA (snoRNA)	18
2.6 Small nuclear RNA (snRNA)	19
2.7 Ribozyme	19
2.8 MicroRNA (miRNA)	20
2.9 Small interfering RNA (siRNA)	21
2.10 Long non-coding RNA (lncRNA)	21
2.11 Circular RNA (circRNA)	23
2.12 Riboswitches	23
2.13 Bacterial small RNA (sRNA)	24
2.14 CRISPR-RNAs	26
2.15 Schlussbemerkung	27
2.16 Literaturverzeichnis	27

	<i>Jörn Walter und Nina Gasparoni</i>	
3.	Die Bedeutung von RNA für die molekulare Diagnostik	30
3.1	Überblick	30
3.2	Methodenspektrum der RNA-Analytik	31
3.3	Anwendungen der RNA-Analytik in der Diagnostik	33
3.4	Verarbeitung diagnostischer RNA-Daten	37
3.5	Herausforderungen RNA-gestützter Diagnostik	38
3.6	Ausblick	39
3.7	Literaturverzeichnis	40
	<i>Anke Sparmann und Jörg Vogel</i>	
4.	Die Erschließung der RNA-Welt für Therapeutika	42
4.1	Einleitung	42
4.2	ASO-Therapeutika	44
4.3	Therapeutika, die auf RNA-Interferenzmechanismen beruhen	46
4.4	Therapeutisches Genome-Editing	47
4.5	mRNA-basierte Impfstoffe	49
4.6	Ausblick	50
4.7	Literaturverzeichnis	51
	<i>Daniela Remus</i>	
5.	Ausblick: Fortschritt und Fragen	52
5.1	Aktueller Stand der RNA-Forschung	52
5.2	Wissenschaftliche Erwartungen und gesellschaftliche Akzeptanz	53
	Autorinnen und Autoren	56
	Mitglieder der Arbeitsgruppe Gentechnologiebericht	57
	Publikationen der Arbeitsgruppe Gentechnologiebericht	58

Vorwort

Die in kürzester Zeit entwickelten mRNA-Impfstoffe gegen SARS-CoV-2 waren einerseits eine von vielen verzweifelt ersehnte medizinische Antwort auf eine der größten bisherigen globalen Pandemien, spalteten aber andererseits auch die Gesellschaft in zwei nahezu unversöhnliche Lager von Impfbefürworterinnen und Impfgegnern. Diese Entwicklung illustrierte einmal mehr, dass Gentechnologien zu den einflussreichsten und zugleich gesellschaftlich umstrittensten Entwicklungen der vergangenen Jahrzehnte zählen. Der „Gentechnologiebericht – Monitoring und interdisziplinärer Dialog“ hat es sich seit 21 Jahren zur Aufgabe gemacht, einen sachlich-informierten Diskurs über die unterschiedlichen Gentechnologien in Deutschland zu fördern, um dadurch Wissenschaft, Gesellschaft und Politik näher zusammenzubringen.

Diesem Ziel verpflichtet widmet sich die vorliegende Broschüre dem schon angesprochenen, bis vor Kurzem in der breiten Öffentlichkeit weitgehend unbekanntem, aber dennoch sehr wichtigen Informationsträger – der seit Corona so heiß diskutierten RNA. Die erste Botschaft mag bereits überraschen – auch wenn mRNA-Impfstoffe in aller Munde sind, repräsentieren sie doch nur einen kleinen Teil eines wachsenden Forschungsgebiets, das bereits seit einem halben Jahrhundert bearbeitet wird und dessen Bedeutung für die Medizin erst in den letzten Jahren breitere Aufmerksamkeit und Anerkennung findet. Die vorliegende Bestandsaufnahme versucht daher, den gesamten Bereich der RNA-Forschung abzubilden und beschreibt den aktuellen Stand der Grundlagenforschung zur RNA-Biologie wie auch die immer wichtigere Rolle unterschiedlicher RNAs in der medizinischen Diagnostik und in der Therapie von Krankheiten. Tatsächlich wurden in den letzten Jahren mannigfaltige Arten von RNAs entdeckt und untersucht, die unterschiedlichste Funktionen innerhalb der Zellen von Bakterien, Pflanzen, Tieren und Menschen übernehmen. Angesichts rasant zunehmenden Wissens hinsichtlich der Rolle verschiedener RNAs bei einer Reihe von Krankheiten, lässt sich ihr Potenzial bei der Entwicklung einer personalisierten Medizin, also einer zielgerichteteren Diagnostik und Therapie kaum überbewerten.

Die als Mitherausgeber der Broschüre fungierende Arbeitsgruppe *Gentechnologiebericht* ist interdisziplinär zusammengesetzt aus Natur-, Geistes- und Sozialwissenschaftlern/-wissenschaftlerinnen.¹ Sie arbeitet die unterschiedlichen Anwendungen der Gentechnolo-

¹ In der vorliegenden Broschüre wird nicht einheitlich gegendert, die Entscheidung darüber wurde den jeweiligen Autoren/Autorinnen der einzelnen Beiträge überlassen.

gien sorgfältig auf und behält deren Entwicklungen langfristig im Blick. Die regelmäßig erscheinenden Veröffentlichungen zu spezifischen Themen richten sich überwiegend an eine breite Öffentlichkeit. Das für die vorliegende Publikation gewählte, neue „Im Fokus“-Format hat es ermöglicht, auf die aktuellen Entwicklungen vergleichsweise schnell zu reagieren. Sehr gefreut haben wir uns, dass wir neben dem AG-Mitglied Jörn Walter vier weitere ausgewiesene Expertinnen und Experten – Nina Gasparoni, Mathias Munschauer, Anke Sparmann und Jörg Vogel – als Autorinnen und Autoren gewinnen konnten, die in ihren Beiträgen verschiedene, zum Teil äußerst komplexe Aspekte der RNA-Biologie und ihrer praktischen Bedeutung anschaulich darstellen. In ihrer Einleitung und dem Ausblick hilft die Wissenschaftsjournalistin Daniela Remus nicht nur, die einzelnen Beiträge einzuordnen, sondern diskutiert auch Fragen der gesellschaftlichen Relevanz der RNA-Forschung.

Die namentlich gekennzeichneten Beiträge geben nicht unbedingt die Meinung der Herausgeber, der Arbeitsgruppe oder des BIH wieder – diese stehen jedoch hinter der Qualität der geleisteten Arbeit.

Herzlich danken möchte ich allen Mitwirkenden an dieser Publikation, insbesondere den Autorinnen und Autoren für Ihre Beteiligung und Frau Hannah Schickl aus der Geschäftsstelle für die Umsetzung des Projekts.

Die Arbeitsgruppe ist der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) für ihre langjährige Unterstützung seit ihrer Gründung im Jahr 2001 und der Friede Springer Stiftung für die finanzielle Förderung von 2019 bis 2021 zu großem Dank verpflichtet. Seit 2022 führt die Arbeitsgruppe ihre Metaforschung am Berlin Institute of Health (BIH) fort. Wir danken dem BIH sehr herzlich für die Weiterführung des Projektes und freuen uns auf die weitere Zusammenarbeit.

Boris Fehse

Sprecher der AG *Gentechnologiebericht* am BIH
Hamburg, im Oktober 2022

1. Einleitung: Die Coronapandemie als Forschungsbeschleuniger

Daniela Remus

1.1 Der mRNA-Impfstoff gegen das Coronavirus

Am 8. Dezember 2020 erhält die 90-Jährige Britin Margret Keenan als erster Mensch der Welt eine mRNA-Impfung gegen SARS-CoV-2. Die internationale Presse drängelt sich mit Kameras, Mikrofonen und Fotoapparaten, als die alte Dame in der englischen Stadt Coventry mit dem neu entwickelten Impfstoff von BioNTech/Pfizer geimpft wird. Lächelnd blickt sie in die Kameras und berichtet, wie glücklich und dankbar sie sei.

Die internationale Zulassung und Verabreichung dieses Impfstoffs, knapp ein Jahr nach Beginn der Coronapandemie, wird als Meilenstein im Kampf gegen das Virus gefeiert. Aber die Entwicklung dieses Vakzins in Rekordzeit ist weitaus mehr: Denn der Impfstoff basiert auf einem neuartigen Prinzip, der sogenannten mRNA-Technologie, die bis dahin nur Insidern bekannt war. Jetzt schreiben Journalisten Erklärstücke zu dieser Technologie, Freundinnen diskutieren die Vor- und Nachteile von Vektor- gegenüber mRNA-Impfstoffen, Forschungsgelder fließen und die Pharmafirmen umwerben die Pioniere dieser Forschung, die rund drei Jahrzehnte lang um Anerkennung und finanzielle Unterstützung kämpfen mussten. Tatsächlich repräsentieren die mRNA-Impfstoffe, die so plötzlich im Fokus des öffentlichen Interesses stehen, aber nur einen Teilbereich der RNA-Technologie, auf deren Basis Wissenschaftlerinnen und Wissenschaftler seit Jahren an innovativen Therapien, neuartigen Medikamenten und an der Entschlüsselung neu auftauchender Pathogene forschen.

1.2 Voraussetzungen der Forschung

In der Mitte des 20. Jahrhunderts gelang in der Zellbiologie ein erster großer Erkenntnis-schritt: Forschende konnten nachweisen, dass sich die Erbinformation einer Zelle, umgangssprachlich auch Erbgut, in der DNA, der Desoxyribonukleinsäure, befindet. Das war der Auftakt zur Entschlüsselung vieler molekularbiologischer Prozesse, die bis dahin nicht bekannt waren – auch weil die Technik dafür fehlte. Erst seit der zweiten Hälfte des 20. Jahrhunderts ermöglichten und ermöglichen spezielle Technologien, z. B. zur Aufdeckung molekularer Strukturen (Elektronenmikroskopie), zur Vermehrung der Erbinformation (rekombinante DNA), zur biochemischen Analyse der Abfolge der einzelnen DNA-Bausteine (Sequenzierung, später auch im Hochdurchsatz) verbunden mit der notwendigen Rechentechnik, diese Erkenntnisse.

Seit den frühen 1960er-Jahren ging es bei den molekularbiologischen Arbeiten auch verstärkt um die Erforschung der RNA. Dieses Kürzel steht für Ribonukleinsäure. Die RNA ist biochemisch ähnlich aufgebaut wie die DNA und für das Funktionieren einer Zelle unverzichtbar. Forschende fanden bald heraus, dass die Gesamt-RNA einer Zelle aus vielen verschiedenen RNA-Arten besteht, die jeweils unterschiedliche Funktionen erfüllen (siehe hierzu Munschauer/Vogel, Kapitel 2). Die unterschiedlichen RNAs sind aber alle gleichermaßen elementare Bausteine jeder einzelnen Zelle, jedes einzelnen Lebewesens, ob Bakterium, Pflanze, Tier oder Mensch. Ohne diese Moleküle fänden keine zellulären Prozesse statt, die Zellen könnten ihre genetische Information nicht umsetzen, sich nicht teilen oder Nährstoffe verarbeiten.

Die Erkenntnis, dass RNA für alle Lebewesen existenziell ist, bewog den Physiker, Biochemiker und Nobelpreisträger Walter Gilbert sogar dazu, 1986 die RNA-Welt-Hypothese aufzustellen, die viele Forschende seither teilen. Sie besagt, dass die Entstehung des Lebens grundlegend auf RNA zurückzuführen ist, weil die Baupläne für die ersten Lebewesen in Form von RNA gespeichert waren. Dass also ihr Erbgut nicht in der doppelsträngigen DNA, wie in heutigen Lebewesen, sondern in der einzelsträngigen RNA aufgehoben war. Evolutionär betrachtet wäre die RNA demnach nicht nur ein Vorläufer der DNA, sondern auch die entscheidende Bedingung für das Leben, wie wir es auf der Erde kennen.

1.3 Die „berühmte“ mRNA

Den nächsten zentralen Schritt der RNA-Forschung machten Wissenschaftlerinnen und Wissenschaftler in den 1990er-Jahren. Verschiedene Teams aus Frankreich, Deutschland und den USA arbeiteten an und mit der mRNA, der „messenger“ oder Boten-RNA. Also mit der RNA, die, wie es der Name schon sagt, als Bote dient, um die im Zellkern gespeicherte Erbinformation zu den Protein- bzw. Eiweißfabriken der Zelle zu bringen. Dafür kopiert sie einzelne DNA-Abschnitte spiegelbildlich. Das schon damals von den Forschenden postulierte Ziel: Die mRNA medizinisch nutzbar zu machen, entweder als neuartigen Impfstoff oder als Medikament.

Voraussetzung dafür war die Möglichkeit, eine synthetische mRNA herstellen zu können. Da diese Moleküle aber extrem kurzlebig sind, manche lösen sich bereits nach wenigen Minuten auf, stellte sich die Herstellung im Labor als ausgesprochen schwierig dar. Die Idee der Forschenden war dabei, die künstliche Boten-RNA als Transporter zu nutzen, ohne auf den Umweg über den Zellkern angewiesen zu sein – und die über die mRNA transportierten Informationen auf diese Weise direkt zu den Proteinfabriken der Zellen zu transportieren. So wie heute beispielsweise die Covid-Impfung funktioniert, bei der die Boten-RNA den Bauplan für das Spike-Protein des SARS-CoV-2-Virus in die Zellen bringt, die es produzieren und den Immunzellen präsentieren, damit diese es bei einer späteren Infektion sofort (wieder)erkennen und somit das Virus sehr schnell bekämpfen können.

Zu den Pionieren auf diesem Gebiet gehören, neben verschiedenen anderen, Katalin Karikó, die heute bei BioNTech forscht, oder Steve Pascolo und Ingmar Hoerr, die das Tübinger Unternehmen Curevac gründeten. Sie alle erarbeiteten, unabhängig voneinander, Stück für Stück die wissenschaftlichen Grundlagen dafür, die Boten-RNA als molekulargenetisches Werkzeug zu nutzen. Allerdings gelang der zweifelsfreie Nachweis für das Funktionieren jahrelang nicht. Technologisch-wissenschaftliche Schwierigkeiten und mangelnde finanzielle Unterstützung führten dazu, dass ihr Forschungsansatz ein Nischenthema blieb, trotz mancher Erfolge in Maus- und Rattenmodellen und trotz vereinzelter klinischer Studien. Er galt als zu kompliziert und zu teuer. Roche und Pfizer zogen sich als letzte große Pharmakonzerne 2011 aus diesem Segment zurück. Erst 2020, nach fast 30 Jahren Forschung, konnten die Wissenschaftlerinnen und Wissenschaftler der Mainzer Firma BioNTech und des US-amerikanischen Unternehmens Moderna durch die Entwicklung eines Impfstoffs gegen SARS-CoV-2 zeigen, dass die RNA-Technologie als Therapeutikum tatsächlich funktioniert, dass sie sicher, effektiv und wirksam ist.

Aber dieser Erfolg war ungeplant und auf gewisse Weise auch für die Wissenschaft überraschend. Denn die mRNA-Pioniere arbeiteten nicht an Impfstoffen gegen eine Infektionskrankheit. Ihr Ziel war es vielmehr, mit dieser Methode Impfungen gegen Krebserkrankungen zu entwickeln. Und zwar nicht als präventive Impfung, so wie beispielsweise eine Masernimpfung vor der Erkrankung schützen soll, sondern als therapeutische Option. Und das bedeutet, erst dann zu impfen, wenn sich ein Tumor bereits gebildet hat. Heutzutage unterscheiden die Forschenden mindestens 300 unterschiedliche Tumorerkrankungen, die umgangssprachlich unter dem Oberbegriff „Krebs“ zusammengefasst werden. Und selbst Krebsarten, die bis vor Kurzem als eine Krankheit zusammengefasst wurden, stellen sich oft anhand der modernen, molekularen Analysen als heterogen heraus. Auch dafür ist die RNA-Technologie nicht unerheblich, denn mit der sogenannten Einzelzellsequenzierung gelingt es Wissenschaftlerinnen und Wissenschaftlern, solche molekularen Unterschiede innerhalb eines Tumors zu identifizieren (siehe dazu ausführlich Walter/Gasparoni, Kapitel 3). Ein Umstand, der für eine erfolgreiche Therapie sehr wichtig ist. Deshalb, so der Gedanke hinter dem Konzept einer Krebsimpfung, müsse das Immunsystem mit den Informationen versorgt werden, die es benötigt, um passgenau die Tumorzellen zu entdecken und zu eliminieren, die den Abwehrzellen des Organismus bis dahin entgangen waren. Die Aufgabe der Informationsübermittlung soll eine im Labor entwickelte mRNA übernehmen, ähnlich wie jetzt bei der Covid-Impfung. Diesen Ansatz verfolgten beispielsweise Özlem Türeci und Ugur Sahin, das Gründerteam von BioNTech, sowie Ingmar Hopp und Steve Pascolo, die Gründer von Curevac.

1.4 Viele verschiedene Arten von RNA

Die anderen Arten von RNA sind außerhalb der Forschungsgemeinschaft weitaus weniger bekannt. Aber auch sie sind unverzichtbar. Diese Moleküle erfüllen zum Teil sehr spezielle Aufgaben: Sie regulieren, transportieren, konstruieren oder katalysieren Zellprozesse, wie ausführlich in Kapitel 2 (Munschauer/Vogel) dargestellt wird. Hier zur Einführung nur wenige Beispiele, auch wenn die Forschenden mittlerweile sehr viel mehr unterschiedliche RNAs entdeckt, analysiert und in ihren Funktionen verstanden haben: Zur Gesamt-RNA

einer Zelle gehören beispielsweise auch die tRNAs, die bereits in den 1960er-Jahren erstmalig beschrieben wurden. Das T steht für Transport. Allein davon kommen in einer menschlichen Zelle 31 verschiedene Variationen vor, die dafür zuständig sind, die in unserem Körper für die Eiweißsynthese benutzten 20 unterschiedlichen Aminosäuren zu den Proteinfabriken (Ribosomen) zu transportieren und in der korrekten Reihenfolge passgenau aneinander zu heften.

Den größten Anteil an der Gesamt-RNA aber haben die rRNAs. Das R steht für ribosomal und bedeutet, dass diese RNAs in den Ribosomen, den sogenannten Proteinfabriken der Zellen, vorkommen. Auch sie konnten bereits in den 1960er-Jahren erstmals beschrieben werden und machen, nach gegenwärtigem Kenntnisstand, rund 90 % der gesamten RNA-Menge einer Zelle aus.

Ende der 1970er-Jahre gelang es Forschenden darüber hinaus zu zeigen, dass zelluläre Prozesse, die von der mRNA angestoßen werden, unterbrochen bzw. gestört werden können. Zum Beispiel dadurch, dass sogenannte Oligonukleotide (kurze RNA- oder DNA-Abschnitte, die aus den Einzelbausteinen der Nukleinsäuren, den Nukleotiden, bestehen), benutzt werden, um über deren komplementäre Bindung an die mRNA zelluläre Abläufe zu beeinflussen. Damals gelang es einem Forschungsteam, ein sogenanntes Antisense-Oligonukleotid, kurz ASO, synthetisch herzustellen und damit die Replikation, also Vervielfältigung von Viren in einer menschlichen Zelle zu verhindern. Ein vielbeachteter Schritt, der erstmalig deutlich machte, dass RNAs das Potenzial haben, die Medikamentenentwicklung nachhaltig zu verändern (mehr dazu in Kapitel 4, Sparmann/Vogel). Das erste Arzneimittel, das auf diesem Prinzip beruht, kam 20 Jahre später auf den Markt. 1998 wurde Fomiviren zugelassen, ein Medikament, das gegen ein spezielles Herpesvirus wirkt, das bei Menschen mit einer Immunschwäche zur Erblindung führen kann.

Seit den 2000er-Jahren erforschen Wissenschaftlerinnen und Wissenschaftler darüber hinaus die siRNA. „Small interfering RNA“, auf Deutsch ungefähr kleine eingreifende RNA. Mit ihrer Hilfe lassen sich gezielt Gene abschalten. Ein Ansatz, der für die Behandlung und Heilung von Erbkrankheiten genutzt werden kann. 2018 erhielt das erste Medikament, das auf diesen Erkenntnissen gründet, die Zulassung. Medizinerinnen und Mediziner setzen Patisiran als Therapeutikum gegen die Polyneuropathie, eine chronische Nervenkrankheit, ein.

1.5 RNA-Technologie und klinische Forschung

Auch wenn längst noch nicht alle zellulären Prozesse verstanden sind, genauso wenig wie alle physiologischen Funktionen der unterschiedlichen RNAs, steht doch fest, dass diesen Molekülen eine Schlüsselfunktion bei der Erkennung und Behandlung vielfältigster Erkrankungen zukommt. Die Bedeutung der RNA-Technologie für die klinisch-medizinische Forschung ist kaum hoch genug einzuschätzen. Denn mit ihrer Hilfe lassen sich gesunde und kranke Zellen voneinander unterscheiden. Ein Ansatz, der sowohl in der Krebsdiagnostik, in der Immunologie und für die Infektionsforschung eine große Rolle spielt – also ein zusätzliches Instrument für die Diagnostik darstellt, wie in Kapitel 3 (Walter/Gasparoni) ausführlich entfaltet wird.

Außerdem können mithilfe der RNA-Sequenzierungen nicht nur Krankheiten diagnostiziert, sondern auch Krankheitserreger wie pathogene Organismen oder Viren identifiziert werden. So wie beispielsweise das SARS-CoV-2-Virus erst durch eine derartige Sequenzierung analysiert wurde. Was letztendlich als Voraussetzung dafür gilt, einen Impfstoff gegen Covid überhaupt entwickeln zu können.

Und da Forscherinnen und Mediziner durch die Erkenntnis der verschiedenen RNA-Arten interzelluläre Prozesse beeinflussen und steuern können, liegt in der RNA-Technologie auch der Schlüssel für die unterschiedlichsten therapeutischen Ansätze, wie in Kapitel 4 (Sparmann/Vogel) erklärt wird. Seien es Impfstoffe wie der gegen SARS-CoV-2 oder Medikamente, die, wie bereits gegen die spinale Muskelatrophie zugelassen, unerwünschte zelluläre Prozesse stoppen können.

RNAs werden permanent gebildet und wieder abgebaut. Diese Instabilität ist eine der zentralen Herausforderungen für die Forschenden, die sie als Medikament, Impfstoff oder Diagnostik-Tool nutzbar machen wollen. Andererseits aber liegt in dieser zeitlich begrenzten Aktivität eine große Chance für den medizinischen Einsatz. Denn synthetisierte RNA kann wichtige Prozesse anstoßen, wie an der Impfung gegen SARS-CoV-2 zu sehen, aber sie verbleibt nicht im Organismus und reichert sich dort deshalb auch nicht an. Auch wenn genau das die Sorge vieler Bürgerinnen und Bürger war, als die neuen mRNA-basierten Impfstoffe gegen Covid zugelassen wurden. Diese Kurzlebigkeit macht die RNAs für die Forschung so interessant, weil dadurch die Wahrscheinlichkeit unerwünschter Spätfolgen verringert werden kann.

Viele Forschende bezeichnen die RNA-Technologie als einen Paradigmenwechsel in der Medizin. Denn das Wissen um die RNA erweitert sich permanent. So konnte das internationale Projekt ENCODE, das seit 2003 läuft, beispielsweise zeigen, dass eine weitaus größere Zahl von RNAs in einer Zelle vorhanden ist, als zuvor bekannt. Auch wenn bis heute nicht verstanden ist, welche Funktion viele von ihnen haben, wird die Komplexität dieser Moleküle Schritt für Schritt sichtbar, wie Kapitel 2 (Munschauer/Vogel) zeigt. Und für die Wissenschaftlerinnen und Wissenschaftler eröffnen sich damit kontinuierlich weitere Möglichkeiten, diese Moleküle zu nutzen. Therapeutisch, beispielsweise als Impfung oder als Medikament, wie in Kapitel 4 (Sparmann/Vogel) ausführlich beschrieben, aber auch diagnostisch, um zu erkennen, welche zellulären Fehlfunktionen spezielle Krankheiten wie beispielsweise Tumoren, hervorrufen, wie in Kapitel 3 (Walter/Gasparoni) entfaltet.

2. Bekannte Typen von RNA und ihre biologischen Funktionen

Mathias Munschauer und Jörg Vogel

Die Bandbreite der heute bekannten RNA-Klassen und ihrer wichtigen und vielfältigen Funktionen in biologischen Systemen ist sehr hoch und es ist nicht ausgeschlossen, dass in Zukunft weitere Klassen beschrieben werden, insbesondere in Mikroorganismen. Dass über die RNA-Sequenz spezifische Interaktionen mit anderen RNA- oder DNA-Molekülen „programmiert“ werden können, ist für die Funktion verschiedener nicht-kodierender RNAs, aber auch für das Design von Antisense-RNA oder therapeutischen mRNAs von fundamentaler Bedeutung. Auch als therapeutisches Molekül ist RNA spätestens seit der erfolgreichen Entwicklung der COVID-19-mRNA-Impfstoffe in den Fokus gerückt und stellt eine neue Wirkstoffklasse mit immensem Potenzial für die Therapie verschiedenster Krankheiten dar.

2.1 Hintergrund

Desoxyribonukleinsäure oder kurz DNA fungiert in allen uns bekannten Domänen des Lebens als Träger der genetischen Information. Bei der Umsetzung der genetischen Information wird eine DNA-Sequenz, d. h. die Abfolge der Basen in einem DNA-Strang, in eine Aminosäuresequenz übersetzt, aus der sich wiederum Proteine zusammensetzen. Ein auf diese Weise in Form von DNA kodiertes Gen (ein DNA-Abschnitt) enthält daher den molekularen Bauplan zur Herstellung eines bestimmten Proteins, dessen Sequenz über die Abfolge der vier verschiedenen DNA-Basen (Adenin, Cytosin, Guanin oder Thymin) vorgegeben ist. Nach der Entdeckung von DNA als genetischer Informationsträger wurde jedoch schnell klar, dass Proteine nicht direkt von der berühmten DNA-Doppelhelix abgelesen und synthetisiert werden. Stattdessen ist für diesen Schritt die Synthese eines weiteren genetischen Informationsträgers notwendig: Der Ribonukleinsäure oder kurz RNA. Ein Gen wird also von der DNA abgelesen und in RNA umgeschrieben. Im Gegensatz zur doppelsträngigen DNA ist RNA ein einzelsträngiges Molekül, welches sich aber ebenfalls aus vier Grundbausteinen zusammensetzt: Adenin, Cytosin, Guanin oder Uracil. Die Sequenz dieser RNA-Grundbausteine definiert den Proteinbauplan. In eukaryotischen Systemen¹ liegt die DNA im Zellkern vor und wird dort in sogenannte Boten-RNA oder auch mRNA („messenger RNA“) umgeschrieben. Schon während der RNA-Synthese im Zellkern durchläuft ein mRNA-

¹ Als Eukaryoten bezeichnet man Lebewesen, deren Zellen im Unterschied zu Prokaryoten einen Zellkern aufweisen, z. B. Pilze, Pflanzen, Tiere oder Menschen.

Molekül verschiedene Prozessierungsschritte, die im Transport der fertigen oder auch „reifen“ mRNA aus dem Zellkern heraus in das Zytoplasma ihren Abschluss finden (siehe Abschnitt 2.2). Im Zytoplasma angekommen wird die mRNA dann von den Ribosomen der Zelle als Vorlage für die Proteinsynthese genutzt. Der Fluss der genetischen Information von DNA über mRNA hin zu Proteinen, nicht aber von Proteinen zurück zu mRNA, wird hierbei als das zentrale Dogma der molekularen Biologie bezeichnet (Crick, 1970). Im zentralen Dogma fungiert DNA sowohl als Träger der genetischen Information als auch als Matrize für die Synthese von mRNA. Die primäre Funktion von RNA schien zunächst auf die eines Botenmoleküls beschränkt.

Bereits in den frühen Tagen der Molekularbiologie zeigte sich jedoch, dass nicht jedes Gen im Genom auch für ein Protein kodiert. Die Entdeckung von ribosomaler RNA (rRNA) und sogenannter „transfer RNA“ (tRNA) in den 1960er-Jahren markierte den Beginn der Forschung an nicht-kodierenden RNAs. Sowohl rRNA als auch tRNA sind hochkonserviert, d. h. sie haben sich im Laufe der Evolution kaum verändert, und spielen eine fundamentale Rolle in der Proteinbiosynthese aller Organismen. Ribosomen sind die makromolekularen Komplexe, die die Proteinherstellung in Zellen ausführen, und setzen sich zu etwa zwei Dritteln aus rRNA und einem Drittel aus Proteinen zusammen. Der rRNA-Anteil ist hierbei sowohl strukturell als auch funktionell von entscheidender Bedeutung. Ebenso fundamental ist die Bedeutung der tRNA. Dieses RNA-Molekül ordnet jedem Basentriplett einer mRNA eine spezifische Aminosäure zu und ist somit für die Übersetzung des genetischen Codes in eine definierte Proteinsequenz verantwortlich.

Eine bahnbrechende Entdeckung im Feld der RNA-Biologie war die Identifizierung eines RNA-basierten Enzyms (Ribonuklease P oder kurz RNase P). Sidney Altman konnte zeigen, dass das Enzym RNase P kein Protein kodiert, sondern als katalytisch aktive nicht-kodierende RNA fungiert und einen Prozessierungsschritt in der Reifung von tRNAs katalysiert (Stark et al., 1978). Für die Entdeckung der katalytischen Eigenschaften von RNA (Stark et al., 1978; Cech et al., 1981), also der Fähigkeit von RNA, chemische Reaktionen zu beschleunigen, wurden Sidney Altman und Tom Cech im Jahre 1989 mit dem Nobelpreis ausgezeichnet. Katalytische Funktionen waren bis dato ausschließlich Proteinen zugeschrieben worden. Während DNA als Informationsträger dient und Proteine (u. a.) chemische Reaktionen katalysieren, zeigte sich also, dass RNA beide Funktionen ausfüllen kann. Dies führte zur Entwicklung der RNA-Welt-Hypothese durch Walter Gilbert im Jahre 1986, welche besagt, dass sich das frühe Leben ausschließlich auf Basis von RNA als Informationsträger und katalytisch aktivem Molekül entwickelt haben könnte, lange bevor die Evolution DNA für die Speicherung von genetischer Information und Proteine als hochspezialisierte und potente Enzyme hervorbrachte (Gilbert, 1986).

Diese frühen Arbeiten wiesen also bereits auf die außerordentlich wichtigen und vielfältigen Funktionen von RNA in biologischen Systemen hin und sollten die Grundlagen für weitere bahnbrechende Entdeckungen im Feld der RNA-Biologie schaffen. Im Zuge der Sequenzierung des humanen Genoms (Lander et al., 2001) zeigte sich dann, dass nur ein unerwartet kleiner Anteil des menschlichen Genoms für Proteine kodiert – etwa 98 % sind hingegen nicht-kodierend. Erstaunlicherweise wird aber ein Großteil (etwa 75 %) dieser

nicht-kodierenden Bereiche des Genoms in RNA umgeschrieben (Djebali et al., 2013), was auf die Existenz einer großen Zahl an nicht näher charakterisierten nicht-kodierenden RNAs hinwies. Heutzutage wissen wir, dass die Zahl an nicht-kodierenden RNAs nicht nur an die Zahl der proteinkodierenden mRNAs heranreicht, sondern diese möglicherweise sogar übersteigt (Iyer et al., 2015). Während für die Mehrheit dieser nicht-kodierenden RNAs noch keine genauen Funktionen bekannt sind, gibt es inzwischen gut charakterisierte Klassen an regulatorischen nicht-kodierenden RNAs. Im Folgenden werden die aktuell bekannten biologischen Funktionen von verschiedenen wichtigen RNA-Typen kurz zusammengefasst und anschaulich erläutert (siehe auch Tabelle 1).

2.2 Messenger RNA (mRNA)

Der allgemein am besten bekannte RNA-Typus ist ohne Zweifel die mRNA. Der Lebenszyklus einer mRNA beginnt mit dem Ablesen des entsprechenden Sequenzbereichs (dem Gen) auf DNA-Ebene durch ein Enzym – die sogenannte RNA-Polymerase. Für die Synthese von proteinkodierenden mRNAs ist in Eukaryoten das Enzym RNA-Polymerase II zuständig. Der Mechanismus der RNA-Synthese wird auch als Transkription bezeichnet und kann in die Stufen Initiation, Elongation und Terminierung unterteilt werden. Jede Phase der Transkription ist präzise reguliert, was die Transkription zu einem äußerst komplexen Vorgang macht (Cramer, 2019). Zunächst wird ein primäres Transkript, die sogenannte prä-mRNA hergestellt. Erstaunlicherweise liegt ein eukaryotisches Gen auf DNA-Ebene aber nicht in Form einer kontinuierlichen proteinkodierenden Sequenz vor. Stattdessen sind kodierende Bereiche, auch Exons genannt, von einer variablen Zahl an nicht-kodierenden Sequenzen, den sogenannten Introns, unterbrochen. Um eine funktionierende mRNA zu erhalten, die von Ribosomen in eine Aminosäuresequenz übersetzt werden kann, müssen diese Introns aus der prä-mRNA herausgeschnitten werden. Der hierfür verantwortliche Prozess wird mRNA-Spleißen genannt und wurde in den 1970er-Jahren von Phillip Sharp beschrieben (Berget et al., 1977). Das mRNA-Spleißen wird von einem makromolekularen Komplex, dem sogenannten Spleißosom ausgeführt. Das Spleißosom ist ein hochdynamischer Komplex, der sich aus einer Vielzahl an Proteinen und fünf evolutionär hochkonservierten nicht-kodierenden RNAs, den sogenannten „small nuclear RNAs“, zusammensetzt (Wahl et al., 2009). Für die katalytische Funktion des Spleißosoms sind wiederum insbesondere diese RNA-Komponenten von großer Bedeutung.

Die Entdeckung des mRNA-Spleißens zeigte letztendlich auch eine völlig neue Ebene der Genregulation auf: Das alternative Spleißen. Aus der gleichen prä-mRNA können durch alternatives Spleißen einzelne Exons in regulierter Weise herausgeschnitten oder integriert werden, sodass anhand alternativ gespleißter mRNAs unterschiedliche Proteinprodukte (sogenannte Spleiß-Isoformen) hergestellt werden können. Dies vergrößert die Diversität des Proteoms² und kann für regulatorische Prozesse von enormer Bedeutung sein. Die genaue molekulare Charakterisierung von Spleißvorgängen ermöglichte auch die Entwicklung von nukleinsäurebasierten Molekülen, die die Bindung des Spleißosoms an spezifische Spleißstellen modulieren und somit beeinflussen, welches Proteinprodukt gebildet

² Als Proteom wird die Gesamtheit aller Proteine in einem Organismus oder einer Zelle bezeichnet.

wird (Cartegni/Krainer, 2003). Dieses Prinzip wird, wie in Kapitel 4 (Sparmann/Vogel) beschrieben, erfolgreich für die Therapie der spinalen Muskelatrophie (SMA) eingesetzt.

Das Spleißen findet bereits co-transkriptional, d. h. während der Transkription, statt und auch bei der Terminierung der mRNA-Synthese kommt es zu weiteren RNA-Reifungsprozessen. Mit Erreichen eines Sequenzsignals in der prä-mRNA wird das neu synthetisierte Transkript am 3'-Ende³ geschnitten und polyadenyliert, d. h. mit einem Schwanz von 20–100 Adenosin-Nukleotiden⁴ versehen (Moore, 2005). Dieser Poly(A)-Schwanz schützt das Transkript vor dem Abbau und stimuliert gleichzeitig die mRNA-Translation. Interessanterweise scheint diese Stimulation der translationalen Aktivität auf einer Interaktion zwischen dem Poly(A)-Schwanz und einem weiteren funktionalen Element am anderen Ende des mRNA-Moleküls zu beruhen: Der 5'-Kappe oder auch 5'-Cap. Schon während der Transkription wird über einen enzymatischen Vorgang mit mehreren Zwischenstufen eine schützende Kappenstruktur, die häufig aus einem modifizierten Guanin-Nukleotid besteht, am 5'-Ende der mRNA angebracht (Moore, 2005). Ähnlich des Poly(A)-Schwanzes schützt auch die 5'-Kappe die RNA vor Abbau, stimuliert die Translation und ist des Weiteren auch für den Export der reifen mRNA aus dem Zellkern ins Zytoplasma wichtig (Moore, 2005). Nach dem mRNA-Export, der von verschiedenen RNA-Bindeproteinen reguliert wird und über kleine Öffnungen in der Kernmembran, dem sogenannten Kernporenkomplex, stattfindet, beginnt der zytoplasmatische Lebenszyklus der reifen mRNA.

Von zentraler Bedeutung für die Genexpression ist natürlich die Translation der mRNA zu Protein durch die Ribosomen. Die Translation ist ähnlich der Transkription unterteilt in Initiation, Elongation und Terminierung. Hierbei wird insbesondere der Schritt der Initiation aktiv von einer Vielzahl an RNA-Bindeproteinen und kleinen nicht-kodierenden RNAs wie etwa mikroRNAs („microRNA“) reguliert (Jackson et al., 2010). Die Translation beginnt am 5'-Ende der mRNA und beruht auf der Assoziation einer Reihe von eukaryotischen Initiationsfaktoren, welche die Bindung der kleinen ribosomalen Untereinheit,⁵ gefolgt von der großen ribosomalen Untereinheit unterstützen, wodurch schließlich ein translationskompetentes Ribosom gebildet wird (Jackson et al., 2010). Das Ablesen des kodierten Proteins beginnt an dem sogenannten Startcodon⁶ (RNA-Sequenz: AUG), wofür eine mit der Aminosäure Methionin beladene tRNA notwendig ist (siehe Abschnitt 2.4). Die Termination der Translation erfolgt wiederum an einem ebenfalls charakteristischen Basentriplett, dem Stopcodon.⁷ Die Sequenz einer mRNA beginnt jedoch nicht mit dem Startcodon und endet mit dem Stopcodon. Stattdessen enthält quasi jede zelluläre mRNA sogenannte nicht-translatierte Bereiche, die UTRs („un-translated regions“) an den 5'- und 3'-Enden. UTRs

3 Die Enden eines RNA-Moleküls werden anhand ihrer chemischen Struktur benannt. RNA setzt sich aus vier verschiedenen Nukleotiden zusammen. Das Rückgrat der RNA besteht aus Ribose, also einem Zucker mit fünf Kohlenstoffatomen, die in nummerierter Anordnung vorliegen. Am sogenannten 5'-Ende einer RNA befindet sich eine Phosphatgruppe am 5. Kohlenstoffatom der Ribose. Am 3'-Ende hingegen befindet sich eine Hydroxylgruppe am 3. Kohlenstoffatom der Ribose.

4 Nukleotide sind die kleinsten Bausteine von Nucleinsäuren (DNA und RNA).

5 Ein Ribosom setzt sich aus zwei Untereinheiten zusammen: die große und die kleine Untereinheit. Beide Untereinheiten bestehen ihrerseits wiederum aus einer Vielzahl an Proteinen und der rRNA. Nur wenn sich beide Untereinheiten an eine mRNA anlagern, kann die Translation, d. h. das Ablesen der mRNA und die Synthese eines Proteins stattfinden.

6 Ein Codon bezeichnet ein RNA-Basentriplett, dem eine spezifische Aminosäure zugeordnet wird. Die Zuordnung erfolgt über die Anlagerung eines zum Codon komplementären Anticodons einer tRNA. Spezialisierte Codons, die den Start bzw. das Ende der proteinkodierenden Sequenz markieren, werden als Start- bzw. Stopcodons bezeichnet. Startcodons bewirken den Einbau der ersten Aminosäure Methionin, wohingegen an Stopcodons keine Aminosäure eingebaut wird, sodass die Proteinsynthese abbricht.

7 Es existieren drei verschiedene Stopcodons mit den RNA-Sequenzen UGA, UAA und UAG.

enthalten regulatorisch wichtige Sequenzbereiche wie etwa Bindungsstellen für RNA-Bindeproteine oder microRNAs und dienen dazu, die Stabilität und translationale Aktivität für jede mRNA genau zu regulieren, worüber die Menge an jeweils abgelesenem Protein gesteuert wird. Neben einer Reihe von Qualitätskontrollmechanismen, die defekte mRNAs erkennen und gezielt abbauen, kommt es im Zytoplasma unweigerlich zu einer Verkürzung des Poly(A)-Schwanzes, was eine Destabilisierung der mRNA zur Folge hat. Mit dem Abbau endet der Lebenszyklus einer mRNA.

Da der molekulare Bauplan einer mRNA immer dem gleichen Schema folgt (5'-Cap, 5'-UTR, proteinkodierende Sequenz inkl. Start- und Stopcodons, 3'-UTR, Poly(A)-Schwanz), ist es mit relativ einfachen Methoden der Molekularbiologie möglich, ein synthetisches mRNA-Molekül herzustellen. Durch die Kombination der molekularen Bausteine lässt sich also der Bauplan eines Proteins erstellen, der, wenn in eine Zelle eingebracht, von dieser interpretiert und in ein Proteinprodukt übersetzt wird. Die Zielzelle übernimmt also die Produktion des gewünschten Proteins, wenn ihr der entsprechende Bauplan zur Verfügung gestellt wird. Dieses Prinzip liegt den vielfältigen Anwendungsmöglichkeiten der mRNA-Therapie (siehe Sparmann/Vogel, Kapitel 4) zugrunde.

2.3 Ribosomale RNA (rRNA)

Ribosomale RNA wurde zusammen mit tRNA bereits in den 1960er-Jahren entdeckt, was sicherlich auch durch ihre enorme zelluläre Häufigkeit begünstigt wurde – ribosomale RNA kann bis zu 90 % der gesamten RNA-Menge in einer Zelle ausmachen. Es ist daher nicht verwunderlich, dass für die Produktion von solch großen Mengen rRNA viele Genkopien benötigt werden. Gene, die für rRNA kodieren, werden auch rDNA genannt und liegen im humanen Genom auf mehreren Chromosomen in tandemartig hintereinanderliegenden Kopien vor. Nicht alle rDNA-Bereiche werden aktiv abgelesen. Die aktiven Bereiche liegen konzentriert in einer als Kernkörperchen bezeichneten Region des Zellkerns vor. Ähnlich der mRNA-Synthese ist für das Ablesen der rDNA ein spezielles Enzym, die RNA-Polymerase I notwendig. Zunächst wird ein langes prä-rRNA-Transkript, die sogenannte 45S-rRNA generiert. Aus dieser prä-rRNA werden dann durch verschiedene aufeinanderfolgende Prozessierungsschritte drei kleinere rRNAs gewonnen, die 28S-, 18S-, und 5,8S-rRNAs. Für die Prozessierung und Modifikation dieser rRNAs ist eine weitere Klasse an nicht-kodierenden RNAs, die sogenannten „small nucleolar RNAs“ (snoRNAs) zuständig (siehe Abschnitt 2.5). Eine weitere rRNA, die 5S-rRNA wird von einem anderen Enzym, nämlich der RNA-Polymerase III, an anderer Stelle des Genoms synthetisiert. Von allen vier rRNAs werden gleiche Mengen für den Zusammenbau der Ribosomen benötigt. Hierzu werden die im Zytoplasma synthetisierten ribosomalen Proteine in das Kernkörperchen transportiert, wo sich dann zusammen mit den verschiedenen ribosomalen RNAs die große und kleine ribosomale Untereinheit zusammenfinden, bevor diese schließlich wieder in das Zytoplasma ausgeschleust werden.

Bereits in den 1990er-Jahren mehrten sich die Hinweise darauf, dass intakte rRNA für die katalytische Aktivität des Ribosoms von besonderer Bedeutung ist (Noller et al., 1992). Dank hochauflösender Strukturaufnahmen (Yan et al., 2015) wissen wir heute, dass das

katalytische Zentrum des Ribosoms von RNA-RNA-Interaktionen dominiert wird, die für die Verknüpfung von Aminosäuren im Rahmen der Proteinsynthese essenziell sind.

Ribosomale RNA gehört zur Grundausstattung einer jeden lebenden Zelle und war wahrscheinlich bereits in den ersten lebenden Organismen vorhanden. Da rRNA in allen Organismen die gleiche molekulare Funktion ausführt, geht man davon aus, dass rRNA nur selten von einem Organismus auf den anderen übertragen wird und die Sequenz eines rRNA-Gens daher nicht nur die Entwicklungsgeschichte dieses einen Gens, sondern des gesamten Organismus widerspiegelt. Aus diesem Grund eignet sich eine Analyse der rDNA-Sequenz hervorragend dafür, die verwandtschaftlichen Beziehungen zwischen verschiedenen Organismen aufzuklären (Woese et al., 1990).

2.4 Transfer RNA (tRNA)

In den 1960er-Jahren isolierte Robert Holley die erste tRNA und löste nicht nur deren Sequenz auf, sondern schlug auch die heute noch bekannte Kleeblattstruktur vor (Holley et al., 1965a; Holley et al., 1965b). Zur gleichen Zeit fanden Marshall Nirenberg und Heinrich Matthaei heraus, dass das Basentriplett UUU der Aminosäure Phenylalanin zugeordnet werden kann (Nirenberg/Matthaei, 1961), was die elementare Rolle von tRNAs bei der Übersetzung des genetischen Codes aufzeigte. Eine tRNA ist im Regelfall etwa 80 Nukleotide lang und besitzt ein spezifisches Basentriplett in der mittleren Schleife, das sogenannte Anticodon. Während der Translation können sich nur solche tRNAs, die ein Anticodon besitzen, welches komplementär zu dem dazugehörigen Basentriplett einer mRNA ist, dort anlagern. Da jedes tRNA-Molekül am 3'-Ende mit einer spezifischen Aminosäure verknüpft ist, wird über die tRNA während der Translation jedem mRNA-Codon eine spezifische Aminosäure zugeordnet. So wird der genetische Code von einer mRNA-Sequenz in eine Aminosäuresequenz übersetzt. Für die Arbeiten, die diesen Mechanismus aufzeigten, wurde Marshall Nirenberg gemeinsam mit Robert Holley und Har Gobind Khorana 1968 mit dem Nobelpreis ausgezeichnet.

2.5 Small nucleolar RNA (snoRNA)

Small nucleolar RNAs sind nicht-kodierende RNAs mit einer Länge von etwa 60 bis 300 Nukleotiden, die oft in Introns von prä-mRNAs kodiert werden und zu den am häufigsten vorkommenden nicht-kodierenden RNAs in Eukaryoten zählen. Die Biosynthese von intronischen snoRNAs beginnt mit dem zuvor bereits beschriebenen mRNA-Spleißen (Abschnitt 2.2), wodurch das Intron in zirkularisierter Form (dem sogenannten Lariat oder Lasso) freigesetzt wird. Es folgen verschiedene Prozessierungsschritte, um das Lariat zu linearisieren und die reife snoRNA auf die benötigte Größe zu trimmen (Kufel/Grzechnik, 2019). Wie die meisten funktionellen RNAs auch, arbeiten snoRNAs mit Proteinen zusammen und bilden sogenannte Ribonukleoproteinkomplexe, die auch als „small nucleolar ribonucleoprotein particle“ oder kurz snoRNP bezeichnet werden (Matera et al., 2007). Die Rolle der snoRNA-Komponente eines solchen Komplexes kommt der eines molekularen „Guides“ gleich: snoRNAs sind dafür zuständig, die entsprechenden Proteine des Komplexes an eine Zielsequenz in einer anderen RNA zu leiten. Wie eingangs bereits dargelegt, spielen

snoRNAs eine wichtige Rolle bei der Modifikation und Prozessierung von anderen RNAs, insbesondere der rRNA, und sind daher auch im Kernkörperchen, dem Ort der rRNA-Synthese und -Prozessierung zu finden. Es gibt zwei wichtige Klassen an snoRNAs, die rRNA modifizieren und aufgrund ihrer Sekundärstruktur und Aktivität in C/D-Box- und H/ACA-Box-snoRNAs unterschieden werden können (Kufel/Grzechnik, 2019).

2.6 Small nuclear RNA (snRNA)

Small nuclear RNAs sind kurze, hochkonservierte nicht-kodierende RNAs von 100 bis 300 Nukleotiden Länge, die essenzieller Bestandteil des Spleißosoms sind. Da das mRNA-Spleißen im Zellkern stattfindet, sind auch snRNAs dort zu finden. Die snRNAs werden von den RNA-Polymerasen II und III synthetisiert und kommen in sehr hohen Kopienzahlen vor. Aufgrund ihres hohen Uracil-Anteils werden snRNAs als U1-snRNA, U2-snRNA, etc. benannt. Das Spleißosom setzt sich aus fünf verschiedenen snRNAs zusammen (U1, U2, U4, U5 und U6) (Matera/Wang, 2014). Jede snRNA interagiert wiederum mit spezifischen Proteinen, insbesondere den sogenannten Sm- und LSm-Proteinen, um funktionale snRNP-Komplexe zu bilden (Wahl et al., 2009). Im Gegensatz zum Ribosom ist das Spleißosom jedoch kein statischer Komplex, sondern durchläuft je nach Fortschritt der Spleißreaktion verschiedene Stadien, die durch genauestens festgelegte Assoziation und Dissoziation von Protein- und RNA-Untereinheiten charakterisiert sind (Wahl et al., 2009). Die Rolle der snRNA liegt in der Erkennung von Sequenzelementen in prä-mRNAs, wie etwa den 5'- und 3'-Spleißstellen. Somit sind snRNAs für die korrekte Erkennung von Exon-Intron-Übergängen zuständig. Auch die Entfernung der nicht-kodierenden intronischen Sequenzen wird wesentlich von RNA-RNA-Interaktionen katalysiert (Matera/Wang, 2014). Wie bereits in Abschnitt 2.2 dargelegt, ist alternatives Spleißen ein wichtiger regulatorischer Vorgang, der nicht nur die Proteinvierfalt erhöht, sondern auch Proteine mit ganz unterschiedlichen Funktionen hervorbringen kann. Durch die gezielte Modulation der RNA-RNA-Interaktionen zwischen snRNAs und einer Ziel-prä-mRNA lässt sich das Ergebnis der Spleißreaktion zielgerichtet beeinflussen, was z. B. in der oben angesprochenen SMA-Therapie erfolgreich angewendet wird.

2.7 Ribozyme

Als Ribozyme werden katalytisch aktive RNA-Moleküle bezeichnet, die chemische Reaktionen katalysieren (Doherty/Doudna, 2001). Die bereits beschriebene RNase P und auch die Gruppe der selbstspleißenden Introns⁸ gehören zu der Klasse der Ribozyme und spielen beide eine wesentliche Rolle bei der Entdeckung der katalytischen Aktivität von RNA (siehe Abschnitt 2.1). Zu den wichtigsten Ribozymen einer Zelle zählen auch das Ribosom und das Spleißosom. Zwar handelt es sich in beiden Fällen um makromolekulare Protein-RNA-Komplexe mit einer Vielzahl an Proteinuntereinheiten, die katalytische Reaktion selbst wird aber im Ribosom von der 28S-rRNA und im Spleißosom von den snRNAs ausgeführt. Ein wesentlicher Unterschied zwischen Ribozymen und proteinbasierten Enzymen liegt in

⁸ Einige RNAs können die Entfernung von Introns selbst katalysieren und benötigen für diese Aufgabe keine großen Proteinkomplexe wie etwa das Spleißosom. Diese RNAs werden auch als autokatalytische oder selbstspleißende Introns bezeichnet und können in Gruppe-I- und Gruppe-II-Introns unterschieden werden.

der Geschwindigkeit, mit der die katalysierte Reaktion stattfinden kann, weniger in der Vielfalt an möglichen Reaktionen (Emilsson et al., 2003). Die Entdeckung der Ribozyme spielte bei der Entwicklung der RNA-Welt-Hypothese eine wichtige Rolle.

2.8 MicroRNA (miRNA)

Die erstmalige Beschreibung einer microRNA im Fadenwurm *Caenorhabditis elegans* (*C. elegans*) im Jahre 1993 (Lee et al., 1993) und die spätere Entdeckung einer weiteren hochkonservierten microRNA in einer Vielzahl von Organismen (Pasquinelli et al., 2000; Reinhart et al., 2000) waren bedeutende Ereignisse in der modernen molekularbiologischen Forschung. Im Vergleich zu anderen regulatorischen nicht-kodierenden RNAs sind reife microRNAs geradezu winzig: Aus einem primären RNA-Polymerase-II-Transkript mit einer charakteristischen Haarnadelstruktur⁹ wird zunächst eine etwa 60 Nukleotide lange Vorläufer-RNA, die sogenannte prä-microRNA herausgeschnitten. Dieser erste Prozessierungsschritt erfolgt im Zellkern und wird von dem sogenannten Mikroprozessorkomplex und seiner Untereinheit Drosha durchgeführt (Bartel, 2004). Mithilfe von Exportproteinen gelangt die prä-microRNA dann ins Zytoplasma, wo es zu einem zweiten Prozessierungsschritt kommt: Das Enzym Dicer schneidet eine etwa 21 Nukleotide lange Duplex-RNA aus dem Vorläufermolekül heraus (Bartel, 2004). Reife microRNAs sind also klein genug, dass sie lange Zeit schlicht übersehen wurden. Die reife microRNA wird dann in den „RNA-induced Silencing Complex“, kurz RISC eingebaut, wobei der RNA-Duplex entwunden wird und nur eine einzelsträngige RNA zurückbleibt. Diese kurze RNA fungiert dann abermals als molekularer „Guide“, der den RISC-Komplex an eine Zielsequenz im 3'-UTR einer mRNA leitet – die microRNA „zeigt“ dem Proteinkomplex also das richtige Ziel. Die Erkennung der Zielsequenz erfolgt über eine wiederum sehr kurze Sequenz von nur etwa 7 Nukleotiden am 5'-Ende der microRNA. Dieser sogenannte „microRNA seed“ lagert sich über Basenpaarungen an die komplementäre Sequenz in der Ziel-mRNA an (Bartel, 2004). Der Rest der microRNA bleibt in eukaryotischen Systemen in aller Regel ungepaart und es kommt vorwiegend zur Hemmung der Translation der Ziel-mRNA. In weniger häufigen Fällen, wenn die Zielsequenz vollständig komplementär zur microRNA ist und die gesamte microRNA eine Basenpaarung eingehen kann, kommt es zum Schnitt der mRNA durch die RISC-Untereinheit AGO2 (Meister et al., 2004), was den Abbau der mRNA zur Folge hat. MicroRNAs fungieren also auf posttranskriptioneller Ebene¹⁰ und dienen zur Feinregulation der Genexpression. In einer Vielzahl von zellulären Prozessen spielen microRNAs eine entscheidende Rolle. Eine einzelne microRNA kann dabei hunderte mRNAs gezielt regulieren. Im menschlichen Genom wurden bereits mehr als 1.000 microRNAs entdeckt. Viele davon zeigen spezifische Expressionsmuster und sind in verschiedenen Krankheiten dereguliert. Dies macht microRNAs auch als Zielstrukturen für Therapeutika, z. B. in der Krebstherapie, interessant.

⁹ Haarnadelstrukturen kommen häufig bei RNA vor, wenn innerhalb eines Einzelstranges in kurzen Abschnitten Doppelstrangstrukturen entstehen.

¹⁰ Die posttranskriptionelle Ebene der Genregulation bezeichnet die Gesamtheit der regulatorischen Mechanismen, die im Anschluss an die RNA-Synthese, also die Transkription, stattfinden.

2.9 Small interfering RNA (siRNA)

Small interfering RNAs weisen viele Gemeinsamkeiten zu microRNAs auf und wurden etwa zur gleichen Zeit entdeckt. Die siRNAs sind ebenfalls kurze RNAs, etwa 21 Nukleotide lang, die im Zytoplasma von dem Enzym Dicer aus langen doppelsträngigen RNAs herausgeschnitten werden (McManus/Sharp, 2002). Im Gegensatz zu den endogenen, d. h. im Genom kodierten microRNAs fällt bei siRNAs die Prozessierung im Zellkern und der anschließende Export in das Zytoplasma weg und die Biosynthese beschränkt sich auf einen einzelnen Prozessierungsschritt. Die siRNAs werden ebenfalls in den RISC-Komplex eingebaut und leiten diesen zu einer durch die siRNA definierten Zielsequenz (McManus/Sharp, 2002). Ein entscheidender Unterschied ist jedoch, dass siRNAs vollständig komplementär zur Zielsequenz sind und somit einen Schnitt der Ziel-RNA auslösen können, was zu deren Abbau führt. Dieser Mechanismus wird auch als RNA-Interferenz (kurz RNAi) bezeichnet und kann neben der Regulation von zelleigenen mRNAs auch fremde RNA, z. B. die eines Virus oder anderen Pathogens, gezielt zerschneiden.

Im Jahre 2001 konnte Thomas Tuschl zeigen, dass der RNA-Interferenz-Mechanismus auch in Säugerzellen funktioniert (Elbashir et al., 2001), was ungeahnte Möglichkeiten eröffnete. Es war nun erstmalig möglich, anhand der Ziel-mRNA-Sequenz ein synthetisches RNA-Molekül von gerade mal 21 Nukleotiden Länge herzustellen, um so die Ziel-mRNA herabzuregulieren. Es wurde also ein quasi universeller Schalter gefunden, um jede beliebige mRNA auszuschalten. Dies war eine Revolution in der Molekularbiologie, da nun systematisch die Funktion von so gut wie jedem beliebigen Gen untersucht werden konnte (McManus/Sharp, 2002). Für die Entdeckung der RNA-Interferenz (Fire et al., 1998) wurden Craig Mello und Andrew Fire im Jahre 2006 mit dem Nobelpreis ausgezeichnet. Auch für die Entwicklung von RNA-basierten Therapeutika war die Entdeckung der RNA-Interferenz ein Schlüsselmoment und wir verdanken viele der Technologien, die in den mRNA-basierten COVID-19-Vakzinen Verwendung finden, der Forschung an siRNAs als Therapeutika (siehe auch Sparmann/Vogel, Kapitel 4).

2.10 Long non-coding RNA (lncRNA)

Die bisher beschriebenen Klassen an kodierenden und nicht-kodierenden RNAs zeichnen sich dadurch aus, dass jeder RNA-Typus eine charakteristische molekulare Struktur aufweist und alle RNAs innerhalb eines Typus nach dem gleichen Prinzip funktionieren. Dies trifft auf die Klasse der langen nicht-kodierenden RNAs nicht zu. Im Vergleich zu den kurzen nicht-kodierenden RNAs wie microRNAs oder siRNAs werden lncRNAs schlicht über eine Länge von mehr als 200 Nukleotiden definiert. Ähnlich den proteinkodierenden mRNAs werden lncRNAs im Zellkern von dem Enzym RNA-Polymerase II transkribiert, erhalten eine 5'-Cap-Struktur, werden co-transkriptional gespleißt und in den meisten Fällen auch polyadenyliert. Biochemisch unterscheidet sich eine lncRNA also nicht grundlegend von einer mRNA. Der fundamentale Unterschied ist jedoch, dass eine lncRNA keinen Leserahmen enthält, der von den Ribosomen abgelesen und in ein Protein übersetzt werden kann. Damit stellt sich quasi unvermittelt die Frage, was denn die molekulare Funktion einer lncRNA sein könnte. Im Vergleich zu mRNAs haben lncRNAs tendenziell etwas weniger

Exons, werden etwa zehnfach weniger stark abgelesen und sind auch weniger konserviert als mRNAs. Als Gruppe scheinen lncRNAs etwas höher konserviert als Introns, reichen in ihrer Konservierung aber nicht an mRNA-Exons heran (Ulitsky/Bartel, 2013). Trotzdem zeichnen sich lncRNAs aber z. B. durch zelltypspezifische Expressions- und Regulationsmuster aus und es lassen sich einzelne hochkonservierte und robust exprimierte lncRNAs im menschlichen Genom finden, was auf funktionelle Relevanz hindeuten kann. Im humanen Genom konnten dank neuester Sequenzieretechnologien, insbesondere dem „next-generation sequencing“ bereits tausende lncRNAs entdeckt und katalogisiert werden. Für die Mehrheit dieser Transkripte sind aktuell noch keine physiologischen Funktionen bekannt, weshalb sich die Frage stellt, ob eine Vielzahl der entdeckten lncRNAs evtl. schlicht Hintergrundaktivität der Transkriptionsmaschinerie sein könnte. Dennoch gibt es gut beschriebene Beispiele für funktionell relevante lncRNAs mit faszinierenden molekularen Mechanismen.

Die am besten verstandene lncRNA ist unter dem Namen XIST bekannt. Anfang der 1990er-Jahre konnten Forscher zeigen, dass das auf dem humanen X-Chromosom¹¹ liegende Gen *XIST* eine ungewöhnlich lange RNA produziert, welche jedoch keine proteinkodierende Sequenz enthält und nicht wie mRNAs im Zytoplasma, sondern ausschließlich im Zellkern zu finden ist (Brockdorff et al., 1992). Interessanterweise wird XIST nur von einem der zwei X-Chromosomen in weiblichen Zellen abgelesen. Hierbei handelt es sich um genau jenes Chromosom, welches während der weiblichen Embryonalentwicklung inaktiviert wird (Augui et al., 2011). Dieser Prozess der X-Inaktivierung findet ausschließlich in weiblichen Zellen statt und ist als Dosiskompensation bekannt. Hierbei handelt es sich um eine von vielen Organismen genutzte Strategie, um sicherzustellen, dass wichtige Gene, die auf dem X-Chromosom liegen, in männlichen und weiblichen Zellen in gleichen Mengen abgelesen werden. Da männliche Zellen ein X- und ein Y-Chromosom haben, weibliche Zellen dagegen zwei X-Chromosome, wird während der weiblichen Embryonalentwicklung eines der zwei X-Chromosome inaktiviert (Augui et al., 2011), um dieses Ziel zu erreichen. Es zeigte sich, dass XIST für die Inaktivierung dieses X-Chromosoms verantwortlich ist (Penny et al., 1996). Hierzu nutzt XIST einen faszinierenden Mechanismus: Die XIST-RNA baut aus verschiedenen Proteinen einen als Inhibitor agierenden Proteinkomplex zusammen, der die Transkription von Genen lokal hemmen kann (Chu et al., 2015; McHugh et al., 2015). XIST bewirkt dann eine Verteilung dieses repressiven Komplexes über das gesamte X-Chromosom hinweg, was zur transkriptionellen Stilllegung und schließlich zur nahezu vollständigen Inaktivierung des betroffenen X-Chromosoms führt.

Ausgehend von den von XIST genutzten molekularen Prinzipien wurden verschiedene Mechanismen für die Funktion von lncRNAs vorgeschlagen. Zu den am häufigsten beschriebenen Mechanismen gehört die Funktion als molekulares „Gerüst“ von Proteinkomplexen oder die sequenzabhängige Anbringung von Proteinkomplexen an eine bestimmte DNA- oder RNA-Region. Oft wird auch eine Funktion als Modulator der Aktivität bestimmter Proteine oder als eine Art molekularer „Köder“ vorgeschlagen. Hierbei könnte die regulato-

¹¹ Jede menschliche Zelle enthält eine Kopie des gesamten Genoms, verteilt auf 23 Chromosomenpaare (46 Einzelchromosomen). In männlichen und weiblichen Zellen sind 22 dieser Chromosomenpaare identisch. Dies sind die sogenannten Autosomenpaare. Das biologische Geschlecht wird von den verbliebenen zwei Chromosomen bestimmt – den Geschlechtschromosomen. Frauen haben zwei X-Chromosome, Männer hingegen haben nur ein X- und ein deutlich kürzeres Y-Chromosom.

rische Aktivität einer lncRNA dadurch zustande kommen, dass die lncRNA ein spezifisches Protein oder auch eine regulatorische RNA wie etwa eine bestimmte microRNA besonders stark bindet und so von der eigentlichen Zielsequenz fernhält. Welcher dieser Mechanismen für eine bestimmte lncRNA in Frage kommt, unterscheidet sich von Fall zu Fall und erfordert eine genaue molekulare Charakterisierung, was oft Jahre an Forschungsarbeit erfordert. In vielen Krankheiten zeigen lncRNAs charakteristische Expressions- und Regulationsmuster, weshalb es berechtigtes Interesse an lncRNAs als therapeutische Zielstrukturen für medizinisch relevante Indikationen oder auch als Marker für krankhafte Veränderungen im Körper gibt.

2.11 Circular RNA (circRNA)

Eine relativ neu entdeckte Klasse an RNAs sind die zirkulären RNAs oder kurz „circRNAs“ (Hansen et al., 2014; Memczak et al., 2014). Wie der Name bereits erahnen lässt, zeichnen sich circRNAs dadurch aus, dass sie eine geschlossene Ringstruktur haben und nicht wie gewöhnliche Transkripte ein definiertes 5'- und 3'-Ende aufweisen. Die Biosynthese dieser zirkulären RNAs erfolgt über ein relativ selten auftretendes Spleißphänomen, das sogenannte „backsplicing“. Hierbei wird das 5'-Ende eines Introns (die Spleißdonorstelle) mit dem 3'-Ende eines davorliegenden Introns (die Spleißakzeptorstelle) verbunden, wodurch ein geschlossener RNA-Ring entsteht. In der Regel sind circRNAs nicht-kodierend, es gibt jedoch vereinzelt Studien, die eine Translation von einzelnen zirkulären RNAs nahelegen. Ob hierbei jedoch stabile und vor allem funktional relevante Proteinprodukte gebildet werden, ist noch nicht abschließend geklärt. Zur Funktion von zirkulären RNAs ist noch wenig bekannt, ihre einzigartige molekulare Struktur verleiht diesen RNA-Ringen aber vermutlich eine ungewöhnlich hohe Stabilität gegenüber Enzymen, die RNA von den 5'- und 3'-Enden her abbauen. Diese Stabilität könnte auch für die Funktion von circRNAs bedeutend sein. Das am besten verstandene Beispiel einer zirkulären RNA mit einer biologischen Funktion ist ein RNA-Ring mit dem Namen CDR1as. Die Entdeckung dieser RNA basierte auf ihrer ungewöhnlich starken Bindung durch eine Komponente des microRNA-RISC-Komplexes (Memczak et al., 2014) (siehe Abschnitt 2.8). Neben dieser starken biochemischen Interaktion weist CDR1as über 60 evolutionär konservierte Bindestellen einer spezifischen microRNA (miR-7) auf. Studien in Zellen und Tiermodellen konnten zeigen, dass CDR1as wahrscheinlich eine große Zahl an miR-7-Molekülen binden kann, was eine Reduktion der funktional verfügbaren miR-7-Moleküle bewirkt und einer miR-7-Hemmung gleichkommt. Die Funktion dieser zirkulären RNA scheint also die eines „Ködners“ zu sein, der viele Kopien eines Regulatormoleküls bindet und diese von anderen Zielen fernhält. Aktuell sind nur wenige physiologisch relevante circRNAs bekannt, vorgeschlagene Mechanismen reichen aber von der Bindung von Proteinen, über den Zusammenbau von Proteinkomplexen, bis hin zum Transport von interagierenden Proteinen oder RNAs. Aufgrund ihrer erhöhten Stabilität könnten sich circRNAs auch als Biomarker eignen.

2.12 Riboswitches

Im Gegensatz zu den bisher beschriebenen RNA-Klassen sind die sogenannten Riboswitches kein eigenständiger RNA-Typus, sondern vielmehr ein regulatorisches RNA-Element

von 30 bis 200 Nukleotiden Länge, welches in den untranslatierten Regionen vorwiegend bakterieller mRNAs zu finden ist. Die herausragende Eigenschaft dieses RNA-Elements ist, dass es niedermolekulare Metabolite¹² ohne Zutun eines Proteins direkt binden kann. In den meisten Fällen hat diese Bindung dann einen regulatorischen Einfluss auf die translationale Aktivität der mRNA oder deren Transkription (Mironov et al., 2002; Winkler et al., 2002). Mechanistisch lässt sich dieser Effekt durch eine Änderung der RNA-Struktur nach Bindung des Metaboliten erklären. Die meisten Riboswitches finden sich in Bakterien. Hier kann durch eine Änderung der RNA-Struktur z. B. die ribosomale Bindungsstelle verborgen werden, was den Start der Translation verhindert (Winkler/Breaker, 2003). Alternativ kann die Metabolitbindung zur Ausbildung einer Haarnadelstruktur führen, was einen Abbruch der Transkription herbeiführen kann.

2.13 Bacterial small RNA (sRNA)

Regulatorische RNAs finden sich nicht nur in Eukaryoten, sondern auch in Bakterien. Eine besonders wichtige Klasse bakterieller RNAs sind die sogenannten kleinen RNAs oder auch sRNAs. Hierbei handelt es sich um nicht-kodierende RNAs mit einer Länge von 50 bis 250 Nukleotiden. Obwohl die Existenz dieser kleinen RNAs in Bakterien bereits seit den 1970er-Jahren bekannt ist, ermöglichte erst die systematische Sequenzierung und Analyse von bakteriellen Genomen die Entdeckung von sRNAs in größerer Zahl (Argaman et al., 2001; Wassarman et al., 2001; Vogel et al., 2003). Im Gegensatz zu den kleinen RNAs der Eukaryoten (insbesondere miRNAs und siRNAs) ist die Gruppe der sRNAs recht divers was ihre Länge und Sekundärstruktur, aber auch ihre regulatorischen Funktionen und Mechanismen angeht. Die sRNAs können als unabhängige Transkriptionseinheiten abgelesen oder aus 3'-UTRs von mRNAs herausgeschnitten werden. Viele sRNAs beinhalten Haarnadelstrukturen und können mit Proteinen oder mRNAs interagieren, um so die bakterielle Genexpression zu kontrollieren. Mechanistisch funktionieren viele sRNAs über eine sequenzspezifische Anlagerung an regulatorische Regionen in bakteriellen mRNAs. Von besonderer Bedeutung ist hierbei etwa die Ribosomenbindungsstelle, die z. B. durch eine sRNA blockiert werden kann, was eine effektive Translation verhindert. Alternativ kann durch die Bindung einer sRNA auch eine Strukturänderung in der Ziel-mRNA vermittelt werden, wodurch die Ribosomenbindungsstelle erst freigelegt wird. Dieser Mechanismus bewirkt dann eine effizientere Translation der Ziel-mRNA. Darüber hinaus können sRNAs auch die Transkription und Stabilität ihrer Ziel-mRNAs beeinflussen. Anhand der sRNA-Sequenz können komplementäre Zielsequenzen in mRNAs bestimmt werden, wodurch sich regulatorische Ziele vorhersagen lassen. Häufig wird dies jedoch dadurch erschwert, dass die Interaktion von sRNAs über relativ kurze Sequenzregionen mit nur unvollständiger Basenanlagerung an die Ziel-mRNA vermittelt wird. Es ist also eine detaillierte molekulare Charakterisierung von einzelnen sRNAs und deren Zielen notwendig. In Gram-negativen¹³ Bakterien arbeiten sRNAs häufig mit speziellen Helferproteinen (sogenannte RNA-Chaperone) wie etwa Hfq (Vogel/Luisi, 2011) zusammen. RNA-Chaperone beeinflussen die Stabili-

¹² Metabolite sind chemische Verbindungen in einer Zelle, die als Zwischenprodukte des biologischen Stoffwechsels gebildet werden.

¹³ Bakterien können anhand der sogenannten Gram-Färbung (benannt nach Hans Christian Gram) in zwei große Klassen unterschieden werden: Gram-positive und Gram-negative Bakterien. Die unterschiedliche Färbbarkeit von Bakterien beruht hierbei auf dem Aufbau der Zellwand, was ein wichtiges systematisches Unterscheidungsmerkmal darstellt.

tät der sRNA und vermitteln darüber hinaus die Anlagerung der sRNA an eine Zielsequenz. Ähnlich zu den Riboswitches (siehe Abschnitt 2.12) können sRNAs auf Metabolite oder Signale aus der Umwelt reagieren und eine entsprechende Genregulation veranlassen. Dies macht sRNAs zu wichtigen Regulatoren in verschiedenen bakteriellen Signalwegen wie etwa der Eisenhomöostase, der chemischen Kommunikation mittels Signalmolekül (bekannt als „Quorum Sensing“), der Biofilmbildung, aber auch der Steuerung von Virulenzprogrammen in pathogenen Bakterien.

Tabelle 1: Übersicht über bekannte RNA-Typen und -Funktionen bei Eukaryoten

RNA-Typ	Größe	Domäne des Lebens	Funktion
mRNA	im Menschen: durchschnittl. 2–3 Kilobasen (kb)	alle	Botenmolekül, das den genetischen Bauplan für die Herstellung von Proteinen durch das Ribosom enthält.
rRNA	im Menschen: 18S: 1.9 kb 28S: 5 kb	alle	Bestandteil des Ribosoms mit strukturellen und funktionellen Aufgaben.
tRNA	70–95 Basen	alle	Übersetzung des genetischen Codes durch die Zuordnung einer Aminosäure zu einem mRNA-Basentriplett.
snoRNA	60–300 Basen	Eukaryoten, Archaeen	Modifikation und Prozessierung von rRNA.
snRNA	100–300 Basen	Eukaryoten	Bestandteil des Spleißosoms. Entfernung von Introns aus prä-mRNA.
miRNA	21 Basen	Eukaryoten	Regulation von mRNA-Stabilität und -Translation auf post-transkriptioneller Ebene.
siRNA	21 Basen	Eukaryoten	RNA-Regulation im Rahmen der RNA-Interferenz.
lncRNA	>200 Basen	Eukaryoten, Bakterien	Häufig unbekannt. Vielfältige genregulatorische Mechanismen möglich.
circRNA	100–4.000 Basen	Eukaryoten	Häufig unbekannt. Vielfältige genregulatorische Mechanismen.

sRNA	50-250 Basen	Bakterien, Archaeen	Regulation von mRNA-Stabilität, -Translation und -Transkription in Bakterien.
Ribozym	verschieden	Eukaryoten, Bakterien, Archaeen	RNA-Molekül mit katalytischen Eigenschaften.
Riboswitch	30-200 Basen	Bakterien, Archaeen, vereinzelt in Eukaryoten	Bildung von Metaboliten und Kontrolle von mRNA-Transkription und -Translation.
crRNA	40-50 Basen	Bakterien, Archaeen	Anlagerung von Cas-Proteinen an DNA-Zielsequenz. Teil des CRISPR-Immunsystems in Bakterien.
tracrRNA	50-150 Basen	Bakterien, Archaeen	Anlagerung von Cas-Proteinen an DNA-Zielsequenz. Teil des CRISPR-Immunsystems in Bakterien.

2.14 CRISPR-RNAs

Bakterien besitzen ein antivirales adaptives Immunsystem bekannt als „Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-associated (Cas)“, oft auch einfach nur CRISPR genannt. Für die Funktion dieses Immunsystems bei dem für die Genomeditierung besonders relevanten Cas9-System sind zwei nicht-kodierende RNAs von zentraler Bedeutung: die CRISPR-RNA (crRNA) und die sogenannte „trans-activating crRNA“ (tracrRNA). Die crRNA wird zunächst als lange prä-crRNA abgelesen. Hieraus werden dann über verschiedene Prozessierungsschritte einzelne crRNAs herausgeschnitten (Deltcheva et al., 2011). Diese crRNAs besitzen einen relativ kurzen Abschnitt von 17-20 Basen der sich an fremde DNA, wie etwa der eines eindringenden Virus, anlagern kann. Diese auch als „Spacer“ bezeichneten Sequenzabschnitte wechseln sich mit konservierten Wiederholungseinheiten („Repeats“) ähnlicher Länge ab, die in allen crRNAs vorhanden sind; an diese lagert sich die tracrRNA an. Die crRNA fungiert also als eine molekulare Brücke zwischen der DNA-Zielsequenz und der tracrRNA. Der RNA-Duplex aus crRNA und tracrRNA wird von CRISPR-assoziierten Proteinen (Cas-Proteine) wie etwa Cas9 gebunden. Gemeinsam leiten die crRNA und die tracrRNA das Cas9-Protein an die virale DNA-Zielsequenz, woraufhin das Cas9-Protein die Zielsequenz schneidet und letztendlich den Abbau solcher Fremd-DNA ermöglicht. Nach einer überstandenen Infektion baut das Bakterium kurze DNA-Fragmente des Eindringlings in den dafür vorgesehenen CRISPR-Lokus¹⁴ seines Genoms ein, um bei zukünftigen Infektionen passende crRNAs zu generieren, die dann die DNA des Eindringlings zerschneiden. Da die Zielsequenz eines Cas-Proteins relativ einfach über die crRNA-Sequenz programmiert werden kann, konnten innerhalb kürzester Zeit

¹⁴ CRISPR-Lokus: Ein Abschnitt in bakteriellen Genomen, der neben den Genen für Cas-Proteine auch die Überbleibsel viraler Infektionen enthält und dort archiviert.

eine Vielzahl an CRISPR/Cas-basierten Technologien und Anwendungen entwickelt werden, die es erlauben, die DNA eines Zielorganismus gezielt zu verändern. Diese einfache Programmierbarkeit liegt auch den vielfältigen Anwendungen von CRISPR/Cas-Systemen insbesondere in der biomedizinischen Forschung und in der Entwicklung von Gentherapien zugrunde.

2.15 Schlussbemerkung

Zusammenfassend lässt sich feststellen, dass die Bandbreite der bekannten RNA-Klassen und -Funktionen sehr hoch ist. Auch ist nicht ausgeschlossen, dass in Zukunft noch weitere Klassen beschrieben werden, insbesondere in Mikroorganismen. Neben der Funktion als Träger genetischer Information kann RNA mit einer Vielzahl an Biomolekülen (DNA, RNA, Proteine), aber auch mit Metaboliten interagieren. Aufgrund dieser Eigenschaft eignet sich RNA hervorragend dazu, Interaktionen zwischen verschiedenen Biomolekülen zu vermitteln, um so regulatorische Komplexe zu bilden. Von besonderer Bedeutung ist hierbei die Tatsache, dass über die RNA-Sequenz spezifische Interaktionen mit anderen RNA- oder DNA-Molekülen „programmiert“ werden können. Dieses Prinzip ist für die Funktion verschiedener nicht-kodierender RNAs, aber auch für das Design von Antisense-RNA oder therapeutischen mRNAs von fundamentaler Bedeutung. Eine weitere herausragende Eigenschaft von RNA ist die Katalyse von chemischen Reaktionen und die Funktion als Sensoren für Metabolite oder andere physikalische Parameter wie z. B. Temperatur. Neben pro- und eukaryotischen Organismen nutzen auch Pathogene wie etwa RNA-Phagen oder RNA-Viren RNA als Träger für genetische Information und als Matrize für die Proteinbiosynthese. Auch als therapeutisches Molekül ist RNA spätestens seit der erfolgreichen Entwicklung der COVID-19-mRNA-Impfstoffe in den Fokus gerückt und stellt eine neue Wirkstoffklasse mit immensem Potenzial für die Therapie verschiedenster Krankheiten dar.

2.16 Literaturverzeichnis

- Argaman, L. et al. (2001):** Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. In: *Curr Biol* 11: 941-950. DOI: 10.1016/s0960-9822(01)00270-6.
- Augui, S. et al. (2011):** Regulation of X-chromosome inactivation by the X-inactivation centre. In: *Nat Rev Genet* 12: 429-442. DOI: 10.1038/nrg2987.
- Bartel, D. P. (2004):** MicroRNAs genomics, biogenesis, mechanism, and function. In: *Cell* 116: 281-297. DOI: 10.1016/s0092-8674(04)00045-5.
- Berget, S. M. et al. (1977):** Spliced segments at the 5' terminus of adenovirus 2 late mRNA. In: *Proc National Acad Sci* 74: 3171-3175. DOI: 10.1073/pnas.74.8.3171.
- Brockdorff, N. et al. (1992):** The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. In: *Cell* 71: 515-526. DOI: 10.1016/0092-8674(92)90519-i.
- Cartegni, L./Krainger, A. R. (2003):** Correction of disease-associated exon skipping by synthetic exon-specific activators. In: *Nature Structural Biology* 10: 120-125. DOI: 10.1038/nsb887.
- Cech, T. R. et al. (1981):** In vitro splicing of the ribosomal RNA precursor of tetrahymena: Involvement of a guanosine nucleotide in the excision of the intervening sequence. In: *Cell* 27: 487-496. DOI: 10.1016/0092-8674(81)90390-1.
- Chu, C. et al. (2015):** Systematic discovery of *Xist* RNA binding proteins. In: *Cell* 161: 404-416. DOI: 10.1016/j.cell.2015.03.025.
- Cramer, P. (2019):** Organization and regulation of gene transcription. In: *Nature* 573: 45-54. DOI: 10.1038/s41586-019-1517-4.
- Crick, F. (1970):** Central dogma of molecular biology. In: *Nature* 227: 561-563. DOI: 10.1038/227561a0.

- Deltcheva, E. et al. (2011):** CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. In: *Nature* 471(7340): 602–607. DOI: 10.1038/nature09886.
- Djebali, S. et al. (2013):** Landscape of transcription in human cells. In: *Nature* 488: 101–108. DOI: 10.1038/nature11233.
- Doherty, E. A./Doudna, J. A. (2001):** Ribozyme structures and mechanisms. In: *Annu Rev Biophys Biom* 30: 457–475. DOI: 10.1146/annurev.biophys.30.1.457.
- Elbashir, S. M. et al. (2001):** Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. In: *Nature* 411: 494–498. DOI: 10.1038/35078107.
- Emilsson, G. M. et al. (2003):** Ribozyme speed limits. In: *Rna* 9: 907–918. DOI: 10.1261/rna.5680603.
- Fire, A. et al. (1998):** Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. In: *Nature* 391: 806–811. DOI: 10.1038/35888.
- Gilbert, W. (1986):** Origin of life: The RNA world. In: *Nature* 319: 618–618. DOI: 10.1038/319618a0.
- Hansen, T. B. et al. (2014):** Natural RNA circles function as efficient microRNA sponges. In: *Nature* 495: 384–388. DOI: 10.1038/nature11993.
- Holley, R. W. et al. (1965a):** Structure of a ribonucleic acid. In: *Science* 147: 1462–1465. DOI: 10.1126/science.147.3664.1462.
- Holley, R. W. et al. (1965b):** Nucleotide sequences in the yeast alanine transfer ribonucleic acid. In: *J Biological Chem* 240: 2122–2128.
- Iyer, M. K. et al. (2015):** The landscape of long noncoding RNAs in the human transcriptome. In: *Nature Genetics* 47: 199–208. DOI: 10.1038/ng.3192.
- Jackson, R. J. et al. (2010):** The mechanism of eukaryotic translation initiation and principles of its regulation. In: *Nature Reviews Molecular Cell Biology* 11: 113–127. DOI: 10.1038/nrm2838.
- Kufel, J./Grzechnik, P. (2019):** Small nucleolar RNAs tell a different tale. In: *Trends Genet* 35: 104–117. DOI: 10.1016/j.tig.2018.11.005.
- Lander, E. S. et al. (2001):** Initial sequencing and analysis of the human genome. In: *Nature* 409: 860–921. DOI: 10.1038/35057062.
- Lee, R. C. et al. (1993):** The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. In: *Cell* 75: 843–854. DOI: 10.1016/0092-8674(93)90529-y.
- Matera, A. G./Wang, Z. (2014):** A day in the life of the spliceosome. In: *Nat Rev Mol Cell Bio* 15: 108–121. DOI: 10.1038/nrm3742.
- Matera, A. G. et al. (2007):** Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. In: *Nat Rev Mol Cell Bio* 8: 209–220. DOI: 10.1038/nrm2124.
- McHugh, C. A. et al. (2015):** The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. In: *Nature* 521: 232–236. DOI: 10.1038/nature14443.
- McManus, M. T./Sharp, P. A. (2002):** Gene silencing in mammals by small interfering RNAs. In: *Nat Rev Genet* 3: 737–747. DOI: 10.1038/nrg908.
- Meister, G. et al. (2004):** Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. In: *Molecular Cell* 15: 185–197. DOI: 10.1016/j.molcel.2004.07.007.
- Memczak, S. et al. (2014):** Circular RNAs are a large class of animal RNAs with regulatory potency. In: *Nature* 495: 333–338. DOI: 10.1038/nature11928.
- Mironov, A. S. et al. (2002):** Sensing small molecules by nascent RNA: A mechanism to control transcription in bacteria. In: *Cell* 111: 747–756. DOI: 10.1016/S0092-8674(02)01134-0.
- Moore, M. J. (2005):** From birth to death: the complex lives of eukaryotic mRNAs. In: *Science* 309: 1514–1518. DOI: 10.1126/science.1111443.
- Nirenberg, M. W./Matthaei, J. H. (1961):** The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. In: *Proc National Acad Sci* 47: 1588–1602. DOI: 10.1073/pnas.47.10.1588.
- Noller, H. F. et al. (1992):** Unusual resistance of Peptidyl transferase to protein extraction procedures. In: *Science* 256: 1416–1419. DOI: 10.1126/science.1604315.
- Pasquinelli, A. E. et al. (2000):** Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. In: *Nature* 408: 86–89. DOI: 10.1038/35040556.
- Penny, G. D. et al. (1996):** Requirement for Xist in X chromosome inactivation. In: *Nature* 379: 131–137. DOI: 10.1038/379131a0.
- Reinhart, B. J. et al. (2000):** The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. In: *Nature* 403: 901–906. DOI: 10.1038/35002607.

Stark, B. C. et al. (1978): Ribonuclease P: an enzyme with an essential RNA component. In: Proc National Acad Sci 75: 3717–3721. DOI: 10.1073/pnas.75.8.3717.

Ulitsky, I./Bartel, D. P. (2013): lincRNAs: genomics, evolution, and mechanisms. In: Cell 154: 26–46. DOI: 10.1016/j.cell.2013.06.020.

Vogel, J./Luisi, B. F. (2011): Hfq and its constellation of RNA. In: Nat Rev Microbiol 9: 578–589. DOI: 10.1038/nrmicro2615.

Vogel, J. et al. (2003): RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria. In: Nucleic Acids Res 31: 6435–6443. DOI: 10.1093/nar/gkg867.

Wahl, M. C. et al. (2009): The Spliceosome: Design principles of a dynamic RNP machine. In: Cell 136: 701–718. DOI: 10.1016/j.cell.2009.02.009.

Wassarman, K. M. et al. (2001): Identification of novel small RNAs using comparative genomics and microarrays. In: Gene Dev 15: 1637–1651. DOI: 10.1101/gad.901001.

Winkler, W. C./Breaker, R. R. (2003): Genetic control by metabolite-binding riboswitches. In: Chembiochem 4: 1024–1032. DOI: 10.1002/cbic.200300685.

Winkler, W. et al. (2002): Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. In: Nature 419: 952–956. DOI: 10.1038/nature01145.

Woese, C. R. et al. (1990): Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. In: Proc National Acad Sci 87: 4576–4579. DOI: 10.1073/pnas.87.12.4576.

Yan, C. et al. (2015): Structure of a yeast spliceosome at 3.6-angstrom resolution. In: Science 349: 1182–1191. DOI: 10.1126/science.aac7629.

3. Die Bedeutung von RNA für die molekulare Diagnostik

Jörn Walter und Nina Gasparoni

RNA-Analysen haben eine rasant wachsende Bedeutung für die klinische Forschung und Diagnostik. Sie erschließen individuelle Zellprogramme und erlauben so die molekulare Unterscheidung gesunder und kranker Zellen. RNA-Analysen ermöglichen darüber hinaus die Diagnose von Pathogenen und bilden die Basis für die Entwicklung und Anwendung moderner, z. T. selbst ebenfalls RNA-gestützter Therapie- und Immunisierungsverfahren. Die rasante Entwicklung im Bereich der Next-Generation-Sequencing(NGS)-Technologien hat für die RNA-Analytik in den vergangenen Jahren zu einer Reihe neuer und bahnbrechender Anwendungsmöglichkeiten geführt. Besonders hervorzuheben ist hier die Entwicklung hochsensitiver Einzelzellanalysen, die ganz neue Perspektiven für die klinische Diagnostik erschließen.

3.1 Überblick

RNAs entstehen durch Umschreiben von DNA in RNA. Die genaue quantitative und qualitative Vermessung der Gesamtheit aller zu einem bestimmten Zeitpunkt in der Zelle vorliegenden verschiedenen RNAs, des Transkriptoms, gibt den aktuellen molekularen Zustand von Zellen oder Geweben wieder. Neben mRNAs, d. h. Kopien von Genen, die mehrheitlich in Proteine umgeschrieben werden, rücken nicht-kodierende RNAs (wie miRNAs, circRNAs, snoRNAs, piRNAs, rRNAs, tRNAs u. v. a.) zunehmend in den Fokus diagnostischer Ansätze (siehe auch Munschauer/Vogel, Kapitel 2, für eine Übersicht der RNAs). Analysen durch Sequenzierung (RNA-Seq, siehe Abschnitt 3.2) bieten besonders tiefe Einblicke in die Menge und Vielfalt der verschiedenen RNAs. Die Breite der Auflösung und die Sensitivität von Sequenzierverfahren ist dabei deutlich höher als die anderer RNA-Analyseverfahren (siehe unten), sodass sogar Transkriptome einzelner Zellen abgebildet werden können. Dies eröffnet vollkommen neue diagnostische Perspektiven. RNA-Sequenzdaten sind ein Schlüssel zur Identifikation pathogener Organismen und Viren. Die große Bedeutung von RNA-Sequenzdaten bei der Bekämpfung viraler Infektionskrankheiten hat uns die Forschung an SARS-CoV-2 (COVID-19) eindringlich verdeutlicht. RNA-Daten sind heute in nahezu allen Bereichen der Medizin ein zentraler Bestandteil der Krankheitsdiagnostik und liefern (nicht nur) begleitende Daten für eine Diagnose, sondern können auch für die Wahl der geeigneten Therapie bedeutsam sein.

3.2 Methodenspektrum der RNA-Analytik

Die in der (Routine-)RNA-Diagnostik eingesetzten Methoden umfassen eine Reihe molekularer Techniken (siehe Tabelle 2): Einzelne RNAs/Gene können mithilfe von PCR-Methoden analysiert und quantifiziert werden. Dabei wird das Vorhandensein oder die Menge der (krankheitsspezifischen) RNA-Kopien (sogenannte Biomarker) durch gezielte Vervielfältigung der RNA mithilfe quantitativer PCR-Methoden untersucht (Bustin/Mueller, 2005; VanGuilder et al., 2005).¹ Diese seit vielen Jahren eingesetzte gezielte Methodik wird genspezifisch angewandt und eignet sich zum schnellen und quantitativen Nachweis einzelner Biomarker in einer großen Anzahl von Proben. In einigen Bereichen ist dieses Verfahren nach wie vor der Goldstandard für die Diagnostik (z. B. von Viren: SARS-CoV-2 u. v. a.). Mit der Entwicklung von Microarrays wurde eine neue Ebene der Komplexität in der Diagnostik erreicht. Arrays ermöglichen die gleichzeitige Analyse vieler bekannter krankheitsrelevanter Gene. Die Analyse der Veränderungen der von Genen abgelesenen RNA erfolgt durch Binden der zu untersuchenden RNA an bekannte „Fängermoleküle“ auf einem Glaträger. Durch optische Verfahren kann dann die relative Stärke der Bindung ermittelt und so Rückschlüsse auf die relative Menge der am Microarray „gefangenen“ RNA gezogen werden (Gorreta et al., 2012).² PCRs und Microarray-Analysen sind allerdings auf die Untersuchung einer begrenzten Zahl vordefinierter Gene beschränkt, bedürfen also eines A-priori-Wissens. Tiefe und neue explorative Analysen können nur mithilfe von NGS-Verfahren („next-generation sequencing“) erreicht werden. Die RNA-Sequenzierung bietet eine umfassende und unvoreingenommene Herangehensweise an das gesamte Transkriptom, ohne mögliche Zielregionen vorab definieren zu müssen (Mortazavi et al., 2008; Wang et al., 2009). RNA-Sequenzierung generiert daher einen viel tieferen Einblick in die RNA-Vielfalt einer einzelnen oder mehrerer Zellen und eröffnet zudem ein breites Anwendungsspektrum von der Einzelzellanalyse bis zur Identifikation neuer, noch unbekannter Viren. Sequenziermethoden der 2. Generation mit kurzen „Leselängen“ und hohem Probendurchsatz sind der gegenwärtige Standard für RNA-Sequenzierung.³

Die gute und reproduzierbare Quantifizierbarkeit und schnelle Anwendbarkeit sind wichtige Kriterien für eine Nutzung in der Diagnostik. In der Durchführung einfache, sensitive und genaue Techniken wie die quantitative PCR (qPCR) oder Array-Verfahren sind daher oft immer noch vielfach angewandte Methoden. Sie bieten Vorteile für eine gezielte und schnelle Diagnostik und eine bessere Vergleichbarkeit und Standardisierung über viele

1 Die PCR („polymerase chain reaction“, Polymerase-Kettenreaktion) ist eine Methode zur Vervielfältigung (Amplifikation) kurzer, definierter DNA-Abschnitte. Die Methode basiert auf sich wiederholenden Zyklen von drei Reaktionsphasen. Nach der Auftrennung der doppelsträngigen DNA in Einzelstränge lagern sich zunächst kurze DNA-Fragmente an, die komplementär zu einem Bereich der zu amplifizierenden Sequenz sind. Diese dienen als Anfangspunkt für die nachfolgende enzymatische DNA-Synthese, bei der der DNA-Abschnitt vervielfältigt wird. Die Produkte der vorangegangenen Zyklen dienen dabei als Vorlagen für weitere Amplifikationen (exponentielle Amplifikation). Im Falle der RNA-Analytik wird die RNA zunächst in DNA umgeschrieben (cDNA, copyDNA bzw. komplementäre DNA). Die Menge des PCR-Produkts gibt Aufschluss über die Expressionsstärke der untersuchten Zielregion.

2 Microarrays sind biotechnologische Werkzeuge zur gleichzeitigen Bestimmung der Genexpression einer großen Anzahl von (ausgewählten) Genen. Ein Träger (z. B. eine Glasplatte) ist mit einzelsträngigen Fragmenten beschichtet, die den zu untersuchenden Zielregionen entsprechen. Die RNA wird zunächst in DNA (cDNA) umgeschrieben, die dann fragmentiert und markiert wird (z. B. mit Fluoreszenzfarbstoffen). Die cDNA-Fragmente binden (hybridisieren) an entsprechende komplementäre Fragmente auf dem Array. Über die Messung der Fluoreszenzintensität wird die Menge der jeweils gebundenen Fragmente und damit die Stärke der Expression jeder Zielregion bestimmt.

3 Die RNA-Sequenzierung (basierend auf Sequenziermethoden der 2. Generation) startet mit der Isolierung der RNA aus Zellen. Dabei wendet man spezifische Verfahren an, um die verschiedenen RNA-Arten zu trennen und die gewünschten RNAs (mRNA, miRNA oder ncRNA) an- oder abzureichern. Nach einem Umschreiben der RNA in eine cDNA (Reverse Transkription, RT) erfolgt dann eine enzymatische Zerkleinerung der cDNA und eine anschließende Vervielfältigung mittels PCR (zusammen: RT-PCR). Man erhält dann eine Bibliothek („library“) dieser Bruchstücke, die mittels Hochdurchsatzsequenzierung ausgelesen werden können. Im PCR-Schritt wird eine eindeutige Kodierung an das Ende der Bruchstücke angefügt, die ebenfalls sequenziert wird und es möglich macht, alle Bruchstücke wieder der korrekten Ursprungsprobe zuzuweisen (Demultiplexing). Die Sequenzierung erfolgt auf sogenannten „short read“-Sequenziermaschinen (kurze „Leselängen“, d. h. es werden in der Regel 50–100 Basen lange einzelne Bruchstücke ermittelt).

Proben hinweg. Letztendlich bauen sie aber stets auf Daten auf, die durch RNA-Sequenzierung gewonnen wurden. Entsprechend werden die drei Methoden oft aufeinanderfolgend in der Diagnostik eingesetzt, wie man am Beispiel der COVID-19-Diagnostik sehr gut nachvollziehen kann.

Die Voraussetzung für die Entdeckung von SARS-CoV-2 und die Entwicklung von breit aufgestellten diagnostischen Verfahren und Impfstoffen war die rasche und umfangreiche Sequenzierung des RNA-Genoms der ersten Virusisolate. Sie ermöglichten eine rasche Bestimmung und Einordnung des Virus. Aufbauend auf der RNA-Sequenz wurden dann erste PCR-gestützte Nachweistestverfahren und nachfolgend eine Reihe weiterer immunologischer und molekularer Tests etabliert (Corman et al., 2020; Dao Thi et al., 2020; Yan et al., 2020; Kellner et al., 2022). Kontinuierlich erhobene RNA-Sequenzdaten von Virusisolaten bilden dabei die Grundlage für die Entwicklung und stetige Anpassung der Verfahren, z. B. an neu entstandene Varianten. Die Kenntnis des Virusgenoms und seiner Varianten ist zugleich die Grundlage für die Entwicklung neuer und (stetig) angepasster RNA-Impfstoffe. Das Monitoring von Virusvarianten durch RNA-Sequenzierung und spezifische PCR-basierte Testverfahren war und ist essenziell, um die Verbreitung von Varianten zu bestimmen. Schließlich nutzt man die Daten vergleichender SARS-Coronavirus-Sequenzierungen, um den Ursprung und die Evolution des RNA-Virus zu ergründen (Andersen et al., 2020).

RNA-Sequenzieranalysen spielen aber nicht nur für die Diagnose von SARS-CoV-2 eine Rolle. Um die Pathologie der COVID-19-Erkrankung besser zu verstehen, nutzt man RNA-Sequenzierungsdaten in einer Vielzahl von Studien, z. B. an aus Speichel-, Blut- und Lungenzellen (z. B. Bronchiallavagen) gewonnenen Virusisolaten von COVID-19-Patienten. Diese Studien – oft Einzelzellstudien – haben erheblich dazu beigetragen, die Pathogenese, den Verlauf und die Langzeitwirkungen einer COVID-19-Infektion besser zu verstehen (Bernardes et al., 2020; Chua et al., 2020; Schulte-Schrepping et al., 2020; Sungnak et al., 2020; Muus et al., 2021; Stephenson et al., 2021; Yang et al., 2021). Insgesamt lässt sich am Beispiel der COVID-19-Forschung und -Diagnostik die immense Bedeutung von RNA-Sequenzierverfahren ermessen.

Die gegenwärtigen RNA-Sequenziermethoden bauen auf NGS-Technologien der 2. Generation auf. Sie sind mittlerweile soweit verbessert worden, dass mit kleinsten Mengen von RNA komplexe Analysen durchführbar sind. Mit dem Aufkommen von NGS-Technologien der 3. Generation⁴ erschließen sich aber eine Reihe neuer, fortgeschrittener quantitativer und qualitativer Möglichkeiten für die RNA-Diagnostik. So sind einige der neuen Sequenzierverfahren deutlich schneller und kostengünstiger. Auch hier ist die Nutzung von neuen Verfahren für die SARS-CoV-2-Sequenzierung ein Beispiel (Brinkmann et al., 2020; Freed et al., 2020). Ihre breite Anwendung wird aber noch Zeit in Anspruch nehmen, da sowohl die neuen Technologien als auch die neuen Herausforderungen für die Datenauswertung erst für eine breite Anwendung implementiert werden müssen.

⁴ Die Sequenzierverfahren der 3. Generation umfassen verschiedene Technologien, bei denen DNA/RNA-Moleküle direkt analysiert werden und keine PCR-Amplifikation bzw. im Falle der RNA-Analyse kein Umschreiben der RNA in cDNA erforderlich ist. Den Technologien liegen dabei unterschiedliche methodische Ansätze zugrunde. Die Einzelmolekül-Echtzeitsequenzierung ermöglicht im Gegensatz zu Systemen der 2. Generation eine direkte Analyse einzelner RNA-Moleküle. Hierzu werden eine Reihe kombinierter biochemischer und physikalischer Verfahren genutzt, die ein Auslesen von langen RNA-Molekülen ermöglichen. Einige Systeme bieten zudem die Möglichkeit, modifizierte Basen in der RNA zu erkennen und somit im Sequenzkontext im Einzelmolekül zielgenau auszulesen.

Sequenzierverfahren der 3. Generation bieten zudem die Möglichkeit, RNA direkt zu sequenzieren, d. h. ohne Umschreiben in cDNA. Hierdurch spart man nicht nur Arbeitsschritte, sondern erhält auch zusätzliche Informationen über die RNA, die durch die „short read“-NGS-Verfahren der 2. Generation nur begrenzt oder nur indirekt zu gewinnen sind. Sequenzierverfahren der 3. Generation benötigen zwar mehr RNA als Ausgangsmaterial, bieten dafür aber längere Leselängen der RNA-Moleküle, d. h. komplexe RNA-Moleküle können nahezu „intakt“ (in voller Länge) „durch“sequenziert werden. Man kann so die native Form der ursprünglich gebildeten RNA diagnostizieren und direkt Längen- und Spleißvarianten⁵ bestimmen (Athanasopoulou et al., 2021). Für krankheitsrelevante Gene sind das wichtige zusätzliche Informationen, die in neue diagnostische Tests einfließen werden. Mit einigen der neuen 3.-Generationsverfahren kann zudem nicht nur die Basenabfolge, sondern auch gleichzeitig die Modifikationen der RNA-Basen mit ausgelesen werden (Athanasopoulou et al., 2021). Inwieweit diese (epitranskriptionelle) Modifikationskartierung eine Rolle in der Diagnostik spielen wird, ist aber noch unklar.

3.3 Anwendungen der RNA-Analytik in der Diagnostik

In der Routinediagnostik leisten etablierte und zertifizierte RT-PCR- und RNA-Microarray Testverfahren einen wichtigen Beitrag, auch da sie (wie oben bereits angemerkt) aufgrund ihrer Standardisierbarkeit einfacher für den Routinegebrauch und damit für Zulassungsverfahren geeignet sind. Im klinisch forschenden Bereich werden sie jedoch aus den oben genannten Gründen zunehmend von Verfahren der 2. und 3. Generation der RNA-Sequenzierung (NGS) abgelöst. Im Folgenden möchten wir kurz auf die möglichen Anwendungsbereiche der verschiedenen Methoden eingehen (siehe Tabelle 2).

Eine Hauptanwendung von RNA-Analytik in Forschung und Diagnostik ist neben der Identifikation von Pathogenen (Viren, Bakterien) die Bestimmung der Expression (Menge und Form von RNAs) katalogisierter kodierender und nicht-kodierender Gene von menschlichen Zellen. Im Bereich der nicht-kodierenden RNAs spielt die Analyse von miRNAs eine bedeutende Rolle als Indikator veränderter Regulation in Zellen. Die miRNA-gestützte Diagnostik wird vor allem im Bereich der Krebsforschung, der Immunologie und bei neurodegenerativen Erkrankungen genutzt, oft ergänzend zu einer mRNA-basierten Genexpressionsanalyse (Kappel/Keller, 2017). Die auf RNA-Sequenzierung gestützten Diagnoseverfahren haben durch ihre größere Sensitivität Vorteile, wenn das zur Verfügung stehende Material sehr limitiert ist. Zu nennen sind hier z. B. Analysen von RNAs aus Blutplasma oder Körperflüssigkeiten, die von Zellen des Körpers abgegeben wurden. Solche „frei rumschwimmenden“ (extrazellulären) RNAs, z. B. miRNAs, sind ein erstes Indiz für eine bestehende Erkrankung (z. B. Krebs) im Körper und sind als eine mögliche erste Maßnahme nicht-invasiver molekularer Diagnostik von großem Interesse (Kappel/Keller, 2017).

⁵ Beim Spleißen handelt es sich um einen wichtigen Prozessierungsschritt der RNA. Dabei werden nach der Transkription aus der sogenannten prä-mRNA nicht-kodierende Bereiche (Introns) entfernt und die Exons (kodierende Bereiche) miteinander verbunden. Die Exons können dabei z. T. unterschiedlich miteinander kombiniert werden (z. B. durch Auslassen oder Umlagern von Exons). Durch solche Variationen im Spleißen derselben RNA entstehen unterschiedliche mRNAs, die dann in unterschiedliche Aminosäuresequenzen (und damit Proteine) übersetzt werden (siehe auch Munschauer/Vogel, Kapitel 2).

Prinzipiell orientieren sich RNA-gestützte Analysen – dies gilt sowohl für Arrays als auch RNA-Sequenzierung – an direkten (oder indirekten) Vergleichen der Genexpressionsunterschiede zwischen den RNAs gesunder und erkrankter Zellen und Geweben. Aus diesen Vergleichen zieht man Rückschlüsse über die Stadien und Schwere von Erkrankungen. Aus diesen Daten lassen sich dann statistisch abgesicherte Informationen extrahieren, die zur Bestimmung krankheitsrelevanter Genveränderungen und so zur begleitenden Eingrenzung und Optimierung von Behandlungsstrategien genutzt werden („companion diagnostics“)⁶ (Sager et al., 2015; Slembrouck et al., 2019).

RNA-Sequenzierungsdaten bieten dabei eine Vielzahl zusätzlicher Informationen. Sie erlauben es, Genveränderungen (Genvarianten) in gesunden und erkrankten Zellen zu bestimmen, und ergänzen und erweitern so die genetische Diagnostik durch funktionelle zellspezifische Daten. Aus RNA-Sequenzierungsdaten lassen sich zudem alternative Transkriptvarianten oder Veränderungen im Spleißen einzelner Gene bestimmen (Wang et al., 2009; Bayega et al., 2018; Murdock, 2020). Diese zusätzlichen genspezifischen Informationen sind z. B. bei der Krebsdiagnostik, aber auch der Diagnose seltener Krankheiten von großer Bedeutung. So wird die RNA-Sequenzierung zunehmend als ergänzendes „Instrument“ zur genombasierten molekularen Diagnose eingesetzt, da die Interpretation einiger Genvarianten allein auf der Grundlage von DNA-Sequenzierungsdaten oft uneindeutig ist. Der Nachweis einer abweichenden Expression (Ableserate) eines Gens wie auch von RNA-Varianten der Krankheitsgene und deren Zuordnung zu bestimmten Geweben oder Zellen liefert wichtige zusätzliche Informationen bezüglich der funktionellen Auswirkungen von genetischen Veränderungen. Die Einbindung von RNA-Sequenzierung trägt so zu einer deutlich verbesserten Diagnose seltener Erkrankungen bei (Peymani et al., 2022; Yépez et al., 2022).

RNA-Sequenzierungsdaten erlauben es zudem, die Folgen einer genomischen Veränderung nachzuweisen. So können z. B. Genfusionen entdeckt werden. In erkrankten (aber auch gesunden) Zellen können Umlagerungen von Chromosomenabschnitten dazu führen, dass Gene auf einem Chromosom miteinander so verknüpft werden, dass neue Fusionsgene entstehen. Diese werden dann auf RNA-Ebene als Fusionstranskript, also wie ein einziges Gen, abgelesen, sodass ein völlig neues (Fusions-)Protein entsteht. Solche Fusionsproteine können funktionelle Konsequenzen haben. Einige von ihnen können direkt zur malignen Entartung von Zellen, d. h. zur Krebsentstehung, beitragen. Das bekannteste Beispiel ist sicher BCR-ABL auf dem Philadelphia-Chromosom, welches ursächlich für bestimmte Leukämien (Blutkrebs) ist. Die Diagnose auf Grundlage der Expression (RNA-Sequenzierung) bietet hier eine effiziente Alternative zum Nachweis der zugrundeliegenden genetischen Veränderungen mittels Ganzgenomsequenzierung (Mertens et al., 2015; Heyer et al., 2020; Kerbs et al., 2022).

Generell ist festzustellen, dass gerade die qualitativen Unterschiede von RNAs eine zunehmende Aufmerksamkeit in der Diagnostik erlangen. So wird man in Zukunft vermehrt RNA-Sequenzierungsdaten im genetischen Kontext der untersuchten Person betrachten und

⁶ Der Begriff „companion diagnostics“ bezieht sich auf bestimmte diagnostische Tests, die vor bzw. begleitend zu einer Therapie durchgeführt werden. Dabei wird auf Grundlage der Analyse spezifischer Biomarker die Anwendbarkeit einer Therapie bzw. eines Arzneimittels auf individuelle Patienten bestimmt (personalisierte Medizin).

interpretieren, d. h. die individuelle genetische Ausstattung wird im Kontext von und Zusammenhang mit Genaktivitätsveränderungen bestimmt und bewertet. Man ermittelt dabei sogenannte „Expressionsquantitätsvarianten“ („expression quantitative trait locus“, eQTL) (Montgomery et al., 2010; Pickrell et al., 2010; Majewski/Pastinen, 2011; Lappalainen et al., 2013). eQTL-Varianten sind bedeutsam für die Analyse und das Verständnis komplexer Erkrankungen, die durch Expressionsveränderungen mehrerer bis vieler Gene verursacht werden.

Die RNA-Analytik wird auch zunehmend zur Erforschung von Krankheitsmodellen eingesetzt. Hierbei werden z. B. *in vitro* etablierte Zelllinien bzw. sogenannte Organoide⁷ als Ersatz für primäre Zellen bzw. Organe eingesetzt, die für eine direkte Analyse nicht oder nur begrenzt zur Verfügung stehen, wie z. B. Nervenzellen. Diese Modellsysteme werden im Labor *in vitro* nach standardisierten Protokollen aus (i. d. R. pluripotenten) Vorläuferzellen generiert, welche sich zu verschiedenen Zelltypen ausdifferenzieren und dabei eine komplexe dreidimensionale Struktur ausbilden. Eine gezielte Stimulierung der Zellen bzw. Organoide erlaubt dann, z. B. Alters- und Krankheitsprozesse oder auch (Neben-)Effekte durch (potenzielle) medikamentöse Wirkstoffe zu simulieren. Der Vergleich der Genexpressionssignaturen mittels RNA-Sequenzierung oder Arrays hilft hier nicht nur bei der Klärung der essenziellen Frage, wie nahe das Modell dem In-vivo-Organ bzw. den es bildenden Zellen kommt. Auch die induzierten Auswirkungen auf die molekularen Signaturen der Zellen oder des Organoids werden über (Einzelzell-)RNA-Sequenzierung hochsensitiv erfasst und über einen Abgleich mit unbehandelten Vergleichsproben interpretierbar. Die daraus gezogenen Erkenntnisse liefern wichtige Informationen für die Diagnose und auch Therapie der untersuchten Erkrankungen (Brancati et al., 2020).

Eine steil ansteigende Anzahl von Anwendungen erfährt in jüngster Zeit die Einzelzellsequenzierung. Hierunter summiert man eine Reihe von NGS-Verfahren, mit denen komplexe Transkriptome einzelner Zellen bestimmt werden.⁸ Einzelzelltranskriptomdaten bieten einen vollkommen neuen Zugang zur Definition von Zelltypen, Zellzuständen und pathologischen Veränderungen in erkrankten Zellen. Dabei können Veränderungen in der Zellzusammensetzung während der Entwicklung bzw. Alterung, aber auch im Zusammenhang mit verschiedenen Erkrankungen wie Autoimmunerkrankungen und Krebs, direkt untersucht werden. Für verschiedene Krebsarten wie Brust- und Lungenkrebs wurden bereits Einzelzell-„Atlanten“ erstellt, die neue Erkenntnisse über die zelluläre Heterogenität in Tumoren liefern und damit eine neue Ebene für diagnostische und therapeutische Anwendungen eröffnen (siehe auch Walter/Schickl, 2019). In Zukunft wird die Einzelzellsequenzierung einzelner Tumore dazu beitragen, die Entstehungsgeschichte und die Evolution individueller Tumoren besser zu verstehen und Diagnose und Therapie entsprechend auszurichten. Die RNA-Analytik von Einzelzellen kann dabei auch noch mit epigenetischen Techniken verbunden werden, wie z. B. mit ATAC-Sequenzierung,⁹ um eine weitere Ebene funktioneller

⁷ Siehe auch: Bartfeld et al., 2020.

⁸ Derzeitige Methoden der Einzelzellsequenzierung stützen sich auf Mikroprozessierungstechniken, um Zellen eines Gewebes zu vereinzeln (voneinander zu trennen) und weiter zu verarbeiten. Dabei werden komplexe Sequenzierungsbibliotheken mit individuellen Kodierungen erstellt, mit denen sich nach paralleler Hochdurchsatzsequenzierung tausender Zellen die Expressionsprofile einzelner Zellen unterscheiden lassen.

⁹ Mit der Methode ATAC-Sequenzierung („Assay for Transposable Accessible Chromatin using sequencing“) kann man regulatorische Regionen im Genom kartieren, in dem die lokale Zugänglichkeit des Chromatins (Komplex aus DNA und Proteinen, der die Chromosomen bildet) analysiert wird.

Dysregulation zu erschließen. Einzelzellsequenzierungsverfahren bieten eine extrem hohe Auflösung. Die räumliche Zuordnung bzw. Verortung der einzelnen sequenzierten Zellen (z. B. im Tumor) geht bei diesem Verfahren allerdings verloren. Diese Informationen werden ergänzt durch neue technische Ansätze, mit deren Hilfe man die Expression von Genen (Genvarianten) auch vor Ort, d. h. direkt im Gewebe auf Gewebeschnitten, studieren kann. Diese Technik wird als „spatial transcriptomics“ bezeichnet. Sie erlaubt die zellgenaue Verortung pathologischer Genexpressionsveränderungen im Gewebe (Moor/Itzkovitz, 2017).

All diese Anwendungsbeispiele verdeutlichen die enorm große Bedeutung der RNA-gestützten Einzelzell Diagnostik. Mit einer zunehmenden Anwendbarkeit im klinischen Bereich, insbesondere aufgrund der rasant fortschreitenden Methoden zur Automatisierung RNA-gestützter Experimente, wird die Einzelzellanalytik in naher Zukunft eine zentrale und führende Rolle in der klinischen Forschung und Diagnostik einnehmen. Die Errichtung von Forschungszentren wie der Zellklinik an der Charité gemeinsam mit dem MDC¹⁰ unterstreicht diese Tendenz sehr deutlich.

Tabelle 2: Überblick über die unterschiedlichen diagnostischen Verfahren und ihre Anwendung

Diagnostisches Verfahren	Nutzbarkeit	Anwendung
PCR	Schneller Nachweis und Quantifizierung von ausgesuchten, bekannten einzelnen RNAs.	<ul style="list-style-type: none"> • Nachweis der RNA pathogener Organismen und Viren (einzelne Gene) • Analysen einzelner, genspezifischer mRNAs und nicht-kodierender RNAs (z. B. miRNAs) • Quantitativer Nachweis bekannter Markergene, Fusionstranskripte • Anwendbarkeit für Einzelzellen
Microarray	Standardisierter Nachweis und relative Quantifizierung mehrerer bis vieler bekannter (annotierter) Gene, bzw. von Biomarkern.	<ul style="list-style-type: none"> • Nachweis von mRNAs (ohne Genvarianten) • Nachweis annotierter miRNAs und anderer nicht-kodierender RNAs
RNA-Sequenzierung (2. Generation) (Leselängen: 50–100 Basenpaare)	Hochsensitiver Nachweis geringster RNA-Mengen. Quantitative und Qualitative Analyse aller in der Zelle vorkommenden RNAs. Identifizierung neuer Transkripte und Transkriptvarianten.	<ul style="list-style-type: none"> • Nachweis von RNA pathogener Organismen und Viren • Quantifizierung von Markergenen (mRNA) oder nicht-kodierenden RNAs (z. B. miRNAs, circRNAs) • Auslesen bekannter und Erkennen neuer Fusionstranskripte • Teilerfassung von Transkript- und Spleißvarianten • Breite Anwendbarkeit für Einzelzellanalysen, z. B. Bestimmung der Zellidentität und Zellzusammensetzung (Heterogenität) im Gewebe („Zellatlanten“)

¹⁰ Berlin Cell Hospital, siehe unter: <https://www.mdc-berlin.de/de/themen/berlin-cell-hospital> [31.08.2022]; sowie Fokusbereich „Einzelzellansätze für die personalisierte Medizin“ des BIH, Charité und MDC, siehe unter: <https://www.mdc-berlin.de/de/content/single-cell-ansaeetze-fuer-die-personalisierte-medizin> [31.08.2022].

RNA-Sequenzierung (3. Generation) (Leselängen: 1–100 Kilobasenpaare)	Direkte Sequenzierung komplexer, langer RNAs. Bestimmung von RNA Modifikationen.	<ul style="list-style-type: none"> • Erfassen von kompletten Transkript- und Spleißvarianten (mRNA) • Darstellung komplexer Virus-Transkriptvarianten (z. B. SARS-CoV-2) • Auslesen modifizierter Basen in der RNA • Bisher keine direkte Anwendbarkeit für Einzelzellanalysen
---	---	--

3.4 Verarbeitung diagnostischer RNA-Daten

Für die Nutzung von Daten, die mithilfe hochspezialisierter Technologien (Arrays, 2. und 3. Generation-Sequenzierungsmethoden) erzeugt werden, spielt die bioinformatische Auswertung eine fundamentale Rolle. Um für die Diagnostik nutzbar zu sein, bedarf die Auswertung und Interpretation komplexer Array- und RNA-Sequenzierungsdaten standardisierter computergestützter Verfahren. Hierzu existieren sehr gute Programmpakete, die von der primären Datenprozessierung über die Qualitätsüberprüfung bis hin zur Interpretation reichen (Slonim/Yanai, 2009; Kukurba/Montgomery, 2015; Conesa et al., 2016; Zoabi/Shomron, 2021; Pavlovich/Cauchy, 2022).

Aktiv verwaltete öffentliche Referenzdatenbanken ermöglichen einen normierten Vergleich, d. h. eine Zuordnung der Proben und Zellen und damit auch eine Vergleichbarkeit der auf die Daten gestützten funktionellen Interpretation. Alle in der Diagnostik erhobenen Daten sollten daher in datengeschützten Räumen (Datenbanken) abgelegt werden, um so im kontrollierten Rahmen der Forschung und Diagnostik zur Verfügung zu stehen. In Europa spielt hier das European Genome Archive (EGA)¹¹ und sein föderierter deutscher Knoten, das German Human Genome-Phenome Archive (GHGA)¹² eine wichtige Rolle als sichere Datenablage, die gleichzeitig eine Zugänglichkeit und Bereitstellung von Vergleichsdaten ermöglicht. Daneben sind Datenportale wie Ensemble (funktionelle Daten),¹³ und die Plattformen des Genotype-Tissue Expression Project (GTEx, Expressions- und genetische Daten)¹⁴ und des Human Cell Atlas (HCA, Einzelzellaten)¹⁵ von unschätzbare Bedeutung für die diagnostische Interpretation.

Im zweiten Schritt erfolgt dann die Ermittlung der genspezifischen RNA-Menge in einer oder mehreren Zellen und deren „normierter“ Vergleich mit anderen Zellen (bzw. mit Referenzdaten). Aus RNA-Sequenzierung kann man in diesen weiterführenden Analysen auslesen, in welchen Mengen Gene in Zellen oder Geweben abgelesen wurden. Zudem kann man bestimmen, in welchem Ausmaß alternative Formen (Spleißvarianten, siehe oben) gen-

¹¹ Siehe unter: <https://ega-archive.org/> [29.08.2022].

¹² Siehe unter: <https://www.ghga.de/> [29.08.2022].

¹³ Siehe unter: <https://www.ensembl.org/index.html> [29.08.2022].

¹⁴ Siehe unter: <https://gtexportal.org/home/> [29.08.2022].

¹⁵ Siehe unter: <https://www.humancellatlas.org/> [29.08.2022].

spezifischer RNAs gebildet werden. Durch Verortung der RNA-Sequenzierungsdaten im Genom kann man feststellen, ob Gene von einer Stelle aus abgelesen werden oder ob (neue) alternative Start- und Endstellen (Isoformen) für die Bildung einer genspezifischen RNA in erkrankten Zellen genutzt werden. Abhängig von der Tiefe und Qualität der Daten kann man auch Rückschlüsse auf genetische Veränderungen ziehen und so eine genetische Diagnostik unterstützen (z. B. in Krebszellen).

Alle diese primären, funktionellen Zuordnungen der RNA-Daten sind wichtig für die Unterstützung von Diagnosen. In Zukunft wird man aber sicher Analysen und Interpretationen auf der nächstfolgenden Ebene mit einbeziehen. Hier untersucht man, welche Gene bzw. RNAs gemeinsam und in welcher Stärke in erkrankten Zellen oder Geweben fehlreguliert sind. Durch diese gleichzeitige Betrachtung vieler Gene und durch Nutzung von (selbstlernenden) Netzwerkmodellierungen kann man in einem zweiten Schritt korrelative Beziehungen der veränderten Expression von Genen bestimmen (Kakati et al., 2019; Chowdhury et al., 2020). Eine auf Netzwerkanalysen gestützte Interpretation ermöglicht eine viel differenziertere Betrachtung der Veränderungen und trägt zur Entwicklung neuer verbesserter Diagnose- und Behandlungsoptionen bei. Erkrankungsspezifische Effekte werden in diesen Analysen mehr im Kontext der regulatorischen Veränderung in erkrankten Zellen betrachtet.

Dies erweitert auch das Spektrum für die Nutzung von genspezifischen „Biomarker“-Testverfahren (z. B. mittels PCR-Tests). Um den funktionellen Veränderungen von erkrankten Zellen näherzukommen, werden zunehmend weitere Omics-Daten¹⁶ in die Interpretation mit einbezogen. Man spricht dann von integrierten (Multi-Omics-)Analysen, in denen gleichzeitig erhobene oder aus Referenzdaten bezogene Omics-Daten, z. B. zum Proteom oder Epigenom, mit genutzt werden, um z. B. zu ergründen, inwieweit Programmabläufe in den Zellen bereits verändert sind und ob und wie diese umkehrbar sind (Argelaguet et al., 2018; Subramanian et al., 2020; Li et al., 2021; Correa-Aguila et al., 2022).

3.5 Herausforderungen RNA-gestützter Diagnostik

Die RNA-sequenzierungsgestützte Diagnostik eröffnet ein differenziertes Bild der Zellfunktion und ist zudem breit „skalierbar“. Analysen sind selbst mit winzigsten Mengen von RNA aus Biopsien und Blut, selbst aus Blutplasma durchführbar. Je nach Quelle und angewandtem Verfahren ergeben sich aber auch Herausforderungen für eine RNA-gestützte Analytik. So bestimmen die Methoden der RNA-Isolierung, ob große oder kleine RNAs analysierbar sind (<200 Basenpaare, >200 Basenpaare) oder ob genspezifische (polyA)¹⁷ oder alle RNAs („total RNA“)¹⁸ angeschaut werden. Ein wichtiger Aspekt der RNA-Analytik ist die Integrität der isolierten RNA für die Diagnostik. RNA hat eine deutlich geringere Stabilität als DNA. Die Qualität RNA-gestützter Daten hängt daher stark von einem schnellen Zugang und der optimalen Gewinnung des Materials (Gewebe, Blut oder Zellen) ab. RNA-Analysen

¹⁶ Der Begriff „Omics“ bezieht sich auf molekularbiologische Studien, bei denen der gesamte zelluläre Gehalt der untersuchten Moleküle im Mittelpunkt steht, z. B. Proteomik (Proteine), Epigenomik (epigenetische Zustände) usw.

¹⁷ Zur Analyse der genspezifischen bzw. kodierenden RNAs (mRNAs) können diese z. B. über die Polyadenylierung (polyA) an ihrem 3'-Ende angereichert werden (3' bezieht sich auf das in Leserichtung hintere Ende der RNA).

¹⁸ Damit gemeint sind alle RNAs (kodierende und nicht-kodierende RNAs).

sollten möglichst an frischen Zellen bzw. Zellmaterial durchgeführt werden und an standardisierte Abläufe gekoppelt sein, um eine hohe Qualität und Vergleichbarkeit zu erzielen. Vergleichbarkeit sollte zudem stets durch Mehrfachwiederholungen der Analysen verbessert werden. Durch computergestützte Normalisierungsverfahren der Rohdaten kann man einige der experimentell inhärenten Varianzprobleme in der RNA-Analytik zu einem gewissen Ausmaß ausgleichen, d. h. technische Abweichungen, die z. B. auf unterschiedliche RNA-Qualität zurückzuführen sind, können „herausgerechnet“ werden.

Die Komplexität von RNA-Daten macht Ihre Interpretation zu einer herausfordernden Aufgabe. RNA-Daten eines Gewebes repräsentieren das sich überlagernde Gesamtspektrum aller Zellen dieses Gewebes. Ein diagnostischer Vergleich von gesundem und krankem Gesamtgewebe bietet daher nur einen gemittelten Wert der im Gewebe befindlichen gesunden und krankhaft veränderten Zellen. Man kann diese Daten durch computergestützte Verfahren (sogenannte Deconvolution [Entwirrung]) verbessern und nutzt hierzu Referenzdaten isolierter Zelltypen aus öffentlichen Datenbanken. Falls die Möglichkeit besteht, kann man die Analysen aber auch experimentell verbessern, indem man vor einer RNA-Analyse die Zellen des Gewebes oder Blutes mithilfe von Zellsortierern auftrennt und die Zelltypen dann separat sequenziert.¹⁹ Dies ist ein gängiges Verfahren, um z. B. die diversen Zellen im Blut voneinander zu trennen. Genaue Einblicke in die Zellzusammensetzung und Zellveränderungen erhält man aber erst durch Einzelzellanalysen. Hier werden zunächst die Zellen des Gewebes oder Blutes vereinzelt und dann tausende von zufällig ausgewählten Zellen des Gewebes einzeln parallel sequenziert und analysiert. Die Kosten und die experimentelle Logistik sowie der Zugang zu hochspezialisierten Instrumenten und die benötigte Expertise nehmen mit dem Auflösungsgrad der Analysen deutlich zu. Es wird daher immer wichtiger werden, die Entscheidungskompetenz zu fördern, welche praxisnahen und realisierbaren Diagnoseverfahren für eine spezifische diagnostische Fragestellung anzuwenden sind. Die Etablierung von Zentren für hochspezialisierte RNA-Analytik an Kliniken ist dabei eine Notwendigkeit, um so eine effiziente Nutzung und Entwicklung moderner Technologien und bioinformatischer Verfahren zu ermöglichen.

3.6 Ausblick

Die auf RNA-Analysen gestützte molekulare Diagnostik ist von größter Bedeutung für die personenbezogene medizinische Anwendung. Insbesondere die neuen Verfahren der Einzelzellanalytik werden in absehbarer Zukunft die Sensitivität und Qualität pathologischer Befunde in hohem Maße ergänzen und bereichern. Die rasant fortschreitende Anwendung neuer technischer Methoden zur RNA-Sequenzierung (Einzelzellen, „spatial transcriptomics“, Direktsequenzierung der 3. Generation) wird Daten generieren, die eine wesentlich differenziertere und genauere Bestimmung individueller Veränderungen in erkrankten Zellen erlauben. Dies wird eine neue Ära personenbezogener molekularer Diagnostik eröffnen.

¹⁹ Die Zellen des Gewebes oder Blutes werden dabei basierend auf Größe, Form und der Anwesenheit bestimmter Moleküle auf der Zelloberfläche nach Typ getrennt und so angereichert.

3.7 Literaturverzeichnis

- Andersen, K. G. et al. (2020):** The proximal origin of SARS-CoV-2. In: *Nat Med* 26(4): 450–452. DOI: 10.1038/s41591-020-0820-9.
- Argelaguet, R. et al. (2018):** Multi-Omics Factor Analysis – a framework for unsupervised integration of multi-omics data sets. In: *Mol Syst Biol* 14(6): e8124. DOI: 10.15252/msb.20178124.
- Athanasopoulou, K. et al. (2021):** A. Third-Generation Sequencing: The spearhead towards the radical transformation of modern genomics. In: *Life* 12(1): 30. DOI: 10.3390/life12010030.
- Bartfeld, S. et al. (2020):** Organoide. Ihre Bedeutung für Forschung und Medizin und Gesellschaft. Baden-Baden: Nomos.
- Bayega, A. et al. (2018):** Current and future methods for mRNA analysis: A drive toward single molecule sequencing. In: *Methods Mol Biol* 1783: 209–241. DOI: 10.1007/978-1-4939-7834-2_11.
- Bernardes, J. P. et al. (2020):** Longitudinal multi-omics analyses identify responses of megakaryocytes, erythroid cells, and plasmablasts as hallmarks of severe COVID-19. In: *Immunity* 53(6): 1296–1314.e9. DOI: 10.1016/j.immuni.2020.11.017.
- Brancati, G. et al. (2020):** Resolving neurodevelopmental and vision disorders using organoid single-cell multi-omics. In: *Neuron* 107(6): 1000–1013. DOI: 10.1016/j.neuron.2020.09.001.
- Brinkmann, A. et al. (2021):** AmpliCoV: Rapid whole-genome sequencing using multiplex PCR amplification and real-time Oxford nanopore MinION sequencing enables rapid variant identification of SARS-CoV-2. In: *Front Microbiol* 12: 651151. DOI: 10.3389/fmicb.2021.651151.
- Bustin, S. A./Mueller, R. (2005):** Real-time reverse transcription PCR (qRT-PCR) and its potential use in clinical diagnosis. In: *Clin Sci (Lond)* 109(4): 365–379. DOI: 10.1042/CS20050086.
- Chowdhury, H. A. et al. (2020):** (Differential) Co-expression analysis of gene expression: A survey of best practices. In: *IEEE/ACM Trans Comput Biol Bioinform* 17(4): 1154–1173. DOI: 10.1109/TCBB.2019.2893170.
- Chua, R. L. et al. (2020):** COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. In: *Nat Biotechnol* 38(8): 970–979. DOI: 10.1038/s41587-020-0602-4.
- Conesa, A. et al. (2016):** A survey of best practices for RNA-seq data analysis. In: *Genome Biol* 17: 13. DOI: 10.1186/s13059-016-0881-8.
- Corman, V. M. et al. (2020):** Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR In: *Euro Surveill* 25(3): 2000045. DOI: 10.2807/1560-7917.ES.2020.25.3.2000045.
- Correa-Aguila, R. et al. (2022):** Multi-omics data integration approaches for precision oncology. In: *Mol Omics* 18(6): 469–479. DOI: 10.1039/d1mo00411e.
- Dao Thi, V. L. et al. (2020):** A colorimetric RT-LAMP assay and LAMP-sequencing for detecting SARS-CoV-2 RNA in clinical samples. In: *Sci Transl Med* 12(556): eabc7075. DOI: 10.1126/scitranslmed.abc7075.
- Freed, N. E. et al. (2020):** Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. In: *Biol Methods Protoc* 5(1): bpaa014. DOI: 10.1093/biomet/bpaa014.
- Gorreta, F. et al. (2012):** Genomic profiling: cDNA arrays and oligoarrays. In: *Methods Mol Biol* 823: 89–105. DOI: 10.1007/978-1-60327-216-2_7.
- Heyer, E. E. et al. (2020):** Diagnosis of fusion genes using targeted RNA sequencing In: *Nat Commun* 10(1): 1388. DOI: 10.1038/s41467-019-09374-9.
- Kakati, T. et al. (2019):** Comparison of methods for differential co-expression analysis for disease biomarker prediction. In: *Comput Biol Med* 113: 103380. DOI: 10.1016/j.combiomed.2019.103380.
- Kappel, A./Keller, A. (2017):** miRNA assays in the clinical laboratory: workflow, detection technologies and automation aspects. In: *Clin Chem Lab Med* 55(5): 636–647. DOI: 10.1515/cclm-2016-0467.
- Kellner, M. J. et al. (2022):** A rapid, highly sensitive and Open-Access SARS-CoV-2 detection assay for laboratory and home testing. In: *Front Mol Biosci* 9: 801309. DOI: 10.3389/fmolb.2022.801309.
- Kerbs, P. et al. (2022):** Fusion gene detection by RNA-sequencing complements diagnostics of acute myeloid leukemia and identifies recurring NRIP1-MIR99AHG rearrangements. In: *Haematologica* 107(1): 100–111. DOI: 10.3324/haematol.2021.278436.
- Kukurba, K. R./Montgomery, S. B. (2015):** RNA Sequencing and Analysis. In: *Cold Spring Harb Protoc* 2015(11): 951–969. DOI: 10.1101/pdb.top084970.
- Lappalainen, T. et al. (2013):** Transcriptome and genome sequencing uncovers functional variation in humans. In: *Nature* 501(7468): 506–511. DOI: 10.1038/nature12531.
- Li, Y. et al. (2021):** Advances in bulk and single-cell multi-omics approaches for systems biology and precision medicine. In: *Brief Bioinform* 22(5): bbab024. DOI: 10.1093/bib/bbab024.

- Majewski, J./Pastinen, T. (2011):** The study of eQTL variations by RNA-seq: from SNPs to phenotypes. In: *Trends Genet* 27(2): 72–79. DOI: 10.1016/j.tig.2010.10.006.
- Mertens, F. et al. (2015):** The emerging complexity of gene fusions in cancer. In: *Nat Rev Cancer* 15(6): 371–381. DOI: 10.1038/nrc3947.
- Montgomery, S. B. et al. (2010):** Transcriptome genetics using second generation sequencing in a Caucasian population. In: *Nature* 464(7289): 773–777. DOI: 10.1038/nature08903.
- Moor, A. E./Itzkovitz, S. (2017):** Spatial transcriptomics: paving the way for tissue-level systems biology. In: *Curr Opin Biotechnol* 46: 126–133. DOI: 10.1016/j.copbio.2017.02.004.
- Mortazavi, A. et al. (2008):** Mapping and quantifying mammalian transcriptomes by RNA-Seq. In: *Nat Methods* 5(7): 621–628. DOI: 10.1038/nmeth.1226.
- Murdock, D. R. (2020):** Enhancing diagnosis through RNA sequencing. In: *Clin Lab Med* 40(2): 113–119. DOI: 10.1016/j.cll.2020.02.001.
- Muus, C. et al. (2021):** Single-cell meta-analysis of SARS-CoV-2 entry genes across tissues and demographics. In: *Nat Med* 27(3): 546–559. DOI: 10.1038/s41591-020-01227-z.
- Pavlovich, P. V./Cauchy, P. (2022):** Sequences to Differences in Gene Expression: Analysis of RNA-Seq Data. In: *Methods Mol Biol* 2508: 279–318. DOI: 10.1007/978-1-0716-2376-3_20.
- Peymani, F. et al. (2022):** RNA sequencing role and application in clinical diagnostic. In: *Pediatr Investig* 6(1): 29–35. DOI: 10.1002/ped4.12314.
- Pickrell, J. K. et al. (2010):** Understanding mechanisms underlying human gene expression variation with RNA sequencing. In: *Nature* 464(7289): 768–772. DOI: 10.1038/nature08872.
- Sager, M. et al. (2015):** Transcriptomics in cancer diagnostics: developments in technology, clinical research and commercialization. In: *Expert Rev Mol Diagn* 15(12): 1589–1603. DOI: 10.1586/14737159.2015.1105133.
- Schulte-Schrepping, J. et al. (2020):** Severe COVID-19 is marked by a dysregulated myeloid cell compartment. In: *Cell* 182(6): 1419–1440.e23. DOI: 10.1016/j.cell.2020.08.001.
- Slembrouck, L. et al. (2019):** Decentralization of next-generation RNA sequencing-based MammaPrint® and Blueprint® kit at University Hospitals Leuven and Curie Institute Paris. In: *Transl Oncol* 12(12): 1557–1565. DOI: 10.1016/j.tranon.2019.08.008.
- Slonim, D. K./Yanai, I. (2009):** Getting started in gene expression microarray analysis. In: *PLoS Comput Biol* 5(10): e1000543. DOI: 10.1371/journal.pcbi.1000543.
- Stephenson, E. et al. (2021):** Single-cell multi-omics analysis of the immune response in COVID-19. In: *Nat Med* 27(5): 904–916. DOI: 10.1038/s41591-021-01329-2.
- Subramanian, I. et al. (2020):** Multi-omics data integration, interpretation, and its application. In: *Bioinform Biol Insights* 14: 1177932219899051. DOI: 10.1177/1177932219899051.
- Sungnak, W. et al. (2020):** SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. In: *Nat Med* 26(5): 681–687. DOI: 10.1038/s41591-020-0868-6.
- VanGuilder, H. D., et al. (2008):** Twenty-five years of quantitative PCR for gene expression analysis. In: *Biotechniques* 44(5): 619–626. DOI: 10.2144/000112776.
- Walter, J./Schickl, H. (2019):** Einzelzellanalyse in Forschung und Medizin – Eine Stellungnahme der interdisziplinären Arbeitsgruppe Gentechnologiebericht. Berlin Brandenburgische Akademie der Wissenschaften, Berlin.
- Wang, Z. et al. (2009):** RNA-Seq: a revolutionary tool for transcriptomics. In: *Nat Rev Genet* 10(1): 57–63. DOI: 10.1038/nrg2484.
- Yan, C. et al. (2020):** Rapid and visual detection of 2019 novel coronavirus (SARS-CoV-2) by a reverse transcription loop-mediated isothermal amplification assay. In: *Clin Microbiol Infect* 26(6): 773–779. DOI: 10.1016/j.cmi.2020.04.001.
- Yang, A. C. et al. (2021):** Dysregulation of brain and choroid plexus cell types in severe COVID-19. In: *Nature* 595(7868): 565–571. DOI: 10.1038/s41586-021-03710-0.
- Yépez, V. A. et al. (2022):** Clinical implementation of RNA sequencing for Mendelian disease diagnostics. In: *Genome Med* 14(1): 38. DOI: 10.1186/s13073-022-01019-9.
- Zoabi, Y./Shomron, N. (2021):** Processing and Analysis of RNA-seq Data from Public Resources. In: *Methods Mol Biol* 2243: 81–94. DOI: 10.1007/978-1-0716-1103-6_4.

4. Die Erschließung der RNA-Welt für Therapeutika

Anke Sparmann und Jörg Vogel

RNA-basierte Therapeutika sind entweder Arzneimittel, die zelluläre RNA als Zielstruktur haben, oder aber in menschliche Zellen eingeschleuste RNA, die die Produktion bestimmter Proteine bedingt. Nach langjähriger Forschungs- und Entwicklungsarbeit gibt es für diese neue Art von Medizin jetzt die ersten großen Erfolge bei der Patientenbehandlung.

4.1 Einleitung

Die Idee RNA-basierter Therapeutika ist fast ein halbes Jahrhundert alt. Die Ursprünge können auf das Jahr 1978 zurückgeführt werden, als Mary Stephenson und Paul Zamecnik ein synthetisches Antisense-Oligonukleotid (kurz ASO)¹ herstellten, um damit die Replikation des Rous-Sarkom-Virus in Gewebekultur zu hemmen (Zamecnik/Stephenson, 1978). Durch sequenzspezifische Bindung an die virale Boten-RNA (mRNA, siehe auch Munschauer/Vogel, Kapitel 2) hemmte dieses ASO die Translation der mRNA zu Protein (Stephenson/Zamecnik, 1978). Diese beiden Studien bereiteten den Weg dafür, die einzigartige Eigenschaft der sequenzspezifischen Bindung von Nukleinsäuren für die Entwicklung eines neuen Arzneimitteltyps zu nutzen. Während die herkömmliche Medikamentenentwicklung oft mühsames Screening nach chemischen Leitsubstanzen erfordert, können RNA- und DNA-Medikamente gezielt entwickelt werden. Das liegt daran, dass der Code, der sowohl die Wahrscheinlichkeit als auch die Spezifität der Oligonukleotid-Bindung bestimmt – die komplementäre Basenpaarung – gut verstanden ist. Solange man also die Sequenz der Ziel-RNA kennt, kann man ein ASO, welches an eine bestimmte Stelle dieser Ziel-RNA bindet praktisch am Bildschirm entwerfen. Dennoch brauchte es viele Jahre, um dieses Konzept aus dem Labor heraus in eine klinische Anwendung zu bringen. Die Nutzung von RNA in Arzneimitteln birgt beträchtliche Herausforderungen, bedingt durch die der RNA eigenen pharmakologischen Eigenschaften, hervorgerufen unter anderem durch die Größe und Ladung des RNA-Moleküls. Auch die geringe Aufnahme von RNA in Zielzellen und eine immunbezogene Toxizität stellen beträchtliche Hürden dar. Es brauchte Jahre an Forschungserfolgen in verschiedenen wissenschaftlichen Disziplinen, um anfängliche Rückschläge zu überwinden.

¹ Ein ASO ist eine kurzkettige Nukleinsäure mit einer frei wählbaren Abfolge von Basen; es bindet über komplementäre Basenpaarung an eine andere Nukleinsäure (RNA oder DNA), deren Basenabfolge dazu exakt passt.

Die Entdeckung des eingangs beschriebenen Antisense-Prinzips legte den Grundstein für die Entwicklung von RNA-basierten therapeutischen Strategien. Bereits vor der vollständigen Aufklärung der zugrundeliegenden Wirkmechanismen wurde dieses Konzept in der Medikamentenentwicklung angewandt. 1998 wurde das erste Oligonukleotid-Arzneimittel, das auf einem Antisense-Wirkmechanismus basiert, auf dem US-amerikanischen Markt zugelassen. Fomivirsen bindet eine Sequenz in der mRNA des Cytomegalovirus (CMV), das bei einem geschwächten Immunsystem schwere Erkrankungen hervorrufen kann, und hemmt die CMV-Replikation. Das ASO-Medikament war zur Behandlung von CMV-Retinitis – einer Infektion der Netzhaut, die zur Erblindung führen kann – bei Personen mit erworbenem Immunschwächesyndrom (AIDS) indiziert. Trotz des therapeutischen Nutzens war der Erfolg von Fomivirsen nur von kurzer Dauer: Aufgrund von Fortschritten in der antiretroviralen HIV-Therapie wurde es bald wieder vom Markt genommen. Dennoch lieferte Fomivirsen den ersten Nachweis des klinischen Werts von ASOs.

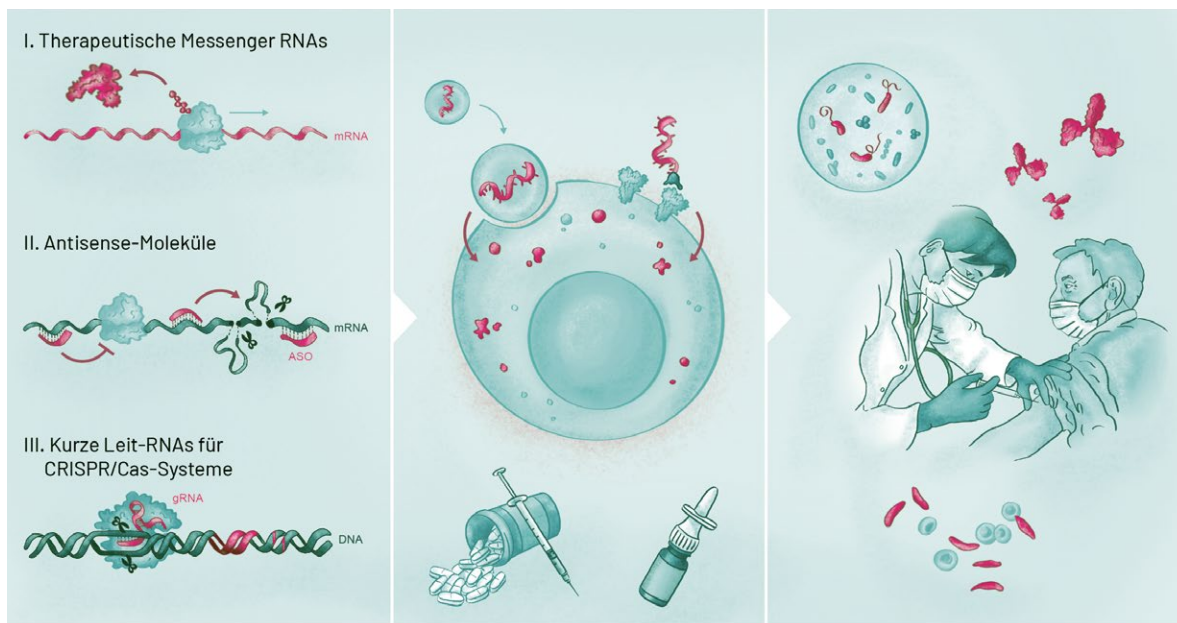
Wissenschaftliche Entwicklungen in den 1990er- und frühen 2000er-Jahren, insbesondere die Entdeckung von microRNAs (miRNAs), RNA-Interferenz (RNAi) und „Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-associated (Cas)“-Systemen, haben unser Verständnis der Genregulation grundlegend verändert und die Art und Weise, in der biologische Forschung betrieben wird, revolutioniert. Wie unten ausgeführt, haben auch diese RNA-basierten molekularen Mechanismen der Genregulation in die Medikamentenentwicklung Eingang gefunden.

Neben den nicht-kodierenden RNA-Therapeutika kann RNA-basierte Therapie aber noch breiter definiert werden, nämlich wenn RNA selbst als Informationsquelle dient, um im Patienten die Produktion eines bestimmten Proteins auszulösen, und zwar mittels verabreichter synthetischer mRNA. Erste Experimente dazu gab es schon in den 1990er-Jahren mit der Erforschung von *in vitro* transkribierter (IVT) mRNA, also im Labor synthetisierter mRNA. Zu solchen ersten Anwendungen zählten zum Beispiel die Produktion von körpereigenen Proteinen in Säugetierzellen sowie der Versuch, Impfungen gegen Krebs und Infektionskrankheiten zu entwickeln (Jirikowski et al., 1992; Mandl et al., 1998; Zhou et al., 1999). Allerdings verhinderten die ineffiziente zelluläre Aufnahme exogener, d. h. von außen in den Organismus eindringender, mRNA, deren geringe Stabilität im Körper und schwache Umsetzung in Protein in den Zielzellen sowie eine durch RNA hervorgerufene Reaktion des körperlichen Immunsystems eine erfolgreiche klinische Anwendung. Wieder waren wissenschaftliche Durchbrüche notwendig: Verabreichungswege mussten verbessert und Ansätze zur spezifischen Anreicherung der mRNA in bestimmten Zelltypen gefunden werden. Eine besonders wichtige Entwicklung war die Synthese der IVT-mRNA mit chemisch veränderten Nukleosiden.² Derartig modifizierte IVT-mRNA war nicht nur stabiler und führte zu einer höheren Produktion des gewünschten Proteins, sie war auch weniger immunstimulierend (Karikó et al., 2005), was wiederum die Verabreichung höherer Mengen erlaubte. Anders ausgedrückt wurde ein Weg gefunden, um synthetische RNAs weitgehend unbemerkt vom Immunsystem in die Zielzellen einzuschleusen (siehe auch Abschnitt 4.5). Dort angekommen wird das modifizierte mRNA-Molekül von der Zelle abgelesen und in

² Nukleoside sind ähnlich aufgebaut wie Nukleotide, die Bausteine der DNA oder RNA, enthalten aber keine Phosphatreste.

Protein umgewandelt – genauso wie das mit zelleigenen mRNAs im Rahmen der regulären Genexpression geschieht (siehe zur Genexpression Munschauer/Vogel, Kapitel 2). Die oben aufgeführten Fortschritte waren eine wichtige Voraussetzung für die außergewöhnlich schnelle Entwicklung von mRNA-basierten Impfstoffen zur Bekämpfung der SARS-CoV-2-Pandemie im Jahr 2020. Dieser enorme Erfolg hat nunmehr größtes Interesse an RNA als eine wichtige Molekülklasse in den Bereichen Diagnostik (siehe Walter/Gasparoni, Kapitel 3), Prävention und Behandlung von Krankheiten geweckt. Die Impfung eines großen Teils der Weltbevölkerung mit mRNA-basierten Vakzinen hat zudem umfangreiche Daten zur Sicherheit und Wirksamkeit von RNA als Impfstoff geliefert, was wiederum für die Zulassung der nächsten RNA-basierten Medikamente wichtig sein wird. Nun gilt es, das Wirkprinzip auf möglichst viele Krankheiten zu übertragen. Im Folgenden werden wir verschiedene Strategien der RNA-basierten Medizin diskutieren, darunter ASOs, therapeutische mRNAs und CRISPR/Cas-Systeme (siehe Abbildung 1).

Abbildung 1: Strategien der RNA-basierten Medizin



© SciGraphix/Sandy Westermann

4.2 ASO-Therapeutika

ASO-Therapeutika sind üblicherweise einzelsträngige, stark modifizierte und stabilisierte Nukleinsäure-Analoga,³ die das Prinzip der komplementären Basenpaarung nutzen, um an zelluläre RNAs zu binden und RNA-Stoffwechselprozesse zielgerichtet zu beeinflussen. ASOs haben verschiedene Wirkmechanismen, die sowohl den sequenzspezifischen Abbau von Ziel-RNAs als auch das durch Bindung bewirkte Unterbinden von Prozessierungsvorgängen wie etwa dem RNA-Spleißen⁴ oder der Translation zu Proteinen, umfassen.

³ Nukleinsäure-Analoga sind synthetisch hergestellte Substanzen, die im chemischen Aufbau natürlichen Nukleinsäuren ähneln.

⁴ Als Spleißen wird ein wichtiger Schritt der mRNA-Verarbeitung bezeichnet, bei dem aus einer zunächst in der Transkription von DNA zu RNA gebildeten, sogenannten prä-mRNA, die „reife“ mRNA entsteht. Die prä-mRNA enthält sowohl kodierende Bereiche (Exons) als auch nicht-kodierende Sequenzen (Introns). Durch das Spleißen werden die Introns entfernt und die angrenzenden Exons miteinander zur fertigen mRNA verknüpft, die dann in Protein translatiert wird (siehe auch Munschauer/Vogel, Kapitel 2).

Der wohl am besten verstandene Wirkmechanismus von ASOs ist der Abbau der Ziel-RNA durch die Ribonuklease⁵ RNase H1. RNase H1 baut RNA in DNA-RNA-Hybriden⁶ ab, benötigt dazu allerdings mindestens fünf, besser 8 bis 10, mit der Ziel-RNA gepaarte Desoxynukleotide, die Bausteine der DNA. Besonders effizient sind sogenannte „Gapmer“-ASOs, bei denen ein zentraler Desoxynukleotidkern von Nukleotiden flankiert wird, die an der 2'-Position chemisch verändert sind. Diese flankierenden Modifizierungen verbessern die Bindung des Gapmers zur Ziel-RNA und verringern zudem den Abbau durch unspezifische Nucleasen, d. h. sie erhöhen die Stabilität des Gapmers im Körper. Da nahezu perfekte Komplementarität zwischen ASOs und der Ziel-RNA innerhalb des Desoxynukleotidkerns erforderlich ist, um RNase-H1-vermittelten Abbau zu aktivieren, zeigen ASOs, die diesen Wirkmechanismus haben, eine hohe Spezifität. Das wiederum reduziert das Potenzial für ungewünschte Bindungen, sogenanntes „off targeting“.

Neben dem RNase-H1-induzierten Abbau gibt es weitere Mechanismen, durch die ASOs den zellulären Spiegel von Ziel-RNAs senken können. Zum einen können ASOs zelluläre Qualitätskontrollmechanismen auslösen, wie zum Beispiel den Nonsense-mediated mRNA Decay (NMD)⁷ oder No-Go Decay (NGD)⁸ von Ziel-mRNA-Molekülen. Zum anderen können ASOs endogene RNA-Stoffwechselprozesse, etwa die Polyadenylierung,⁹ beeinflussen und darüber den natürlichen Abbau von Ziel-mRNA-Molekülen beschleunigen. Des Weiteren können an die Translationsinitiationsstelle von mRNAs bindende ASOs die Initiation der Proteinsynthese durch Ribosomen blockieren und so ganz direkt zu einer Verminderung des kodierten Proteins führen. Dieser Mechanismus ist insbesondere in Bakterien sehr effektiv. ASOs können aber auch die Interaktion der mRNA mit Translationsinitiationsfaktoren¹⁰ hemmen oder die Spaltung von sogenannten 5'-Cap-Strukturen¹¹ auslösen, was ebenfalls zu einer Hemmung der Translation und zu mRNA-Abbau führt.

Das klinisch bisher am erfolgreichsten eingesetzte ASO nutzt jedoch einen Wirkmechanismus, der auf einer Beeinflussung des RNA-Spleißens beruht. Nusinersen ist ein seit 2016 zugelassenes ASO zur Behandlung von Spinaler Muskelatrophie (SMA), einer neuromuskulären Erkrankung, die durch Mutationen im *survival of motor neuron 1 (SMN1)*-Gen verursacht wird. Diese Mutationen führen zum Verlust des SMN-Proteins, ohne welches die Motoneuronen im Rückenmark und im Hirnstamm degenerieren. Das Ergebnis sind Muskelschwäche und Gewebeschwund. SMA ist eine destruktive Erkrankung, die bei 60 % der damit geborenen Säuglinge bereits vor dem sechsten Lebensmonat zu schweren Symptomen

5 Ribonukleasen sind Enzyme, die RNA abbauen.

6 DNA-RNA-Hybride sind Doppelstrangmoleküle, die aus einer DNA- und einer komplementären RNA-Nukleotidkette bestehen.

7 Als Nonsense-mediated mRNA Decay wird ein Kontrollmechanismus in eukaryotischen Zellen bezeichnet, der unerwünschte vorzeitige Stopcodons in mRNAs erkennt und diese mRNAs selektiv abbaut, um deren Synthese als verkürzte Proteine zu verhindern.

8 No-Go Decay ist ein mRNA-Qualitätskontrollmechanismus, bei dem mRNAs, die blockierte Ribosomen aufweisen, abgebaut werden.

9 Als Polyadenylierung bezeichnet man das Anhängen von Adenin-Nukleotiden, den sogenannten Poly(A)-Schwanz, an das 3'-Ende der mRNAs durch das Enzym Poly(A)-Polymerase.

10 Translationsinitiationsfaktoren sind Proteine oder Proteinkomplexe, die eine Funktion während der Initiation, also dem Beginn der Translation haben; sie sind notwendig für eine effiziente Translation der mRNA.

11 Die 5'-Cap-Struktur (aus dem Englischen „cap“ – Kappe) ist häufig ein modifiziertes Guanin-Nukleotid, das über eine seltene 5'-5'-Phosphodiesterbindung an das erste Nukleotid der mRNA geknüpft ist. Diese Struktur schützt mRNAs vor dem Abbau und ist für den Export einer mRNA aus dem Zellkern in das Zytoplasma und für die Translation wichtig (siehe auch Munschauer/Vogel, Kapitel 2).

führt; die mittlere Lebenserwartung liegt unter zwei Jahren. Das *SMN2*-Gen, ein Paralog¹² von *SMN1*, kodiert für ein identisches SMN-Protein. Die mRNA von *SMN2* wird jedoch im Rahmen des RNA-Spleißens alternativ verarbeitet, wodurch 90 % der *SMN2*-Transkripte ein Exon (Exon 7) fehlt. Im Ergebnis wird ein verkürztes, instabiles Protein produziert. Die molekularen Grundlagen dieses anomalen *SMN2*-Exon-7-„Skippings“ (Überspringen) wurden um die Jahrtausendwende aufgeklärt. *SMN2* enthält einen Cytosin-zu-Thymin-Austausch in Exon 7, was die Bindung von Spleißaktivatoren an die RNA schwächt und dadurch die effiziente Erkennung der 3'-Spleißstelle verringert. Schon im Jahr 2003 konstruierten Luca Cartegni und Adrian Krainer bifunktionelle ASOs,¹³ die als synthetische Spleißaktivatoren fungierten: Ein Peptid, das einen Spleißaktivator nachahmt, wurde mit einem ASO verknüpft, der an Exon 7 bindet und somit den Einbau von Exon 7 erhöht (Cartegni/Krainer, 2003).

In Zusammenarbeit mit der Firma Ionis Pharmaceuticals wurde in den darauffolgenden Jahren die Strategie zur Hemmung des Exon-7-Skippings optimiert. Es stellte sich heraus, dass ASOs, die an einer anderen Stelle der *SMN2* RNA binden, ebenfalls den Anteil von „reparierten“ mRNAs mit Exon 7 erhöhen, und dies sogar ohne angehängte Peptideinheit. Hier wird die Bindung von Spleißinhibitoren verhindert. Darüber hinaus konnten durch chemische Veränderungen der ASOs deren pharmakologische Eigenschaften erheblich verbessert werden. Vielversprechende Ergebnisse in präklinischen Studien sowie die nachfolgenden, außerordentlich erfolgreichen klinischen Studien führten 2016 zur Zulassung von Nusinersen auf dem US-amerikanischen Markt (in Europa: 2017).

Inzwischen ist Nusinersen in mehr als 40 Ländern zur Behandlung von SMA verfügbar. Dieser Wirkstoff hat nicht nur die Behandlung von SMA-Patienten revolutioniert, er ist wohl auch das erste Antisense-Medikament mit einem nennenswerten kommerziellen Erfolg. Der Erfolg von Nusinersen und die Zulassung von aktuell neun weiteren ASOs, meist zur Behandlung von seltenen Erbkrankheiten ohne andere Therapieoptionen, lassen hoffen, dass die ASO-Technologien die einst in sie gesetzten Hoffnungen erfüllen werden.

4.3 Therapeutika, die auf RNA-Interferenzmechanismen beruhen

Anfang der 1990er-Jahre machten Wissenschaftler in verschiedenen experimentellen Systemen, insbesondere auch in Pflanzen, rätselhafte Beobachtungen, die auf RNA-induzierte Genregulierung hindeuteten. Zunächst wurde dahinter ein Antisense-Mechanismus vermutet, der ähnlich den ASOs auf einer Bindung zwischen der regulierenden RNA und zellulärer mRNA beruht. Dann aber zeigten Andrew Fire und Craig Mello im Jahr 1998, dass doppelsträngige RNA in Fadenwürmern einen Vorgang auslöst, der zu sequenzspezifischem RNA-„Silencing“¹⁴ führt. Dieser Prozess wurde „RNA-Interferenz“, kurz RNAi, getauft (Fire et al., 1998). In nachfolgenden Studien wurde der zugrundeliegende molekulare Mechanismus

¹² Paraloge entstehen durch Genduplikation und haben eine hohe Übereinstimmung in ihrer Basensequenz. Die Genkopien sind anfangs identisch und haben oft ähnliche oder redundante Funktionen, können sich aber mit der Zeit unterschiedlich entwickeln.

¹³ Als bifunktionelle Verbindungen bezeichnet man Verbindungen, die zwei funktionelle Gruppen enthalten, in diesem Fall der Spleißaktivator und das ASO.

¹⁴ Silencing [engl.]: zum Schweigen bringen, ruhigstellen, abstellen.

erschlossen: Doppelsträngige RNA-Moleküle werden in „small interfering RNAs“ (siRNAs) umgewandelt, die in einen „RNA-induced Silencing Complex“ (RISC) eingebaut werden, der aus verschiedenen zellulären Proteinen besteht. Die siRNAs leiten dann diesen Komplex zu komplementären Ziel-RNAs, um Letztere enzymatisch zu spalten (Hammond, 2000; Zamore, 2000) (siehe auch Munschauer/Vogel, Kapitel 2). Diese Entdeckungen führten dann zur Entwicklung von synthetischen siRNAs, die nach Zugabe von außen spezifische Gene in Säugetierzellen abschalten konnten (Caplen et al., 2001; Elbashir et al., 2001). Die damit verbundene RNAi-Technologie wurde sehr schnell zu einem wichtigen Werkzeug der biologischen Grundlagenforschung, konnten doch damit zum ersten Mal in großem Umfang die Funktionen von Genen analysiert werden.

Bis zur Zulassung des ersten RNAi-Medikaments sollten dennoch fast 20 Jahre vergehen. 2018 wurde Patisiran für die Behandlung von Polyneuropathie bei Patienten mit Hereditärer Transthyretin-Amyloidose (hATTR) zugelassen. Diese seltene und verheerende neurodegenerative Erkrankung wird durch die Ablagerung von Amyloidfibrillen verursacht, die durch instabile Transthyretin (TTR)-Proteine gebildet werden. Patisiran ist eine aus modifizierten Oligonukleotiden bestehende doppelsträngige siRNA, die den Abbau der TTR-mRNA in der Leber induziert, darüber den Blutspiegel des Proteins senkt und somit mögliche Amyloidablagerungen reduziert. Ein Nachfolger von Patisiran ist seit Juni 2022 auf dem Markt: Vutrisiran nutzt den gleichen RNAi-Mechanismus, beruht aber auf einer verbesserten Stabilisierungsschemie. Zudem ist es mit dem Zuckerderivat N-Acetylgalactosamin (kurz GalNAc)¹⁵ gekoppelt, welches die Aufnahme der siRNA in Leberzellen erhöht und eine einfachere Dosierung ermöglicht. Während Patisiran alle drei Wochen intravenös injiziert werden muss, reicht bei Vutrisiran eine subkutane Injektion pro Quartal.

Inzwischen wurden bereits fünf RNAi-Medikamente durch die US-amerikanische Food and Drug Administration (FDA), die US-Behörde für Lebens- und Arzneimittel, zugelassen. Von RNAi-Therapeutika verspricht man sich ein noch größeres Anwendungsspektrum, vorausgesetzt, dass die körperweite Verabreichung und zielgerichtete zelluläre Aufnahme des Wirkstoffes außerhalb des Leber- und Nierengewebes verbessert werden kann. Unabhängig davon ist die 20-jährige Entwicklungsgeschichte auf dem Gebiet der RNAi-Therapeutika ein klarer Beweis, dass sich Hartnäckigkeit bei der Entwicklung solch programmierbarer Medikamente, die sich nach Bedarf entwerfen und anpassen lassen, auszahlt.

4.4 Therapeutisches Genome-Editing

Die Entdeckung eines von RNA gelenkten Immunabwehrsystems in Bakterien und Archaeabakterien¹⁶ gilt als ein Wendepunkt in der RNA-Biologie und Genomforschung. Diese sogenannten CRISPR/Cas-Systeme sind heute weitläufig bekannt, auch über die wissenschaftliche Gemeinschaft hinaus, hauptsächlich durch ihren großen Erfolg als programmierbare

¹⁵ N-Acetylgalactosamin ist ein Derivat der Galactose, das an den Asialoglycoproteinrezeptor bindet, der selektiv an der Oberfläche von Leberzellen vorkommt. Die Kopplung an GalNAc wird verwendet, um Oligonukleotide spezifisch an die Leber zu liefern.

¹⁶ Archaea sind kleine einzellige Mikroorganismen, die neben Bakterien und Eukaryoten eine der drei Domänen bilden, in die alle zellulären Lebewesen eingeteilt werden. Wie Bakterien gehören sie zu den Prokaryoten, besitzen also keinen Zellkern.

Genome-Editing-Tools¹⁷ (aus dem Englischen „editing“ – bearbeiten und „tools“ – Werkzeuge). Nach dem experimentellen Nachweis, dass CRISPR/Cas-Systeme eine adaptive Immunität gegen fremde DNA ermöglichen (Barrangou et al., 2007), und der Aufklärung des zugrundeliegenden Abwehrmechanismus (Garneau et al., 2010; Deltcheva et al., 2011; Gasiunas et al., 2012; Jinek et al., 2012) wurde bald auch das enorme Potenzial der RNA-gesteuerten Cas-Nukleasen für die Gentechnik erkannt. Dieses Konzept wurde dann in einer Reihe von Studien etabliert, die zeigten, dass durch das Cas9-Protein vermittelte Genomeditierung, also die gezielte Veränderung von DNA-Sequenzen, sowohl in Maus- als auch in menschlichen Zellen und sogar in komplexen Modellorganismen möglich ist (Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013). Diese neue Technologie revolutionierte in einem bis dato unbekanntem Tempo die biomedizinische Grundlagenforschung und stieß ähnlich der RNA-Interferenz unmittelbar auf großes Interesse der Pharmaindustrie hinsichtlich einer klinischen Anwendbarkeit.

Therapeutische Ansätze, die Genomeditierung zur Behandlung von Erbkrankheiten nutzen, befinden sich derzeit in verschiedenen Phasen der klinischen Erprobung, einige davon mit beachtlichem Erfolg. Zum Beispiel wurden mittels CRISPR/Cas-Gentherapie vielversprechende erste Behandlungsergebnisse bei Sichelzellanämie erzielt. Sichelzellanämie ist eine erblich bedingte Bluterkrankung, bei der rote Blutkörperchen eine sichelförmige Form annehmen, was zu verschlossenen Blutgefäßen, starken Schmerzen und lebensbedrohlichem Organversagen führen kann. Diese Krankheit tritt bei Menschen auf, die zwei defekte Kopien des *β-Globin*-Gens tragen. *β-Globin* ist ein Protein, das für die Bildung von Hämoglobin in Blutzellen und somit für den Sauerstofftransport im Körper erforderlich ist. Zwar kann die Krankheit mit einer Knochenmarktransplantation geheilt werden, aber eine solche erfordert immer einen passenden Stammzellenspender. Solche Spender sind selten und können nicht immer gefunden werden. Zudem ist eine Knochenmarktransplantation mit Risiken behaftet.

Der RNA-basierte, gentherapeutische neue Ansatz zur Heilung der Sichelzellanämie besteht darin, dass Patienten eigene Blutstammzellen entnommen werden, um in diesen dann mit der CRISPR/Cas-„Genschere“ die krankheitsverursachende Mutation im *β-Globin* zu korrigieren. Alternativ wird versucht, die Synthese von fötalem Hämoglobin zu reaktivieren. Bei Sichelzellanämiepatienten ist das Gen für fötales Hämoglobin in der Regel intakt. Das von diesem Gen kodierte Protein könnte also das defekte *β-Globin* ersetzen, doch die Synthese von fötalem Hämoglobin wird normalerweise früh nach der Geburt abgeschaltet. Die auf die eine oder andere Art korrigierten Stammzellen werden dann wieder in den Patienten zurücktransplantiert, um dort fortan gesunde rote Blutkörperchen zu produzieren. Erste Studien zeigen, dass dieser Ansatz die Lebensqualität schwerkranker Menschen verbessern kann.

Bei der Sichelzellanämie und anderen Blutkrankheiten können vom Patienten entnommene Zellen mit CRISPR/Cas im Labor bearbeitet werden (dieser Ansatz wird auch als *ex vivo* bezeichnet). Hingegen wird die Mehrzahl der genetischen Krankheiten eine Genomkorrek-

17 CRISPR/Cas-Editing-Tools setzen sich zusammen aus einer Variante des Cas-Enzyms, welches einen präzisen Schnitt in die zelluläre DNA einführt, und einer sogenannten „Leit-RNA“, die das Cas-Enzym mithilfe komplementärer Basenpaarung an die richtige Stelle des Genoms lenkt.

tur im Körper (*in vivo*) erfordern. Als Beispiel dafür sei hier Duchenne Muskeldystrophie (DMD) genannt, eine Erkrankung, die mit zunehmendem und irreversiblen Muskelschwund einhergeht. Diese bisher unheilbare Erkrankung manifestiert sich im Kindesalter, schreitet langsam voran und verkürzt die Lebenserwartung erheblich. DMD wird verursacht durch Mutationen im *Dystrophin*-Gen, das für ein wichtiges Gerüstprotein in Muskelzellen kodiert. Im Gegensatz zu den bisherigen Therapien, die den Krankheitsverlauf nur verzögern können, verspricht eine Korrektur des defekten *Dystrophin*-Gens über Genomeditierung, das fehlende Dystrophin-Protein dauerhaft wiederherzustellen. Es ist bekannt, dass die Wiederherstellung eines kleinen Prozentsatzes (etwa 15 %) des normalen zellulären Dystrophins bereits einen klinischen Vorteil brächte. Allerdings müssen noch einige wissenschaftliche und technische Hürden genommen werden, bevor diese Therapien die Klinik erreichen. Hierzu zählen insbesondere die erfolgreiche Einbringung der CRISPR/Cas-Genschere in das Zielgewebe, die zielgenaue und effiziente Genombearbeitung in bestimmten Zelltypen, die Kontrolle möglicher Nebeneffekte durch Veränderungen an anderen Stellen im Genom und die Immunogenität¹⁸ von bakteriellen Cas-Proteinen. Dennoch birgt die CRISPR/Cas-Technologie große Hoffnung für Patienten mit bisher unheilbaren Krankheiten.

4.5 mRNA-basierte Impfstoffe

Im Jahr 2020 katalysierte die COVID-19-Pandemie die schnellste Impfstoffentwicklung in der Geschichte. Die von den Firmen BioNTech/Pfizer und Moderna innerhalb von 10 Monaten zur Zulassung gebrachten mRNA-Impfstoffe gegen das SARS-CoV-2-Virus haben nunmehr erhebliches Interesse an der Anwendung von mRNA sowohl für prophylaktische als auch für therapeutische Zwecke geweckt. Allerdings beruht dieser enorme Erfolg auch auf jahrzehntelanger, langwieriger Grundlagenforschung, die oft mit erheblichen Widerständen zu kämpfen hatte.

Auf mRNA basierende Impfstoffe wurden vor mehr als 30 Jahren in der Hoffnung konzipiert, sichere und vielseitige Vakzine zu entwickeln, die obendrein leicht herzustellen sind. Grundsätzlich haben mRNA-Impfstoffe gegenüber herkömmlichen Impfstoffen mehrere Vorteile. Im Gegensatz zu einigen auf DNA-Viren basierenden Impfstoffen besteht kein Risiko einer Integration der mRNA in das zelluläre Genom und damit verbundener möglicher unerwünschter Spätfolgen. Außerdem können mRNA-Impfstoffe im Labor synthetisiert werden, was eine schnelle, skalierbare und kostengünstige Produktion ermöglicht. Im Prinzip kann ein einzelner mRNA-Impfstoff gleich konventionellen Totimpfstoffen auch mehrere Antigene¹⁹ kodieren, wodurch die Immunantwort gegen diverse virale Varianten mit einer einzigen Formulierung ermöglicht wird. Dies ist zum Beispiel bei den neu angepassten Impfstoffen für die Omikron-Variante von SARS-CoV-2 der Fall. Ursprünglich stieß mRNA als mögliches Therapeutikum aufgrund von Bedenken hinsichtlich ihrer geringen Stabilität im Körper bei gleichzeitiger, potenziell gefährlicher Immunstimulation auf viel Skepsis. Wir haben es den Bemühungen einiger hartnäckiger Forscher zu verdanken, dass in den letzten 10 Jahren das Interesse an klinischen Anwendungen durch ein verbessertes

¹⁸ Als Immunogenität wird die Eigenschaft eines Stoffes bezeichnet, im Körper eine Reaktion des Immunsystems auszulösen.

¹⁹ Ein Antigen ist ein artfremder Eiweißstoff, der im Körper die Bildung von Antikörpern bewirkt.

Verständnis der mRNA-Pharmakologie, der Entwicklung wirksamer RNA-Formulierungen und der Kontrolle der mRNA-Immunogenität wieder zugenommen hat.

Als therapeutisches Molekül ist mRNA sehr groß und dazu noch negativ geladen, was den Transport über Zellmembranen enorm erschwert. Darüber hinaus ist mRNA recht instabil und wird im menschlichen Gewebe schnell von Nukleasen abgebaut. Um eine zielgerichtete Zustellung und effiziente zelluläre Aufnahme von mRNA in Patienten zu ermöglichen, wurden eine Reihe innovativer mRNA-Transportvehikel entwickelt, wie zum Beispiel Lipid- oder Polymer-Nanopartikel. Einmal im Zytosol einer menschlichen Zelle angekommen, wird unmodifizierte RNA, ähnlich wie die RNA von Viren, von unserem angeborenen Immunsystem erkannt, was eine überschießende Immunantwort hervorrufen kann. Durch die Idee, modifizierte Nukleoside, insbesondere modifiziertes Uridin, in synthetische mRNAs einzubauen, wurde dieses Problem weitgehend gelöst. Außerdem zeigte derart modifizierte mRNA auch noch eine erhöhte biologische Stabilität sowie eine verbesserte Translation zu Protein (Karikó et al., 2008) – eine bahnbrechende Entwicklung.

Die mRNA-Impfstoffe gegen SARS-CoV-2 sind allerdings nicht die ersten ihrer Art. Schon vor der COVID-19-Pandemie hatte eine Reihe von vorklinischen und klinischen Studien vielversprechende Ergebnisse für mRNA-Impfstoffe gegen mehrere Erreger erbracht, darunter Zikavirus, Ebolavirus, Denguevirus, Papillomavirus und das Virus, das Tollwut auslöst. Ein weiterer wichtiger Bereich, in dem man sich von mRNA-Vakzinen erhebliche Therapiefortschritte verspricht, ist die Behandlung von Krebserkrankungen. In mehreren laufenden klinischen Studien wird versucht, bei Patienten eine Immunantwort gegen spezifische Neoantigene²⁰ ihres jeweiligen Tumors hervorzurufen. Diese werden vor der Behandlung bestimmt, um dann personalisierte mRNA-Vakzine herzustellen, die das Immunsystem des Patienten anleiten, verstärkt gegen die Tumorzellen des eigenen Körpers vorzugehen. Allerdings stellt die eingeschränkte Zugänglichkeit vieler Tumoren für Immunzellen eine größere Hürde für mRNA-Impfstoffe dar, als zum Beispiel die Bekämpfung eines Erregers in der Blutbahn und im Gewebe. Dennoch versprechen mRNA-Impfstoffe und auch mRNA-basierte Medikamente, mit denen etwa ein fehlendes, für den Stoffwechsel wichtiges Protein ersetzt werden kann, eine völlig neue Art von Medizin mit großem Potenzial.

4.6 Ausblick

Die Sicht, dass die Rolle von zellulärer RNA überwiegend darin besteht, Proteine zu kodieren, ist seit Langem überholt. Nach jahrzehntelanger Forschung auf dem Gebiet der RNA-Biologie und bahnbrechenden technologischen Entwicklungen werden RNA-Therapien nun Realität. Die innerhalb von nur 10 Monaten entwickelten und zugelassenen mRNA-Impfstoffe haben diesem Gebiet einen großen Schub verliehen. Es ist zu erwarten, dass wir in naher Zukunft neue Therapeutika aus allen Bereichen der RNA-Medizin sehen werden – das Potenzial und der mögliche Nutzen für Patienten ist zu groß, um ignoriert zu werden.

20 Neoantigene sind neu auftretende Antigene. Sie können durch DNA-Mutationen, die während der Krebsentwicklung vorkommen, hervorgerufen werden und sind daher charakteristisch für Tumorzellen.

4.7 Literaturverzeichnis

- Barrangou, R. et al. (2007):** CRISPR provides acquired resistance against viruses in prokaryotes. In: *Science* 315: 1709–1712.
- Caplen, N. J. et al. (2001):** Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. In: *Proceedings of the National Academy of Sciences of the United States of America* 98: 9742–9747.
- Cartegni, L./Kraimer, A. R. (2003):** Correction of disease-associated exon skipping by synthetic exon-specific activators. In: *Nature structural biology* 10: 120–125.
- Cong, L. et al. (2013):** Multiplex genome engineering using CRISPR/Cas systems. In: *Science* 339: 819–823.
- Deltcheva, E. et al. (2011):** CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. In: *Nature* 471: 602–607.
- Elbashir, S. M. et al. (2001):** Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. In: *Nature* 411: 494–498.
- Fire, A. et al. (1998):** Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. In: *Nature* 391: 806–811.
- Garneau, J. E. et al. (2010):** The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. In: *Nature* 468: 67–71.
- Gasiunas, G. et al. (2012):** Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. In: *Proceedings of the National Academy of Sciences of the United States of America* 109: E2579–E2586.
- Hammond, S. M. et al. (2000):** An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. In: *Nature* 404: 293–296.
- Jinek, M. et al. (2012):** A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. In: *Science* 337: 816–821.
- Jinek, M. et al. (2013):** RNA-programmed genome editing in human cells. In: *eLife* 2: e00471.
- Jirikowski, G. F. et al. (1992):** Reversal of diabetes insipidus in Brattleboro rats: intrahypothalamic injection of vasopressin mRNA. In: *Science* 255: 996–998.
- Karikó, K. et al. (2005):** Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. In: *Immunity* 23: 165–175.
- Karikó, K. et al. (2008):** Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. In: *Molecular therapy: the journal of the American Society of Gene Therapy* 16: 1833–1840.
- Mali, P. et al. (2013):** RNA-guided human genome engineering via Cas9. In: *Science* 339: 823–826.
- Mandl, C. W. et al. (1998):** In vitro-synthesized infectious RNA as an attenuated live vaccine in a flavivirus model. In: *Nature medicine* 4: 1438–1440.
- Stephenson, M. L./Zamecnik, P. C. (1978):** Inhibition of Rous sarcoma viral RNA translation by a specific oligodeoxyribonucleotide. In: *Proceedings of the National Academy of Sciences of the United States of America* 75: 285–288.
- Zamecnik, P. C./Stephenson, M. L. (1978):** Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. In: *Proceedings of the National Academy of Sciences of the United States of America* 75: 280–284.
- Zamore, P. D. et al. (2000):** RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. In: *Cell* 101: 25–33.
- Zhou, W. Z. et al. (1999):** RNA melanoma vaccine: induction of antitumor immunity by human glycoprotein 100 mRNA immunization. In: *Human gene therapy* 10: 2719–2724.

5. Ausblick: Fortschritt und Fragen

Daniela Remus

5.1 Aktueller Stand der RNA-Forschung

Im Jahr 2002 hat das wissenschaftliche Fachmagazin *Science* die RNA als Molekül des Jahres bezeichnet. Damals hatten Forschende mehrere Arbeiten veröffentlicht, die zeigten, dass jede Zelle über eine Vielzahl unterschiedlicher RNA-Arten verfügt und dass diese sehr viel mehr können, als ausschließlich die genetischen Informationen der DNA in Proteine umzuwandeln. Aber das war erst der Anfang. Seither wächst das Wissen um diese elementaren Bausteine des Lebens in hohem Tempo an. Und es wird immer deutlicher, dass die verschiedenen RNA-Arten extrem komplexe und unterschiedliche Funktionen in den Zellen erfüllen.

Wie Mathias Munschauer und Jörg Vogel in Kapitel 2 gezeigt haben, tragen manche RNAs genetische Informationen, andere interagieren mit Biomolekülen und Metaboliten. RNAs stoßen zelluläre Entwicklungen an oder beschleunigen chemische Prozesse.

Noch haben Wissenschaftlerinnen und Wissenschaftler längst nicht alle Funktionen dieser lebenswichtigen Moleküle verstanden. Von einigen RNAs ist bis jetzt nicht klar, warum sie existieren und welche physiologische Relevanz ihnen zukommt. Und es ist davon auszugehen, dass die Forschenden auch längst noch nicht alle RNAs identifizieren konnten. Dennoch herrscht Aufbruchsstimmung. Denn es ist absehbar, dass das Wissen um die RNA nicht nur weitere bedeutsame Erkenntnisse der Grundlagenforschung ermöglichen, sondern auch die Medikamentenentwicklung und die Diagnostik grundlegend verändern wird.

In Kapitel 3 haben Nina Gasparoni und Jörn Walter ausgeführt, dass und wie RNAs bereits heute als diagnostisches Werkzeug in der Forschung und der klinischen Praxis dienen. Vor allem in der Krebsdiagnostik, in der Immunologie und in der Infektionsforschung. Denn es ist möglich, mithilfe von RNA-Analysen durch spezielle Sequenzierungsverfahren, kranke von gesunden Zellen mit einem zuvor nicht erreichbaren Grad von Differenzierung zu unterscheiden. Und auch bei der Identifizierung von Pathogenen und Viren sind diese Verfahren ausgesprochen nützlich. Mit ihrer Hilfe ist es beispielsweise gelungen, das SARS-CoV-2-Virus Anfang 2020 schon kurz nach den ersten Krankheitsfällen im chinesischen Wuhan zu bestimmen. Nur durch die genaue Identifizierung und Beschreibung dieses Virus konnten Forschende in der Folgezeit so schnell Impfstoffe dagegen entwickeln.

Darüber hinaus erlaubt es die RNA-basierte Einzelzellanalyse, innerhalb eines Gewebes, beispielsweise eines Tumors, Unterschiede in der genetischen Ausstattung festzustellen. Die Forschenden erwarten deshalb, dass die RNA-Diagnostik in Zukunft standardisiert pathologische Befunde ergänzen wird, sodass schon in wenigen Jahren individuellere Behandlungen als bisher für weitaus mehr Patientinnen und Patienten möglich sein werden. Dass also irgendwann tatsächlich jeder Patient, jede Patientin eine personalisierte Therapie bekommt.

Bereits seit den 1970er-Jahren arbeiten Forschende daran, das Wissen um die Funktionen der unterschiedlichen RNAs zu nutzen, um neue Therapeutika zu entwickeln, wie Anke Sparmann und Jörg Vogel in Kapitel 4 dargestellt haben. Erste Medikamente auf RNA-Basis sind seit einigen Jahren auf dem Markt, für andere laufen klinische Studien.

Dass die RNA-Technologie das Potenzial hat, die Arzneimittelentwicklung zu revolutionieren, ist durch die Entwicklung der mRNA-Impfstoffe gegen das SARS-CoV-2-Virus weltweit bekannt.

Ein enormer Durchbruch, nicht nur für die RNA-Pioniere, die sich in der Wissenschaftsgemeinschaft lange gegen skeptische Einwände behaupten mussten, sondern auch für den gesamten Forschungsbereich. Auch die großen Pharmaunternehmen, die sich aus der RNA-Forschung zurückgezogen hatten, sind infolge des Erfolgs der mRNA-Impfstoffe in das Forschungsfeld zurückgekehrt und investieren jetzt viel Geld in diesen Bereich. Deshalb ist absehbar, dass die RNA-Technologie viele Therapeutika schon sehr bald grundlegend verändert haben wird. Denn durch sie sind sehr viel präzisere Behandlungen möglich, ob gegen Tumoren, Infektionskrankheiten oder genetisch bedingte Erkrankungen. Und da mittlerweile bekannt ist, dass die genetische Ausstattung jedes Menschen besonders ist, kann die RNA-Technologie einen entscheidenden Beitrag dazu leisten, die von Wissenschaftlerinnen und Wissenschaftlern angestrebte personalisierte Medizin schon bald in sehr viel mehr Bereichen als bisher zur Realität werden zu lassen.

5.2 Wissenschaftliche Erwartungen und gesellschaftliche Akzeptanz

Kein Zweifel: Das Potenzial dieses Forschungsfelds ist enorm. Aber die Entwicklung der mRNA-Impfstoffe in Rekordzeit hat auch gezeigt, dass diese Technologie nicht von allen Menschen begeistert und als Chance aufgenommen wird. An der Einführung der mRNA-Impfstoffe lässt sich exemplarisch beobachten, wie groß der Abstand zwischen der gegenwärtigen molekularbiologischen und biotechnologischen Forschung und dem Wissen in der Gesellschaft ist. Auf die Fragen „Was ist RNA?“ und „Welche Bedeutung kann ihr zellbiologisch zukommen?“ können nur sehr wenige Menschen außerhalb der Forschungsgemeinschaft antworten. Und das kann verunsichern – selbst diejenigen, die keinen Verschwörungsideologien anhängen –, allein schon deshalb, weil das biologische Grundwissen zur sachlichen Einordnung der neuen Technologie fehlt. Nicht selten kommen Ängste auf, ob die Forschenden auch tatsächlich alle Faktoren mitbedacht haben, sodass Menschen weder kurz- noch langfristig zu Schaden kommen. Und gerade im Fall der Covid-Impfungen zeigte sich, dass sich die Ängste vor dem bis dahin unbekanntem Impfstoff mit einem all-

gemeinen Misstrauen gegen „die“ Pharmaindustrie verbunden, der unterstellt wurde, Gewinnmaximierung über Gesundheit zu stellen.

Diese Skepsis gegenüber innovativen wissenschaftlichen Entdeckung ist weder neu noch von vornherein zu verurteilen, aber sie kann zu einem Auseinanderdriften von Wissenschaft und Gesellschaft führen. Und dann von gesellschaftlichen Gruppen missbraucht werden, denen es nicht um Wissenschaftsverständnis geht, sondern darum, ihre politische Agenda zu verfolgen und durchzusetzen, wie in der Coronapandemie deutlich zu beobachten war.

Ob genterapeutische Eingriffe mit der Genschere CRISPR/Cas, Embryonenforschung oder regenerative Medizin: Viele gentechnologische Forschungsbereiche sind heutzutage nahezu entkoppelt vom Wissen in der Gesellschaft. Das verursacht Unsicherheit, Widerstände und Sorgen – wie aktuell beispielsweise den Verdacht, mRNA-Impfstoffe könnten das Erbgut dauerhaft verändern.

Hinzu kommt, dass Aspekte der RNA-Technologie, wie bereits aus anderen Bereichen der Gentechnologie bekannt, ethische Folgefragen aufwerfen, die nicht trivial sind. Beispielsweise die Frage, was mit denjenigen Informationen geschieht, nach denen gar nicht primär gesucht wird, die aber bei den Sequenzierungen quasi als Beifang mitgeliefert werden. Zwar regelt das Gendiagnostikgesetz den Umgang mit diesen Zufallsbefunden durch das sogenannte Recht auf Nichtwissen (informationelle Selbstbestimmung), aber Kritikerinnen und Kritiker bemängeln, dass nicht alle möglichen Zufallsbefunde im Vorfeld thematisiert und somit ausgeschlossen werden können. Dass also u. U. Befunde zutage treten, von denen nicht klar ist, ob die betreffende Person diese doch wissen wollen würde, weil sie ernsthafte Konsequenzen für ihre Lebensführung zur Folge hätten. Und umgekehrt, wenn jemand über Zufallsbefunde informiert werden möchte, welche Informationen sollten dann übermittelt werden? Alle? Zumal der genetische Status eines Menschen nicht zwangsläufig Auskunft gibt über sich im Verlaufe des Lebens manifestierende Krankheiten. Und: Gebietet nicht in manchen Fällen vielleicht auch die medizinische Dringlichkeit, sich über das Recht des Individuums auf Nichtwissen hinwegzusetzen?

Außerdem: Was soll mit diesem rapide anwachsenden Wissen geschehen? Wie sicher ist die Datenaufbewahrung, kann ein Zugriff Unbefugter und ein potenzieller Missbrauch der Daten mit an Sicherheit grenzender Wahrscheinlichkeit ausgeschlossen werden?

Darüber hinaus zeigt der Umgang mit der RNA-basierten Genschere CRISPR/Cas anschaulich, wie brüchig internationale Konsense sein können und wie schnell ethische Grenzen überschritten werden. Trotz großer Einigkeit unter Forschenden, dass Keimbahneingriffe noch nicht sicher genug für die Anwendung im therapeutischen Kontext sind, dauerte es lediglich sechs Jahre von der Entdeckung der Genschere CRISPR/Cas bis zu ihrer Anwendung in der Keimbahn durch einen chinesischen Wissenschaftler und der Geburt der ersten „editierten Babys“.

Allein diese nur kurz skizzierten Beispiele machen deutlich, wie notwendig Wissensvermittlung durch die Forschenden ist. Breite gesellschaftliche Diskussionen sind erforder-

lich, um Klarheit darüber zu gewinnen, wie mit diesen neuen, potenten Werkzeugen in Zukunft umgegangen werden sollte. Einen Beitrag dazu möchte die AG *Gentechnologiebericht* mit dieser Broschüre leisten.

Autorinnen und Autoren/Authors

Dr. Nina Gasparoni

Wissenschaftliche Mitarbeiterin/Research Associate, Arbeitsgruppe für Genetik/Epigenetik, Universität des Saarlandes

Dr. Mathias Munschauer

Leiter/Head, unabhängige Helmholtz-Nachwuchsgruppe, Helmholtz Institut für RNA-basierte Infektionsforschung (HIRI), Würzburg

Dr. Daniela Remus

Freie Wissenschaftsjournalistin/Freelance Journalist

Dr. Anke Sparmann

Wissenschaftliche Lektorin/Scientific Writer, Helmholtz Institut für RNA-basierte Infektionsforschung (HIRI), Würzburg

Prof. Dr. Jörg Vogel

Direktor/Director, Helmholtz-Institut für RNA-basierte Infektionsforschung (HIRI) und Institut für Molekulare Infektionsbiology, Julius-Maximilians-Universität, Würzburg

Prof. Dr. Jörn Walter

Professor für Genetik/Professor of Genetics, Universität des Saarlandes

Mitglieder der Arbeitsgruppe *Gentechnologiebericht*/ Members of the Working Group *Gene Technology Report*

Prof. Dr. Boris Fehse

Sprecher der AG *Gentechnologiebericht*/Spokesperson of the WG *Gene Technology Report*,
Universitätsklinikum Hamburg-Eppendorf, Klinik für Stammzelltransplantation

Prof. Dr. Jörn Walter

Stellvertretender Sprecher der AG *Gentechnologiebericht*/Deputy Spokesperson of the
WG *Gene Technology Report*, Universität des Saarlandes, Institut für Biowissenschaften

Prof. Dr. Sina Bartfeld

Technische Universität Berlin, Institut für Biotechnologie

Prof. Dr. Stephan Clemens

Universität Bayreuth, Lehrstuhl Pflanzenphysiologie

Prof. Dr. Tobias J. Erb

Max-Planck-Institut für terrestrische Mikrobiologie, Marburg; Philipps-Universität
Marburg, Mikrobiologie

Prof. Dr. Dr. h. c. Heiner Fangerau

Heinrich-Heine-Universität Düsseldorf, Institut für Geschichte, Theorie und Ethik der
Medizin

Dr. Jürgen Hampel

Universität Stuttgart, Institut für Sozialwissenschaften

Prof. Dr. Martin Korte

Technische Universität Braunschweig, Zoologisches Institut, Abt. für Zelluläre Neuro-
biologie und Helmholtz-Institut für Infektionsforschung Braunschweig, AG NIND

Prof. Dr. Ralf Müller-Terpitz

Universitäten Heidelberg und Mannheim, Institut für Deutsches, Europäisches und
Internationales Medizinrecht, Gesundheitsrecht und Bioethik

Prof. Dr. Stefan Mundlos

Charité Berlin, Institut für Medizinische Genetik und Humangenetik

Prof. Dr. Jens Reich

Max-Delbrück-Centrum für Molekulare Medizin, Berlin

Prof. Dr. Silke Schicktanz

Universitätsmedizin Göttingen, Institut für Ethik und Geschichte der Medizin

Prof. Dr. Dr. Eva C. Winkler

Ruprecht Karls Universität Heidelberg, Nationales Centrum für Tumorerkrankungen,
Universitätsklinikum, Sektion Translationale Medizinethik

Prof. Dr. Martin Zenke

RWTH Aachen, Universitätsklinikum, Institut für Biomedizinische Technik

Publikationen der Arbeitsgruppe Gentechnologiebericht/ Publications of the Working Group Gene Technology Report

BÜCHER/BOOKS

- Fehse, B. et al. (Hrsg.) (2021):** Fünfter Gentechnologiebericht. Sachstand und Perspektiven für Forschung und Anwendung. Nomos, Baden-Baden. Unter: <https://www.nomos-elibrary.de/10.5771/9783748927242/fuenfter-gentechnologiebericht> [26.09.2022].
- Bartfeld, S. et al. (Hrsg.) (2020):** Organoide. Ihre Bedeutung für Forschung, Medizin und Gesellschaft. Nomos, Baden-Baden. Unter: <https://www.nomos-elibrary.de/10.5771/9783748908326/organoide> [26.09.2022].
- Hucho, F. et al. (Hrsg.) (2018):** Viertes Gentechnologiebericht. Bilanzierung einer Hochtechnologie. Nomos, Baden-Baden. Unter: <https://www.nomos-elibrary.de/10.5771/9783845293790/viertes-gentechnologiebericht> [26.09.2022].
- Zenke, M. et al. (Hrsg.) (2018):** Stammzellforschung. Aktuelle wissenschaftliche und gesellschaftliche Entwicklungen. Nomos, Baden-Baden. Unter: <https://www.nomos-elibrary.de/10.5771/9783845287720/stammzellforschung> [26.09.2022].
- Walter, J./Hümpel, A. (Hrsg.) (2017):** Epigenetik. Implikationen für die Lebens- und Geisteswissenschaften. Nomos, Baden-Baden. Unter: <https://www.nomos-elibrary.de/10.5771/9783845270838/epigenetik> [26.09.2022].
- Müller-Röber, B. et al. (Hrsg.) (2015):** Dritter Gentechnologiebericht. Analyse einer Hochtechnologie. Nomos, Baden-Baden. Unter: <https://www.nomos-elibrary.de/10.5771/9783845246956/dritter-gentechnologiebericht> [26.09.2022].
- Müller-Röber, B. et al. (Hrsg.) (2013):** Grüne Gentechnologie. Aktuelle wissenschaftliche, wirtschaftliche und gesellschaftliche Entwicklungen. 3. neubearb. u. erg. Aufl. Forum W, Limburg.
- Köchy, K./Hümpel, A. (Hrsg.) (2012):** Synthetische Biologie. Entwicklung einer neuen Ingenieurbiologie? Forum W, Dornburg.
- Fehse, B./Domasch, S. (Hrsg.) (2011):** Gentherapie in Deutschland. Eine interdisziplinäre Bestandsaufnahme. 2. akt. u. erw. Aufl. Forum W, Dornburg.
- Müller-Röber, B. et al. (Hrsg.) (2009):** Zweiter Gentechnologiebericht. Analyse einer Hochtechnologie in Deutschland. Forum W, Dornburg.
- Engelhard, M. et al. (2009):** Genetic Engineering in Livestock. Springer, Berlin, Heidelberg.
- Hucho, F. et al. (2008):** Gentherapie in Deutschland. Eine interdisziplinäre Bestandsaufnahme. Forum W, Dornburg.
- Schmidtke, J. et al. (Hrsg.) (2007):** Gendiagnostik in Deutschland. Status quo und Problemerkundung. Supplement zum Gentechnologiebericht. Forum W., Limburg.
- Müller-Röber, B. et al. (Hrsg.) (2007):** Grüne Gentechnologie. Aktuelle Entwicklungen in Wissenschaft und Wirtschaft. Spektrum, München.
- Wobus, A. M. et al. (Hrsg.) (2006):** Stammzellforschung und Zelltherapie. Stand des Wissens und der Rahmenbedingungen in Deutschland. Supplement zum Gentechnologiebericht. Spektrum, München.
- Hucho, F. et al. (Hrsg.) (2005):** Gentechnologiebericht. Analyse einer Hochtechnologie in Deutschland. Spektrum, München.
- Hucho, F./Köchy, K. (2003):** Materialien für einen Gentechnologiebericht. Grundlagenforschung, Medizinische Anwendung, ökonomische Bedeutung. Spektrum, Heidelberg.
- Köchy, K. et al. (Hrsg.) (2002):** Gentechnologie als Wirtschaftsfaktor. Spektrum, Heidelberg, Berlin.

BROSCHÜREN/BOOKLETS

- IAG Gentechnologiebericht (Hrsg.) (2021):** Fünfter Gentechnologiebericht. Sachstand und Perspektiven für Forschung und Anwendung. Kurzfassung. BBAW, Berlin. Unter: https://edoc.bbaw.de/files/3609/BBAW_Gentechnologiebericht_V_Kurzfassung.pdf [26.09.2022].
- IAG Gentechnologiebericht/German Stem Cell Network (Hrsg.) (2020):** Organoide – von der Stammzelle zur zukunftsweisenden Technologie/Organoids – from stem cells to future technologies. White Paper. Berlin. Unter: https://edoc.bbaw.de/files/3436/BBAW_Whitepaper_GSCN_11_2020.pdf [26.09.2022].
- Walter, J./Schickl, H. (Hrsg.) (2019):** Einzelzellanalyse in Forschung und Medizin. Eine Stellungnahme der interdisziplinären Arbeitsgruppe Gentechnologiebericht. BBAW, Berlin. Unter: https://edoc.bbaw.de/files/3278/BBAW_Einzelzellanalyse_Walter_Schickl.pdf [26.09.2022].
- IAG Gentechnologiebericht (Hrsg.) (2018):** Viertes Gentechnologiebericht. Bilanzierung einer Hochtechnologie. Kurzfassung. BBAW, Berlin. Unter: https://edoc.bbaw.de/files/2991/BBAW_Broschuere_Gentechnologiebericht_4_Kurzfassung.pdf [26.09.2022].

IAG Gentechnologiebericht (Hrsg.) (2018): Stammzellforschung. Aktuelle wissenschaftliche und gesellschaftliche Entwicklungen. Kurzfassung. BBAW, Berlin. Unter: https://edoc.bbaw.de/files/2990/BBAW_Broschuere_Stammzellforschung_Kurzfassung.pdf [26.09.2022].

IAG Gentechnologiebericht (Hrsg.) (2017): Epigenetik. Implikationen für die Lebens- und Geisteswissenschaften. Kurzfassung. BBAW, Berlin. Unter: https://edoc.bbaw.de/files/2885/2017_BBAW_Epigenetik_Kurzfassung.pdf [26.09.2022].

IAG Gentechnologiebericht (Hrsg.) (2015): Dritter Gentechnologiebericht. Analyse einer Hochtechnologie. Kurzfassung. BBAW, Berlin. Unter: https://edoc.bbaw.de/files/2329/BBAW_DritterGentechnologiebericht_KF_2015.pdf [26.09.2022].

Reich, J. et al. (Hrsg.) (2015): Genomchirurgie beim Menschen. Zur verantwortlichen Bewertung einer neuen Technologie. Analyse der Interdisziplinären Arbeitsgruppe Gentechnologiebericht der Berlin-Brandenburgischen Akademie der Wissenschaften. BBAW, Berlin. Unter: https://edoc.bbaw.de/files/2483/2015_Analyse_GenomchirurgieBeimMenschen.pdf [26.09.2022].

IAG Gentechnologiebericht (Hrsg.) (2013): Grüne Gentechnologie. Aktuelle wissenschaftliche, wirtschaftliche und gesellschaftliche Entwicklungen. Kurzfassung. BBAW, Berlin. Unter: https://www.gentechnologiebericht.de/fileadmin/user_upload/Webseitendateien/Publikationen/Gruene_Gentechnologie_2013.pdf [26.09.2022].

IAG Gentechnologiebericht (Hrsg.) (2012): Synthetische Biologie. Entwicklung einer neuen Ingenieurbiologie? Kurzfassung. BBAW, Berlin. Unter: https://www.gentechnologiebericht.de/fileadmin/user_upload/Webseitendateien/Publikationen/Synthetische_Biologie_2012.pdf [26.09.2022].

IAG Gentechnologiebericht (Hrsg.) (2011): Gentherapie in Deutschland. Eine interdisziplinäre Bestandsaufnahme. Kurzfassung. BBAW, Berlin. Unter: https://www.gentechnologiebericht.de/fileadmin/user_upload/Webseitendateien/Publikationen/Gentherapie_in_Deutschland_2011.pdf [26.09.2022].

IAG Gentechnologiebericht (Hrsg.) (2009): Zweiter Gentechnologiebericht. Analyse einer Hochtechnologie in Deutschland. Kurzfassung. BBAW, Berlin. Unter: https://www.gentechnologiebericht.de/fileadmin/user_upload/Webseitendateien/Publikationen/Zweiter_Gentechnologiebericht_2009.pdf [26.09.2022].

Beier, H. et al. (2009): Neue Wege der Stammzellforschung. Reprogrammierung von differenzierten Körperzellen. BBAW, Berlin. Unter: https://www.gentechnologiebericht.de/fileadmin/user_upload/Webseitendateien/Publikationen/Stammzellforschung_und_Zelltherapie_2006.pdf [26.09.2022].

IAG Gentechnologiebericht (Hrsg.) (2007): Gendiagnostik in Deutschland. Status quo und Problemerkundung. Zusammenfassung. BBAW, Berlin. Unter: https://www.gentechnologiebericht.de/fileadmin/user_upload/Webseitendateien/Publikationen/Gendiagnostik_in_Deutschland_2007.pdf [26.09.2022].

IAG Gentechnologiebericht (Hrsg.) (2005): Gentechnologiebericht. Analyse einer Hochtechnologie in Deutschland. Kurzfassung. BBAW, Berlin. Unter: https://www.gentechnologiebericht.de/fileadmin/user_upload/Webseitendateien/Publikationen/Gentechnologiebericht_2005.pdf [26.09.2022].

SONSTIGE PUBLIKATIONEN/OTHER PUBLICATIONS

Bartfeld, S. et al. (2020): Special Issue: 3D Organoids. In: Journal of Molecular Medicine 99(4). Unter: <https://link.springer.com/journal/109/volumes-and-issues/99-4> [26.09.2022].

Fehse, B. et al. (2018): Debatte 19 – Die Gentechnologie in der Gesellschaft: Von großen Versprechungen, hohen Erwartungen und Missverständnissen. Streitgespräche in den Wissenschaftlichen Sitzungen der Versammlung der Berlin-Brandenburgischen Akademie der Wissenschaften am 01. Dezember 2017. Hg. v. Grötschel, M., Berlin. Unter: https://edoc.bbaw.de/files/2926/BBAW_Debatte_19.pdf [26.09.2022].

Zenke, M. (Hrsg.) (2017): Special Issue: Stem cells. From biomedical research towards clinical applications. In: Journal of Molecular Medicine 95(7). Unter: <https://link.springer.com/journal/109/95/7/page/1> [26.09.2022].

Ropers, H. H. et al. (2013): Stellungnahme zu den neuen Sequenzierungstechniken und ihren Konsequenzen für die genetische Krankenversorgung. Hg. v. Berlin-Brandenburgische Akademie der Wissenschaften, Berlin. Unter: https://www.gentechnologiebericht.de/fileadmin/user_upload/Webseitendateien/Dokumente/Stellungnahmen-Gendiagnostik-1.pdf [26.09.2022].

IAG Gentechnologiebericht (Hrsg.) (2003): Positionen der philosophischen Ethik zur Frage des Klonens. Infoblatt. Berlin.

IAG Gentechnologiebericht (Hrsg.) (2002): Datenbanken zur Molekularbiologie und Genetik. Infoblatt. Berlin.

Contents

Foreword	62
<i>Daniela Remus</i>	
1. Introduction: The corona pandemic as a research accelerator	64
1.1 The mRNA vaccine against the coronavirus	64
1.2 Preconditions for research	64
1.3 The “famous” mRNA	65
1.4 Many different types of RNA	66
1.5 RNA technology and clinical research	67
<i>Mathias Munschauer and Jörg Vogel</i>	
2. Known types of RNA and their biological functions	69
2.1 Background	69
2.2 Messenger RNA (mRNA)	71
2.3 Ribosomal RNA (rRNA)	72
2.4 Transfer RNA (tRNA)	73
2.5 Small nucleolar RNA (snoRNA)	74
2.6 Small nuclear RNA (snRNA)	74
2.7 Ribozymes	75
2.8 MicroRNA (miRNA)	75
2.9 Small interfering RNA (siRNA)	76
2.10 Long non-coding RNA (lncRNA)	76
2.11 Circular RNA (circRNA)	78
2.12 Riboswitches	78
2.13 Bacterial small RNA (sRNA)	79
2.14 CRISPR RNAs	79
2.15 Final remarks	81
2.16 References	82

	<i>Jörn Walter and Nina Gasparoni</i>	
3.	The importance of RNA for molecular diagnostics	84
3.1	Overview	84
3.2	Spectrum of RNA analysis methods	85
3.3	Applications of RNA analytics in diagnostics	87
3.4	Processing of diagnostic RNA data	90
3.5	Challenges of RNA-based diagnostics	92
3.6	Outlook	93
3.7	References	93
	<i>Anke Sparmann and Jörg Vogel</i>	
4.	Opening up the world of RNA for therapeutics	96
4.1	Introduction	96
4.2	ASO therapeutics	98
4.3	Therapeutics based on RNA interference mechanisms	100
4.4	Therapeutic genome editing	101
4.5	mRNA-based vaccines	102
4.6	Outlook	104
4.7	References	104
	<i>Daniela Remus</i>	
5.	Outlook: Progress and questions	106
5.1	Current status of RNA research	106
5.2	Scientific expectations and social acceptance	107
	Authors	56
	Members of the Working Group Gene Technology Report	57
	Publications of the Working Group Gene Technology Report	58

Foreword

The mRNA vaccines that were developed in the shortest space of time to combat SARS-CoV-2, were on the one hand a desperately yearned for medical answer to one of the largest global pandemics to date, but on the other they also split society into two almost irreconcilable camps of pro- and anti-vaxxers. This phenomenon illustrated once again that gene technologies are among the most influential and at the same time socially controversial developments of the last few decades. For 21 years, the "Gene Technology Report – Monitoring and Interdisciplinary Dialog" has made it its task to promote a factually informed discourse on the various gene technologies in Germany to thereby bringing science, society and politics closer together.

In keeping with this objective, this brochure is devoted to the aforementioned information carrier, largely unknown among the wider public until recently but nevertheless of crucial importance – RNA, which has been the subject of fierce debate since the advent of the coronavirus. The first message to convey may already come as something of a surprise: even if mRNA vaccines are all the rage, they only represent a small part of a growing field of research that has been well ploughed for half a century, and the importance of which for medicine has only won broader attention and recognition in the last few years. The following tour d'horizon therefore attempts to reflect the entire field of RNA research, describing the current status of basic research in RNA biology as well as the ever more important role of various RNAs in medical diagnostics and in developing therapeutics for diseases. The last few years have indeed seen the discovery and examination of many different types of RNA, which take on wide-ranging functions within the cells of bacteria, plants, animals and humans. In view of the rapid growth of knowledge regarding the role played by various RNAs in a series of diseases, their potential in the development of personalized medicine, i.e. more targeted diagnostics and therapy, can scarcely be overestimated.

The Working Group *Gene Technology Report* acting as co-publisher of the brochure is composed along interdisciplinary lines of natural scientists, humanists and social scientists.¹ It works through the various applications of gene technologies with care and keeps a watchful eye on their long-term evolution. The regularly released publications on specific topics are mainly aimed at the wider public. The new "In focus" format selected for the

¹ The present brochure is not uniformly gendered, and decisions in this regard were left to the individual authors of the articles.

present publication makes it possible to react comparatively quickly to new developments. We were delighted that besides Working Group member Jörn Walter, we were able to enlist four further proven experts – Nina Gasparoni, Mathias Munschauer, Anke Sparmann and Jörg Vogel – as authors to illuminate various, in some cases extremely complex aspects of RNA biology and their practical significance in their articles. In her introduction and outlook, science journalist Daniela Remus not only helps to categorize the individual articles but also discusses questions regarding the social relevance of RNA research.

Articles attributed by name do not necessarily reflect the views of the publishers, the Working Group or the BIH – but the latter back the quality of the work performed.

I would like to extend my sincere thanks to all those who have worked on this publication, especially to the authors for their participation and to Ms. Hannah Schickl from the business office for implementing the project.

The Working Group owes a debt of gratitude to the Berlin-Brandenburg Academy of Sciences and Humanities (BBAW) for their long-standing support since its inception in 2001 and to the Friede Springer Foundation for its financial support from 2019 to 2021. Since 2022, the Working Group has continued its meta research at the Berlin Institute of Health (BIH). We would like to offer our heartfelt thanks to the BIH for continuing the project and we look forward to maintaining this collaboration.

Boris Fehse

Spokesperson for the Working Group *Gene Technology Report* at the BIH
Hamburg, October 2022

1. Introduction: The corona pandemic as a research accelerator

Daniela Remus

1.1 The mRNA vaccine against the coronavirus

On December 8, 2020, in the English city of Coventry, 90-year-old British woman Margret Keenan receives an mRNA vaccine against SARS-CoV-2, making her the first person in the world to receive this vaccine. Surrounded by international press, hovering around her with cameras and microphones, the old lady confirms how happy and grateful she is to be receiving BioNTech/Pfizer's newly developed vaccine.

The international approval and administration of this vaccine, barely a year after the start of the corona pandemic, is celebrated as a milestone in the fight against the virus. But the development of this vaccine in record time means much more than that. The vaccine is based on a novel principle: mRNA technology, until then only known to insiders. Suddenly, journalists all over the world are trying to explain this technology, friends are discussing the advantages and disadvantages of vector or mRNA vaccines, research funds are flowing and pharmaceutical companies are courting the pioneers of this research, who have previously had to fight for recognition and financial support for about three decades. In fact, the mRNA vaccines, suddenly the focus of public interest, only represent one part of RNA technology. This part being the base for research of innovative therapies, novel drugs and the decoding of emerging pathogens for years.

1.2 Preconditions for research

In the middle of the 20th century, cell biology achieved its first major breakthrough: researchers were able to prove that the genetic information of a cell, colloquially known as genetic material, is contained in the DNA, the deoxyribonucleic acid. This was the start for deciphering many molecular biological processes that had not been known until then – due to the fact that technological possibilities were limited. It was not before the second half of the 20th century that special technologies, e.g. for revealing molecular structures (electron microscopy), for multiplying genetic information (recombinant DNA), for biochemical analysis of the sequence of the individual DNA building blocks (sequencing, later also in high-throughput) combined with the necessary computer technology, made it possible to come up with these findings.

Since the early 1960s, molecular biology work has also been increasingly focused on RNA research. This abbreviation stands for ribonucleic acid. RNA has a biochemical structure similar to that of DNA and is indispensable to the functioning of a cell. Researchers soon discovered that the complete RNA of a cell consists of many different types of RNA, each of these performing different functions (see Munschauer/Vogel, Chapter 2). However, all of the different RNAs constitute elementary building blocks for each individual cell, each individual living being, be it bacterium, plant, animal or human. Without these molecules, no cellular process would take place, cells would not be able to convert their genetic information, divide or process nutrients.

The realisation that RNA is existential for all living things led physicist, biochemist and Nobel Prize winner Walter Gilbert to propose the RNA world hypothesis in 1986 – a hypothesis shared by many scientists ever since. It states that the origin of life can fundamentally be traced back to RNA, since the blueprints for the first living beings were stored in the form of RNA. In other words, their genetic material was not contained in double-stranded DNA, as in today's organisms, but in single-stranded RNA. In evolutionary terms, RNA would therefore not only be a precursor of DNA, but also the decisive condition for life as we know it on earth.

1.3 The “famous” mRNA

In the 1990s, various teams of scientists from France, Germany and the US were working with mRNA, the “messenger” RNA. This RNA serves as a messenger, to transport genetic information stored in the cell nucleus to the protein factories of the cell. To do this, it copies individual DNA sections in mirror image. The goal postulated by the researchers at the time was: to make mRNA medically useful, either as a new type of vaccine or as a drug.

The possibility to produce a synthetic mRNA was a precondition for this. However, since these mRNA molecules are very short-lived, some dissolving after only a few minutes, their production in a laboratory was proven to be extremely difficult. So, the scientists came up with the idea of using the artificial messenger RNA as a transporter, without having to rely on diversions via the cell nucleus. The mRNA would transport the information directly to the protein factories of the cells. Today, this is how the Covid vaccination works: the messenger RNA delivers the blueprint for the spike protein of the SARS-CoV-2 virus directly into the cells that produce it. Here, it is exposed to the immune cells, so that they can immediately recognise it in the event of a subsequent infection and thus fight the virus very quickly.

Katalin Karikó, who is now a researcher at BioNTech, and Steve Pascolo and Ingmar Hoerr, who founded the Tübingen-based company Curevac, are among the pioneers of this field. Independently, they all worked out the scientific basis for using messenger RNA as a molecular genetic tool. However, for years they were unable to prove beyond doubt that it actually worked. Technological and scientific difficulties and the lack of financial support meant that their research approach remained a niche topic, despite some success with mouse or rat models and a few clinical studies. Their field of research was considered too complicated and too expensive. Roche and Pfizer were the last large pharmaceutical com-

panies to withdraw from this segment in 2011. It was not until 2020, after almost 30 years of research, that scientists from the Mainz-based company BioNTech and the US company Moderna were able to show that RNA technology actually works as a therapeutic agent, and that it is a safe, effective and efficient way of developing a vaccine against SARS-CoV-2.

But this success was unplanned and, in a way, surprising for science. The mRNA pioneers had not actually been working on vaccines against infectious diseases. Rather, their goal had been to develop vaccines against cancer as therapeutic options, not as preventive vaccines, supposed to protect from a disease – like the vaccine against measles. Vaccines as therapeutic options were developed for use only after a tumour had already formed. Nowadays, researchers can distinguish between at least 300 different tumour diseases, which are colloquially summarised under the generic term “cancer”. And even cancers that until recently were grouped together, often turn out to be heterogeneous based on modern molecular analyses. RNA technology plays a significant role in research about this: the so-called single-cell sequencing helps scientists to identify molecular differences within a tumour (see in detail Walter/Gasparoni, Chapter 3). This knowledge is very important for successful therapy. Therefore, the concept of a cancer vaccination is based on the idea that the immune system must be provided with the information it needs, in order to precisely detect and eliminate any tumour cells that could have previously eluded the organism's defence cells. Transmitting this information is done by mRNA developed in a laboratory, similar to mechanism exploited by the Covid vaccine. This approach was pursued, for example, by Özlem Türeci and Ugur Sahin, the founding team of BioNTech, as well as Ingmar Hopp and Steve Pascolo, the founders of Curevac.

1.4 Many different types of RNA

Other types of RNA are far less known outside of the scientific community. But they, too, are indispensable. Some of these molecules fulfil very specific tasks: They regulate, transport, construct or catalyse cell processes, as described in detail in Chapter 2 (Munschauer/Vogel). Here, just a few examples are provided, even though researchers have discovered, analysed and understood the functions of many more different RNAs as of today.

The complete RNA of a cell includes, for example, the tRNAs, which were first discovered in the 1960s. The T stands for transport. There are 31 different variations of these alone in a human cell, which are responsible for transporting the 20 different amino acids used in our bodies for protein synthesis to the protein factories (ribosomes) and aligning them in the correct order.

However, the largest proportion of total RNA is made up of rRNAs. The R stands for ribosomal and means that these RNAs are present in the ribosomes, the so-called protein factories of the cells. They, too, were first described in the 1960s and, according to current knowledge, account up to 90 % of the total RNA amount of a cell.

By the end of the 1970s, researchers also succeeded in showing that cellular processes triggered by mRNA can be interrupted or disturbed. For instance: by using so-called oligo-

nucleotides (short RNA or DNA segments consisting of individual building blocks of nucleic acids, the nucleotides) to influence cellular processes through their complementary binding to the mRNA. At that time, a team of researchers succeeded in synthetically producing a so-called antisense oligonucleotide, ASO for short, and using it to prevent the replication, i.e. proliferation, of viruses in a human cell. This attracted a lot of attention, because this discovery showed for the first time that RNAs can have the potential to change drug development permanently (more on this in Chapter 4, Sparmann/Vogel). 20 years later, the first drug based on this principle was introduced onto the market: In 1998, Fomivirsen was approved, a drug that acts against a specific herpes virus, which can cause blindness in people with immunodeficiencies.

Since the 2000s, scientists have also been researching siRNA, "Small interfering RNA". siRNA can help to switch off specific genes – an approach used to treat and cure hereditary diseases. In 2018, the first drug based on these findings was approved. Medical professionals have since been using Patisiran as a therapeutic for polyneuropathy, a chronic nerve disease.

1.5 RNA technology and clinical research

Although neither all cellular processes nor all physiological functions of the different RNAs are understood yet, it is clear that these molecules play a key role in the recognition and treatment of a wide range of diseases. The importance of RNA technology for clinical-medical research should not be underestimated. With its help, healthy and diseased cells can be distinguished from each other, an approach that plays a major role in cancer diagnostics, immunology and infection research. In other words, it represents an additional instrument for diagnostics, as will be developed in detail in Chapter 3 (Walter/Gasparoni).

Additionally, RNA sequencing can be used not only to diagnose diseases, but also to identify pathogens such as pathogenic organisms or viruses. More so, it was only possible to analyse the SARS-CoV-2 virus through such sequencing, ultimately making it the key to the development of the Covid vaccine.

RNA technology is also the central element to many diverse therapeutic approaches, since researchers can influence and control intercellular processes based on the understanding of the different types of RNA and their functions, as explained in Chapter 4 (Sparmann/Vogel). This can help with developing vaccines, such as the one against SARS-CoV-2, or specific drugs to stop unwanted cellular processes, as one that was recently approved for spinal muscular atrophy.

RNAs are permanently forming and then degrading again. This instability is one of the central challenges for scientists trying to turn them into a drug, vaccine or diagnostic tool. Yet, this temporary activity offers a great opportunity for medical use. Synthesised RNA can trigger important processes, as can be seen in the vaccination against SARS-CoV-2, but it does not remain in the organism and therefore does not accumulate there. Although, this is what many citizens were concerned about when the new mRNA-based vaccines against

Covid were first approved. Their short-lived nature is what makes RNAs so interesting for research, because this can reduce the likelihood of unwanted long-term effects.

Many researchers describe RNA technology as a shift of paradigm in medicine – mainly due to the fact that knowledge about RNA is constantly expanding. For example, the international ENCODE project, which has been running since 2003, was able to show that a far greater number of RNAs are present in a cell than what was previously known. Though not all functions are understood yet, step by step the complexity of these molecules is becoming more visible, as Chapter 2 (Munschauer/Vogel) shows. For scientists, this continuously opens up further possibilities of using these molecules: therapeutically, e.g. as a vaccine or a drug, as described in detail in Chapter 4 (Sparmann/Vogel), but also diagnostically, in order to recognise which cellular malfunctions cause specific diseases such as tumours, as explained in Chapter 3 (Walter/Gasparoni).

2. Known types of RNA and their biological functions

Mathias Munschauer and Jörg Vogel

A very wide spectrum of RNA classes have already been identified along with their important and varying functions in biological systems and it cannot be excluded that further classes will be described in future, especially in microorganisms. The RNA sequence makes it possible to “program” specific interactions with other RNA and DNA molecules, a process of fundamental significance to the function of various non-coding RNAs and also for the design of antisense RNA and therapeutic mRNAs. Moreover, since the successful development of mRNA vaccines against COVID-19, RNA has also come under increased focus as a therapeutic molecule. It represents a new class of drugs with tremendous therapeutic potential for an extremely wide range of different diseases.

2.1 Background

Desoxyribonucleic acid, or DNA for short, serves as the carrier of genetic information in all known domains of life. The process of encoding this genetic information involves a DNA sequence, i.e. the sequence of bases in a strand of DNA, being converted into an amino acid sequence that makes up proteins. A gene (DNA segment) therefore contains the molecular blueprint for producing a specific protein, the sequence of which is determined by the order of four different DNA bases (adenine, cytosine, guanine and thymine). However, following the discovery of DNA as the carrier of genetic information, it soon became clear that proteins are not transcribed and synthesized directly from the famous DNA double helix. Instead, this stage requires the synthesis of another carrier of genetic information: ribonucleic acid, or RNA for short. A gene is therefore read from the DNA and copied into RNA. In contrast to double-stranded DNA, RNA is a single-stranded molecule. However, it is also made up of four basic building blocks: adenine, cytosine, guanine and uracil. The sequence of these RNA building blocks defines the protein blueprint. The DNA in eukaryotic systems¹ is located in the cell’s nucleus, where it is translated into messenger RNA, or mRNA for short. During RNA synthesis in the nucleus, an mRNA molecule passes through different processing stages that conclude with the finished or “mature” mRNA being transported out of the nucleus and into the cytoplasm (see Section 2.2). Once it arrives in the

¹ Eukaryotes are organisms that, unlike prokaryotes, have a nucleus in their cells. They include fungi, plants, animals and humans.

cytoplasm, the mRNA is used as the template for protein synthesis by the cell's ribosomes. The flow of genetic information from DNA to mRNA to proteins, but not from proteins back to mRNA, is known as the central dogma of molecular biology (Crick, 1970). In this central dogma, DNA serves as both the carrier of genetic information and the template for the synthesis of mRNA. Initially, the primary function of RNA appeared to be limited to that of a messenger molecule

However, already in the early days of molecular biology, it became apparent that not every gene in a genome also codes for a protein. The discovery of ribosomal RNA (rRNA) and transfer RNA (tRNA) in the 1960s heralded the dawn of research into non-coding RNAs. Both rRNA and tRNA are highly conserved, meaning that they have hardly changed through the course of evolution and play a fundamental role in protein biosynthesis in all organisms. Ribosomes are the macromolecular complexes that produce protein in cells and are made up of two-thirds rRNA and one-third proteins. The rRNA portion is of decisive importance for structural and functional reasons. An equally fundamental role is carried out by tRNA. This RNA molecule allocates each mRNA base triplet to a specific amino acid and is therefore responsible for translating the genetic code into a defined protein sequence.

The identification of an RNA-based enzyme (ribonuclease P, or RNase P for short) was a groundbreaking discovery in the field of RNA biology. Sidney Altman was able to demonstrate that the enzyme RNase P does not encode a protein and instead serves as a catalytically active non-coding RNA and catalyzes a processing step in the maturation of tRNAs (Stark et al., 1978). In 1989, Sidney Altman and Tom Cech received the Nobel Prize for the discovery of the catalytic properties of RNA (Stark et al., 1978; Cech et al., 1981), i.e. RNA's ability to accelerate chemical reactions. Until then, catalytic functions had been attributed exclusively to proteins. While DNA serves as an information carrier and proteins catalyze chemical reactions, RNA was discovered to be capable of performing both functions. This led to the development of the RNA World hypothesis by Walter Gilbert in 1986, which states that early life may have developed from RNA alone as an information carrier and catalytically active molecule, long before evolution led to the emerge of DNA for the storage of genetic information and proteins as highly specialized, potent enzymes (Gilbert, 1986).

These early studies indicated the extraordinary importance and diverse functions of RNA in biological systems and laid the foundations for further groundbreaking discoveries in the field of RNA biology. In the course of the human genome project (Lander et al., 2001), it was discovered that an unexpectedly small proportion of the human genome codes for proteins and that a staggering 98 % is non-coding. Astonishingly, however, a large proportion (around 75 %) of these non-coding regions are transcribed into RNA (Djebali et al., 2013), which suggested the existence of a large number of uncharacterized non-coding RNAs. Today, we know that the number of non-coding RNAs is not only close to the number of protein-coding mRNAs but might even exceed it (Iyer et al., 2015). Although no precise functions are currently known for the majority of these non-coding RNAs, we now have many well characterized classes of regulatory non-coding RNAs. The following is a brief summary of the currently known biological functions of important types of RNA (see also Table 1).

2.2 Messenger RNA (mRNA)

The most well-known type of RNA is undoubtedly mRNA. The life cycle of an mRNA starts with the gene being read by an enzyme – the RNA polymerase. The enzyme RNA polymerase II is responsible for the synthesis of protein-coding mRNAs in eukaryotes. The mechanism of RNA synthesis is known as transcription and can be divided into three stages: initiation, elongation and termination. Each stage is precisely regulated, which makes transcription an exceptionally complex process (Cramer, 2019). First, a primary transcript (pre-mRNA) is produced. Astonishingly, at the DNA level, a eukaryotic gene does not exist in form of a continuous protein-coding sequence. Instead, coding regions called exons are interrupted by a variable number of non-coding sequences called introns. In order to produce a functioning mRNA that ribosomes can translate into an amino acid sequence, these introns must be removed from the pre-mRNA. The process responsible for this is known as mRNA splicing and was first described by Phillip Sharp in the 1970s (Berget et al., 1977). mRNA splicing is carried out by a macromolecular complex called the spliceosome. This is a highly dynamic complex made up of a number of proteins and five highly conserved, non-coding RNAs known as small nuclear RNAs (Wahl et al., 2009). These RNA components are once again of particular significance for the catalytic function of this macromolecular complex known as the spliceosome.

The discovery of mRNA splicing ultimately revealed an entirely new level of gene regulation: alternative splicing. Alternative splicing can control the removal or inclusion of individual exons into the same pre-mRNA, meaning that alternatively spliced mRNAs can produce different protein products (known as splice isoforms). This increases the diversity of the proteome² and can be enormously important for regulatory processes. The precise molecular characterization of splicing processes also facilitates the development of nucleic acid-based molecules that modulate the binding of the spliceosome at specific splice sites and thereby influence which protein product is formed (Cartegni/Krainer, 2003). This principle has been successfully applied as a treatment for spinal muscular atrophy, as described in Chapter 4 (Sparmann/Vogel).

Splicing takes place co-transcriptionally, i.e. during transcription, with further RNA maturation processes also taking place during termination of mRNA synthesis. When pre-mRNA synthesis reaches a sequence signal, the newly synthesized transcript is cleaved at the 3' end³ and polyadenylated, which involves the addition of a tail of 20–100 adenosine nucleotides⁴ (Moore, 2005). This poly(A) tail protects the transcript against degradation and simultaneously stimulates mRNA translation. Interestingly, this stimulation of translational activity appears to be based on an interaction between the poly(A) tail and a further functional element at the other end of the mRNA molecule: the 5' cap. During transcription, an enzymatic process with numerous steps is responsible for adding a protective cap struc-

² Proteome is the term used to describe the entire set of all proteins in an organism or cell.

³ The names given to the ends of an RNA molecule are based on their chemical structure. RNA is composed of four different nucleotides. The backbone of RNA is made of ribose, a sugar with five carbon atoms in numbered order. At the 5' end of an RNA, there is a phosphate group attached to the fifth carbon atom in the ribose. By contrast, at the 3' end of an RNA, there is a hydroxyl group attached to the third carbon atom in the ribose.

⁴ Nucleotides are the smallest building blocks in nucleic acids (DNA and RNA).

ture, often made up of a modified guanine nucleotide, to the 5' end of the mRNA (Moore, 2005). Like the poly(A) tail, the 5' cap protects the RNA against degradation, stimulates translation, and is also important for the export of mature mRNA out of the nucleus and into the cytoplasm (Moore, 2005). Following the export of mRNA, which is regulated by different RNA binding proteins and takes place via small openings in the cell membrane called the nuclear pore complex, the cytoplasmic life cycle of mature mRNA begins.

The translation of mRNA into protein by ribosomes is obviously of vital importance for gene expression. Like transcription, translation is divided into initiation, elongation and termination. The initiation step in particular is actively regulated by a number of RNA binding proteins and small, non-coding RNAs such as microRNAs (Jackson et al., 2010). Translation begins at the 5' end of an mRNA and is based on the association of a series of eukaryotic initiation factors that support the binding of the small ribosomal subunit⁵ followed by the binding of the large ribosomal subunit, which ultimately results in the formation of a translation-competent ribosome (Jackson et al., 2010). The encoded protein is read starting from the start codon⁶ (RNA sequence: AUG), which requires a tRNA with the amino acid methionine (see Section 2.4). Translation is terminated at a similarly characteristic base triplet, the stop codon.⁷ However, an mRNA sequence does not start with a start codon and end with a stop codon. Instead, each cellular mRNA contain untranslated regions (UTRs) at the 5' and 3' end. UTRs contain important sequence regions for regulation, such as binding sites for RNA binding proteins or microRNAs, and serve to precisely regulate the stability and translational activity for each mRNA, which fine-tune protein production. In addition to a series of quality control mechanisms that identify and specifically degrade defective mRNAs, the poly(A) tail is inevitably shortened in the cytoplasm, which results in destabilization of the mRNA. The life cycle of an mRNA ends with its decay.

As the molecular blueprint of an mRNA always follows the same pattern (5' cap, 5' UTR, protein-coding sequence incl. start and stop codons, 3' UTR, poly(A) tail), it is relatively straightforward to produce a synthetic mRNA molecules. The combination of the required molecular building blocks forms the blueprint for a protein that, when introduced into a cell, is interpreted and translated into a protein product. The target cell undertakes production of the desired protein when presented with the corresponding blueprint. This principle underpins the numerous potential applications of mRNA therapy (see Sparmann/Vogel, Chapter 4).

2.3 Ribosomal RNA (rRNA)

Ribosomal RNA was discovered along with tRNA in the 1960s. Ribosomal RNA can account for up to 90 % of the total RNA in a cell, making it the most abundant RNA type. It is, therefore, hardly surprising that many gene copies are required to produce such large quantities

⁵ A ribosome is composed of two subunits: the large subunit and the small subunit. Both subunits are made up of a number of proteins and rRNA. Only when both subunits assemble on an mRNA can translation – i.e. the reading of mRNA and synthesis of a protein – take place.

⁶ A codon is an RNA base triplet that specifies an amino acid. This involves base-pair interactions between an anticodon in a tRNA and the corresponding codon in mRNA. Specialized codons that mark the start and end of a protein-coding sequence are called start codons and stop codons. Start codons trigger the integration of the first amino acid, methionine, while no amino acids are added after stop codons, and protein synthesis is terminated.

⁷ There are three different stop codons with RNA sequences UGA, UAA and UAG.

of rRNA. Genes that code for rRNA are also called rDNA and appear on numerous chromosomes in the human genome in tandem repeats. Not all regions of rDNA are actively transcribed. The active regions are concentrated in a region of the nucleus known as the nucleolus. Similar to mRNA synthesis, a specific enzyme – RNA polymerase I – is required to read rDNA. The first step is generation of a long pre-rRNA transcript called the 45S rRNA. In various subsequent processing stages, this pre-rRNA is used to produce three smaller rRNAs: the 28S, 18S and 5.8S rRNAs. Another class of non-coding RNAs, the small nucleolar RNAs (snoRNAs), are responsible for the processing and modification of these smaller rRNAs (see Section 2.5). Another rRNA, the 5S rRNA, is synthesized by a different enzyme, namely RNA polymerase III, at a different location in the genome. Equal proportions of all four rRNAs are required for the assembly of ribosomes. For this purpose, the ribosomal proteins synthesized in the cytoplasm are transported into the nucleolus, where they assemble with the various ribosomal RNAs to form the large and small ribosomal subunits before they are finally transported back into the cytoplasm.

As early as the 1990s, there was growing evidence that intact rRNA is particularly important for the catalytic activity of the ribosome (Noller et al., 1992). Thanks to high-resolution structural data (Yan et al., 2015), we now know that the catalytic center of the ribosome is dominated by RNA-RNA interactions, which are essential for the formation of peptide bonds connecting during protein synthesis.

Ribosomal RNA is part of the basic configuration of every living cell and was probably present in the first living organisms. As rRNA performs the same molecular function in all organisms, it is assumed that rRNA is rarely transferred from one organism to another, meaning that the sequence of an rRNA gene reflects the evolutionary history not only of that individual gene but of the entire organism. This is why analyzing rDNA sequences is an exceptionally effective way to cast light on the relationships between different organisms (Woese et al., 1990).

2.4 Transfer RNA (tRNA)

In the 1960s, Robert Holley isolated the first tRNA and not only identified its sequence but also proposed a structure, known today as the cloverleaf model, that is still in use today (Holley et al., 1965a; Holley et al., 1965b). At the same time, Marshall Nirenberg and Heinrich Matthaei discovered that the base triplet UUU codes for the amino acid phenylalanine (Nirenberg/Matthaei, 1961), which established the fundamental role of tRNAs in the translation of the genetic code. A tRNA is usually around 80 nucleotides in length and has a specific base triplet in the central loop called the anticodon. During translation, only tRNAs with an anticodon that is complementary to the corresponding base triplet of an mRNA can be incorporated at the respective site in the ribosome. As each tRNA molecule is joined with a specific amino acid at the 3' end, the tRNA assigns a specific amino acid to each mRNA codon during translation. This is how the genetic code of an mRNA sequence is translated into an amino acid sequence. In 1968, Marshall Nirenberg, Robert Holley and Har Gobind were jointly awarded the Nobel Prize for their work shedding light on this mechanism.

2.5 Small nucleolar RNA (snoRNA)

Small nucleolar RNAs are non-coding RNAs of 60 to 300 nucleotides in length, often encoded in introns of pre-mRNAs, and are among the most common non-coding RNAs in eukaryotes. The biosynthesis of intronic snoRNAs begins with mRNA splicing (as described in Section 2.2), which releases the intron in circular form (known as the lariat or lasso). This is followed by various processing steps to linearize the lariat and trim the mature snoRNA to the required length (Kufel/Grzechnik, 2019). Like most functional RNAs, snoRNAs interact with proteins and form ribonucleoprotein complexes, also known as small nucleolar ribonucleoprotein particles or snoRNPs (Matera et al., 2007). The role of the snoRNA components in such complexes is effectively to act a molecular guide, because snoRNAs are responsible for directing the complex to a target sequence in another RNA. As outlined in the introduction, snoRNAs play an important role in the modification and processing of other RNAs, especially rRNA, and are therefore present in the nucleolus, the location of rRNA synthesis and rRNA processing. There are two important classes of snoRNAs that modify rRNA and are classified based on their secondary structure and activity as C/D box snoRNAs and H/ACA box snoRNAs (Kufel/Grzechnik, 2019).

2.6 Small nuclear RNA (snRNA)

Small nuclear RNAs are short, highly conserved, non-coding RNAs of 100 to 300 nucleotides in length, and are an essential component of the spliceosome. snRNAs are present in the nucleus, where mRNA splicing takes place. RNA polymerases II and III are responsible for synthesizing snRNAs, which are produced in very high copy numbers. Due to their high uracil content, snRNAs are known as U1-snRNA, U2-snRNA, and so on. The spliceosome is composed of five different snRNAs (U1, U2, U4, U5 and U6) (Matera/Wang, 2014). Each snRNA interacts with specific proteins, in particular the Sm and LSm proteins, to form functional snRNP complexes (Wahl et al., 2009). Unlike the ribosome, the spliceosome is not a static complex; instead, it passes through different stages depending on the progress of the splicing reaction, with each stage being characterized by the precisely defined association and dissociation of protein and RNA subunits (Wahl et al., 2009). The role of snRNA is to recognize sequence elements in pre-mRNAs, such as the 5' and 3' splice sites. snRNAs are therefore responsible for correctly identifying the exon-intron junctions. The removal of non-coding intronic sequences is catalyzed by RNA-RNA interactions (Matera/Wang, 2014). As outlined previously in Section 2.2, alternative splicing is an important regulatory process that not only increases protein diversity but can also produce proteins with different functions. The targeted modulation of RNA-RNA interactions between snRNAs and a pre-mRNA makes it possible to precisely influence the outcome of a splicing reaction, which has been applied successfully in the aforementioned SMA therapy and other cases.

2.7 Ribozymes

Ribozymes are catalytically active RNA modules that catalyze chemical reactions (Doherty/Doudna, 2001). The aforementioned RNase P and the group of self-splicing introns⁸ are classified as ribozymes and both played a key role in the discovery of the catalytic activity of RNA (see Section 2.1). Among the most important ribozymes in a cell are the ribosome and the spliceosome. Both are macromolecular protein-RNA complexes with a number of protein subunits; however, the actual catalytic reaction is executed by 28S rRNA in the ribosome and by snRNAs in the spliceosome. A significant difference between ribozymes and protein-based enzymes lies in the speed at which the catalyzed reaction can occur rather than in the variety of possible reactions (Emilsson et al., 2003). The discovery of ribozymes played a crucial role in the development of the RNA World hypothesis.

2.8 MicroRNA (miRNA)

The discovery of the first microRNA in the roundworm *Caenorhabditis elegans* (*C. elegans*) in 1993 (Lee et al., 1993) and the subsequent discovery of additional highly conserved microRNAs in a number of organisms (Pasquinelli et al., 2000; Reinhart et al., 2000), was a significant moment in modern molecular biology research. In comparison to other regulatory non-coding RNAs, mature microRNAs are tiny. First, a precursor RNA of around 60 nucleotides in length called a pre-microRNA is cut out of a primary RNA polymerase II transcript that contains a characteristic hairpin structure.⁹ This first processing step takes place in the nucleus and is performed by the microprocessor complex and its Drosha subunit (Bartel, 2004). With the help of export proteins, the pre-microRNA then moves into the cytoplasm. A second processing step then takes place, in which the Dicer enzyme cleaves and releases an RNA duplex of around 21 nucleotides in length from the precursor molecule (Bartel, 2004). Due to their tiny size, mature microRNAs were simply overlooked for a long time. The mature microRNA is then incorporated into the RNA-induced silencing complex (RISC) where the RNA duplex is unwound and only a single-stranded RNA remains. This short RNA then serves as a molecular guide that directs the RISC complex to a target sequence in the 3' UTR of an mRNA, with the microRNA "showing" the protein complex its target. The microRNA target sequence is defined by a short sequence of seven to eight nucleotides at the 5' end of the microRNA. This "microRNA seed" uses base pairing to bind to the complementary sequence in the target mRNA (Bartel, 2004). In eukaryotic systems, the remainder of the microRNA usually remains unpaired, leading to the inhibition of the target mRNA at the translational level. In less common cases, when the target sequence is completely complementary to the microRNA and the entire microRNA can engage in base pairing, the mRNA is cleaved by the RISC subunit AGO2 (Meister et al., 2004), which results in degradation of the mRNA. MicroRNAs therefore operate at the post-transcriptional level¹⁰ and fine-tune gene expression outcomes. MicroRNAs play an important regulatory role in numerous cellular processes.

⁸ Some RNAs can catalyze the removal of introns independently and do not require large protein complexes like the spliceosome for this task. These RNAs are also known as autocatalytic or self-splicing introns and are divided into group I and group II introns.

⁹ Hairpin structures commonly appear in RNA when double-stranded structures form in short sections within a double strand.

¹⁰ The post-transcriptional level of gene regulation includes all regulatory mechanisms that occur following RNA synthesis (transcription).

A single microRNA can precisely regulate hundreds of mRNAs. More than 1,000 microRNAs have already been discovered in the human genome. Many of them display specific expression patterns and are deregulated in various diseases. This makes microRNAs an interesting target structures for therapeutics, for instance in cancer treatment.

2.9 Small interfering RNA (siRNA)

Small interfering RNAs have many similarities with microRNAs and were discovered at around the same time. siRNAs are small RNAs of about 21 nucleotides in length, that are cleaved from long double-stranded RNAs by the Dicer enzyme (McManus/Sharp, 2002). In contrast the endogenous microRNAs – i.e. microRNAs encoded in the genome – siRNAs do not undergo processing in the nucleus and their biosynthesis is limited to a single cytoplasmic processing step. The siRNAs are also incorporated into the RISC complex and guide it to a target sequence defined by the siRNA (McManus/Sharp, 2002). A decisive distinction, however, is that siRNAs are usually fully complementary to the target sequence and can therefore trigger cleavage of the target RNA, which leads to its degradation. This mechanism is also known as RNA interference (RNAi) and, in addition to regulating the cell's own mRNAs, can degrade foreign RNA, e.g. RNA of a virus or other form of pathogen.

In 2001, Thomas Tuschl showed that the RNA interference mechanism also functions in mammalian cells (Elbashir et al., 2001), which opened up unforeseen possibilities. It became possible, for the first time, to use the target mRNA sequence to produce a synthetic RNA molecule of 21 nucleotides in length in order to downregulate the target mRNA. In effect, a universal switch had been found to deactivate any given mRNA. This was a revolution in molecular biology, enabling researchers to systematically examine the function of any gene of their choosing (McManus/Sharp, 2002). Craig Mello and Andrew Fire were awarded the Nobel Prize in 2006 in recognition of their discovery of RNA interference (Fire et al., 1998). This discovery was also a turning point in the development of RNA-based therapeutics. Indeed, we have to acknowledge that the research into the use of siRNAs as therapeutics has brought upon many of the technologies used in the mRNA-based COVID-19-vaccines (see also Sparmann/Vogel, Chapter 4).

2.10 Long non-coding RNA (lncRNA)

The classes of coding and non-coding RNAs described thus far all share a characteristic molecular structure and function according to principles shared by all members within each RNA class. This is not the case with long non-coding RNAs. Unlike short non-coding RNAs, such as microRNAs or siRNAs, lncRNAs are simply defined by a length of more than 200 nucleotides. Similar to protein-coding mRNAs, lncRNAs are transcribed in the nucleus by the RNA polymerase II enzyme, given a 5' cap structure, co-transcriptionally spliced and, in most cases, are also polyadenylated. From a biochemical perspective, an lncRNA is not radically different from an mRNA. The fundamental difference, however, is that an lncRNA does not contain an open reading frame that can be read by ribosomes and translated into a protein. This immediately raises the question of what exactly the molecular function of an lncRNA could be. Compared to mRNAs, lncRNAs tend to have fewer exons, are around

ten times less abundantly transcribed and are also less conserved than mRNAs. As a group, lncRNAs appear to be more conserved than introns but less conserved than mRNA exons (Ulitsky/Bartel, 2013). Nevertheless, lncRNAs are characterized by cell type-specific expression and regulation patterns. In addition, individual highly conserved and robustly expressed lncRNAs can be found in the human genome, which may indicate functional relevance of these lncRNA. Thanks to state-of-the-art sequencing technologies, in particular next-generation sequencing, thousands of lncRNAs have already been discovered and cataloged in the human genome. At present, no physiological functions have been identified for the majority of these transcripts. The question is, therefore, whether many of the lncRNAs discovered to date might simply emerge from a background activity of the transcription machinery, frequently described as transcriptional noise. However, there are well described examples of functionally relevant lncRNAs with fascinating molecular mechanisms.

The best understood lncRNA is known by the name XIST. In the early 1990s, researchers were able to demonstrate that the *XIST* gene on the human X chromosome¹¹ produces a long RNA that does not contain a protein-coding sequence and, unlike mRNAs, is not found in the cytoplasm but rather in the nucleus (Brockdorff et al., 1992). Interestingly, XIST is only read from one of the two X chromosomes in female cells. This turned out to be the exact chromosome that is inactivated during female embryonic development (Augui et al., 2011). This process of X inactivation occurs exclusively in female cells and is known as dosage compensation. This is a strategy used by many organisms to ensure that important genes on the X chromosome occur in equal quantities in male and female cells. As male cells have one X chromosome and one Y chromosome, while female cells have two X chromosomes, one of the two X chromosomes is inactivated during female embryonic development (Augui et al., 2011) to achieve this. It was discovered that the XIST lncRNA is responsible for inactivating one of the X chromosomes (Penny et al., 1996) and it uses a fascinating mechanism to do so: the XIST RNA assembles different proteins to form a protein complex that acts as an inhibitor capable of locally silencing gene transcription (Chu et al., 2015; McHugh et al., 2015). XIST then distributes this repressive complex over the entire X chromosome, which stops transcriptional activity and ultimately leads to almost complete inactivation of the targeted X chromosome.

Based on the molecular principles used by XIST, various mechanisms have been put forward to explain the function of lncRNAs. The most commonly described mechanisms include the role of a molecular “scaffold” of protein complexes or the sequence-depending recruitment of protein complexes to specific regions of DNA or RNA. Another common suggestion is that lncRNAs may modulate the activity of certain proteins or serve as a type of molecular “decoy”. The regulatory activity of an lncRNA could also be due to the fact that the lncRNA binds a specific protein or a regulatory RNA – such as a specific microRNA – particularly strongly and thereby keeps it away from its actual target sequence. Which mechanism is responsible for the regulatory activity of a specific lncRNA differs from

¹¹ Each human cell contains a copy of the entire genome, distributed across 23 pairs of chromosomes (46 chromosomes in total). In male and female cells, 22 of these pairs of chromosomes are identical. These are called the autosome pairs. Biological sex is determined by the two remaining chromosomes – the sex chromosomes. Women have two X chromosomes, while men have only one X chromosome and a significantly shorter Y chromosome.

case to case and requires a precise molecular characterization, a process that often involves many years of research. In many diseases, lncRNAs display characteristic patterns of expression and regulation. There is therefore a steady interest in lncRNAs as therapeutic target structures for medically relevant indications or as markers for pathological changes in certain tissues.

2.11 Circular RNA (circRNA)

Circular RNAs, or circRNAs for short (Hansen et al., 2014; Memczak et al., 2014) are a relatively newly discovered class of RNAs. As the name suggests, circRNAs are characterized by a closed ring structure, so they do not have defined 5' and 3' ends like ordinary linear transcripts. The biosynthesis of these circular RNAs involves a relatively rare splicing phenomenon called "backsplicing". In this process, the 5' end of an intron (the splice donor site) is joined with the 3' end of a preceding intron (the splice acceptor site), which creates a closed RNA ring. circRNAs are generally non-coding, though a small number of studies have identified translation of certain circular RNAs. However, whether this translation produces stable or, more importantly, functionally relevant protein products has not yet been conclusively determined. At present, little is known about the function of circular RNAs. Nevertheless, their unique molecular structure presumably gives these RNA rings an unusually high level of stability against enzymes that degrade RNA from the 5' and 3' ends. This stability could also be significant for the function of circRNAs. The best understood example of a circular RNA with a biological function is an RNA ring called CDR1as. The discovery of this RNA was based on its unusually strong binding by a component in the microRNA RISC complex (Memczak et al., 2014) (see Section 2.8). In addition to this strong biochemical interaction, CDR1as has over 60 evolutionarily conserved binding sites corresponding to a specific microRNA (miR-7). Studies in cells and animal models have shown that CDR1as can bind to a large number of miR-7 molecules, which reduces the number of functionally available miR-7 molecules and thereby inhibit miR-7. The function of this circular RNA appears to be that of a molecular "decoy", that binds numerous copies of a regulator molecule and keeping them away from other targets. Only a few physiologically relevant circRNAs have been identified to date. However, proposed mechanisms range from binding proteins to assembling protein complexes to transporting interacting proteins or RNAs. Due to their enhanced stability, circRNAs may also be suitable for use as biomarkers.

2.12 Riboswitches

In contrast to the RNA classes described so far, a riboswitch is not a type of RNA in its own right but rather a regulatory segment of RNA of 30 to 200 nucleotides in length that appears in untranslated regions of predominantly bacterial mRNAs. The outstanding quality of this RNA element is its ability to bind metabolites¹² without the assistance of a protein. In most cases, this binding has a regulatory influence on the translational activity or the transcription rates of the mRNA that contains the riboswitch (Mironov et al., 2002; Winkler et al., 2002). This effect can be mechanistically explained by a change in RNA structure

¹² Metabolites are chemical compounds in a cell that are formed as intermediary products in biological metabolism.

after binding the metabolite. A change in the RNA structure can obscure, for example, the ribosome binding site, which inhibits the start of translation (Winkler/Breaker, 2003). Alternatively, metabolite binding can lead to the formation of a hairpin structure, which can in turn induce the termination of transcription.

2.13 Bacterial small RNA (sRNA)

Regulatory RNAs not only appear in eukaryotes but also in bacteria. One particularly important class of bacterial RNAs are small RNAs, or sRNAs for short. These are non-coding RNAs of between 50 and 250 nucleotides in length. Although the existence of these small RNAs in bacteria was first identified in the 1970s, it was only the systematic sequencing and analysis of bacterial genomes that facilitated the discovery of large numbers of sRNAs (Argaman et al., 2001; Wassarman et al., 2001; Vogel et al., 2003). Unlike the small RNAs in eukaryotes (in particular miRNAs and siRNAs), this group of sRNAs is very diverse in terms of their length and secondary structure as well as their regulatory functions and mechanisms. sRNAs can be synthesized as independent transcription units or cleaved from 3' UTRs of mRNAs. Many sRNAs contain hairpin structures and can interact with proteins or mRNAs to control bacterial gene expression. Mechanistically, many sRNAs function by using sequence-specific attachment to regulatory regions in bacterial mRNAs. The ribosome binding site is particularly important in this context and can be blocked, for example, by an sRNA, which inhibits effective translation. Alternatively, binding of an sRNA can also induce a structural change in the target mRNA, which can expose the ribosome binding site in the first place. This mechanism can achieve more efficient translation of the target mRNA. sRNAs can also influence the transcription and stability of their target mRNAs. The sRNA sequence can be used to determine complementary target sequences in mRNAs, which makes it possible to predict regulatory targets. This is often complicated, however, by the fact that the interaction of sRNAs is mediated via relatively short sequence regions with only incomplete base attachment to the target mRNA. A detailed molecular characterization of individual sRNAs and their targets is therefore required. In gram-negative¹³ bacteria, sRNAs often work with special helper proteins (known as RNA chaperones) such as Hfq (Vogel/Luisi, 2011). RNA chaperones influence the stability of sRNA and mediate the attachment of the sRNA to a target sequence. Like riboswitches (see Section 2.12), sRNAs can react to metabolites or signals from their environment and induce corresponding gene regulation changes. This makes sRNAs important regulators in various bacterial signaling pathways such as iron homeostasis, chemical communication using signaling molecules (known as "quorum sensing"), biofilm formation, and the control of virulence programs in pathogenic bacteria.

2.14 CRISPR RNAs

Bacteria have an antiviral adaptive immune system known as "clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas)", often referred to simply as CRISPR. When it comes to the function of Cas9-based immune system, which is particularly

¹³ The Gram staining method (named after Hans Christian Gram) allows scientists to classify bacteria into two large groups: gram-positive bacteria and gram-negative bacteria. The different staining of bacteria is based on the structure of the cell wall, which represents an important systematic distinguishing feature.

important for genome editing, there are two non-coding RNAs of central importance: the CRISPR RNA (crRNA) and the trans-activating crRNA (tracrRNA). The crRNA is initially read as long pre-crRNA. Then, in a series of processing steps, individual crRNAs are cleaved off (Deltcheva et al., 2012). These crRNAs have a relatively short region of 17 to 20 bases that can attach to foreign DNA, such as an infiltrating virus. These sequence sections, also known as “spacers”, alternate with conserved repeat units of similar lengths that are shared among all crRNAs; the tracrRNA attaches to these repeat units. The crRNA therefore serves as a molecular bridge between the DNA target sequence and the tracrRNA. The RNA duplex made up of crRNA and tracrRNA is bound by CRISPR-associated proteins (Cas proteins), such as Cas9. The crRNA and tracrRNA duplex guides the Cas9 protein to the viral DNA target sequence, where the Cas9 protein cleaves the target sequence and ultimately facilitates the degradation of foreign DNA. After clearance of the infection, the bacterium integrates small fragments of the invader’s DNA into the relevant CRISPR locus¹⁴ of its own genome in order to generate suitable crRNAs in the event of future infections, which will enable cleavage of the invader’s DNA. As the target sequence of a Cas proteins can be programmed relatively easily via the crRNA sequence, a number of CRISPR/Cas-based technologies have been developed in a very short period of time, enabling targeted changes the DNA sequence of many organisms. The ease of programming CRISPR-based gene editing technologies is also the basis for applying the various CRISPR/Cas systems in the fields of biomedical research and gene therapy.

Table 1: Overview of known RNA types and functions

RNA type	Size	Domains of life	Function
mRNA	In humans: 2–3 kilobases (kb) on average	All	Messenger molecule that contains the genetic blueprint for the production of proteins by the ribosome.
rRNA	In humans: 18S: 1.9 kb 28S: 5 kb	All	Component of the ribosome; performs structural and functional tasks.
tRNA	70–95 bases	All	Translation of the genetic code by allocation of an amino acid to an mRNA base triplet.
snoRNA	60–300 bases	Eukaryotes, archaea	Modification and processing of rRNA.
snRNA	100–300 bases	Eukaryotes	Component of the spliceosome. Removal of introns from pre-mRNA.
miRNA	21 bases	Eukaryotes	Regulation of mRNA stability and mRNA translation at the post-transcriptional level.

¹⁴ CRISPR locus: A section in bacterial genomes that, in addition to the genes for Cas proteins, also contains and archives the remains of viral infections.

siRNA	21 bases	Eukaryotes	RNA regulation in the context of RNA interference.
lncRNA	>200 bases	Eukaryotes, bacteria	Widely unknown. Various gene-regulating mechanisms are possible.
circRNA	100–4,000 bases	Eukaryotes	Widely unknown. Various gene-regulating mechanisms are possible, e.g. binding miRNAs.
sRNA	50–250 bases	Bacteria, archaea	Regulation of mRNA stability, translation and transcription in bacteria.
Ribozyme	Variable	Eukaryotes, bacteria, archaea	RNA molecule with catalytic properties.
Riboswitch	30–200 bases	Bacteria, archaea, only rarely in eukaryotes	Binding of metabolites, control of mRNA transcription and translation.
crRNA	40–50 bases	Bacteria, archaea	Guiding Cas proteins to the DNA target sequence. Part of the CRISPR immune system in bacteria.
tracrRNA	50–150 bases	Bacteria, archaea	Guiding Cas proteins to the DNA target sequence. Part of the CRISPR immune system in bacteria.

2.15 Final remarks

In summary, it is clear that there is a wide spectrum of known RNA classes and functions. We also cannot exclude the possibility that further classes will be discovered in the future, especially in microorganisms. In addition to their function as a carrier of genetic material, RNA can interact with a number of biomolecules (such as DNA, RNA, and proteins) as well as metabolites. It is this property that makes RNA so exceptionally well suited to mediate interactions between different biomolecules and thereby form regulatory complexes. The fact that the RNA sequence can be used to “program” specific interactions with other RNA or DNA molecules is particularly significant. This principle is of fundamental importance for the function of various non-coding RNAs as well as for the design of antisense RNAs or therapeutic mRNAs. Another outstanding property of RNA is the catalysis of chemical reaction and its function as sensors for metabolites and other physical parameters, e.g. temperature. In addition to eukaryotic and prokaryotic organisms, pathogens such as RNA phages and RNA viruses also use RNA as the carrier of genetic information and as a template for protein biosynthesis. Moreover, since the successful development of mRNA vaccines against COVID-19, RNA has also come into focus as a therapeutic molecule. It represents a new molecule class with tremendous potential for the treatment of a wide range of diseases.

2.16 References

- Argaman, L. et al. (2001):** Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. In: *Curr Biol* 11: 941–950. DOI: 10.1016/s0960-9822(01)00270-6.
- Augui, S. et al. (2011):** Regulation of X-chromosome inactivation by the X-inactivation centre. In: *Nat Rev Genet* 12: 429–442. DOI: 10.1038/nrg2987.
- Bartel, D. P. (2004):** MicroRNAs genomics, biogenesis, mechanism, and function. In: *Cell* 116: 281–297. DOI: 10.1016/s0092-8674(04)00045-5.
- Berget, S. M. et al. (1977):** Spliced segments at the 5' terminus of adenovirus 2 late mRNA. In: *Proc National Acad Sci* 74: 3171–3175. DOI: 10.1073/pnas.74.8.3171.
- Brockdorff, N. et al. (1992):** The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. In: *Cell* 71: 515–526. DOI: 10.1016/0092-8674(92)90519-i.
- Cartegni, L./Kraimer, A. R. (2003):** Correction of disease-associated exon skipping by synthetic exon-specific activators. In: *Nature Structural Biology* 10: 120–125. DOI: 10.1038/nsb887.
- Cech, T. R. et al. (1981):** In vitro splicing of the ribosomal RNA precursor of tetrahymena: Involvement of a guanosine nucleotide in the excision of the intervening sequence. In: *Cell* 27: 487–496. DOI: 10.1016/0092-8674(81)90390-1.
- Chu, C. et al. (2015):** Systematic discovery of *Xist* RNA binding proteins. In: *Cell* 161: 404–416. DOI: 10.1016/j.cell.2015.03.025.
- Cramer, P. (2019):** Organization and regulation of gene transcription. In: *Nature* 573: 45–54. DOI: 10.1038/s41586-019-1517-4.
- Crick, F. (1970):** Central dogma of molecular biology. In: *Nature* 227: 561–563. DOI: 10.1038/227561a0.
- Deltcheva, E. et al. (2011):** CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. In: *Nature* 471(7340): 602–607. DOI: 10.1038/nature09886.
- Djebali, S. et al. (2013):** Landscape of transcription in human cells. In: *Nature* 488: 101–108. DOI: 10.1038/nature11233.
- Doherty, E. A./Doudna, J. A. (2001):** Ribozyme structures and mechanisms. In: *Annu Rev Biophys Biom* 30: 457–475. DOI: 10.1146/annurev.biophys.30.1.457.
- Elbashir, S. M. et al. (2001):** Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. In: *Nature* 411: 494–498. DOI: 10.1038/35078107.
- Emilsson, G. M. et al. (2003):** Ribozyme speed limits. In: *Rna* 9: 907–918. DOI: 10.1261/rna.5680603.
- Fire, A. et al. (1998):** Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. In: *Nature* 391: 806–811. DOI: 10.1038/35888.
- Gilbert, W. (1986):** Origin of life: The RNA world. In: *Nature* 319: 618–618. DOI: 10.1038/319618a0.
- Hansen, T. B. et al. (2014):** Natural RNA circles function as efficient microRNA sponges. In: *Nature* 495: 384–388. DOI: 10.1038/nature11993.
- Holley, R. W. et al. (1965a):** Structure of a ribonucleic acid. In: *Science* 147: 1462–1465. DOI: 10.1126/science.147.3664.1462.
- Holley, R. W. et al. (1965b):** Nucleotide sequences in the yeast alanine transfer ribonucleic acid. In: *J Biological Chem* 240: 2122–2128.
- Iyer, M. K. et al. (2015):** The landscape of long noncoding RNAs in the human transcriptome. In: *Nature Genetics* 47: 199–208. DOI: 10.1038/ng.3192.
- Jackson, R. J. et al. (2010):** The mechanism of eukaryotic translation initiation and principles of its regulation. In: *Nature Reviews Molecular Cell Biology* 11: 113–127. DOI: 10.1038/nrm2838.
- Kufel, J./Grzechnik, P. (2019):** Small nucleolar RNAs tell a different tale. In: *Trends Genet* 35: 104–117. DOI: 10.1016/j.tig.2018.11.005.
- Lander, E. S. et al. (2001):** Initial sequencing and analysis of the human genome. In: *Nature* 409: 860–921. DOI: 10.1038/35057062.
- Lee, R. C. et al. (1993):** The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. In: *Cell* 75: 843–854. DOI: 10.1016/0092-8674(93)90529-y.
- Matera, A. G./Wang, Z. (2014):** A day in the life of the spliceosome. In: *Nat Rev Mol Cell Bio* 15: 108–121. DOI: 10.1038/nrm3742.
- Matera, A. G. et al. (2007):** Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. In: *Nat Rev Mol Cell Bio* 8: 209–220. DOI: 10.1038/nrm2124.

- McHugh, C. A. et al. (2015):** The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. In: *Nature* 521: 232–236. DOI: 10.1038/nature14443.
- McManus, M. T./Sharp, P. A. (2002):** Gene silencing in mammals by small interfering RNAs. In: *Nat Rev Genet* 3: 737–747. DOI: 10.1038/nrg908.
- Meister, G. et al. (2004):** Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. In: *Molecular Cell* 15: 185–197. DOI: 10.1016/j.molcel.2004.07.007.
- Memczak, S. et al. (2014):** Circular RNAs are a large class of animal RNAs with regulatory potency. In: *Nature* 495: 333–338. DOI: 10.1038/nature11928.
- Mironov, A. S. et al. (2002):** Sensing small molecules by nascent RNA: A mechanism to control transcription in bacteria. In: *Cell* 111: 747–756. DOI: 10.1016/s0092-8674(02)01134-0.
- Moore, M. J. (2005):** From birth to death: the complex lives of eukaryotic mRNAs. In: *Science* 309: 1514–1518. DOI: 10.1126/science.1111443.
- Nirenberg, M. W./Matthaei, J. H. (1961):** The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. In: *Proc National Acad Sci* 47: 1588–1602. DOI: 10.1073/pnas.47.10.1588.
- Noller, H. F. et al. (1992):** Unusual resistance of Peptidyl transferase to protein extraction procedures. In: *Science* 256: 1416–1419. DOI: 10.1126/science.1604315.
- Pasquinelli, A. E. et al. (2000):** Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. In: *Nature* 408: 86–89. DOI: 10.1038/35040556.
- Penny, G. D. et al. (1996):** Requirement for Xist in X chromosome inactivation. In: *Nature* 379: 131–137. DOI: 10.1038/379131a0.
- Reinhart, B. J. et al. (2000):** The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. In: *Nature* 403: 901–906. DOI: 10.1038/35002607.
- Stark, B. C. et al. (1978):** Ribonuclease P: an enzyme with an essential RNA component. In: *Proc National Acad Sci* 75: 3717–3721. DOI: 10.1073/pnas.75.8.3717.
- Ulitsky, I./Bartel, D. P. (2013):** lincRNAs: genomics, evolution, and mechanisms. In: *Cell* 154: 26–46. DOI: 10.1016/j.cell.2013.06.020.
- Vogel, J./Luisi, B. F. (2011):** Hfq and its constellation of RNA. In: *Nat Rev Microbiol* 9: 578–589. DOI: 10.1038/nrmicro2615.
- Vogel, J. et al. (2003):** RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. In: *Nucleic Acids Res* 31: 6435–6443. DOI: 10.1093/nar/gkg867.
- Wahl, M. C. et al. (2009):** The Spliceosome: Design principles of a dynamic RNP machine. In: *Cell* 136: 701–718. DOI: 10.1016/j.cell.2009.02.009.
- Wassarman, K. M. et al. (2001):** Identification of novel small RNAs using comparative genomics and microarrays. In: *Gene Dev* 15: 1637–1651. DOI: 10.1101/gad.901001.
- Winkler, W. C./Breaker, R. R. (2003):** Genetic control by metabolite-binding riboswitches. In: *ChemBiochem* 4: 1024–1032. DOI: 10.1002/cbic.200300685.
- Winkler, W. et al. (2002):** Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. In: *Nature* 419: 952–956. DOI: 10.1038/nature01145.
- Woese, C. R. et al. (1990):** Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. In: *Proc National Acad Sci* 87: 4576–4579. DOI: 10.1073/pnas.87.12.4576.
- Yan, C. et al. (2015):** Structure of a yeast spliceosome at 3.6-angstrom resolution. In: *Science* 349: 1182–1191. DOI: 10.1126/science.aac7629.3. The importance of RNA for molecular diagnostics.

3. The importance of RNA for molecular diagnostics

Jörn Walter and Nina Gasparoni

RNA analyses play a unique and outstanding role for clinical research and molecular diagnostics. They reveal individual cell programs, thereby enabling the differentiation between healthy and diseased cells. RNA analyses also facilitate the diagnosis of pathogens and provide the basis for the development and application of modern RNA-assisted treatment and immunization methods. The rapid development of next-generation sequencing (NGS) technologies has led to a series of new applications for RNA analytics in recent years. One such particularly groundbreaking technology is single cell sequencing, which opens up a whole new spectrum of possibilities for the field of RNA based clinical diagnostics.

3.1 Overview

RNAs are produced through the transcription of DNA to RNA. The precise quantitative and qualitative measurement of all the different RNAs present in the cell at a given moment in time, known collectively as the transcriptome, reflects the current molecular status of cells or tissue. In addition to mRNAs, i.e. copies of genes that are mostly translated to proteins, non-coding RNAs (such as miRNAs, circRNAs, snoRNAs, piRNAs, rRNAs, tRNAs and many others) are increasingly becoming the focus of diagnostics methods (see also Munschauer/Vogel, Chapter 2, for an overview of RNAs). Analyses using sequencing (RNA-Seq, see Section 3.2) offer particularly detailed insights into the volume and variety of different RNAs. Sequencing methods have a far broader range of resolution and sensitivity than other RNA analysis methods (see below), to the extent that they can even map the transcriptomes of individual cells. This opens up entirely new possibilities for diagnostic processes. RNA sequencing data is key to the identification of pathogenic organisms and viruses. Research into SARS-CoV-2 (COVID-19) has made it abundantly clear just how important RNA sequencing data is for countering viral infectious diseases. Today, RNA data has become a central component of disease diagnostics in almost all areas of medicine, providing accompanying data to facilitate diagnosis while also playing a significant role in the selection of appropriate treatment.

3.2 Spectrum of RNA analysis methods

The methods used in (routine) RNA diagnostics comprise a series of molecular techniques (see Table 2). Individual RNAs/genes can be analyzed and quantified by way of PCR methods. The presence or quantity of (disease-specific) RNA copies (known as biomarkers) is investigated through targeted amplification of RNA by quantitative PCR methods (Bustin/Mueller, 2005; VanGuilder et al., 2005).¹ This targeted methodology, which has been utilized for many years, is gene-specific and suitable for the rapid, quantitative detection of individual biomarkers in a large number of samples. This method remains the gold standard for diagnostics in certain areas (e.g. in terms of viruses, such as SARS-CoV-2 and many others). The development of microarrays ushered in a new level of complexity in diagnostics. Arrays make it possible to simultaneously analyze numerous known disease-relevant genes. The changes in RNA transcribed from genes are analyzed by binding the RNA in question to known capture molecules on a glass substrate. Optical techniques can determine the relative strength of the bond and thereby draw conclusions as to the relative quantity of RNA “captured” on the microarray (Gorreta et al., 2012).² However, PCRs and microarray analyses are restricted to examining a limited number of pre-defined genes, and therefore require a priori knowledge. New and in-depth exploratory analyses can only be achieved with the help of next-generation sequencing (NGS). RNA sequencing offers a comprehensive and unbiased way to examine the entire transcriptome without the need to define possible target regions in advance (Mortazavi et al., 2008; Wang et al., 2009). RNA sequencing therefore generates a far more detailed insight into the variety of RNA in one or multiple cells and also opens up a wide spectrum of applications, ranging from single cell analysis to the identification of new, as-yet unknown viruses. Second-generation sequencing methods with “short reads” and high sample throughput are currently setting the standard for RNA sequencing.³

Solid, reproducible quantifiability and rapid applicability are key criteria for diagnostic applications. Easy-to-implement, sensitive and precise techniques such as quantitative PCR (qPCR) and array methods are therefore still often in widespread use. They offer advantages for targeted, rapid diagnostics as well as enhanced comparability and standardization across numerous samples. Ultimately, though, they always build on data acquired through RNA sequencing. The three methods are therefore often applied consecutively in diagnostic processes, as illustrated by the example of COVID-19 diagnostics.

¹ PCR (polymerase chain reaction) is a method for amplifying short, defined sections of DNA. It is based on a repeating cycle of three reaction phases. After the double-stranded DNA is separated into individual strands, the short DNA fragments complementary to the region of the sequence to be amplified begin to attach themselves. These fragments serve as a starting point for the subsequent enzymatic DNA synthesis in which the DNA section is amplified. The products of previous cycles serve as templates for further amplification (known as exponential amplification). In the case of RNA analysis, the RNA is initially converted into DNA (cDNA, copyDNA or complementary DNA). The quantity of the PCR product provides an indication of the examined target region's strength of expression.

² Microarrays are biotechnological tools used to determine gene expression of a large number of (selected) genes simultaneously. A substrate (e.g. a glass slide) is coated with single-stranded fragments that correspond to the target regions under examination. The RNA is first transcribed into DNA (cDNA), which is then fragmented and marked (e.g. using fluorescent dyes). The cDNA fragments bind (hybridize) to corresponding complementary fragments on the array. Measuring the fluorescence intensity determines the quantity of bound fragments and therefore the strength of expression of each target region.

³ RNA sequencing (based on second-generation sequencing methods) starts with isolating RNA from cells. Specific processes are used to separate the different types of RNA and either enrich or deplete the desired RNAs (mRNA, miRNA or ncRNA). After transcribing the RNA into a cDNA (reverse transcription, RT), the cDNA is subjected to enzymatic degradation and subsequent PCR amplification (steps collectively known as: RT-PCR). This produces a library of these fragments, which can then be read in high-throughput sequencing. In the PCR step, a unique code is added to the end of the fragments; this code is also sequenced and makes it possible to allocate all segments to the correct original sample (demultiplexing). The sequencing is carried out using short-read sequencing machines (which usually examine fragments of between 50 and 100 bases in length).

The discovery of SARS-CoV-2 and the development of broad-based diagnostic methods and vaccines initially required the rapid and comprehensive sequencing of the RNA genome in the first virus isolates. This made it possible to swiftly define and classify the virus. Building on the RNA sequence, the first PCR-based detection test methods were developed, followed by a series of further immunological and molecular tests (Corman et al., 2020; Dao Thi et al., 2020; Yan et al., 2020; Kellner et al., 2022). The continuous collection of RNA sequencing data for virus isolates therefore lays the foundation for developing and constantly adapting the process, such as when new variants emerge. Knowledge of the virus genome and its variants is also the basis for the development of new and (continuously) adapted RNA vaccines. Monitoring virus variants with RNA sequencing and specific PCR-based test methods has been, and remains, essential to identify the spread of variants. Ultimately, data from comparative SARS coronavirus sequencing is used to fathom the origins and evolution of the RNA virus (Andersen et al., 2020).

However, RNA sequencing analyses are used for far more than simply diagnosing SARS-CoV-2. In order to gain a better understanding of the pathology of the COVID-19 disease, RNA sequencing data is being harnessed in a number of studies, including studies examining virus isolates obtained from the salivary, blood and lung cells (e.g. bronchoalveolar lavage) of COVID-19 patients. These studies – often single-cell studies – have made a significant contribution to improving our understanding of the pathogenesis, progression and long-term effects of a COVID-19 infection (Bernardes et al., 2020; Chua et al., 2020; Schulte-Schrepping et al., 2020; Sungnak et al., 2020; Muus et al., 2021; Stephenson et al., 2021; Yang et al., 2021). Overall, the example of COVID-19 research and diagnostics underscores the vital importance of RNA sequencing methods for molecular diagnostics.

Current RNA sequencing methods build on second-generation NGS technologies. They have now been refined to the extent that complex analyses can be conducted with even minute quantities of RNA. However, the arrival of third-generation NGS technologies⁴ presents a series of new and advanced quantitative and qualitative possibilities for RNA diagnostics. For instance, some of the new sequencing methods are markedly faster and more cost-effective. Once again, the use of new methods for SARS-CoV-2 sequencing is an example (Brinkmann et al., 2020; Freed et al., 2020). However, it will take time for such methods, as both the new technologies and new challenges for data analysis must first be implemented for widespread use.

Third-generation sequencing methods also offer the ability to sequence RNA directly, i.e. to avoid transcribing them into cDNA. This not only reduces the number of work steps, it also provides additional information about the RNA that is only available indirectly or to a limited extent using short-read second-generation NGS methods. Although third-generation sequencing methods require more RNA as starting material, they are able to read longer stretches of RNA molecules, meaning that complex and long RNA molecules can be

⁴ The third generation of sequencing methods comprises different technologies that directly analyze DNA/RNA molecules and do not require PCR amplification or, in the case of RNA analysis, transcription of RNA into cDNA. These technologies are based on various methodological approaches. In contrast to second-generation systems, single-molecule real-time sequencing facilitates the direct analysis of individual RNA molecules. It uses a series of combined biochemical and physical processes that make it possible to read long RNA molecules. Some systems also have the ability to identify modified bases in the RNA and thereby read them precisely in single molecules in the context of sequencing.

sequenced in one reaction. This makes it possible to diagnose the native form of the originally formed RNA and directly determine length and splicing variants⁵ (Athanasopoulou et al., 2021) and represents valuable additional information for the diagnosis of disease-relevant genes. In addition, some new third-generation methods can be used to not only read the base sequence but also the modifications to the RNA bases (Athanasopoulou et al., 2021). The extent to which this (epitranscriptional) modification mapping will play a role in diagnostics, however, remains unclear.

3.3 Applications of RNA analytics in diagnostics

Established and certified RT-PCR and RNA microarray testing methods make an important contribution to routine diagnostics – also because (as noted previously) the ability to standardize such methods facilitates their implementation in routine use, which in turns makes them suitable for approval procedures. In clinical research, however, they are increasingly being replaced by second-generation and third-generation RNA sequencing (NGS) for the reasons outlined above. At this point, we would like to briefly present the potential areas of application for the various methods (see Table 2).

Besides identifying pathogens (viruses, bacteria), another main application of RNA analysis in research and diagnostics is determining the expression (quantity and form of RNAs) of cataloged coding and non-coding genes in human cells. In relation to non-coding RNAs, the analysis of miRNAs plays a significant role as an indicator of altered cellular regulation. MiRNA-aided diagnostics are primarily used in cancer research, immunology and in connection with neurodegenerative disorders, often supplementing mRNA-based gene expression analysis (Kappel/Keller, 2017). The high sensitivity of a diagnostic process incorporating RNA sequencing presents advantages when the material available for analysis is very limited. Such cases include analysis of RNAs from healthy and diseased cells emitted into body fluids (serum, saliva, urine etc.). These “freely swimming” (extra-cellular) RNAs, e.g. miRNAs, provide an initial indication of a disease (e.g. cancer) in the body and are of great interest as a potential “first line” of non-invasive molecular diagnostics (Kappel/Keller, 2017).

Generally, RNA-assisted analysis – in the cases of both arrays and RNA sequencing – is oriented towards direct (or indirect) comparisons of differences in gene expression in the RNAs of healthy and diseased cells and tissue. Such comparisons make it possible to draw conclusions regarding the presence, stage and severity of diseases. Statistically valid information from such comparative data can then be used to determine disease-relevant changes and to support and optimize treatment strategies (a process known as companion diagnostics)⁶ (Sager et al., 2015; Slembrouck et al., 2019).

⁵ Splicing is an important step in RNA processing. After transcription, non-coding regions (introns) are removed from the pre-mRNA and the coding regions (exons) are joined together. The exons can sometimes be combined in different ways (e.g. by omitting or rearranging exons). These variations in the splicing of the same RNA create different mRNAs, which are then transcribed into different amino acid sequences (and therefore proteins) (see also Munschauer/Vogel, Chapter 2).

⁶ The term “companion diagnostics” relates to certain diagnostic tests that are carried out before or concomitant with a therapy. The analysis of specific biomarkers is used to determine the suitability of a therapy or drug for a specific patient (personalized medicine).

RNA sequencing data offers a wealth of additional information. For example, it enables to determine the consequences (expression) of genetic changes (gene variants) in healthy and diseased cells. This supplements and expands genetic diagnostics with functional, cell-specific data. Moreover, RNA sequencing data enables the determination of alternative transcript variants and changes in the splicing of individual genes (Wang et al., 2009; Bayega et al., 2018; Murdock, 2020). This additional gene-specific information is particularly important in areas such as cancer diagnostics as well as the diagnosis of rare diseases. Consequently, RNA sequencing is increasingly being used as a supplementary instrument for genome-based molecular diagnosis, as the interpretation of certain gene variants solely on the basis of DNA sequencing data is often inconclusive. Detection of variant expression (transcription) of a gene and of its RNA variants in the disease and their allocation to specific tissues or cells provides important additional information regarding the functional effects of genetic changes. Integrating RNA sequencing thereby leads to significantly better diagnosis of rare diseases (Peymani et al., 2022; Yépez et al., 2022).

RNA sequencing data also makes it possible to demonstrate the consequences of a given genomic change. For example, it can detect gene fusions. In diseased cells (but also in healthy cells), the rearranging of sections of chromosomes can lead to genes being connected on a chromosome in a way that gives rise to new fusion genes. This is then read at the RNA level as a fusion transcript, i.e. as a single gene, thereby creating an entirely new (fusion) protein. These fusion proteins can have functional consequences. Some can directly contribute to the malignant degeneration of cells, i.e. carcinogenesis. The most well-known example is BCR-ABL fusion of the "Philadelphia" chromosome, which is the cause of certain forms of leukemia (blood cancer). In summary the diagnosis of transcriptomes by RNA sequencing offers an efficient corner stone to understand underlying genetic changes in diseases (Mertens et al., 2015; Heyer et al., 2020; Kerbs et al., 2022). It is therefore noticeable that the qualitative differences of RNAs are attracting growing attention in the field of diagnostics. In the future RNA sequencing data will increasingly be used to interpret the genetic context of the examined person, i.e. their individual genetic make-up will be determined and assessed in the context of, and in relation to, changes in gene activity (Montgomery et al., 2010; Pickrell et al., 2010; Majewski/Pastinen, 2011; Lappalainen et al., 2013). The comprehensive determination of such individual quantitative variants will be significant for the analysis and understanding of complex diseases as they are caused by changes in the expression of multiple or many genes.

RNA analysis is increasingly being used to research disease models. It involves using cell lines and organoids⁷ established *in vitro* as a substitute for primary cells or organs that are either unavailable or only available to a limited extent for direct analysis, such as nerve cells. These model systems are generated *in vitro* in laboratories in accordance with standardized protocols using (usually pluripotent) precursor cells that differentiate into different cell types and thereby form a complex, three-dimensional structure such as organoids. Targeted stimulation of aggregated cells or organoids then makes it possible to simulate, for example, aging and pathological processes and also the (side) effects of (potential) medicinal sub-

⁷ See also: Bartfeld et al., 2020.

stances. Using RNA sequencing (or arrays) to compare gene expression signatures does more than help to clarify the fundamental question of how closely the model recreates the in vivo organ or the cells that comprise it. In addition, the induced effects on the molecular signatures of the cells or organoid can also be recorded with a high degree of sensitivity using (single cell) RNA sequencing and then interpreted through comparison with untreated reference samples. The insights gained through this process provide important information for the diagnosis and therapy of the investigated diseases (Brancata et al., 2020).

The number of applications for single cell sequencing has risen sharply in recent times. This includes a series of NGS methods capable of determining complex transcriptomes of individual cells.⁸ Single cell transcriptome data offers an entirely new way to define cell types, cell states and pathological changes in diseased cells. Single cell “atlases” are currently produced for many “normal” healthy cell types. They provide deep insights in the cell composition of tissues and the changes over the course of development or with aging. Most importantly they allow to rationalize changes in the composition and function of cells with regard to various diseases such as autoimmune diseases and cancer. Single-cell analyses produce deep insights into cell changes found in cancer, i.e. the cellular heterogeneity of tumors. This will take diagnostic and therapeutic applications to a new level (see also Walter/Schickl, 2019). In the future, single cell sequencing of individual tumors will help to improve the understanding of the origins and evolution of individual tumors, and adjust diagnosis and treatment accordingly. RNA analysis of single cells can also be combined with epigenomic techniques, e.g. ATAC sequencing,⁹ to reach and consider another level of functional regulation and interpretation.

Single-cell sequencing methods offer an extremely high resolution. However, in this process the ability to spatially assign or map the individual sequenced cells (e.g. in the tumor) is lost. This information is provided by new technical approaches that can help to study the expression of genes (gene variants) in situ, i.e. directly in the tissue in tissue sections. This method is known as “spatial transcriptomics” and enables the mapping of pathological changes in gene expression in tissue to precise cells (Moor/Itzkovitz, 2017).

These examples of applications illustrate the tremendous significance of RNA-based single cell diagnostics. Due to its increasing applicability in clinically relevant settings, in particular thanks to the rapidly advancing methods to automate RNA-aided experiments, single cell analysis will assume a central and leading role in clinical research and diagnostics in the near future. The establishment of research centers – such as the Berlin Cell Hospital at the Charité, set up in partnership with the MDC¹⁰ – emphatically underscores this trend.

⁸ Current single cell sequencing methods are based on microprocessing techniques in order to isolate (single out) cells in tissue and then process them. This involves creating complex sequencing libraries with individual codings so that, after parallel high-throughput sequencing of thousands of cells, it is possible to differentiate between the expression profiles of individual cells

⁹ The ATAC-Seq (assay for transposable accessible chromatin using sequencing) method makes it possible to map regulatory regions in the genome by analyzing the local accessibility of chromatin (complex of DNA and proteins that form chromosomes).

¹⁰ Berlin Cell Hospital, cf. <https://www.mdc-berlin.de/de/themen/berlin-cell-hospital> [31.08.2022]; and “Single cell approaches for personalized medicine”, a research focus of the BIH, Charité and MDC, cf. <https://www.mdc-berlin.de/content/single-cell-approaches-personalized-medicine> [31.08.2022].

Table 2: Overview of different diagnostic methods and their application

Diagnostic method	Use	Application
PCR	Rapid detection and quantification of selected, known, individual RNAs.	<ul style="list-style-type: none"> • Detection of pathogenic organisms and viruses in RNA (single gene) • Analyses of individual, gene-specific mRNAs and non-coding RNAs (e.g. miRNAs) • Quantitative detection of known marker genes, fusion transcripts • Applicability for single cells
Microarray	Standardized detection and relative quantification of several to many known (annotated) genes or biomarkers.	<ul style="list-style-type: none"> • Detection of mRNAs (without gene variants) • Detection of annotated miRNAs and other non-coding RNAs
RNA sequencing (second generation) (Read length: 50–100 base pairs)	Highly sensitive detection of minute quantities of RNA. Quantitative and qualitative analysis of all RNAs present in a cell. Identification of new transcripts and transcript variants.	<ul style="list-style-type: none"> • Detection of pathogenic organisms and viruses in RNA • Quantification of marker genes (mRNA) or non-coding RNAs (e.g. miRNAs, circRNAs) • Reading known fusion transcripts and identifying new transcripts • Partial detection of transcript and splicing variants • Widely applicable for single cell analyses, e.g. determination of cell identify and cell composition (heterogeneity) in tissue (“cell atlases”)
RNA sequencing (third generation) (Read length: 1–100 kilobase pairs)	Direct sequencing of complex, long RNAs. Determination of RNA modifications.	<ul style="list-style-type: none"> • Recording of complete transcript and splicing variants (mRNA) • Mapping of complex virus transcript variants (e.g. SARS-CoV-2) • Reading modified bases in RNA • To date, not directly applicable for single cell analyses

3.4 Processing of diagnostic RNA data

Bioinformatic information plays a fundamental role in the use of data produced with the help of highly specialized technologies (arrays, second- and third-generation sequencing method). The evaluation and interpretation of complex array and RNA sequencing data needs standardized computer-aided processes that are expedient for diagnostic purposes. Excellent software packages are available, performing tasks ranging from primary data processing to quality assessments to interpretation (Slonim/Yanai, 2009; Kukurba/Montgomery, 2015; Conesa et al., 2016; Zoabi/Shomron, 2021; Pavlovich/Cauchy, 2022).

Actively managed public reference databases facilitate standardized comparisons, i.e. allocation of samples and cells, and thereby also comparability of the data-based functional interpretation. All data collected in diagnostics should therefore be stored in protected locations (databases) in order to make it available for research and diagnostics through a controlled framework. In Europe, the European Genome Archive (EGA)¹¹ and its federated

¹¹ Available at: <https://ega-archive.org/> [29.08.2022].

German branch, the German Human Genome-Phenome Archive (GHGA),¹² play an important role by providing secure data storage while also facilitating access to and provision of comparison data. In addition, data portals such as Ensemble (functional data)¹³ and platforms such as the Genotype-Tissue Expression Project (GTEx, expression and genetic data)¹⁴ and Human Cell Atlas (HCA, single cell data)¹⁵ are of inestimable value for diagnostic interpretation.

The second step involves determining the gene-specific RNA quantity in one or more cells and conducting a “standardized” comparison with other cells (or with reference data). In these further analyses, RNA sequencing can be used to identify the quantities in which genes have been transcribed in cells or tissues. It is also possible to determine the extent to which alternative forms (splicing variants, see above) of gene-specific RNAs have been composed. By locating RNA sequencing data in the genome, we can determine whether genes are read from one site or whether (new) alternative start and end points (isoforms) are used to form a gene-specific RNA in diseased cells. Depending on the data depth and quality, conclusions can be drawn about genetic changes and thereby support genetic diagnostics (e.g. in cancer cells).

All of this primary, functional mapping of RNA data is important for supporting diagnoses. However, next-level analyses and interpretations need to be included in the future. This involves examining which genes and/or RNAs are dysregulated together in diseases cells or tissues and to what degree. By simultaneously examining numerous genes in this manner and using (self-learning) network modeling, it is possible in a second step to determine correlative relationships in changed gene expression (Kakati et al., 2019; Chowdhury et al., 2020). Interpretation based on network analyses facilitates a far more differentiated view of the changes and contributes to the development of new and improved diagnosis and treatment options. In such analyses, disease-specific effects are viewed more in the context of the regulatory change in diseased cells.

This also expands the spectrum of uses for gene-specific biomarker testing (e.g. with PCR tests). The incorporation of additional omics data¹⁶ is increasingly prevalent as a way to better understand the functional changes in diseased cells. This is known as integrated, multi-omics analysis, in which different omics data (e.g. proteome or epigenome) collected from the same probe or comparable reference data sets, are combined to determine the extent to which programming processes have already changed in cells, as well as on which molecular level and to what extent these changes can be reversed (Argelaguet et al., 2018; Subramanian et al., 2020; Li et al., 2021; Correa-Aguila et al., 2022).

¹² Available at: <https://www.ghga.de/> [29.08.2022].

¹³ Available at: <https://www.ensembl.org/index.html> [29.08.2022].

¹⁴ Available at: <https://gtexportal.org/home/> [29.08.2022].

¹⁵ Available at: <https://www.humancellatlas.org/> [29.08.2022].

¹⁶ The term “omics” relates to moleculobiological studies focusing on the entire cellular content of the examined molecules, e.g. proteomics (proteins), epigenomics (epigenetic state), etc.

3.5 Challenges of RNA-based diagnostics

The use of RNA sequencing in diagnostics facilitates a differentiated view of cell functions. It is broadly scalable such that analyses can be conducted with minute quantities of RNA from biopsies and blood, and even blood plasma. However, depending on the source and the method used, RNA-based analysis is subject to certain challenges. RNA isolation methods determine whether large or small RNAs can be analyzed (<200 base pairs, >200 base pairs) or whether gene-specific (polyA)¹⁷ or all RNAs (“total RNA”)¹⁸ should be examined. An important aspect of RNA analysis is the integrity of isolated RNA for diagnostics. RNA is far less stable than DNA. The quality of RNA-based data is therefore heavily dependent on rapid access and optimal extraction of materials (tissue, blood or cells). If possible, RNA analyses should be conducted on fresh cells or cell material and coupled with standardized procedures in order to achieve high quality and comparability. Furthermore, comparability should always be improved by repeating the analyses multiple times. Computerized methods of standardizing the raw data can, to a certain extent, compensate for inherent variance problems in RNA analytics, thereby minimizing the impact of technical variations attributable to issues such as varying RNA quality.

The complexity of RNA data makes interpreting it a challenging proposition. A tissue’s RNA data is representative of the total spectrum of all cells in this tissue. A diagnostic comparison of healthy and diseased total tissue therefore only offers an average value of the healthy and pathologically changed cells in the tissue. This data can be improved with computerized methods (known as deconvolution) harnessing reference data on isolated cell types from public databases. If possible, these analyses can also be improved experimentally by using cell sorting to separate the cells in the tissue or blood before an RNA analysis and then sequencing the different cell types separately.¹⁹ This is a common method for separating the different cells in blood. Single cell analyses, however, are the only way to gain precise insights into cell composition and cell changes. In this case, the cells in the tissue or blood are isolated before thousands of randomly selected cells from the tissue are sequenced and analyzed individually and in parallel. The costs, experimental logistics, the need to for highly specialized instruments and the expertise required all increase significantly in line with the degree of resolution in the analyses. It will therefore become increasingly important to promote the ability to make decisions regarding which practical and feasible diagnostic methods should be used for a specific diagnostic issue. Establishing centers for highly specialized RNA analysis in hospitals is therefore a necessity to facilitate the development of efficient workflows and the constant adaptation to modern technologies and bioinformation methods.

¹⁷ In order to analyze gene-specific or coding RNAs (mRNAs), these can be enriched, e.g. through polyadenylation (polyA) at their 3' end (3' refers to the rear end of the RNA as read).

¹⁸ This refers to all RNAs (coding and non-coding RNAs).

¹⁹ The cells in the tissue or blood are separated into different types based on their size, shape and the presence of certain molecules on the cell surface, thereby enriching them.

3.6 Outlook

Molecular diagnostics based on RNA analysis is exceptionally important for personalized medical applications. New methods, particularly in relation to single cell analysis, will supplement and considerably enhance the sensitivity and quality of pathological findings in the near future. The rapidly advancing application of new technical methods of RNA sequencing (single cells, spatial transcriptomics, third-generation direct sequencing) will generate data that will facilitate significantly more differentiated and precise detection of individual changes in diseased cells. This will usher in a new era in personalized molecular diagnostics.

3.7 References

- Andersen, K. G. et al. (2020):** The proximal origin of SARS-CoV-2. In: *Nat Med* 26(4): 450–452. DOI: 10.1038/s41591-020-0820-9.
- Argelaguet, R. et al. (2018):** Multi-Omics Factor Analysis – a framework for unsupervised integration of multi-omics data sets. In: *Mol Syst Biol* 14(6): e8124. DOI: 10.15252/msb.20178124.
- Athanasopoulou, K. et al. (2021):** A. Third-Generation Sequencing: The spearhead towards the radical transformation of modern genomics. In: *Life* 12(1): 30. DOI: 10.3390/life12010030.
- Bartfeld, S. et al. (2020):** Organoide. Ihre Bedeutung für Forschung und Medizin und Gesellschaft. Baden-Baden: Nomos.
- Bayega, A. et al. (2018):** Current and future methods for mRNA analysis: A drive toward single molecule sequencing. In: *Methods Mol Biol* 1783: 209–241. DOI: 10.1007/978-1-4939-7834-2_11.
- Bernardes, J. P. et al. (2020):** Longitudinal multi-omics analyses identify responses of megakaryocytes, erythroid cells, and plasmablasts as hallmarks of severe COVID-19. In: *Immunity* 53(6): 1296–1314.e9. DOI: 10.1016/j.immuni.2020.11.017.
- Brancati, G. et al. (2020):** Resolving neurodevelopmental and vision disorders using organoid single-cell multi-omics. In: *Neuron* 107(6): 1000–1013. DOI: 10.1016/j.neuron.2020.09.001.
- Brinkmann, A. et al. (2021):** AmpliCoV: Rapid whole-genome sequencing using multiplex PCR amplification and real-time Oxford nanopore MinION sequencing enables rapid variant identification of SARS-CoV-2. In: *Front Microbiol* 12: 651151. DOI: 10.3389/fmicb.2021.651151.
- Bustin, S. A./Mueller, R. (2005):** Real-time reverse transcription PCR (qRT-PCR) and its potential use in clinical diagnosis. In: *Clin Sci (Lond)* 109(4): 365–379. DOI: 10.1042/CS20050086.
- Chowdhury, H. A. et al. (2020):** (Differential) Co-expression analysis of gene expression: A survey of best practices. In: *IEEE/ACM Trans Comput Biol Bioinform* 17(4): 1154–1173. DOI: 10.1109/TCBB.2019.2893170.
- Chua, R. L. et al. (2020):** COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. In: *Nat Biotechnol* 38(8): 970–979. DOI: 10.1038/s41587-020-0602-4.
- Conesa, A. et al. (2016):** A survey of best practices for RNA-seq data analysis. In: *Genome Biol* 17: 13. DOI: 10.1186/s13059-016-0881-8.
- Corman, V. M. et al. (2020):** Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR In: *Euro Surveill* 25(3): 2000045. DOI: 10.2807/1560-7917.ES.2020.25.3.2000045.
- Correa-Aguila, R. et al. (2022):** Multi-omics data integration approaches for precision oncology. In: *Mol Omics* 18(6): 469–479. DOI: 10.1039/d1mo00411e.
- Dao Thi, V. L. et al. (2020):** A colorimetric RT-LAMP assay and LAMP-sequencing for detecting SARS-CoV-2 RNA in clinical samples. In: *Sci Transl Med* 12(556): eabc7075. DOI: 10.1126/scitranslmed.abc7075.
- Freed, N. E. et al. (2020):** Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. In: *Biol Methods Protoc* 5(1): bpaa014. DOI: 10.1093/biomet/bpaa014.
- Gorreta, F. et al. (2012):** Genomic profiling: cDNA arrays and oligoarrays. In: *Methods Mol Biol* 823: 89–105. DOI: 10.1007/978-1-60327-216-2_7.
- Heyer, E. E. et al. (2020):** Diagnosis of fusion genes using targeted RNA sequencing In: *Nat Commun* 10(1): 1388. DOI: 10.1038/s41467-019-09374-9.

- Kakati, T. et al. (2019):** Comparison of methods for differential co-expression analysis for disease biomarker prediction. In: *Comput Biol Med* 113: 103380. DOI: 10.1016/j.compbio.2019.103380.
- Kappel, A./Keller, A. (2017):** miRNA assays in the clinical laboratory: workflow, detection technologies and automation aspects. In: *Clin Chem Lab Med* 55(5): 636–647. DOI: 10.1515/cclm-2016-0467.
- Kellner, M. J. et al. (2022):** A rapid, highly sensitive and Open-Access SARS-CoV-2 detection assay for laboratory and home testing. In: *Front Mol Biosci* 9: 801309. DOI: 10.3389/fmolb.2022.801309.
- Kerbs, P. et al. (2022):** Fusion gene detection by RNA-sequencing complements diagnostics of acute myeloid leukemia and identifies recurring NRIP1-MIR99AHG rearrangements. In: *Haematologica* 107(1): 100–111. DOI: 10.3324/haematol.2021.278436.
- Kukurba, K. R./Montgomery, S. B. (2015):** RNA Sequencing and Analysis. In: *Cold Spring Harb Protoc* 2015(11): 951–969. DOI: 10.1101/pdb.top084970.
- Lappalainen, T. et al. (2013):** Transcriptome and genome sequencing uncovers functional variation in humans. In: *Nature* 501(7468): 506–511. DOI: 10.1038/nature12531.
- Li, Y. et al. (2021):** Advances in bulk and single-cell multi-omics approaches for systems biology and precision medicine. In: *Brief Bioinform* 22(5): bbab024. DOI: 10.1093/bib/bbab024.
- Majewski, J./Pastinen, T. (2011):** The study of eQTL variations by RNA-seq: from SNPs to phenotypes. In: *Trends Genet* 27(2): 72–79. DOI: 10.1016/j.tig.2010.10.006.
- Mertens, F. et al. (2015):** The emerging complexity of gene fusions in cancer. In: *Nat Rev Cancer* 15(6): 371–381. DOI: 10.1038/nrc3947.
- Montgomery, S. B. et al. (2010):** Transcriptome genetics using second generation sequencing in a Caucasian population. In: *Nature* 464(7289): 773–777. DOI: 10.1038/nature08903.
- Moor, A. E./Itzkovitz, S. (2017):** Spatial transcriptomics: paving the way for tissue-level systems biology. In: *Curr Opin Biotechnol* 46: 126–133. DOI: 10.1016/j.copbio.2017.02.004.
- Mortazavi, A. et al. (2008):** Mapping and quantifying mammalian transcriptomes by RNA-Seq. In: *Nat Methods* 5(7): 621–628. DOI: 10.1038/nmeth.1226.
- Murdock, D. R. (2020):** Enhancing diagnosis through RNA sequencing. In: *Clin Lab Med* 40(2): 113–119. DOI: 10.1016/j.cll.2020.02.001.
- Muus, C. et al. (2021):** Single-cell meta-analysis of SARS-CoV-2 entry genes across tissues and demographics. In: *Nat Med* 27(3): 546–559. DOI: 10.1038/s41591-020-01227-z.
- Pavlovich, P. V./Cauchy, P. (2022):** Sequences to Differences in Gene Expression: Analysis of RNA-Seq Data. In: *Methods Mol Biol* 2508: 279–318. DOI: 10.1007/978-1-0716-2376-3_20.
- Peymani, F. et al. (2022):** RNA sequencing role and application in clinical diagnostic. In: *Pediatr Investig* 6(1): 29–35. DOI: 10.1002/ped4.12314.
- Pickrell, J. K. et al. (2010):** Understanding mechanisms underlying human gene expression variation with RNA sequencing. In: *Nature* 464(7289): 768–772. DOI: 10.1038/nature08872.
- Sager, M. et al. (2015):** Transcriptomics in cancer diagnostics: developments in technology, clinical research and commercialization. In: *Expert Rev Mol Diagn* 15(12): 1589–1603. DOI: 10.1586/14737159.2015.1105133.
- Schulte-Schrepping, J. et al. (2020):** Severe COVID-19 is marked by a dysregulated myeloid cell compartment. In: *Cell* 182(6): 1419–1440.e23. DOI: 10.1016/j.cell.2020.08.001.
- Slembrouck, L. et al. (2019):** Decentralization of next-generation RNA sequencing-based MammaPrint® and Blueprint® kit at University Hospitals Leuven and Curie Institute Paris. In: *Transl Oncol* 12(12): 1557–1565. DOI: 10.1016/j.tranon.2019.08.008.
- Slonim, D. K./Yanai, I. (2009):** Getting started in gene expression microarray analysis. In: *PLoS Comput Biol* 5(10): e1000543. DOI: 10.1371/journal.pcbi.1000543.
- Stephenson, E. et al. (2021):** Single-cell multi-omics analysis of the immune response in COVID-19. In: *Nat Med* 27(5): 904–916. DOI: 10.1038/s41591-021-01329-2.
- Subramanian, I. et al. (2020):** Multi-omics data integration, interpretation, and its application. In: *Bioinform Biol Insights* 14: 1177932219899051. DOI: 10.1177/1177932219899051.
- Sungnak, W. et al. (2020):** SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. In: *Nat Med* 26(5): 681–687. DOI: 10.1038/s41591-020-0868-6.
- VanGuilder, H. D., et al. (2008):** Twenty-five years of quantitative PCR for gene expression analysis. In: *Biotechniques* 44(5): 619–626. DOI: 10.2144/000112776.
- Walter, J./Schickel, H. (2019):** Einzelzellanalyse in Forschung und Medizin – Eine Stellungnahme der interdisziplinären Arbeitsgruppe Gentechnologiebericht. Berlin Brandenburgische Akademie der Wissenschaften, Berlin.

Wang, Z. et al. (2009): RNA-Seq: a revolutionary tool for transcriptomics. In: *Nat Rev Genet* 10(1): 57–63. DOI: 10.1038/nrg2484.

Yan, C. et al. (2020): Rapid and visual detection of 2019 novel coronavirus (SARS-CoV-2) by a reverse transcription loop-mediated isothermal amplification assay. In: *Clin Microbiol Infect* 26(6): 773–779. DOI: 10.1016/j.cmi.2020.04.001.

Yang, A. C. et al. (2021): Dysregulation of brain and choroid plexus cell types in severe COVID-19. In: *Nature* 595(7868): 565–571. DOI: 10.1038/s41586-021-03710-0.

Yépez, V. A. et al. (2022): Clinical implementation of RNA sequencing for Mendelian disease diagnostics. In: *Genome Med* 14(1): 38. DOI: 10.1186/s13073-022-01019-9.

Zoabi, Y./Shomron, N. (2021): Processing and Analysis of RNA-seq Data from Public Resources. In: *Methods Mol Biol* 2243: 81–94. DOI: 10.1007/978-1-0716-1103-6_4.

4. Opening up the world of RNA for therapeutics

Anke Sparmann and Jörg Vogel

RNA-based therapeutics are pharmaceuticals that target cellular RNA or, alternatively, RNA introduced into human cells that induce the production of specific proteins. After many years of research and development, this new form of medicine has now recorded its first major clinical successes.

4.1 Introduction

The quest for RNA-based therapeutics is almost half a century old. Its origins can be traced back to 1978, when Mary Stephenson and Paul Zamecnik produced a synthetic antisense oligonucleotide (ASO)¹ to inhibit the replication of the Rous sarcoma virus in tissue culture (Zamecnik/Stephenson, 1978). This ASO inhibited the translation of messenger RNA (mRNA, cf. Munschauer/Vogel, Chapter 2) into protein through sequence-specific binding to the viral mRNA (Stephenson/Zamecnik, 1978). These two studies paved the way for the use of this unique quality of nucleic acids to interact in a sequence-specific fashion in the development of a new type of pharmaceutical. While conventional drug design often requires laborious screening to identify lead chemical compounds, RNA and DNA-based medicines can be developed with a targeted approach. This is because the code that determines both the affinity and the specificity of the oligonucleotide binding – complementary base pairing – is well understood. If the sequence of the target RNA is known, an ASO capable of binding at a specific location on this target RNA can effectively be drafted on a computer screen. Nevertheless, it took many years to transfer this concept from the lab to a clinical application. The use of RNA in pharmaceuticals faces significant challenges due to the RNA's inherent pharmacological properties, which are determined by factors including the size and charge of the RNA molecule. The poor uptake of RNA in target cells and immune-related toxicity present additional hurdles. It took years of research in various scientific disciplines to overcome initial setbacks.

The discovery of the antisense concept described above laid the foundations for the devel-

¹ An ASO is a short-chain nucleic acid with a freely configurable base sequence. It uses complementary base pairing to bind to another nucleic acid (RNA or DNA) with a matching base sequence.

opment of RNA-based therapeutic strategies. In fact, this concept was applied in drug development even before the underlying molecular mechanisms were fully understood. The first oligonucleotide therapeutic based on an antisense mechanism of action was approved for use on the US market in 1998. Fomivirsen binds to a sequence in the mRNA of cytomegalovirus (CMV), which can cause severe illness in people with a weakened immune system, and inhibits CMV replication. The ASO therapeutic was indicated for use in the treatment of CMV retinitis – an infection of the retina that can result in blindness – in people with acquired immune deficiency syndrome (AIDS). Yet, despite its therapeutic benefit, the success of Fomivirsen was short-lived, and it was soon removed from the market due to advances in antiretroviral therapy for HIV. Nevertheless, Fomivirsen provided initial evidence of the clinical value of ASOs.

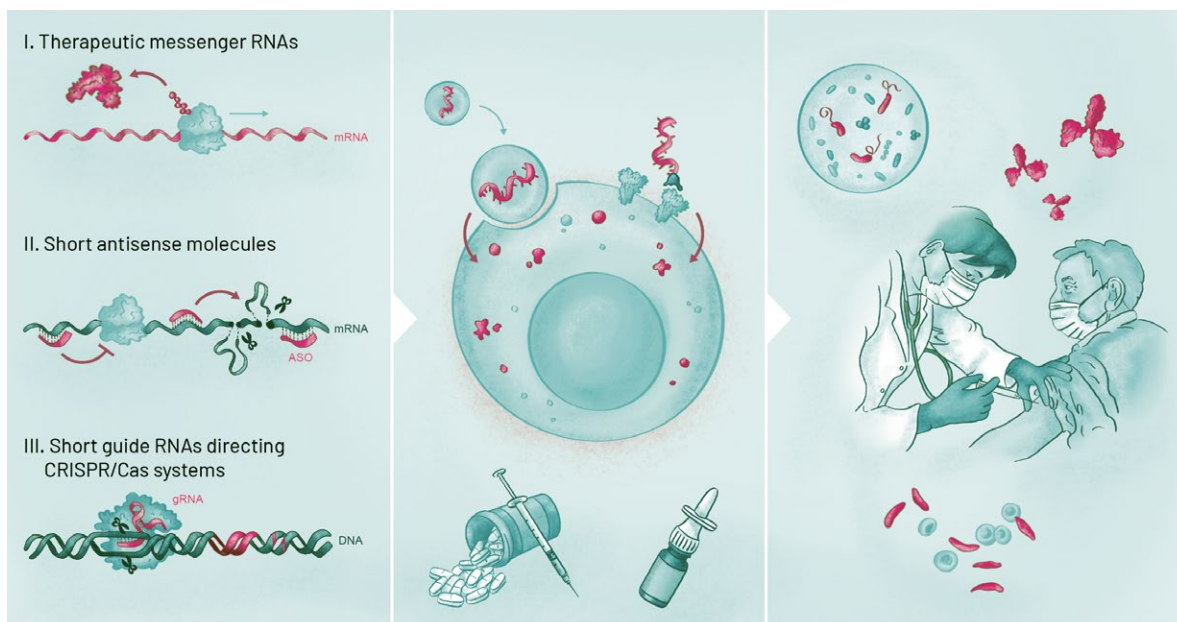
Scientific discoveries in the 1990s and early 2000s, in particular the discovery of micro-RNAs (miRNAs), RNA interference (RNAi) and clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) systems, have fundamentally transformed our understanding of gene regulation and revolutionized how biological research is conducted. As detailed below, these RNA-based molecular mechanisms in gene regulation have also found their way into drug development.

In addition to non-coding RNA therapeutics, RNA-based therapy can be even more broadly defined, namely when RNA itself serves as a source of information to trigger the production of a specific protein in the patient through administration of synthetic mRNA. The first experiments in this regard took place in the 1990s with studies focused on *in vitro* transcribed (IVT) mRNA, i.e. mRNA synthesized in the lab. Its first applications included the expression of autologous proteins in mammalian cells and efforts to develop vaccines against cancer and infectious diseases (Jirikowski et al., 1992; Mandl et al., 1998; Zhou et al., 1999). However, inefficient cellular uptake of exogenous mRNA (i.e. mRNA introduced into the organism from outside), its poor stability within the body and weak translation into protein in target cells, as well as an immune reaction triggered by the RNA were issues that hampered development of a successful clinical application. Once again, scientific breakthroughs were required. Administration routes needed to be improved and approaches for the specific enrichment of mRNA in certain cell types had to be found. The synthesis of IVT mRNA using chemically modified nucleosides marked a particularly significant development.² This type of modified IVT mRNA was not only more stable and led to increased production of the desired protein, it also proved to be less immunostimulatory (Karikó et al., 2005), which in turn made it possible to administer higher doses. In other words, researchers found a way to smuggle synthetic RNAs into target cells largely unnoticed by the immune system (see also Section 4.5). Once there, the cell reads the modified mRNA molecule and translates it into protein – just as happens with the cell's own mRNAs in the course of normal gene expression (see Munschauer/Vogel, Chapter 2, in relation to gene expression). The advances described above were essential for the exceptionally swift development of mRNA-based vaccines to combat the SARS-CoV-2 pandemic in 2020. This huge success has stirred massive interest in RNA as a major molecule class for the diagno-

² Nucleosides have a similar structure to nucleotides, the building blocks of DNA and RNA, but do not contain phosphate residues.

sis (see Walter/Gasparoni, Chapter 3), prevention and treatment of diseases. Vaccinating a large proportion of the global population with mRNA-based vaccines has also provided extensive data on the safety and efficacy of RNA as a vaccine, which will in turn play an important role in the approval of future RNA-based drugs. The challenge now is to transfer this principle to as many diseases as possible. In the following, we will discuss various strategies of RNA-based medicine, including ASOs, therapeutic mRNAs and CRISPR/Cas systems (see Figure 1).

Figure 1: Strategies for RNA-based medicine



© SciGraphix/Sandy Westermann

4.2 ASO therapeutics

ASO therapeutics are usually single-stranded, heavily modified and stabilized nucleic acid analogs³ that rely on the principle of complementary base pairing to bind to cellular RNAs and modulate RNA metabolic processes. ASOs have different mechanisms of action, including the sequence-specific degradation of target RNAs as well as binding to RNA to prevent processing steps such as RNA splicing⁴ and translation into proteins.

A well-established mechanism of action of ASOs is the degradation of target RNA by the ribonuclease⁵ RNase H1. RNase H1 degrades RNA within DNA:RNA hybrids⁶ but requires at least five and ideally eight to ten target RNA-paired desoxynucleotides (the building blocks

³ Nucleic acid analogs are synthetically produced substances that are similar in chemical structure to natural nucleic acids.

⁴ Splicing is an important step in mRNA processing in which the pre-mRNA, synthesized during the transcription of DNA to RNA, is transformed into a "mature" mRNA. The pre-mRNA contains both coding regions (exons) and non-coding sequences (introns). Splicing removes the introns and connects the adjacent exons together to form mature mRNA, which is then translated into protein (see also Munschauer/Vogel, Chapter 2).

⁵ Ribonucleases are enzymes that degrade RNA.

⁶ DNA/RNA hybrids are double-stranded molecules composed of a DNA chain and a complementary RNA chain.

of DNA) for enzymatic activity. Gapmer ASOs, in which a central desoxynucleotide core is flanked by nucleotides that are chemically modified at the 2'-position, are particularly efficient in triggering RNase H1-mediated degradation. These flanking modifications improve the binding of the Gapmer ASO to the target mRNA and also reduce degradation of the ASO by non-specific nucleases, which increases the Gapmer's stability in the body. Given the need for near-perfect complementarity between the ASOs and the target RNA within the desoxynucleotide core to activate RNase H1-mediated degradation, ASOs with this mechanism of action exhibit a high degree of specificity. This reduces the potential for undesired interactions, also known as off-targeting.

In addition to RNase H1-induced degradation, there are other mechanisms by which ASOs can lower the cellular level of target RNAs. Firstly, ASOs can trigger cellular quality control mechanisms, such as nonsense-mediated mRNA decay (NMD)⁷ or no-go decay (NGD)⁸ of target mRNA molecules. Secondly, ASOs can influence endogenous RNA metabolic processes such as polyadenylation⁹ and thereby accelerate the natural decay of target mRNA molecules. In addition, ASOs bound to the translation initiation sites of mRNAs can block the initiation of protein synthesis by ribosomes and thereby directly reduce protein levels. This mechanism is particularly effective in bacteria. Moreover, ASOs can also inhibit the interaction between mRNA and translation initiation factors¹⁰ or trigger the cleavage of 5' cap structures,¹¹ which also inhibits translation and leads to mRNA decay.

The ASO with arguably the most substantial clinical success to date uses a mechanism of action based on RNA splicing. Approved in 2016, Nusinersen is an ASO used to treat spinal muscular atrophy (SMA), a neuromuscular disorder caused by mutations in the *survival of motor neuron 1 (SMN1)* gene. These mutations lead to the loss of the SMN protein, without which the motor neurons in the bone marrow and brain stem degenerate. The result is muscle weakness and tissue loss. SMA is a destructive condition: 60 % of infants born with the disorder experience the onset of severe symptoms within six months, and the average life expectancy is less than two years. The *SMN2* gene, a paralog¹² of *SMN1*, codes for an identical SMN protein. However, the mRNA of *SMN2* undergoes alternative RNA splicing, with 90 % of *SMN2* transcripts lacking an exon (exon 7). This results in the production of a truncated, unstable protein. The molecular basis of this anomalous splicing, known as *SMN2* exon 7 skipping, was discovered around the turn of the millennium. *SMN2* contains a cytosine-to-thymine substitution in exon 7, which weakens the binding of splicing activators to the RNA and thereby impairs efficient recognition of the 3' splice site. Back in 2003,

7 Nonsense-mediated mRNA decay is a control mechanism in eukaryotic cells. It identifies mRNAs with premature stop codons and selectively degrades these mRNAs in order to prevent their translation into truncated proteins.

8 No-go decay is an mRNA quality control mechanism that degrades mRNAs with stalled ribosomes.

9 Polyadenylation is the addition of adenine nucleotides, known as the Poly(A) tail, to the 3'-end of mRNAs by the enzyme Poly(A) polymerase.

10 Translation initiation factors are proteins or protein complexes that function during translation initiation, the first step in translation. They are required for the efficient translation of mRNA.

11 The 5' cap structure is often a modified guanine nucleotide linked to the first nucleotide of the mRNA by a rare 5'-5'-phosphodiester bond. The cap structure protects mRNAs against decay and is important for mRNA export from the cell nucleus into the cytoplasm as well as for translation (see also Munschauer/Vogel, Chapter 2).

12 Paralogs are produced through gene duplication and exhibit a high level of sequence similarity. The gene copies are initially identical and often have redundant functions but they can develop differently over time.

Luca Cartegni and Adrian Krainer constructed bifunctional ASOs¹³ that functioned as synthetic splicing activators: a peptide imitating a splicing activator is linked to an ASO that binds to exon 7, thereby increasing the inclusion of exon 7 (Cartegni/Krainer, 2003).

In cooperation with Ionis Pharmaceuticals, work continued over the next years to optimize the strategy for inhibiting exon 7 skipping. It was discovered that ASOs that bind to another site on the SMN2 RNA also increase the proportion of “repaired” mRNAs including exon 7, even without the added peptide unit. Here, the ASO inhibits the binding of splicing inhibitors. In addition, chemical modification of the ASOs improved their pharmacological properties. Promising results in pre-clinical studies and subsequent, extraordinarily successful clinic studies led to Nusinersen receiving approval for use on the US market in 2016 (and in Europe in 2017).

Since then, Nusinersen has become available as a treatment for SMA in more than 40 countries. Not only has this substance revolutionized the treatment of SMA patients, it is also the first antisense drug to achieve appreciable commercial success. Indeed, the success of Nusinersen and the approval of nine additional ASOs to date, mostly for the treatment of rare genetic diseases without other therapy options, provide reason for optimism that ASO technologies may realize the hopes placed in them.

4.3 Therapeutics based on RNA interference mechanisms

In the early 1990s, scientists made puzzling observations in various experimental systems, especially in plants, that were indicative of RNA-induced gene regulation. At first, it was suspected that this was due to an antisense mechanism that, similar to ASOs, depends on hybridization between regulatory RNA and cellular mRNA. In 1998, however, Andrew Fire and Craig Mello demonstrated that double-stranded RNA in roundworms triggers a process that leads to sequence-specific RNA silencing. This process was dubbed “RNA interference”, or RNAi for short (Fire et al., 1998). Subsequent studies identified the underlying molecular mechanism: double-stranded RNA is processed into small interfering RNAs (siRNAs), which are integrated into an RNA-induced silencing complex (RISC) composed of different cellular proteins. The siRNAs guide this complex to complementary target RNAs to facilitate their enzymatic cleavage (Hammond, 2000; Zamore, 2000) (see also Munschauer/Vogel, Chapter 2). These discoveries then led to the development of synthetic siRNAs, which can silence specific genes in mammalian cells (Caplen et al., 2001; Elbashir et al., 2001). The associated RNAi technology soon became a ubiquitous tool in basic biology research, which in turn made it possible to analyze gene functions on a global scale for the first time.

However, it took almost 20 years for the first RNAi drug to receive approval. In 2018, Patisiran was approved for the treatment of polyneuropathy in patients with hereditary transthyretin amyloidosis (hATTR). This rare and fatal neurodegenerative disorder is caused by the accumulation of amyloid fibrils, which are formed by unstable transthyretin (TTR) pro-

¹³ Bifunctional compounds contain two functional groups, in this case the splicing activator and the ASO.

teins. Patisiran, a double-stranded siRNA composed of modified oligonucleotides, induces the decay of TTR mRNA in the liver, thereby decreasing serum level of the protein and reducing potential amyloid deposits. Vutrisiran, a successor to Patisiran, came on the market in June 2022. It uses the same RNAi mechanism but is based on enhanced stabilization chemistry. It is also coupled with the sugar derivative N-acetylgalactosamin (GalNAc),¹⁴ which increases the uptake of the siRNA in liver cells and allows the administration of lower doses. While Patisiran has to be injected intravenously every three weeks, treatment with Vutrisiran only involves one subcutaneous injection every three months.

In the meantime, five RNAi-based drugs have been approved by the US Food and Drug Administration (FDA). It is hoped that RNAi therapeutics can be used in an even broader spectrum of applications, provided that systemic administration and targeted delivery of the agent to non-liver and non-kidney tissues can be improved. Nevertheless, the 20-year story of the development of RNAi therapeutics is clear proof that, when it comes to developing programmable medicines that can be designed and adapted as needed, persistence pays off.

4.4 Therapeutic genome editing

The discovery of an RNA-guided immune defense system in bacteria and archaeobacteria¹⁵ is considered a turning point in RNA biology and genome research. Today, these CRISPR/Cas systems are widely known, also beyond the scientific community, primarily thanks to their considerable success as programmable genome-editing tools.¹⁶ Following the experimental proof that CRISPR/Cas systems provide adaptive immunity against foreign DNA (Barrangou et al., 2007) and elucidation of the underlying defense mechanism (Garneau et al., 2010; Deltcheva et al., 2011; Gasiunas et al., 2012; Jinek et al., 2012), it was not long before researchers realized the enormous potential that RNA-mediated Cas nucleases hold for genetic engineering. This concept was then established in a series of studies that showed that Cas9-mediated genome editing, i.e. the targeted manipulation of DNA sequences, is possible in both mouse and human cells and even in complex model organisms (Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013). This new technology revolutionized basic biomedical research at an unprecedented pace and, like RNA interference, attracted considerable interest from the pharmaceutical industry with regard to potential clinical applications.

Therapeutic approaches that use genome editing to treat hereditary diseases are currently in clinic trials, with some recording considerable success. For example, CRISPR/Cas gene therapy has achieved promising initial results in the treatment of sickle cell anemia. Sickle cell anemia is an inherited blood disorder in which red blood cells take on a sickle-type

¹⁴ N-acetylgalactosamin is a derivative of galactose and binds to the asialoglycoprotein receptor, which is selectively expressed on the surface of liver cells. GalNAc is used to deliver oligonucleotides to the liver.

¹⁵ Archaea are small, single-cell microorganisms. In addition to bacteria and eukaryotes, they constitute one of the three domains of life. Like bacteria, they are prokaryotes, i.e. they lack a nucleus.

¹⁶ CRISPR/Cas editing tools consist of a variant of the Cas enzyme, which precisely cuts the cellular DNA, and a guide RNA, which leads the Cas enzyme to the correct genomic site through complementary base pairing.

shape, which can result in occluded blood vessels, severe pain and life-threatening organ failure. This disorder occurs in people with two abnormal copies of the *β-globin* gene. *β-globin* is a protein required for the formation of hemoglobin in blood cells and therefore for the transport of oxygen within the body. Although the disorder can be treated by means of a bone marrow transplant, such treatment always requires a suitable stem cell donor. Such donors are rare and cannot always be found. In addition, bone marrow transplantations are fraught with risks.

The new approach to healing sickle cell anemia through RNA-based gene therapy involves removing a patient's own blood stem cells in order to use the CRISPR/Cas "genetic scissors" to revert the underlying mutation in the *β-globin* gene. An alternative method aims to reactivate the synthesis of fetal hemoglobin. The gene for fetal hemoglobin is usually intact in sickle cell anemia patients. The protein coded by this gene could replace the abnormal *β-globin*, yet the synthesis of fetal hemoglobin is normally deactivated shortly after birth. Following correction using either method, the stem cells are retransplanted into the patient, where they then produce healthy red blood cells. Initial studies indicate that this approach can improve the quality of life of seriously ill patients.

In the case of sickle cell anemia and other blood disorders, cells extracted from patients can be processed in the lab using CRISPR/Cas (an approach also referred to as *ex vivo*). In contrast, the majority of genetic diseases will require gene editing within the body (*in vivo*). One such example is Duchenne muscular dystrophy (DMD), a disorder characterized by progressive and irreversible muscle wasting. This condition, which remains incurable, manifests itself during childhood, progresses slowly and has a significant impact on life expectancy. DMD is caused by mutations in the *dystrophin* gene, which codes for an important scaffold protein in muscle cells. Unlike other therapies developed to date, which are only able to slow the progress of the disorder, using gene editing to correct the defective *dystrophin* gene promises to restore the missing dystrophin protein with lasting effect. It has already been shown that restoring a small percentage (around 15 %) of normal cellular dystrophin would provide a clinical benefit. There are still a number of scientific and technical hurdles to overcome, however, before these therapies reach the clinic. These include the ability to successfully introduce the CRISPR/Cas genetic scissors into target tissue, ensuring precise and efficient genome editing in specific cell types, controlling possible side effects caused by making changes in other parts of the genome and the immunogenicity¹⁷ of bacterial Cas proteins. Nevertheless, CRISPR/Cas technology offers considerable hope for patients with conditions currently considered incurable.

4.5 mRNA-based vaccines

In 2020, the COVID-19 pandemic catalyzed the fastest vaccine development in history. The mRNA vaccines against the SARS-CoV-2 virus, for which BioNTech/Pfizer and Moderna received approval within just 10 months, once again aroused significant interest in the use of mRNA for prophylactic as well as therapeutic purposes. Yet, this tremendous success was

¹⁷ Immunogenicity is the property of a substance to trigger a reaction in the body's immune system.

based on decades of painstaking basic research, which often involved contending with considerable resistance.

The first concepts for mRNA-based vaccines were created more than 30 years ago in the hope of developing safe, versatile vaccines – that were easy to produce as well. mRNA vaccines have a number of inherent advantages over conventional vaccines. Unlike with certain viral DNA vaccines, there is no risk of integration of the mRNA into the cellular genome, which might lead to undesirable side-effects. Furthermore, mRNA vaccines can be synthesized in the lab, thereby enabling cost efficient and scalable production. In principle, a single mRNA vaccine can encode multiple antigens,¹⁸ similar to conventional inactivated vaccines, and thereby facilitate an immune response against various viral variants with a single formulation. This is the case, for example, in the newly adapted vaccines targeting the Omicron variant of SARS-CoV-2. The concept of using mRNA as a potential therapeutic was originally met with a great deal of skepticism due to concerns regarding its poor stability in the body and its potentially dangerous immunostimulation. We have a number of persistent researchers to thank for the fact that, in the last 10 years, interest in clinical applications has risen again due to an improved understanding of mRNA pharmacology, the development of effective RNA formulations and the control of mRNA immunogenicity.

As a therapeutic molecule, mRNA is large and negatively charged, making it extremely difficult for it to cross cell membranes. Furthermore, mRNA is very unstable and is quickly degraded in human tissue by nucleases. A series of innovative mRNA transport vehicles have been developed to facilitate the targeted delivery and efficient cellular uptake of mRNA in patients, including lipid and polymer nanoparticles. Once it reaches the cellular cytosol, unmodified RNA – like the RNA in viruses – is identified by our innate immune system, which can trigger an excessive immune response. This problem has largely been solved by integrating modified nucleosides, especially modified uridine, into synthetic mRNAs. In fact, mRNA modified in this way also exhibits increased stability and improved translation into protein (Karikó et al., 2008) – a ground-breaking development.

The mRNA vaccines against SARS-CoV-2 are by no means the first of their kind. Even before the COVID-19 pandemic, a series of pre-clinical and clinical studies had shown promising results with mRNA vaccines developed to protect against various pathogens, including the Zika virus, Ebola virus, dengue virus, papilloma virus and the virus that causes rabies. Another important area in which mRNA vaccines are expected to make significant progress is cancer treatment. In several ongoing clinical studies, researchers are seeking to activate an immune response in patients against specific neoantigens¹⁹ expressed on their respective tumor. These neoantigens are determined before treatment in order to produce personalized mRNA vaccines that instruct the patient's immune system to attack their own body's tumor cells more aggressively. For many tumors, however, the limited ability of immune cells to access the tumor represents a more substantial barrier for mRNA

¹⁸ An antigen is a foreign protein that triggers the production of antibodies in the body.

¹⁹ Neoantigens are newly occurring antigens. They can be caused by DNA mutations that arise during cancer development and are therefore characteristic for tumor cells.

vaccines than fighting a pathogen in the blood or in tissue. Nevertheless, mRNA vaccines and mRNA-based drugs, which can replace a missing protein vital to metabolic processes, hold tremendous potential and promise to open up an entirely new field of medicine.

4.6 Outlook

The notion that the role of cellular RNA is predominantly to code for proteins has been outdated for some time. After decades of research in the field of RNA biology and a number of ground-breaking technological advances, RNA therapies have now become a reality. The mRNA vaccines developed and approved within just 10 months provided a strong push forward in this area. We can expect to see new therapeutics from all areas of RNA medicine in the near future. Ultimately, its potential and the possible benefits for patients are too great to ignore.

4.7 References

- Barrangou, R. et al. (2007):** CRISPR provides acquired resistance against viruses in prokaryotes. In: *Science* 315: 1709–1712.
- Caplen, N. J. et al. (2001):** Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. In: *Proceedings of the National Academy of Sciences of the United States of America* 98: 9742–9747.
- Cartegni, L./Kraimer, A. R. (2003):** Correction of disease-associated exon skipping by synthetic exon-specific activators. In: *Nature structural biology* 10: 120–125.
- Cong, L. et al. (2013):** Multiplex genome engineering using CRISPR/Cas systems. In: *Science* 339: 819–823.
- Deltcheva, E. et al. (2011):** CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. In: *Nature* 471: 602–607.
- Elbashir, S. M. et al. (2001):** Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. In: *Nature* 411: 494–498.
- Fire, A. et al. (1998):** Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. In: *Nature* 391: 806–811.
- Garneau, J. E. et al. (2010):** The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. In: *Nature* 468: 67–71.
- Gasiunas, G. et al. (2012):** Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. In: *Proceedings of the National Academy of Sciences of the United States of America* 109: E2579–E2586.
- Hammond, S. M. et al. (2000):** An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. In: *Nature* 404: 293–296.
- Jinek, M. et al. (2012):** A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. In: *Science* 337: 816–821.
- Jinek, M. et al. (2013):** RNA-programmed genome editing in human cells. In: *eLife* 2: e00471.
- Jirikowski, G. F. et al. (1992):** Reversal of diabetes insipidus in Brattleboro rats: intrahypothalamic injection of vasopressin mRNA. In: *Science* 255: 996–998.
- Karikó, K. et al. (2005):** Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. In: *Immunity* 23: 165–175.
- Karikó, K. et al. (2008):** Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. In: *Molecular therapy: the journal of the American Society of Gene Therapy* 16: 1833–1840.
- Mali, P. et al. (2013):** RNA-guided human genome engineering via Cas9. In: *Science* 339: 823–826.
- Mandl, C. W. et al. (1998):** In vitro-synthesized infectious RNA as an attenuated live vaccine in a flavivirus model. In: *Nature medicine* 4: 1438–1440.
- Stephenson, M. L./Zamecnik, P. C. (1978):** Inhibition of Rous sarcoma viral RNA translation by a specific oligodeoxyribonucleotide. In: *Proceedings of the National Academy of Sciences of the United States of America* 75: 285–288.

Zamecnik, P. C./Stephenson, M. L. (1978): Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. In: Proceedings of the National Academy of Sciences of the United States of America 75: 280–284.

Zamore, P. D. et al. (2000): RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. In: Cell 101: 25–33.

Zhou, W. Z. et al. (1999): RNA melanoma vaccine: induction of antitumor immunity by human glycoprotein 100 mRNA immunization. In: Human gene therapy 10: 2719–2724.

5. Outlook: Progress and questions

Daniela Remus

5.1 Current status of RNA research

In 2002, the world-renowned scientific journal *Science* named RNA the “molecule of the year”. At that time, researchers had published several papers showing that every cell has a large number of different types of RNA and that the latter can do much more than just convert the genetic information in DNA into proteins. But that was just the beginning. Since then, knowledge of these elementary building blocks of life has been growing at a rapid pace. And it is becoming increasingly clear that the various types of RNA perform extremely complex and different functions in cells.

As Mathias Munschauer and Jörg Vogel have shown in Chapter 2, some RNAs carry genetic information, others interact with biomolecules and metabolites; RNAs trigger cellular developments or accelerate chemical processes.

Scientists still have a long way to go before they can understand all functions of these vital molecules. Indeed, for some RNAs, it is not clear yet why they exist and what physiological relevance they have. Also, it can be assumed that researchers have to date not been able to identify all RNAs. Nevertheless, scientists are optimistic: the knowledge of RNAs will not only enable further significant findings in basic research, but will also fundamentally change the development of therapies and diagnostics in general.

In Chapter 3, Nina Gasparoni and Jörn Walter explained how RNAs already serve as diagnostic tools in research and clinical practice – especially in cancer diagnostics, in immunology and in infection research. Using special sequencing methods for single-cell RNA analyses, it is possible to distinguish diseased cells from healthy ones on a previously unknown level of discrimination. And these methods are also extremely useful in identifying pathogens and viruses. With their help, for example, it has been possible to detect the SARS-CoV-2 virus in early 2020, close in time to the first known cases of the disease in Wuhan, China. Only by accurately identifying and describing this virus could researchers subsequently develop a vaccine against it.

In addition, RNA-based single-cell analysis makes it possible to detect differences in genetic makeup within given tissues, including, most importantly, tumors. The researchers

therefore expect that in the future RNA diagnostics will enrich and supplement pathological findings in a standardized manner and thus enable more individualized treatments than before, personalized therapies, in just a few years.

Researchers have been working since the 1970s on using knowledge of the function of different RNAs to develop new therapeutics, as Anke Sparmann and Jörg Vogel have shown in Chapter 4. The first RNA-based drugs have been on the market for several years now, and clinical trials are underway for others.

Thanks to the development of the mRNA vaccines against the SARS-CoV-2 virus, the huge potential of RNA technology for drug development has become common knowledge. This was an enormous breakthrough, not only for the RNA pioneers, who had to stand their ground against skeptical objections in the scientific community for a long time, but also for the entire research field. In view of the success of mRNA vaccines, even the big pharmaceutical companies, which had previously withdrawn from RNA research, have now returned to the field and are investing large amounts of money in this area. It is believed, that RNA technology will very soon change many therapeutics fundamentally, as it will enable much more precise and targeted treatments, whether against tumors, infectious diseases or genetic disorders. And since it is known that each person's genetic makeup is highly individual, RNA technology can contribute tremendously to making personalized medicine a reality in the near future.

5.2 Scientific expectations and social acceptance

There is no doubt that the potential of this field of research is enormous. But the fact that mRNA vaccines were developed in record time has also caused worries, not everyone has been enthusiastic about this technology and embraces it as an opportunity. The introduction of mRNA vaccines thus also serves an example of the huge gap between current molecular biology and biotechnology research on one hand, and the knowledge of today's society on the other hand. What is RNA? What significance can it have in cell biology? Very few people outside of the research community can answer these questions adequately. And this can be unsettling, even for those who do not adhere to any conspiracy ideologies, simply due to a lack of biological knowledge needed to classify this new technology. Indeed, quite a number of people worry whether the researchers have actually taken all potential risk factors into account to avoid harm – in the short as well as long-term. And especially in the case of Covid vaccinations, it has become apparent that fears of the unfamiliar vaccine were aligning with a general mistrust of “the” pharmaceutical industry that has been assumed to be putting profit maximization before health.

The skepticism towards innovative scientific discovery is not new, neither should it be condemned a priori, but it can lead to science and society drifting apart. Social groups, not interested in understanding science, but rather in pursuing and pushing their political agenda can abuse this fear, as could clearly be observed during the Covid pandemic.

Many gene technology research areas are almost detached from knowledge in society: Be

it gene therapy interventions with gene scissors CRISPR/Cas, embryonic research or regenerative medicine. This causes uncertainty, resistance and concerns amongst the general public, and leads to beliefs such as the current suspicion that mRNA vaccines could permanently alter ones' genomes.

In addition, aspects of RNA technology, as already known from other areas of genetic engineering, raise ethical follow-up questions which are not trivial. For example – what happens to those findings that are not explicitly sought after, but which are supplied as by-products of sequencing? Although the Genetic Diagnostics Act governs the treatment of chance findings through the right to ignorance (informational self-determination), critics nevertheless object that not every possible chance finding has been specified in advance and can thereby be excluded. In other words, findings may be revealed, where it is not clear whether the person in question would want to know them as they would entail serious consequences for their lifestyle. And conversely, if a person would like to be informed of any chance findings, what information should be passed on in that case? All of it? Especially, as the genetic status of a person does not necessarily predict the manifestation of certain diseases during life. And does the medical urgency in some cases not override the individual's right to ignorance?

In addition, what is to happen with this rapidly growing knowledge? How safe is the data storage, and is it virtually certain that unauthorized access and any potential misuse of the data can be ruled out?

Furthermore, the handling of RNA-based gene scissors CRISPR/Cas vividly shows how fragile international consent can be and how quickly ethical boundaries are crossed. Despite the agreement amongst researchers and scientists to refrain from germline interventions, after their discovery it took only six years for Chinese scientists to use the CRISPR/Cas gene scissors to modify the germline of embryos resulting in the birth of the first gene-edited babies.

These examples highlight the need for education and knowledge transfer by researchers. Broad societal discussions are necessary to gain clarity on how these new, potent tools should be handled in the future. The Working Group *Gene Technology Report* contributes to this objective with this brochure.

Diese Bestandsaufnahme ist online abrufbar unter:
<https://www.gentechnologiebericht.de/publikationen>

This stocktaking is available online at:
<https://www.gentechnologiebericht.de/en/publications>

ISBN: 978-3-00-074152-4