



**Jörn Walter, Hannah Schickl (Hrsg.)**

---

## **Einzelzellanalyse in Forschung und Medizin : eine Stellungnahme der interdisziplinären Arbeitsgruppe *Gentechnologiebericht***

Berlin-Brandenburgische Akademie der Wissenschaften, 2019  
ISBN: 978-3-939818-84-7

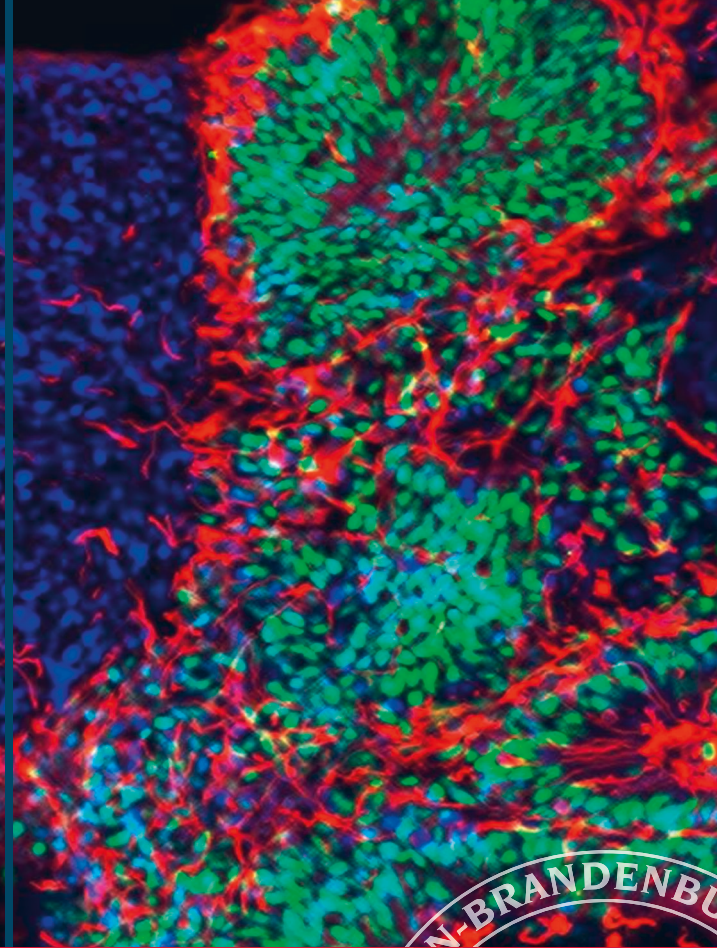
Persistent Identifier: [urn:nbn:de:kobv:b4-opus4-32789](https://nbn-resolving.org/urn:nbn:de:kobv:b4-opus4-32789)

---

Die vorliegende Datei wird Ihnen von der Berlin-Brandenburgischen Akademie der Wissenschaften unter einer Creative Commons Namensnennung 4.0 International Lizenz zur Verfügung gestellt.



Jörn Walter und  
Hannah Schickl  
(Herausgeber)



# EINZELZELLANALYSE IN FORSCHUNG UND MEDIZIN

Eine Stellungnahme der interdisziplinären Arbeitsgruppe  
*Gentechnologiebericht*

BERLIN-BRANDENBURGISCHE AKADEMIE DER WISSENSCHAFTEN





Berlin-Brandenburgische Akademie der Wissenschaften (BBAW)

EINZELZELLANALYSE IN FORSCHUNG UND MEDIZIN

Eine Stellungnahme der interdisziplinären

Arbeitsgruppe *Gentechnologiebericht*





**EINZELZELLANALYSE  
IN FORSCHUNG UND MEDIZIN**  
EINE STELLUNGNAHME DER INTERDISZIPLINÄREN  
ARBEITSGRUPPE *GENTECHNOLOGIEBERICHT*

---

Jörn Walter und Hannah Schickl  
Herausgeber

Herausgeber: Jörn Walter und Hannah Schickl für die interdisziplinäre Arbeitsgruppe *Gentechnologiebericht*, ein Drittmittelprojekt der Berlin-Brandenburgischen Akademie der Wissenschaften

Übersetzung: Übersetzungsagentur24

Grafik: angenehme gestaltung/Thorsten Probst

Coverfoto: Agnieszka Rybak-Wolf

Druck: PIEREG Druckcenter Berlin GmbH

© Berlin-Brandenburgische Akademie der Wissenschaften, 2019

Jägerstraße 22–23, 10117 Berlin, [www.bbaw.de](http://www.bbaw.de)

Die Publikation erscheint mit Unterstützung der Friede Springer Stiftung und des vom Bundesministerium für Bildung und Forschung geförderten Netzwerkes *Single Cell Omics Germany* (SCOG).

friede springer stiftung

Nachdruck, auch auszugsweise, nur mit ausdrücklicher Genehmigung der Herausgeber.

ISBN: 978-3-939818-84-7

## INHALTSVERZEICHNIS

Vorwort .....	8
<b>1. Einleitung (Jörn Walter, Nina Gasparoni) .....</b>	<b>10</b>
1.1 Vom komplexen Gewebe zur Einzelzellsignatur – neue Horizonte für die moderne Zellbiologie .....	10
1.2 Technologische Entwicklungen und die Entstehung der modernen Einzelzellbiologie .....	11
1.3 Initiativen zur Entwicklung des Bereichs der Einzelzell-Omics .....	13
1.4 Aktuelle Entwicklungen im Bereich der Einzelzell-Omics .....	14
1.5 Einzelzellanalysen und Entwicklungsbiologie .....	16
1.6 Einzelzellanalysen in der Krankheitsforschung .....	17
1.7 Herausforderungen und Grenzen der Einzelzelltechnologien .....	18
1.8 Schlussbemerkung .....	19
1.9 Literatur .....	20
<b>2. Einzelzellgenomik verändert die Entwicklungsbiologie (J. Philipp Junker, Christian Popp, Nikolaus Rajewsky) .....</b>	<b>23</b>
2.1 Einführung .....	23
2.2 Störungsanalyse in Entwicklungsmodellssystemen .....	24
2.3 Räumliche Informationen .....	26
2.4 Zeitliche Informationen – pseudotemporale Ordnung .....	27
2.5 Zeitliche Informationen – Abstammungsanalyse mit hohem Durchsatz .....	29
2.6 Messung anderer Parameter als RNA .....	30
2.7 LifeTime: eine neue Initiative basierend auf Einzelzellanalyse .....	32
2.8 Literatur .....	34



<b>3. Einzelzell-Omics in der Biomedizin gestern, heute und morgen</b> (Anna C. Aschenbrenner, Elvira Mass, Joachim L. Schultze) .....	38
3.1 Neue Wege der Einzelzell-Omics für eine neue Wahrnehmung und Behandlung verbreiteter Erkrankungen.....	38
3.2 Einzelzell-Omics gestern .....	38
3.3 Einzelzell-Omics heute .....	40
3.4 Einzelzell-Omics morgen .....	42
3.5 Literatur .....	45
<b>4. Analyse von Einzelzellgenomik-Daten mit Methoden des maschinellen Lernens</b> (Hananeh Aliee, Anna Sacher, Fabian J. Theis).....	48
4.1 Einführung .....	48
4.2 Maschinelles Lernen in der Einzelzell-Transkriptomik .....	50
4.3 Ausblick.....	54
4.4 Literatur .....	54
<b>5. Einzelzell-Transkriptomanalyse in Pflanzen</b> (Bernd Müller-Röber) .....	55
5.1 Pflanzliche Einzelzell-Transkriptome .....	55
5.2 Transkriptomanalysen mittels isolierter Zellkerne.....	56
5.3 Zukünftige Forschung .....	57
5.4 Fazit .....	58
5.5 Literatur .....	59
<b>6. Einzelzellanalysen und Überlegungen zur Ethik</b> (Heiner Fangerau, Lilian Marx-Stölting, Angela Osterheider) .....	60
6.1 Einführung .....	60
6.2 Ethische Themenfelder.....	61
6.3 Schluss .....	64
6.4 Literatur .....	64

<b>7. Problemfelder und Indikatoren zum Thema Einzelzellanalyse</b> (Angela Osterheider, Yaroslav Koshelev, Marlen Reinschke, Lilian Marx-Stölting) .....	66
7.1 Einführung: Motivation und Zielsetzung .....	66
7.2 Problemfelder .....	66
7.3 Indikatoren .....	70
7.4 Zusammenfassung .....	75
7.5 Literatur .....	76
<b>8. Kernaussagen und Handlungsempfehlungen zur Einzelzellanalytik</b> (Interdisziplinäre Arbeitsgruppe <i>Gentechnologiebericht</i> ) .....	77
8.1 Kernaussagen zur Einzelzellanalytik .....	77
8.2 Handlungsempfehlungen für den Umgang mit der Einzelzelltechnologie und Einzelzelldaten .....	82
<b>9. Anhang</b>	
9.1 Autorinnen und Autoren .....	84
9.2 Mitglieder der interdisziplinären Arbeitsgruppe <i>Gentechnologiebericht</i> .....	87
9.3 Publikationen der interdisziplinären Arbeitsgruppe <i>Gentechnologiebericht</i> .....	88

## VORWORT

Die interdisziplinäre Arbeitsgruppe (IAG) *Gentechnologiebericht* an der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) beobachtet und begleitet seit 2001 neue Entwicklungen der Gentechnologien und ihre Relevanz für Wissenschaft und Gesellschaft. Zu ihren Aufgaben an der Schnittstelle zwischen Wissenschaft, Politik, Wirtschaft und Öffentlichkeit gehört es, aktuelle Themen möglichst frühzeitig aufzugreifen und umfassend zu beleuchten, um so einen objektiven öffentlichen Diskurs anzustoßen bzw. zu fördern.

Bei Einzelzellanalysen handelt es sich um eine Vielzahl von Analysemethoden, denen gemeinsam ist, dass sie nicht wie bisher nur an Zellverbänden, Gruppen von Zellen, Geweben und Organen, sondern auch an einzelnen Zellen durchgeführt werden. Das Feld birgt großes Potenzial nicht nur für die Grundlagenforschung, sondern auch für medizinische und biotechnologische Anwendungen, da neue Ebenen kontext- und personenbezogener Interpretation biologischer Zusammenhänge erschlossen werden. Die vorliegende Broschüre zur Einzelzellanalyse gibt einen Überblick über die neuen Möglichkeiten aus den Perspektiven der Entwicklungsbiologie, Biomedizin und Bioinformatik und greift mögliche gesellschaftliche Implikationen und Konsequenzen auf.

Namentlich gekennzeichnete Beiträge geben nicht unbedingt die Meinung der Herausgeber oder der Arbeitsgruppe wieder. Die IAG verantwortet jedoch gemeinsam das Kapitel „Kernaussagen und Handlungsempfehlungen zur Einzelzellanalytik“. Die vorgestellten Handlungsempfehlungen wurden gemeinsam von den Mitgliedern der IAG erarbeitet, werden aber nicht notwendigerweise von allen Mitgliedern der Akademie vertreten; die BBAW steht jedoch uneingeschränkt hinter der Qualität der geleisteten Arbeit.

Ein herzlicher Dank gilt der Friede Springer Stiftung für die Förderung der Arbeit der IAG an der BBAW. Dank gebührt auch den Autorinnen und Autoren der Beiträge sowie dem Herausgaberteam und der Geschäftsstelle der IAG.

Diese Broschüre wurde auf Initiative der IAG *Gentechnologiebericht* an der BBAW erstellt. Wir freuen uns sehr, dass dies in enger Zusammenarbeit mit dem Netzwerk *Single Cell Omics Germany* stattgefunden hat.

*Boris Fehse*

Sprecher der interdisziplinären Arbeitsgruppe *Gentechnologiebericht*  
an der Berlin-Brandenburgischen Akademie der Wissenschaften

Hamburg, im August 2019

## 1. EINLEITUNG

### 1.1 VOM KOMPLEXEN GEWEBE ZUR EINZELZELLSIGNATUR – NEUE HORIZONTE FÜR DIE MODERNE ZELLBIOLOGIE

Seit dem Aufkommen der modernen Zellbiologie Anfang des 20. Jahrhunderts suchen Wissenschaftler nach Techniken zur Erfassung der molekularen Mechanismen, die biologische Programme in einzelnen Zellen des Organismus steuern. Mit der Einzelzellanalytik steht der Wissenschaft seit Kurzem ein Mittel zur Verfügung, das umfassende und präzise Daten über den molekularen Charakter und die Funktionsweise *einzelner* Zellen generiert. Die Interpretation dieser Daten eröffnet der Forschung tiefgreifende neue Möglichkeiten des Verständnisses komplexer biologischer Vorgänge in einer Zelle, von komplexen Entwicklungsprozessen und Alterung bis hin zur Anpassung an Umweltbedingungen, und von komplexen Vorgängen der Organentwicklung bis hin zu Ursachen und Folgen von Erkrankungen. Mithilfe der neuen Technologien lassen sich diese Prozesse zellgenau erfassen und zwar simultan für tausende bis zu Millionen einzelner Zellen. Neue Verfahren ermöglichen es, die so gewonnenen Daten für die Modellierung der räumlichen Zuordnung der Zelle im Gewebe und in ihrer entwicklungsbiologischen Dynamik zu nutzen. Damit bringt die Einzelzellanalytik Biologen ihrem ursprünglichen Ziel, Eigenschaften und Funktionsweisen von Zellen im Organismus präzise zu verorten, zu verstehen und zu beeinflussen, erheblich näher.

Bisherige funktionelle Konzepte zellulärer Programme gründen sich auf genetische, biochemische und molekulare Daten. Vor den Möglichkeiten der Einzelzellanalyse mussten umfassende Untersuchungen an Zellgemischen oder vermeintlich homogenen Zellpopulationen durchgeführt werden, die zuvor aus Geweben oder Körperflüssigkeiten isoliert wurden. Die molekularen Signaturen (wie bspw. Genexpressionsprofile) solcher Zellpopulationen spiegeln jedoch immer die Summe der Eigenschaften der einzelnen Zellen wider und erlauben es nicht, die individuelle Variations- und Funktionsbreite einzelner Zellen zu erfassen. Veränderungen, die während der Entwicklung, im Verlauf des Zellzyklus und bei Prozessen des individuellen Zellalters auftreten oder die als Auswirkungen ihrer räumlichen Anordnung entstehen, sind mit den herkömmlichen Methoden nur bedingt analysierbar. Derartige individuelle Eigenschaften lassen sich durch die Analyse von

Zellpopulationen nicht adäquat bestimmen. Hinzu kommt das Problem, dass die Analyse von Zellpopulationen einer Selektion oder Vorsortierung nach bestimmten Eigenschaften bedarf, sodass nicht alle Zellen eines Gewebes gleichzeitig erfasst werden können.

Mit der Entwicklung umfassender Einzelzell-Omics<sup>1</sup>-Technologien bieten sich nun tiefgreifende Lösungen für viele dieser Probleme. In Kombination mit neuen verzerrungsfreien Sortiertechniken, erweiterten Mikroskopie-Verfahren wie Multi-RNA-FISH<sup>2</sup> und neuartigen bioinformatischen Ansätzen werden Einzelzellanalysen Antworten auf bislang unlösbare Fragen liefern und neue systemische Einblicke in die Funktion einzelner Zellen in einer komplexen biologischen Umgebung eröffnen.

## 1.2 TECHNOLOGISCHE ENTWICKLUNGEN UND DIE ENTSTEHUNG DER MODERNEN EINZELZELLBIOLOGIE

Die Grundlage der modernen „Einzelzell-Omics“ wurde durch die funktionelle Annotation des menschlichen Genoms und der Genome aller wichtigen Modellorganismen gelegt. Die Lokalisierung von Genen und anderen funktionellen bzw. regulatorischen Teilen des Genoms hat zahlreiche Studien zur funktionellen Genomik beflügelt, die zum Ziel haben, verschiedenen Zelltypen ihre spezifischen molekularen Programme zuzuordnen. Diese „funktionelle Genomrevolution“ wurde durch die schnelle Entwicklung neuartiger technologischer Fortschritte im Bereich der Hochdurchsatz-Sequenzierung ermöglicht, den sogenannten Next-Generation-Sequencing-Technologien (NGS-Technologien). NGS-Methoden, die ursprünglich für die Genomsequenzierung entwickelt worden sind, wurden schnell für Funktionsanalysen von Zellen nutzbar gemacht, zum Beispiel für die umfassende Erstellung von Genexpressionsprofilen (zur Darstellung der jeweils in einer Zelle exprimierten Gene) mit NGS-basierten RNA-Sequenzierungsmethoden (RNA-seq)<sup>3</sup>.

- 1 „Omics“ ist ein Neologismus, der verschiedene Forschungsfelder in den Lebenswissenschaften beschreibt, die die Wortendung „omics“ enthalten, z. B. Genomik, Transkriptomik, Metabolomik und Proteomik. Die Wortendung zeigt an, dass der Fokus der Studien auf dem gesamten zellulären Gehalt an untersuchten Molekülen liegt (d. h. auf der Gesamtheit der Gene, Gentranskripte, Stoffwechselprodukte oder Proteine der Zelle).
- 2 FISH (Fluoreszenz-in-situ-Hybridisierung) ist eine Methode, bei der bestimmte Moleküle einer Probe mit fluoreszierenden Markern versehen werden, um sie nachweisen zu können. „In situ“ bringt zum Ausdruck, dass die Moleküle in ihrer natürlichen Position (also in der Zelle, im Gewebe) untersucht werden können.
- 3 Die RNA-Sequenzierung nutzt NGS, um die Anzahl und Anwesenheit von RNA-Transkripten in einer Probe zum Zeitpunkt der Untersuchung nachzuweisen (auch „Gesamttranskriptom-Shotgun-Sequenzierung“ genannt). Da diese Methode genutzt wird, um das zelluläre Transkriptom zu analysieren, gehört sie zur Transkriptomik.

Vor etwa 10 Jahren wurden die ersten manuellen Versuche unternommen, umfassende mRNA<sup>4</sup>-seq-Signaturen aus einigen wenigen sortierten einzelnen Säugetierzellen zu erhalten (Tang et al., 2009). Diese ersten erfolgreichen Anwendungen haben das neue Forschungsgebiet beflügelt und sehr schnell wurden neue Hochdurchsatzmethoden für die Einzelzellisolierung und die NGS-Prozessierung entwickelt, um umfassende RNA-seq-basierte Genexpressionsprofile für *vieler* Einzelzellen auf einmal zu erhalten. Alle diese neuartigen Techniken kombinieren Mikroprozessierungs-Techniken mit anspruchsvollen Protokollen zur Erstellung komplexer Sequenzier-Bibliotheken einzelner Zellen.<sup>5</sup> Die neuen Technologien erweitern die Forschungsmöglichkeiten in zwei wichtige Richtungen: i) eine immer umfassendere Darstellung molekularer Signaturen in einzelnen Zellen und ii) die Möglichkeit, eine große Anzahl von Einzelzellen in kostengünstigeren Hochdurchsatzsystemen gleichzeitig zu analysieren.<sup>6</sup>

Das Forschungsfeld begann mit der Erstellung einer Reihe von hochauflösenden Einzelzellkarten für Blutzellen, die neue tiefe Einblicke in die Entwicklung sowie die Anpassung der verschiedenen Immunzelltypen im Verlauf von Erkrankungen gewähren (Kowalczyk et al., 2015; Wilson et al., 2015). Bald darauf wurden die ersten umfassenden Einzelzellkarten von Modellorganismen erstellt, gefolgt von jüngsten Veröffentlichungen zu Geweben und Organen bei Menschen und Mäusen (Han et al., 2018), wie Leber (Halpern et al., 2017; MacParland et al., 2018, Aizarani et al., 2019), Gehirn (Darmanis et al., 2015; Lake et al., 2018; Rosenberg et al., 2018), Niere (Magella et al., 2018), Lunge (Treutlein et al., 2014; Xu et al.,

- 4 mRNA ist die Abkürzung für messenger- (deutsch „Boten-“) RNA und bezeichnet das Molekül, das Produkt der Genexpression ist. Im Zuge eines Prozesses, der „Transkription“ genannt wird, dient die DNA als Vorlage für den Aufbau der RNA, die wiederum zu mRNA prozessiert wird. Die mRNA verlässt den Zellkern und wird in eine Aminosäuresequenz übersetzt und bildet so ein Protein. mRNA ist daher das Transkript (Abschrift) der korrespondierenden DNA, und das Studium des RNA-Gehaltes einer Zelle wird Transkriptomik genannt. Das Transkriptom einer Zelle enthält ihre gesamte RNA und ermöglicht Rückschlüsse darauf, welche Gene in einer bestimmten Zelle (zu einem bestimmten Zeitpunkt) exprimiert sind.
- 5 Vom Transkriptom werden DNA-Kopien erstellt (mRNA wird in DNA umgeschrieben, sogenannte cDNA für komplementäre DNA), d. h. es wird eine „Bibliothek“ der in einer Zelle vorhandenen einzelnen mRNA Moleküle erstellt, die dann mit Hilfe von Hochdurchsatz-Sequenzierung (NGS) ausgelesen werden kann. Man bestimmt so das Vorhandensein und die Anzahl der mRNA-Kopien eines Gens.
- 6 Aktuelle RNA-seq-Methoden nutzen entweder Mikrofluidik-Systeme, wie die gängigsten „Drop-seq“-Methoden, oder sie verwenden „zellcontainerähnliche“ Lösungen in Mikrotiterplatten. In beiden Systemen werden die RNA-Bibliotheken einzelner Zellen mittels ausgefeilter individueller (Adapter-) Kodierungstechniken markiert, um die Expressionsprofile einzelner Zellen unterscheiden zu können. Einige der neuen Technologien wie „MARS-seq“ (Jaitin et al., 2014; Keren-Shaul et al., 2019), „Drop-seq“ (Macosko et al., 2015), „Seq-Well“ (Gierahn et al., 2017), „SPLiT-seq“ (Rosenberg et al., 2018) – um nur einige der aktuellen „Spitzenreiter“ zu nennen – haben einen hohen Durchsatz erreicht, der die Generierung von RNA-seq-Signaturen für Millionen von Zellen zu angemessenen Sequenzierungskosten ermöglicht.

2016), oder sogar zu ganzen Tieren (Drosophila, Embryonalstadien der Maus: Karaiskos et al., 2017; Mohammed et al., 2017). Diese ersten umfassenden Analysen von Geweben und sich entwickelnden Organismen zeigen, dass es neben der Identifizierung neuer (bisher nicht spezifizierter) Zelltypen oder Zellzustände möglich ist, Ähnlichkeiten und Veränderungen von Zellfunktionen über Gewebe hinweg zu identifizieren. So kann die Dynamik von Zellpopulationen, d. h. deren Auftreten bzw. Verschwinden, während der Entwicklung ebenso erfasst werden wie die Heterogenität von Zelltypen in erkrankten Geweben (z. B. Krebs) oder die Variationsbreite der Zellzusammensetzung in alternden Geweben – um nur einige der faszinierendsten Facetten zu nennen.

### 1.3 INITIATIVEN ZUR ENTWICKLUNG DES BEREICHS DER EINZELZELL-OMICS

Mit dem Aufkommen von Einzelzellmessungen wurde schnell klar, dass für vergleichende Analysen eine Art Standardisierung sowohl auf experimenteller als auch auf der Ebene der Dateninterpretation erforderlich sein wird. Internationale Forschungskonsortien wie der Human Cell Atlas (HCA)<sup>7</sup> und auch LifeTime<sup>8</sup> wurden gegründet, um das sich schnell entwickelnde Forschungsfeld in diesen Aufgaben voranzubringen durch die Bereitstellung von qualitativ hochwertigen Einzelzelldaten mit definierten Standards und Kontrollen in Referenzdatenbanken. Erste Datenbanken wurden für Maus, Mensch und Drosophila eingerichtet, von denen solche zellspezifischen Einzelzelldaten abgerufen werden können.<sup>9</sup> Der Human Cell Atlas war das erste internationale Konsortium, das 2017 mit dem Ziel gegründet wurde, einen umfassenden Einzelzellatlas aller menschlichen Zellen zu erstellen und neue „cloud“-basierte Informatiklösungen zur Datenspeicherung und -analyse zu entwickeln. Die im Jahr 2018 gestartete europäische Initiative LifeTime ergänzt diese Bemühungen, indem sie sich auf medizinische Anwendungen in verschiedenen krankheitsrelevanten Bereichen konzentriert. Ein Hauptziel ist die Entwicklung und Analyse krankheitsrelevanter Modelle sowie die Entwicklung neuartiger Ansätze, die auf die klinische Verwendung von Einzelzelldaten übertragen werden können. In Deutschland wurde darüber hinaus 2018 das vom BMBF geförderte Netzwerk Single Cell Omics Germany (SCOG)<sup>10</sup> mit dem Ziel gegründet, ein erstes Netzwerk von Laboren aufzubauen,

7 Siehe: <https://www.humancellatlas.org/> [13.08.2019].

8 Siehe: <https://lifetime-fetflagship.eu/> [13.08.2019]. Siehe auch: Junker, Popp, Rajewsky, Kapitel 2.

9 Eine umfassende Datenbankliste ist abrufbar unter: <https://www.singlecell.de/index.php/resources/databases/> [13.08.2019].

10 Siehe: <https://www.singlecell.de/> [16.08.2019].



die Einzelzellanalysen durchführen, und Weiterbildungen in Einzelzelltechnologien anzubieten, insbesondere in aufstrebenden Bereichen wie der Einzelzell-Multi-Omics, der umfassenden Datenanalyse und -interpretation.

Diese gemeinsamen Bemühungen sind darauf ausgerichtet, eine wissenschaftliche Gemeinschaft aufzubauen und zu fördern, die zusammen einen vollständigen Atlas aller Zelltypen des menschlichen Körpers in Einzelzellauflösung ausarbeitet, welcher auch komplexe Szenarien wie Gewebe, Organalterung und Krankheiten umfasst. Diese Daten sollen der Forschungsgemeinschaft frei zugänglich gemacht und neue informatische Ansätze für ihre tiefgehende biologische Interpretation entwickelt werden. Die Komplexität der Einzelzelldaten und die vielen neuen Fragestellungen, die mit dieser Art von hochauflösenden Daten beantwortet werden können, erfordern die Entwicklung neuer bioinformatischer Ansätze, die weit über die für NGS-Daten entwickelten Anwendungen hinausgehen. Dieser Aspekt wird in Kapitel 4 von Aliee, Sacher und Theis näher erläutert.

#### 1.4 AKTUELLE ENTWICKLUNGEN IM BEREICH DER EINZELZELL-OMICS

Gegenwärtig konzentriert sich die Mehrzahl der Einzelzell-Omics-Messungen auf RNA-seq, wobei hauptsächlich Expressionssignaturen des letzten Exons<sup>11</sup> eines Gens erfasst werden (z. B. Chromium (Zheng et al., 2017), Drop-seq). Solche Ansätze sind recht robust und eignen sich gut für die Erzeugung von groben Zellsignaturen, die eine Unterscheidung von Hauptzelltypen ermöglicht. Allerdings erlauben sie nicht die Untersuchung komplexerer Veränderungen in genregulatorischen Programmen, wie alternative Transkript- oder Spleiß-Varianten von Genen, die häufig eine andere und wichtige funktionelle Rolle spielen. Um das Spektrum der Gentranskription in größerer Tiefe zu erfassen, werden umfassendere RNA-seq-Methoden zur Herstellung von Bibliotheken entwickelt, wie zum Beispiel Smart-seq2 (Picelli et al., 2013) und andere (Chen et al., 2019). Solche Ansätze sind zwar noch recht kostspielig, werden aber durch sinkende Sequenzierungskosten immer mehr zur Routine, da die Fülle solcher Daten eine bessere und tiefergehende Interpretation ermöglicht.

Die Erzeugung hochauflösender Signaturen einzelner Zellen geht zu Lasten der räumlichen Orientierung der einzelnen Zellen. Das bedeutet, dass die Forscher

<sup>11</sup> Exons sind Teile des ursprünglichen RNA-Transkripts, die die mRNA infolge eines als „Spleißen“ (Splicing) bezeichneten Prozesses bilden. Die Detektion des letzten Exons dient als Nachweis dafür, dass das Gen transkribiert worden ist.

nicht wissen, welche Umgebung und Position die Zelle innerhalb des Gewebes hatte. Dieses Wissen ist jedoch wichtig, um die Einzelzelldaten richtig interpretieren und in das vorhandene Wissen über Gewebe und den gesamten Organismus integrieren zu können. Zur Lösung dieses Problems werden derzeit Methoden entwickelt, die eine räumliche Anordnung der analysierten Einzelzellen in „virtuelle Gewebe“ erlauben. Die wichtigsten gängigen Verfahren verwenden hochauflösende Bildgebungsverfahren, um Expressionssignaturen in Geweben für eine ausreichende Anzahl von Genen und Zellen zu erstellen, zum Beispiel durch mehrfarbigen RNA-FISH. Derartige *in situ* quantifizierte Expressionsbilder von Genen in einzelnen Zellen von Gewebeschnitten nutzt man dann als „Ankerpunkte“ für eine räumliche Rekonstruktion der Einzelzell-RNA-seq-Signaturen, um anschließend aus den Einzelzelldaten eine Art „virtuell“ rekonstruiertes Gewebe zu erzeugen. Damit wissen die Forscher genau, wann und wo ein Gen in einer Zelle des Organismus angeschaltet ist. Auf diese Verfahren gehen Junker, Popp und Rajewsky in Kapitel 2 genauer ein. Ein alternativer Ansatz, um Ankerpunkte zu generieren, ist das Sammeln weniger oder einzelner Zellen aus definierten Regionen innerhalb von Gewebeschnitten mittels Laser-Capture-Mikroskopie, gefolgt von einer tiefen (Einzel-)Zellsequenzierung (Nichterwitz, 2016; Chen et al., 2017).

Die umfassende und mechanistische Interpretation von Einzelzell-Omics-Daten ist zudem angewiesen auf den Vergleich mit anderen (idealerweise ebenfalls Einzelzell-)Omics-Daten wie zum Beispiel genomischen und funktionellen Epigenomik-Daten (Chromatin und DNA-Modifikationen). Bisher wurden Referenzdaten zur Genomik und funktionellen Genomik nur für Zellgemische von Konsortien wie dem Internationalen Krebs-Genom-Konsortium (ICGC),<sup>12</sup> 4DNucleome,<sup>13</sup> der Enzyklopädie der DNA-Elemente (ENCODE),<sup>14</sup> dem Internationalen Human-Epigenom-Konsortium (IHEC)<sup>15</sup> und anderen generiert. Für einige Methoden, zum Beispiel ATAC-seq, wurden einzelzellbasierte Protokolle entwickelt (Buenrostro et al., 2015), die auch bereits kommerzialisiert sind. Die meisten der anderen genomischen und epigenomischen NGS-basierten Methoden sind jedoch technisch anspruchsvoll und auf Einzelzellebene nur bedingt anwendbar. Bis vor Kurzem erschien es unmöglich, die Genexpression einzelner Zellen direkt mit deren epigenetischen Profilen zu verknüpfen. Pilotversuche von Clark et al. (2018) haben nun gezeigt, dass auch Einzelzelldaten zur Genexpression,

12 Siehe: <https://icgc.org/> [13.08.2019].

13 Siehe: <https://www.4dnucleome.org/> [13.08.2019].

14 Siehe: <https://www.encodeproject.org/> [13.08.2019].

15 Siehe: <http://ihec-epigenomes.org/> [13.08.2019].

DNA-Methylierung und Chromatinzugänglichkeit<sup>16</sup> gleichzeitig von derselben Zelle erhalten werden können. Darüber hinaus wurden die ersten hochtechnischen Ansätze entwickelt, um die dreidimensionale Konfiguration von Chromosomen in Einzelzellen zu bestimmen. Man gewinnt daraus Erkenntnisse über die räumliche Anordnung von Genen im Zellkern und die Bedeutung für die Regulierung der Genaktivität (Nagano et al., 2017).

Die integrierte Interpretation solcher Multi-Omics-Einzelzelldaten stellt ein wichtiges Zukunftsfeld in der Einzelzellbiologie dar. Sie schlägt Brücken zwischen (deskriptiven) transkriptionellen Signaturen einzelner Zellen und den Mechanismen, wie diese Genprogramme festgelegt und ausgeführt werden. Funktionelle Multi-Omics-Daten ermöglichen die Betrachtung biomedizinischer Fragen mit einer nie zuvor erreichten Auflösung und ebnen damit den Weg für ein präzises Verständnis der Mechanismen der Genregulation. Die Komplexität der generierten Daten stellt die Einzelzell-Bioinformatik jedoch vor große Herausforderungen, da unterschiedliche Datentypen (mit unterschiedlichen Dynamikbereichen) integriert und analysiert werden müssen.

## 1.5 EINZELZELLANALYSEN UND ENTWICKLUNGSBIOLOGIE

Studien im Bereich der Entwicklungsbiologie werden in hohem Maße von der Verwendung der Einzelzell-(Multi-)Omics profitieren. Die Auflösung auf Einzelzellebene bietet einen neuartigen umfassenden Überblick über Änderungen des Zellprogramms und deren dynamische Anpassung während der Entwicklung. Es ist wahrscheinlich, dass Einzelzelldaten unsere derzeitige Sicht auf (stochastische und gerichtete) Mechanismen, die Differenzierungsprozesse vorantreiben, sowie möglicherweise auch unsere (eher statische und auf Vorwissen basierende) Sicht auf die Definition von Zelltypen verändern. Sie werden sicherlich auch unser Verständnis dafür verbessern, wie Zellen sich an veränderte Umweltbedingungen anpassen.

Für Planarien (Strudelwürmer; Cao et al., 2017), frühe Mausembryonen (Peng et al., 2016; Mohammed et al., 2017) und Drosophila-Larven (Taufliegenlarven;

<sup>16</sup> Chromatin ist ein Komplex aus DNA und begleitenden Proteinen. Dabei ist die DNA um die sogenannten Histon-Proteine gewickelt und in sich verdrillt. Der Grad der Verdrillung (Kondensation) beeinflusst die Zugänglichkeit des Chromatins für weitere Proteine, die etwa Gene aktivieren oder inaktivieren können. Die DNA-Methylierung ist eine biochemische Modifizierung der DNA, die sich auf das Bindeverhalten regulatorischer Proteine und auf die Chromatin-Konformation auswirkt und so die Genexpression beeinflusst.

Karaiskos et al., 2017) wurden erste eindrucksvolle Beispiele für „Karten des Entwicklungsschicksals“ erstellt. Diese Daten liefern Einblicke in dynamische Veränderungen zellulärer Programme, die in kurzen Phasen selbst-reorganisierender Prozesse auftreten (z. B. während der Gastrulation). Sie ermöglichen es, die Bildung und Organisation von Zellen während der Entwicklung zum Beispiel des Gehirns, der Niere, des Herzens usw. räumlich und zeitlich zu verfolgen und die vorprogrammierten und stochastischen Mechanismen zu identifizieren, die die Diversifikation und Differenzierung der Zellen (z. B. bei der frühen Säugetierentwicklung) antreiben. Es wurden komplexe Rechenmodelle etabliert, um die Dynamik der Entwicklungstrajektorien von Zelllinien abzuleiten und zelluläre Veränderungsabläufe, die die Zelltypentwicklung festlegen, zu verfolgen. In Kombination mit genetischer Markierung oder einer Überlagerung mit mikroskopischen Referenzdaten ermöglichen solche hochauflösenden Omics-Daten ein tiefgehendes Verständnis der Mechanismen, die die räumliche und zeitliche Organisation von Entwicklungsübergängen in verschiedenen Organismen regulieren. Diese Aspekte werden im Kapitel 2 von Junker, Popp und Rajewsky sowie im Kapitel 4 von Aliee, Sacher und Theis näher erläutert.

## 1.6 EINZELZELLANALYSEN IN DER KRANKHEITSFORSCHUNG

Die Zusammensetzung und relative Lokalisierung von Zellen in einem Gewebe bzw. Organ sind wichtige Parameter, um deren physiologische Funktionen in Zusammenhang mit der natürlichen Organfunktion, der Homöostase, dem Altern, der Regeneration, aber auch mit Krankheiten zu verstehen. Einzelzell-Omics bieten einen unvoreingenommenen<sup>17</sup> Ansatz, um den genauen Zusammenhang zwischen der Zellzusammensetzung und der Organbiologie zu untersuchen. Darüber hinaus lassen sich die Folgen einer lokalen Funktionsstörung von Zellen im Organ verfolgen, zum Beispiel bei Prozessen, die zu Verletzungen, Narbenbildung, Fibrose (Vermehrung von Bindegewebe), Steatose (Verfettung) usw. führen. Die Einzelzellanalyse wird es ermöglichen, Veränderungen in der Zellzusammensetzung, die in pathologischen Situationen wie abnormaler Organentwicklung, Autoimmunerkrankungen, chronischen Krankheiten oder Krebs auftreten, direkt zu untersuchen. Die Bestimmung der zellulären Heterogenität in Tumoren oder in einzelnen Leukämiezellen eröffnet eine neue diagnostische Ebene zur Bestimmung der Herkunft, Progression und Heterogenität des Tumors für eine genauere,

17 „Unvoreingenommen“ bedeutet in diesem Zusammenhang, dass Forschende Ergebnisse erlangen können, ohne die Bandbreite möglicher Ergebnisse durch eigene, der Durchführung des Experiments vorangehende, Hypothesen einschränken zu müssen.

tumorspezifische Diagnose und bessere Vorhersagen für den erfolgreichen Einsatz therapeutischer Maßnahmen. Aschenbrenner, Mass und Schultze gehen in Kapitel 3 näher hierauf ein. Die umfangreichen medizinischen Anwendungsmöglichkeiten der Einzelzellanalyse lassen dabei auch die Frage aufkommen, inwiefern dieses Arbeitsfeld ethische Fragen aufwirft. Dieser Aspekt wird in Kapitel 6 von Fangerau, Marx-Stöltzing und Osterheider untersucht.

## 1.7 HERAUSFORDERUNGEN UND GRENZEN DER EINZELZELLTECHNOLOGIEN

Wie bei allen neuen technologischen Entwicklungen sind auch Einzelzellanalysen mit technischen und konzeptionellen Herausforderungen verbunden. Herausforderungen für die breite und routinemäßige Anwendung stellen insbesondere Schwierigkeiten der Vorbereitung von Einzelzellsuspensionen aus komplexen Geweben dar sowie Einschränkungen, RNA oder DNA (aber auch Lipide und Proteine) in ausreichender Menge und qualitativ hochwertig aus Einzelzellen gewinnen zu können (siehe Müller-Röber, Kapitel 5, zur weiteren Erörterung dieses Problems in Bezug auf Pflanzenzellen). Physikalische Einschränkungen wie Schwierigkeiten, bestimmte Zelltypen zum Beispiel in Gehirngeweben zu vereinzelnen, können Experimente stark beeinträchtigen. Neue Ansätze zur Verwendung isolierter Kerne und zur Analyse neugebildeter RNA können einige dieser Probleme überwinden (Krishnaswami et al., 2016) und sogar die Untersuchung von Krankheiten wie Alzheimer in konservierten Post-mortem-Geweben ermöglichen (siehe auch Aschenbrenner, Mass, Schultze, Kapitel 3).

Für die Interpretation von Multi-Omics-Einzelzelldaten und deren Modellierung wird es wesentlich sein, Parameter auf Einzelzellebene zu erschließen, die nicht ausschließlich auf Nukleinsäuren basieren, sondern auch das Vorhandensein von Proteinen, Lipiden und Stoffwechselprodukten in einer Zelle im Blick zu haben. Erste experimentelle Erfolge, die zeigen, dass dies leistbar ist, wurden bereits veröffentlicht, jedoch bleiben die technischen Möglichkeiten einer umfassenden Darstellung von Proteom- und Metabolom- sowie Lipidomspektren aus Einzelzellen noch immer sehr eingeschränkt (Duncan et al., 2019; Marx, 2019; Pasarelli et al., 2019).

Da sich Einzelzellanalysen auf Methoden stützen, die Zellen nach ihren Ähnlichkeiten in Expressionsprofilen gruppieren, ist es wichtig zu kontrollieren, ob die erhaltenen Gruppen und Muster das erwartete und vollständige Zellspektrum in

der Ausgangs-Einzelzellsuspension repräsentieren. Bei vielen Anwendungen, insbesondere beim Nachweis seltener Zellen, muss auch die Effizienz bei der Generierung einer Einzelzellbibliothek und bei der Sequenzierungstiefe berücksichtigt werden, da diese Anwendungen üblicherweise mit hohen Vorbereitungs- und Sequenzierungskosten verbunden sind. Darüber hinaus müssen experimentelle und bioinformatische Standards festgelegt werden, um eine Überinterpretation der NGS-Daten einzelner Zellen zum Beispiel aufgrund von technisch bedingten Ausfällen<sup>18</sup> einzelner Gensignaturen zu vermeiden (Van den Berge et al., 2018). Schließlich erfordert die Identifizierung und biologische Interpretation gruppierter Zellen bestimmte Vorkenntnisse, für die Schätzung der Zellzusammensetzung, zum Beispiel eine ungefähre Kenntnis der Anzahl der Zelltypen, und für die räumliche Rekonstruktion eine Orientierung durch zellspezifische „Marker- bzw. Anker-Gene“ (siehe Aliee, Sacher, Theis, Kapitel 4). Neue und alte Ansätze müssen (weiter-) entwickelt werden, um diesem Bedarf gerecht zu werden.

## 1.8 SCHLUSSBEMERKUNG

Einzelzell-Omics sind ein schnell wachsender und äußerst wichtiger Bereich der funktionellen Genomik. Ihr breites Anwendungs- und Datennutzungsspektrum wird die moderne Biologie und Medizin revolutionieren, in vielen Aspekten bereichern und in eine neue, tiefgreifend molekulare Richtung führen. Konzepte der Zell- und Systembiologie werden tiefer ergründet und teilweise in neuem Licht betrachtet werden. NGS-basierte Einzelzelldaten werden nahezu jedes biologische Forschungsfeld beeinflussen: von der grundlegenden Zellbiologie bis zur Entwicklungsbiologie, von der Physiologie bis zur Pathologie, von der Taxonomie bis zur Ökologie. Für die Medizin ist die Einzelzelldiagnostik eines der heißesten aufstrebenden Gebiete in der personalisierten Medizin mit einem hohen Potenzial, die Präzisionsdiagnostik auf eine neue Ebene zu heben. Der Erfolg der Einzelzellanalyse hängt dabei in hohem Maße von der Entwicklung neuartiger experimenteller und bioinformatischer Lösungen ab. Die Kernstrukturen für eine solche Entwicklung sind gegeben, erfordern aber eine ständige Investition und Anpassung.

<sup>18</sup> Ausfall bedeutet in diesem Zusammenhang, dass Transkripte aus technischen Gründen nicht nachgewiesen werden, z. B. unter anderem durch ineffizientes Umschreiben von RNA in cDNA.

## 1.9 LITERATUR

- Aizarani, N. et al. (2019): A human liver cell atlas reveals heterogeneity and epithelial progenitors. In: *Nature* 572: 199–204.
- Buenrostro, J. D. et al. (2015): Single-cell chromatin accessibility reveals principles of regulatory variation. In: *Nature* 523(7561): 486.
- Cao, J. et al. (2017): Comprehensive single-cell transcriptional profiling of a multi-cellular organism. *Science* 357(6352): 661–667.
- Chen, J. et al. (2017): Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. In: *Nature protocols* 12(3): 566.
- Chen, G./Shi, T. (2019): Single-cell RNA-seq technologies and related computational data analysis. In: *Frontiers in genetics* 10: 317.
- Clark, S. J. et al. (2018): scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. In: *Nature communications* 9(1): 781.
- Darmanis, S. et al. (2015): A survey of human brain transcriptome diversity at the single cell level. In: *Proceedings of the National Academy of Sciences* 112(23): 7285–7290.
- Duncan, K. D. et al. (2019): Advances in mass spectrometry based single-cell metabolomics. In: *Analyst* 144(23): 7285–7290.
- Gierahn, T. M. et al. (2017): Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. In: *Nature methods* 14(4): 395.
- Halpern, K. B. et al. (2017): Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. In: *Nature* 542(7641): 352.
- Han, X. et al. (2018): Mapping the mouse cell atlas by microwell-seq. In: *Cell* 172(5): 1091–1107.
- Jaitin, D. A. et al. (2014): Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. In: *Science* 343(6172): 776–779.
- Karaiskos, N. et al. (2017): The *Drosophila* embryo at single-cell transcriptome resolution. In: *Science* 358(6360): 194–199.
- Keren-Shaul, H. et al. (2019): MARS-seq2. 0: an experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing. In: *Nature protocols* 14(6): 1841.

Kowalczyk, M. S. et al. (2015): Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. In: *Genome research* 25(12): 1860–1872.

Krishnaswami, S. R. et al. (2016): Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. In: *Nature protocols* 11(3): 499.

Lake, B. B. et al. (2018): Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. In: *Nature biotechnology* 36(1): 70.

Macosko, E. Z. et al. (2015): Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. In: *Cell* 161(5): 1202–1214.

MacParland, S. A. et al. (2018): Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. In: *Nature communications* 9(1): 4383.

Magella, B. et al. (2018): Cross-platform single cell analysis of kidney development shows stromal cells express Gdnf. In: *Developmental biology* 434(1): 36–47.

Marx, V. (2019): A dream of single-cell proteomics. In: *Nature methods*, online publication 12.08.2019. doi: 10.1038/s41592-019-0540-6.

Mohammed, H. et al. (2017): Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation. In: *Cell reports* 20(5): 1215–1228.

Nagano, T. et al. (2017): Cell-cycle dynamics of chromosomal organization at single-cell resolution. In: *Nature* 547(7661): 61.

Nichterwitz, S. et al. (2016): Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. In: *Nature communications* 7: 12139.

Passarelli, et al. (2013): Single-cell lipidomics: characterizing and imaging lipids on the surface of individual *Aplysia californica* neurons with cluster secondary ion mass spectrometry. In: *Analytical chemistry* 85(4): 2231–2238.

Peng, G. et al. (2016): Spatial transcriptome for the molecular annotation of lineage fates and cell identity in mid-gastrula mouse embryo. In: *Developmental cell* 36(6): 681–697.

Picelli, S. et al. (2013): Smart-seq2 for sensitive full-length transcriptome profiling in single cells. In: *Nature methods* 10(11): 1096.

Rosenberg, A. B. et al. (2018): Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. In: *Science* 360(6385): 176–182.

Tang, F. et al. (2009): mRNA-Seq whole-transcriptome analysis of a single cell. In: *Nature methods* 6(5): 377.



Treutlein, B. et al. (2014): Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. In: *Nature* 509(7500): 371.

Van den Berge, K. et al. (2018): Observation weights unlock bulk RNA-seq tools for zero inflation and single-cell applications. In: *Genome biology* 19(1): 24.

Wilson, N. K. et al. (2015): Combined single-cell functional and gene expression analysis resolves heterogeneity within stem cell populations. In: *Cell stem cell* 16(6): 712–724.

Xu, Y. et al. (2016): Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. In: *JCI insight* 1(20).

Zheng, G. X. Y. et al. (2017): Massively parallel digital transcriptional profiling of single cells. In: *Nat. Commun.* 8(14049), Online-Publikation 16.01.2017. DOI: 10.1038/ncomms14049.

## 2. EINZELZELGENOMIK VERÄNDERT DIE ENTWICKLUNGSBIOLOGIE

### 2.1 EINFÜHRUNG

Neue Entwicklungen in der Einzelzellgenomik verändern seit einigen Jahren die Entwicklungsbiologie grundlegend. Forscher haben schnell das Potenzial der Einzelzell-Transkriptomik<sup>1</sup> für eine systematische Identifizierung von Zelltypen erkannt, was eine große Verbesserung gegenüber den zuvor verwendeten Ansätzen darstellt, die beispielsweise auf der Morphologie (also auf der Struktur und Form) von Zellen basierten oder mit einer nur geringen Zahl von Markergenen durchgeführt wurden. Neue, umfassende Zelltypatlanten basierend auf Einzelzellgenomik sind eine überaus wertvolle Ressource für die Wissenschaft, da sie beispielsweise eine systematischere Analyse der Auswirkungen von Mutationen ermöglichen, indem sie zeigen, in welchem Zelltyp das mutierte Gen exprimiert wird (Projekte: Human Cell Atlas, Tabula Muris oder Fly Cell Atlas, die zum Ziel haben, alle Zelltypen des jeweils untersuchten Organismus basierend auf Einzelzell-Transkriptomik zu identifizieren).<sup>2</sup> Allerdings bewegen sich aktuelle Arbeiten im Bereich der Einzelzellgenomik, wie im Folgenden näher erläutert wird, über die Zelltypidentifizierung hinaus in Richtung funktionale Informationen über die Auswirkungen von Störungen, den Ursprung von Zelltypen, Differenzierungspfade sowie räumliche Architektur von Geweben und Mechanismen der Genregulierung (Griffiths et al., 2018). Aufgrund ihrer genetischen Zugänglichkeit, ihrer hochgradigen experimentellen Reproduzierbarkeit und des detaillierten Wissens über wichtige Entwicklungsmechanismen, das in jahrzehntelanger Forschung angesammelt wurde, dienen die entwicklungsbiologischen Modellorganismen aktuell häufig als Startpunkt für neue experimentelle und rechnerische Methoden, die dann später auf Krankheitsmodelle oder Proben menschlicher Patienten angewendet werden können. Die Einzelzell-Transkriptomik ist bei Weitem die fortgeschrittenste der

1 Transkriptomik ist die Untersuchung der Gesamtheit aller Transkripte (RNA-Abschriften der DNA) innerhalb einer Zelle und ermöglicht Rückschlüsse darüber, welche Gene in einer bestimmten Zelle exprimiert sind.

2 Bei allen diesen Projekten geht es um die Erstellung von Referenzkarten aller Zellen des untersuchten Organismus (z. B. Mensch, Maus, Fliege). So ist es das Ziel des Human Cell Atlas, „umfassende Referenzkarten aller menschlichen Zellen (der grundlegenden Einheiten des Lebens) zu erstellen, als Grundlage für ein Verständnis für die menschliche Gesundheit und für die Diagnostik, Überwachung und Behandlung von Krankheiten“. Siehe unter: <https://www.humancellatlas.org/> [24.06.2019].

sogenannten Einzelzell-Omics-Technologien und wird somit den größten Teil dieses Kapitels ausmachen, das einen Überblick über die aktuellen Entwicklungen zur Einzelzellgenomik in der Grundlagenforschung liefert. Allerdings entwickeln sich auch Einzelzellmessungen anderer Parameter, insbesondere die Proteomik und DNA-Methylierung sowie die Messung offenen Chromatins,<sup>3</sup> schnell und werden hier ebenfalls behandelt.

## 2.2 STÖRUNGSANALYSE IN ENTWICKLUNGSMODELLSYSTEMEN

Entwicklungsmodelle, die zur Einzelzellanalyse häufig herangezogen werden, sind klassische Tiermodellorganismen wie Fruchtfliegen, Zebrafische und Mäuse, die sich besonders gut für genetische Studien eignen.<sup>4</sup> Eine wichtige neuere Ergänzung hierzu sind aus humanem Material von Patienten gewonnene Organoide,<sup>5</sup> die erstmals menschliches Gewebe für Genveränderungen zugänglich machen (Camp/Treutlein, 2017). Dies ist etwa bei humanen Hirnorganoiden von besonderer Bedeutung, da die einzigartigen Eigenschaften des menschlichen Gehirns im Tiermodell nur schlecht wiedergeben werden können. Die Einzelzellanalyse ermöglicht Studien zur Auswirkung genetischer Störungen auf Entscheidungen über das Zellschicksal. Dies wird bereits regelmäßig in Vergleichen zwischen Wildtypen und mutierten Tieren durchgeführt, oder auch in einer aktuellen Publikation zur frühen mesodermen<sup>6</sup> Spezifizierung durch die Analyse von Mosaiktieren (Pijuan-Sala et al., 2019). In dieser Studie haben die Autoren einen chimären Mausembryo aus Wildtyp- und *Tal1*<sup>-/-</sup>-Zellen geschaffen, der einen direkten Vergleich des

- 3 Proteomik bezieht sich auf die Untersuchung der Gesamtzahl der in einer einzelnen Zelle nachgewiesenen verschiedenen Proteine. Die DNA-Methylierung ist eine biochemische Modifizierung von DNA, die sich auf den Zustand des Chromatins (Komplex aus DNA und den zugehörigen Proteinen) auswirkt, der wiederum die Genexpression reguliert und beeinflusst. Offenes Chromatin meint die Teile des Genoms, die regulatorische Proteine binden können. Diese Regionen sind typischerweise an der Kontrolle der Genexpression beteiligt.
- 4 In sog. Störungsstudien („perturbation studies“) werden spezifische Gene unterbrochen, um die Auswirkung dieser Störung auf die Entwicklung der jeweiligen Zelle untersuchen zu können. Daraus lässt sich die Funktion des Gens ableiten.
- 5 Organoide sind dreidimensionale Stammzellkulturen, die auf der funktionalen Ebene Organen gleichen. Sie sind mehrzellige Einheiten, können dreidimensionale Strukturen bilden und zeigen Funktionen, die typisch sind für das nachgeahmte Organ (Bartfeld/Clevers, 2018).
- 6 Keimblätter bestehen aus verschiedenen Zellschichten im Embryoblasten, die sich fortlaufend ausdifferenzieren und während der Embryonalentwicklung unterschiedliche Gewebe und Organe hervorbringen. Die drei Keimblätter nennt man Endoderm (Innenschicht), Ektoderm (Außenschicht) und Mesoderm (Mittelschicht). Aus dem Mesoderm bilden sich unter anderem Blut, Muskeln und Knochen.

Differenzierungspotenzials des Wildtyps und der mutierten Zelle in demselben Tier ermöglichte.<sup>7</sup>

Über die klassischen genetischen Störungen hinaus kann die Einzelzellanalyse auch in idealer Weise die molekularen und zellulären Auswirkungen anderer Störungen analysieren. Ein besonders beeindruckendes Beispiel für diese Art der Anwendung ist die Regeneration von Gliedmaßen beim Axolotl nach deren Amputation. In einer diesbezüglichen Studie konzentrieren sich Gerber et al. (2018) auf die Identifikation der Zelltypen, die vorübergehend an der Verletzungsstelle erscheinen, um die Skelettregeneration in Gang zu bringen. Durch die Kombination der Einzelzellanalyse mit der Verfolgung der genetischen Abstammung nach Brainbow (einer Methode, bei der einzelne Zellen mit fluoreszierenden Proteinen gefärbt werden) fanden sie heraus, dass es keine bereits existierenden Stammzellen gab. Stattdessen beobachteten sie, dass eine heterogene Zellpopulation (sogenannte Fibroblasten) stammzellartige Eigenschaften annahm und sich in Vorläuferzellen für die neu zu bildenden Skelettelemente entwickelte.

Seit einigen Jahren ist ein vermehrtes Forschungsinteresse an neuen Modellorganismen zu beobachten, was vor allem daher rührt, dass die Einzelzell-Transkriptomik die Identifizierung von Zelltypen und Differenzierungspfaden deutlich erleichtert hat. Daneben steht mit dem Aufkommen von Gene-Editing durch CRISPR/Cas9<sup>8</sup> ein einfaches Hilfsmittel zur Verfügung, um transgene Tiere vieler Arten zu produzieren. Neben der oben genannten Arbeit am Axolotl sind weitere interessante Beispiele Evolutionsstudien an Ringelwürmern und an der Seeanemone *Nematostella* (Achim et al., 2018; Sebé-Pedrós et al., 2018). Diese Projekte beginnen, zusammen mit neuartigen Computermethoden zum Vergleich von Einzelzell Datensätzen unterschiedlicher Spezies, zunehmend interessante Einblicke in die Entwicklung von Zelltypen zu gewähren.

Die Kombination aus Einzelzellgenomik und CRISPR/Cas9-Genome-Editing hat nicht nur zu gesteigertem Interesse an neuen Modellsystemen geführt, sondern inspiriert auch die Methodenentwicklung in anderen Bereichen wie zum Beispiel

7 Mosaiktiere bestehen aus mindestens zwei unterschiedlichen Zellpopulationen mit unterschiedlichem Chromosomengehalt. Der beschriebene chimäre Mausembryo ist so ein Mosaik aus normalen Zellen („Wildtyp“) und sogenannten  $Tal1^{-/-}$ -Zellen, denen der Transkriptionsfaktor Tal1 auf beiden Chromosomen fehlt. Tal1 (die Abkürzung steht für „T-cell acute lymphocytic leukemia protein 1“) spielt eine Rolle in der Regulierung von Genen, die mit Leukämie in Zusammenhang stehen.

8 CRISPR/Cas ist eine Methode zur Bearbeitung von Genen durch Einschnitte an bestimmten vordefinierten Stellen im Genom, die von Guide-RNA identifiziert werden.

Perturbation Screening und Lineage Tracing<sup>9</sup> (wobei auf Letztere im Folgenden näher eingegangen werden soll). Bei CRISPR-Screens wird Cas9 in kultivierte Zellen eingebracht, zusammen mit einer Vielzahl sogenannter „single guide RNAs“ (sgRNAs),<sup>10</sup> die auf unterschiedliche Gene abzielen. Eine Auslesung durch Einzelzell-Transkriptomik ermöglicht dann eine Verbindung der Aktivität einer spezifischen sgRNA mit Veränderungen in der Genexpression (Dixit et al., 2016; Jaitin et al., 2016; Datlinger et al., 2017). Diese Methoden sind zwar aktuell noch auf kultivierte Zellen beschränkt und noch nicht anwendbar auf Entwicklungsmodellsysteme, aber sie versprechen viel für die systematische Identifizierung genregulierender Netzwerke.

### 2.3 RÄUMLICHE INFORMATIONEN

In Geweben und Organen sind Zellen in speziellen räumlichen Strukturen angeordnet, die für deren ordnungsgemäße Funktion notwendig sind. Außerdem werden Zellen stark von ihrer Umgebung beeinflusst (z. B. die Umgebung von Stammzellen, sog. Stammzellnischen) sowie von den Signalen, die sie sich untereinander zusenden. Die Einzelzellgenomik erfordert typischerweise allerdings eine Trennung der Proben in eine Einzelzellsuspension, sodass bei den meisten Ansätzen alle Informationen über die räumliche Anordnung verloren gehen. Aktuell konzentriert man sich in der akademischen Welt wie auch in der industriellen Forschung stark darauf, die räumlichen Informationen bei der Einzelzellanalyse zu erhalten. Entsprechende Ansätze können grob in drei Kategorien unterteilt werden:

1. Methoden, die zusätzliche räumliche Informationen nutzen, die in der Mikroskopie erfasst wurden. Wenn die räumlichen Expressionsmuster zellspezifischer Markergene bekannt sind, dann können Zellen nach diesen Orientierungsgenen ausgerichtet werden (Satija et al., 2015; Karaiskos et al., 2017). Interessanterweise haben die Labore von Rajewsky und Friedman kürzlich gezeigt, dass die räumlichen Expressionsmuster größtenteils aus ersten Prinzipien abgeleitet werden können, auch ohne zusätzliche Informationen,

9 Perturbation Studies untersuchen systematisch den Einfluss von Störungen der Genexpression (z. B. durch gezieltes Editieren, Ausschalten bzw. Überexprimieren von Genen) auf den Phänotyp und die Entwicklung von Zellen. Als Lineage Tracing bezeichnet man die Untersuchung aller Nachkommen einer Zelle. Die beiden Methoden können auch miteinander kombiniert werden.

10 „Single guide RNA“ („ribonucleic acid“; Ribonukleinsäure) lenkt Cas9 an die Stelle, an der geschnitten werden soll.

da die meisten Gene in einfachen Mustern mit glatten Übergängen exprimiert werden (novoSpaRc<sup>11</sup>).

2. Andere Methoden nutzen die Mikroskopie-basierte Detektion von Gen-expression mit räumlicher Auflösung (FISH<sup>12</sup>). Da diese mikroskopischen Methoden nur eine geringe Zahl von Genen nachweisen können, werden mehrere Runden von Experimenten benötigt, um Hunderte von Genen zu detektieren. Kürzlich wurde dieser Ansatz zum ersten Mal auf die Ebene des gesamten Transkriptoms gehoben (Eng et al., 2017). Ein wichtiger Vorteil dieser bildbasierten Techniken ist deren viel höhere Transkript-Erkennungsrate als bei sequenzbasierten Ansätzen, da sie auf Hybridisierung und nicht auf der oft ineffizienten reversen Transkriptionsreaktion beruhen. Dennoch sind diese Methoden noch immer arbeitsaufwendig im Aufbau und zeitaufwendig in der Durchführung.
3. Zudem gibt es neuartige Methoden, bei denen molekulare Barcodes (in Form kurzer DNA-Sequenzen) räumliche Informationen direkt in Gewebeschnitten codieren. Diese Ansätze basieren entweder auf Abfolgen von Primern<sup>13</sup> mit Barcodes für die reverse Transkription, die auf eine Oberfläche aufgebracht werden (Stahl et al., 2016), oder auf Kügelchen mit Barcodes, die auf einer Oberfläche positioniert werden (Rodrigues et al., 2019).

## 2.4 ZEITLICHE INFORMATIONEN – PSEUDOTEMPORALE ORDNUNG

Neben der räumlichen Information gibt es eine weitere große Herausforderung in der Einzelzellgenomik, die Aufnahme zeitlicher Informationen. Da Zellen während der Sequenzierung zerstört werden, ist es nicht möglich, Veränderungen in ihrer Expression sowie Entscheidungen über ihr Schicksal in Echtzeit nachzuverfolgen. Wenn aber die Anzahl Zellen in einer Probe groß genug ist, dann können auch

- 11 novoSpaRc ist eine Rechenmethode zur Vorhersage von Lokalisierungen einzelner Zellen im Raum allein durch die Nutzung von Daten aus Einzelzell-RNA-Sequenzierung. Abstände einzelner Zellen im Expressionsraum werden dabei in deren physische Abstände im Gewebe übertragen.
- 12 FISH ist eine Technik, bei der die spezifische Bindung bestimmter Zielgene bzw. ihrer Transkripte an Nukleinsäuresequenzen genutzt und durch fluoreszierende Marker sichtbar gemacht wird. Diese Fluoreszenz erlaubt Rückschlüsse auf das Vorhandensein und ggf. die räumliche Verteilung der Nukleinsäuren in der Probe.
- 13 Primer sind Oligonukleotide, also kurze Schnipsel aus DNA oder RNA (bis ca. 30 Basen lang), die als Ausgangspunkt für die Synthese von DNA oder RNA durch bestimmte Enzyme (Polymerasen) dienen.

vorübergehende (und somit seltene) Zustände im Datensatz erkannt werden. Somit ist es möglich, Zellen entlang einer abgeleiteten, sogenannten pseudotemporalen Trajektorie<sup>14</sup> anzuordnen (Moignard et al., 2015; Setty et al., 2016; Haghverdi et al., 2016; Kester/van Oudenaarden, 2018). Für kurzfristige Prozesse, die kontinuierlich ablaufen (wie z. B. Hämatopoese), kann so der gesamte Vorgang der transkriptionellen Veränderungen in einem einzigen Experiment erfasst und per Computer rekonstruiert werden (siehe Aliee, Sacher, Theis, Kapitel 4).

Während mit den Methoden der pseudotemporalen Ordnung von Einzelzell-Transkriptomen Zellen effizient entlang kontinuierlicher Trajektorien geordnet werden, geht die Richtung des Differenzierungsvorgangs aus den Daten allein nicht eindeutig hervor. La Manno et al. (2018) haben kürzlich die RNA-Geschwindigkeitsanalyse vorgestellt, eine Rechenmethode, bei der die Richtung im Genexpressionsraum, in die Zellen sich bewegen, basierend auf ungespliceten vs. gesplíceten<sup>15</sup> (d. h. „neuen“ vs. „alten“) Transkriptmolekülen abgeleitet wird. Eine weitere neue Methode, um in die unmittelbare Zukunft von Zellen zu blicken, ist die metabolische Markierung von RNA (Hendriks et al., 2018; Erhard et al., 2019), durch die, basierend auf Markierungen, die experimentell in einem vorgegebenen Zeitfenster in RNA-Moleküle eingebracht werden, alte von neuen Molekülen getrennt werden können.

Trotz der relativen Neuartigkeit der pseudotemporalen Ordnung gibt es bereits zahlreiche biologische Anwendungsbereiche. Dazu gehören eine vollständige Differenzierungstrajektorie für Planarien (Plass et al., 2018) sowie eine Studie, die Übergänge zwischen Venen und Arterien in der koronaren Entwicklung bei der Maus aufgezeigt hat (Su et al., 2018). In erweiterten Ansätzen werden Einzelzell-Transkriptome in unterschiedlichen Entwicklungsstadien gemessen und anschließend mit dem Computer die verschiedenen Zeitpunkte zu kontinuierlichen Trajektorien verbunden (Farrell et al., 2018; Wagner et al., 2018).

14 Für die pseudotemporale Analyse (Pseudozeitanalyse) werden die sequenzierten Zellen nach Ähnlichkeit ihres Transkriptoms geordnet. Die daraus resultierende Folge an Einzelzell-Transkriptomen wird als „Trajektorie“ (Entwicklungsverlauf) bezeichnet und als zeitliche Abfolge von Zellzuständen interpretiert, d. h. als gradueller Übergang vom Stammzellenstadium hin zur differenzierten Zelle.

15 Nach der Transkription durchläuft die RNA Modifizierungen, die zu deren Reifung führen. „Splicing“ ist der Vorgang, bei dem bestimmte Teile der Original-RNA (Introns) ausgeschnitten und entsorgt werden, während die verbleibenden Teile (Exons) zu einer reifen RNA verbunden werden.

## 2.5 ZEITLICHE INFORMATIONEN – ABSTAMMUNGSANALYSE MIT HOHEM DURCHSATZ

Während die pseudotemporale Ordnung und die metabolische RNA-Markierung kurzfristige zeitliche Informationen liefern, wären Informationen über Verhältnisse zwischen Zellen über längere Zeiträume wünschenswert – über Tage, Monate, ja sogar Jahre. Der Bereich der Abstammungsverfolgung macht sich schon lange visuelle Marker (z. B. Fluorophore) zunutze, um Zellen zu markieren und nachzuvollziehen. In neuerer Zeit ist es mit dem Aufkommen der Einzelzellgenomik möglich geworden, die enorme Informationsspeicherkapazität des Genoms zu nutzen, um die Abstammungsverbindungen von Zellen zu bestimmen. Sequenzbasierte Methoden zur Abstammungsanalyse fallen grob in zwei Kategorien: solche, bei denen natürlich vorkommende Mutationen genutzt werden, und solche, bei denen das Genom aktiv modifiziert werden soll.

In der Theorie sind natürlich vorkommende somatische Mutationen (wie z. B. Einzelnukleotidvariationen oder Variationen in der Anzahl Kopien) leistungsstarke Abstammungsmarker, die in der Sequenzierung ausgelesen werden können. Da die Abstammungsnachverfolgung anhand somatischer Mutationen nicht invasiv ist und keine kontinuierliche Beobachtung erfordert, ist sie ideal zur Untersuchung humaner Proben geeignet. In den letzten Jahren wurde diese Strategie in Pionierstudien für frühe Entscheidungen über embryonische Abstammung übernommen. An Organoiden aus einzelnen Mauszellen (Behjati et al., 2014) und in großen Mengen analysierten Humanblutproben (Lodato et al., 2015) haben Analysen somatischer Mutationen eine Rekonstruktion früher embryonischer Abstammungsbäume ermöglicht. In einem neueren richtungsweisenden Paper des Walsh-Labors haben die Autoren nach Genom-Sequenzierung einzelner Zellen Neuronen aus postmortalen menschlichen Gehirnen in einen entwicklungsbezogenen Abstammungsbaum geordnet (Ju et al., 2017). Die generelle Anwendung dieses Ansatzes wird aktuell aber noch durch die hohen Kosten der Sequenzierung des gesamten Genoms einer großen Anzahl Einzelzellen gebremst. Die Abstammungsnachverfolgung basierend auf Mutationen in Mitochondrien (die eine viel höhere Mutationsrate haben) ist eine vielversprechende Alternative zur Abstammungsnachverfolgung mit hohem Durchsatz bei Menschen (Ludwig et al., 2019).

Solche Ansätze sind zwar ideal für menschliche Proben geeignet, aber für Modellorganismen sind Techniken der Abstammungsnachverfolgung basierend auf experimenteller Modifikation normalerweise die bessere Wahl, weil dabei eine bessere



Kontrolle möglich ist. Experimentell kontrollierte Genommodifikationen zur Abstammungsnachverfolgung sind per Rekombination durch synthetische Cre/lox-Kassetten<sup>16</sup> (Pei et al., 2017) oder mit der CRISPR/Cas9-Technologie möglich. Abstammungsnachverfolgung mit hohem Durchsatz basierend auf CRISPR/Cas9 in Kombination mit Zelltypidentifizierung per Einzelzell-RNA-Sequenzierung wurde kürzlich an Zebrafischen (Alemany et al., 2018; Raj et al., 2018; Spanjaard et al., 2018) und Mäusen (Kalhor et al., 2018; Chan et al., 2019) durchgeführt. Es bleiben zwar viele experimentelle und rechnerische Herausforderungen bestehen, aber die CRISPR/Cas-Methode zur Abstammungsnachverfolgung ist als genereller Ansatz zur Identifizierung des entwicklungsbezogenen Ursprungs von Zelltypen und für Einblicke in Mechanismen zelltypabhängiger Erkrankungen besonders vielversprechend.

## 2.6 MESSUNG ANDERER PARAMETER ALS RNA

Wie bereits erwähnt, ist die RNA-Sequenzierung die am weitesten entwickelte Technologie in der Einzelzellgenomik. Aber Messungen anderer Parameter sind auf dem Vormarsch, insbesondere die Messung offenen Chromatins, DNA-Methylierung und Einzelzellprotein-detektion. Einzelzell-ATAC-Sequenzierung (scATAC-seq), eine neue Methode zur Messung offenen Chromatins, kann inzwischen dank neuer Protokolle zum Einfügen von DNA-Barcodes in einzelne Zellen routinemäßig an Tausenden von Zellen durchgeführt werden. Zu den Anwendungsmöglichkeiten gehören Atlanten zur Zugänglichkeit von Chromatin bei Mäusen (Cusanovich/Hill et al., 2018) und in der Entwicklung von *Drosophila* (Cusanovich/Reddington et al., 2018). In einer bemerkenswerten neueren Publikation haben Yoshida et al. (2019) Epigenom- und Transkriptommessungen an 86 primären Immun-Zelltypen aus der Maus generiert und dadurch wichtige Erkenntnisse darüber gewonnen, welche Mechanismen der Genregulation für die Steuerung von Zelldifferenzierung im Immunsystem wichtig sind.

Während scATAC-seq von der Wissenschaft schnell angenommen wird, ist der Einsatz der Einzelzell-DNA-Methylierung bisher beschränkt auf relativ wenige Labore, was wahrscheinlich vor allem an den hohen Kosten der DNA-Methylierungsanalyse liegt. Dabei hat sie bereits wichtige Einblicke geliefert, insbesondere in die frühe Entwicklung. Rulands et al. (2018) haben beispielsweise

<sup>16</sup> Cre/lox-Kassetten sind ein System, das Genveränderungen in bestimmten Zelllinien in lebenden Tieren ermöglicht.

unerwartete Schwankungen im Genombereich in der DNA-Methylierung beim Austritt aus der Pluripotenz festgestellt. Wichtig ist hier, dass die DNA-Methylierung bereits erfolgreich mit RNA-Messungen aus denselben Einzelzellen kombiniert werden konnte (Clark et al., 2018).

Während die Proteindetektion in Einzelzellen im Bereich des vollen Proteoms bisher nicht erfolgreich etabliert werden konnte, gibt es bereits äußerst vielversprechende Ansätze zum Nachweis von Protein-Panels: Die Einzelzell-Massenzytometrie (Bendall et al., 2011) ermöglicht durch den Einsatz von mit Schwermetallen markierten spezifischen Antikörpern den parallelen Nachweis zahlreicher Proteine in Einzelzellen. Dafür binden Antikörper gepaart mit bestimmten Isotopen von Übergangselementen an ihre Epitope, also an ihre Bindungsstellen an den zu untersuchenden Proteinen. Einzelne Zellen werden dann vaporisiert und zu Plasma ionisiert, und elementare Ionen werden mithilfe von sogenannter Flugzeitmassenspektrometrie erkannt.<sup>17</sup> Ein weiterer Sequenzierungsansatz ist CITE-seq (Stoeckius et al., 2017), die einen gleichzeitigen Nachweis von mit Oligonukleotiden markierten Antikörpern und Transkriptom-Messungen in einer effizienten Einzelzellauslesung ermöglicht.

Mit mehr und mehr Datensätzen auf der Basis unterschiedlicher Messtechniken wird in diesem Bereich immer deutlicher, dass neuartige Computermethoden zur Datenintegration erforderlich sind. Dazu gehört u.a. die Abstimmung der Zelltypen, die durch scRNA-seq and scATAC-seq identifiziert wurden, aber auch die Entfernung von Unterschieden zwischen scRNA-seq-Datensätzen, die auf technische Artefakte zurückzuführen sind (z.B. Masseneffekte durch Dissoziationstechniken). In letzter Zeit wurden mehrere vielversprechende Computerverfahren vorgestellt (Barkas et al., 2018; Butler et al., 2018; Haghverdi et al., 2018).

Die Einzelzellanalyse verändert unser Verständnis von biologischer Entwicklung. Mithilfe neuer Methoden und Ansätze haben Entwicklungsbiologen Hilfsmittel an die Hand bekommen, mit denen sie lang verborgene Geheimnisse in der räumlichen und zeitlichen Gewebeorganisation lüften können. Dieses fundamentale Wissen bietet nicht nur eine bessere Sicht auf biologische Prozesse im gesunden Zustand, es bietet auch Gelegenheit, sich auf Abweichungen von der Norm zu konzentrieren, die zu Krankheiten führen. Im letzten Abschnitt wird die neue europäische Initiative LifeTime vorgestellt, die die Einzelzellanalyse in die

<sup>17</sup> In der Flugzeitmassenspektrometrie wird das Verhältnis der Masse eines Ions zu dessen Ladung mit Flugzeitmessung bestimmt. Ionen werden durch ein elektrisches Feld von bekannter Stärke beschleunigt.

Lage versetzen soll, unser Verständnis, die frühe Diagnostik, Überwachung und Behandlung verschiedenster Erkrankungen in Richtung Innovation und personalisierte Medizin voranzutreiben.

## 2.7 LIFETIME: EINE NEUE INITIATIVE BASIEREND AUF EINZELZELLANALYSE

Die Aussagekraft von Einzelzellanalysetechnologien wurde nicht nur kürzlich von der Fachzeitschrift *Science* anerkannt, die diese als „Durchbruch des Jahres 2018“ feierte (Pennisi, 2018), sondern hat auch die Gründung der europäischen Forschungsinitiative LifeTime inspiriert. Das Konsortium besteht aus Hunderten von Forschern in 18 europäischen Ländern und wird von allen wichtigen europäischen Wissenschaftsakademien, von vielen nationalen Regierungen sowie von über 70 Firmen unterstützt. Ihr Ziel besteht darin, menschliche Zellen zu begreifen und zu kartieren, um dann in der Behandlung von Krankheiten bei Patienten direkt ansetzen zu können. Durch das Ausschöpfen des vollen Potenzials von Einzelzelltechnologien, künstlicher Intelligenz und individuell gestalteten experimentellen Krankheitsmodellen (wie Organoiden) möchten die LifeTime-Forscher den Krankheitsbeginn besser vorhersagen und Krankheiten durch gezielte Analyse des patienteneigenen Gewebes behandeln. Um dies erreichen zu können, muss erst verstanden werden, wie Genome innerhalb von Zellen funktionieren, um zu entziffern, wie Zellen Gewebe bilden, und die Dynamiken zu identifizieren, die von einer gesunden Zelle oder vom gesunden Gewebe zu einem pathologischen Zustand führen.

Einzelzelltechnologien bieten eine gute Möglichkeit zur Überwindung einiger der grundlegenden Defizite in unseren aktuellen wissenschaftlichen Ansätzen, zum Beispiel durch die Lösung der räumlichen zellulären Heterogenität oder die Erfassung zellulärer Veränderungen im Laufe der Zeit. Wichtig ist, dass die Erfindung und Nutzung neuer Computer-Tools und künstlicher Intelligenz von entscheidender Bedeutung für das Erreichen der Ziele von LifeTime sein werden. Dadurch wird es erst möglich sein, die generierten Daten zu integrieren und nicht nur den gesunden Zustand, sondern auch die Ursache und Biologie von Krankheiten zu verstehen. Die Verbesserung experimenteller Krankheitsmodelle durch die Anwendung neuer Technologien in der Genommodifikation und Zellumprogrammierung wird es den LifeTime-Forschern möglich machen, Genome und Zellen aus Patientengeweben zu verändern.

Zur Erleichterung der Profilierung mehrerer Schichten der Genomregulierung – ein wichtiger Schritt hin zu den Zielen von LifeTime – müssen auch Einzelzell-Multi-Omics und Bildgebung weiterentwickelt und integriert werden (z. B. Transkriptom, Epigenom, Metabolom, Proteom etc.). Außerdem werden starke Bemühungen in der experimentellen Skalierung nötig sein, um den erforderlichen Probendurchlauf mit der richtigen analytischen Auflösung zu erreichen. Angesichts der Erfahrung mit der Entwicklung anderer Technologien in der Vergangenheit (z. B. DNA-Sequenzierung) können innerhalb von einigen Jahren adäquate Fortschritte und damit einhergehende signifikante Kostensenkungen erwartet werden. Tatsächlich ist dies bei einigen Einzelzelltechnologien bereits heute der Fall, die aktuell mit Millionen von Zellen pro Probe arbeiten.

Weitere Entwicklungen werden auch bei anderen Technologiesäulen erforderlich sein, auf denen LifeTime aufgebaut ist – nicht zuletzt an der Schnittstelle zwischen technologischen Bereichen. Manche sind naturgemäß eng miteinander verbunden wie zum Beispiel die Anpassung rechnerischer und statistischer Techniken an die Skalierung und Integration von Einzelzell-Multi-Omics. Auch werden mit dem Aufkommen detaillierter molekularer und räumlicher Zellreferenzkarten neue Machine-Learning-Tools nötig sein, um die Integration von Patiententrajektorien zu erleichtern und den Krankheitsweg des Patienten anhand elektronischer Krankenakten vorhersagen zu können. Außerdem wird es notwendig sein, neue Computermethoden zu entwickeln, um mechanistisch zu verstehen, was den Transkriptionszustand einer Zelle vorantreibt und letztlich, was dessen präzise Funktion im Zusammenhang mit einem spezifischen Gewebe ist (siehe Aliee, Sacher, Theis, Kapitel 4). Dieses verbesserte Wissen um Ursache und Wirkung in zellulären Regulierungsprofilen wird ein Sprung in Richtung personalisierte Medizin sein. Mit der Anwendung von patientenabgestimmten Organoiden und Organ-on-a-Chip-Modellen, die ebenfalls weiterentwickelt und verbessert werden, möchten die LifeTime-Forschenden die Translation in die Klinik erreichen. Dies wird den Übergang von personalisierten experimentellen Krankheitsmodellen zu aussagekräftigen Vorhersagesystemen vorantreiben.

Das LifeTime-Toolkit aus Methoden und Technologien (LifeTime Technology Platform) wird für eine große Bandbreite an Erkrankungen nutzbar sein. Dazu gehören neurologische Erkrankungen, Infektionskrankheiten, Krebserkrankungen und viele andere Krankheitsfelder. Die Auswahl der Erkrankungen, die mit der LifeTime Technology Platform untersucht werden sollen, erfolgt über einen interaktiven, transparenten und von Kollegen geprüften Prozess, genannt „LifeTime

Launchpad“. Darin wird eine Reihe von Parametern berücksichtigt, wie zum Beispiel die Auswirkungen auf die Gesellschaft, die Heterogenität auf der Zellebene, die Verfügbarkeit von Zellmodellen, die klinische Machbarkeit etc. Der Mechanismus bleibt auch während der Umsetzungsphase von LifeTime erhalten, damit neue Ideen und Gelegenheiten untersucht werden können, sobald sie aufkommen.

## 2.8 LITERATUR

Achim, K. et al. (2018): Whole-Body Single-cell sequencing reveals transcriptional domains in the annelid larval body. In: *Molecular biology and evolution* 35(5): 1047–1062.

Aleman, A. et al. (2018): Whole-organism clone tracing using single-cell sequencing. In: *Nature*, Online-Publikation 28.03.2018. DOI: 10.1038/nature25969.

Barkas, B. et al. (2018): Wiring together large single-cell RNA-seq sample collections. In: *BioRxiv*, Online-Publikation 02.11.2018. DOI: 10.1101/460246.

Bartfeld, S./Clevers, H. (2018): Aus Stammzellen abgeleitete Organoide und ihre Bedeutung für die biomedizinische Forschung und Therapie. In: Zenke, M. et al.: *Stammzellforschung*. Nomos, Baden-Baden.

Behjati, S. et al. (2014): Genome sequencing of normal cells reveals developmental lineages and mutational processes. In: *Nature* 513: 422–425.

Bendall, S. C. et al. (2011): Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. In: *Science* 332(6030): 687–696.

Butler, A. et al. (2018): Integrating single-cell transcriptomic data across different conditions, technologies, and species. In: *Nature Biotechnology* 36: 411–420.

Camp, J. G./Treutlein, B. (2017): Human organomics: a fresh approach to understanding human development using single-cell transcriptomics. In: *Development* 144: 1584–1587.

Chan, M. M. et al. (2019): Molecular recording of mammalian embryogenesis. In: *Nature* 570: 77–82.

Clark, S. J. et al. (2018): scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. In: *Nature Communication*, Online-Publikation 22.02.2018. DOI: 10.1038/s41467-018-03149-4.

Cusanovich, D. A. et al. (2018): A single-cell atlas of in vivo mammalian chromatin accessibility. In: *Cell* 174(5): 1309–1324.

Cusanovich, D. A. et al. (2018): The cis-regulatory dynamics of embryonic development at single-cell resolution. In: *Nature*, Online-Publikation 14.03.2018. DOI: 10.1038/nature25981.

Datlinger, P. et al. (2017): Pooled CRISPR screening with single-cell transcriptome readout. In: *Nature Methods* 14: 297–301.

Dixit, A. et al. (2016): Perturb-Seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. In: *Cell* 167(7): 1853–1866.e17.

Eng, C. H. L. J. et al. (2017): Profiling the transcriptome with RNA SPOTs. In: *Nature Methods* 14: 1153–1155.

Erhard, F. et al. (2019): scSLAM-seq reveals core features of transcription dynamics in single cells. In: *BioRxiv*, Online-Publikation 21.01.2019. DOI: 10.1101/486852.

Farrell, J. A. et al. (2018): Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. In: *Science* 360(6392): eaar3131.

Gerber, T. et al. (2018): Single-cell analysis uncovers convergence of cell identities during axolotl limb regeneration. In: *Science* 362(6413): eaaq0681.

Griffiths, J. A. (2018): Using single-cell genomics to understand developmental processes and cell fate decisions. In: *Molecular Systems Biology*, Online-Publikation 16.04.2018. DOI: 10.15252/msb.20178046.

Haghverdi, L. et al. (2016): Diffusion pseudotime robustly reconstructs lineage branching. In: *Nature Methods* 13: 845–848.

Haghverdi, L. et al. (2018): Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. In: *Nature Biotechnology* 36: 421–427.

Hendriks, G. et al. (2018): NASC-seq monitors RNA synthesis in single cells. In: *BioRxiv*, Online-Publikation 17.12.2018. DOI: 10.1101/498667.

Jaitin, D. A. et al. (2016): Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. In: *Cell* 167(7): 1883–1896.e15.

Ju, Y. S. et al. (2017): Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. In: *Nature* 543: 714–718.

Kalhor, R. et al. (2018): Developmental barcoding of whole mouse via homing CRISPR. In: *Science* 361(6405): eaat9804.

Karaiskos, N. et al. (2017): The drosophila embryo at single-cell transcriptome resolution. In: *Science* 358(6360): 194–199.

- Kester, L./van Oudenaarden, A. (2018): Single-cell transcriptomics meets lineage tracing. In: *Cell Stem Cell* 23(2): 166–179.
- La Manno, G. et al. (2018): RNA velocity of single cells. In: *Nature* 560: 494–498.
- Locato, M. A. et al. (2015): Somatic mutation in single human neurons tracks developmental and transcriptional history. In: *Science* 360(6256): 94–98.
- Ludwig, L. S. et al. (2019): Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. In: *Cell* 176(6): 1325–1339.
- Moignard, V. et al. (2015): Decoding the regulatory network of early blood development from single-cell gene expression measurements. In: *Nature Biotechnology* 33: 269–276.
- Pei, W. et al. (2017): Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. In: *Nature* 548: 456–460.
- Pennisi, E. (2018): Development cell by cell. In: *Science* 362(6421): 1344–1345.
- Pijuan-Sala, B. et al. (2019): A single-cell molecular map of mouse gastrulation and early organogenesis. In: *Nature* 566: 490–495.
- Plass, M. et al. (2018): Cell type atlas and lineage tree of a whole complex animal by single cell transcriptomics. In: *Science* 360(6391): eaaq1723.
- Raj, B. et al. (2018): Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. In: *Nature Biotechnology* 36: 442–450.
- Rodrigues, S. G. et al. (2019): Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. In: *Science* 363(6434): 1463–1467.
- Rulands, S. et al. (2018): Genome-scale oscillations in DNA methylation during exit from pluripotency. In: *Cell Systems* 7(1): 63–76.e12.
- Satija, R. et al. (2015): Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology* 33: 495–502.
- Sebé-Pedrós, A. et al. (2018): Cnidarian cell type diversity and regulation revealed by whole-organism single-cell RNA-seq. In: *Cell* 173(6): 1520–1534.e20.
- Setty, M. et al. (2016): Wishbone identifies bifurcating developmental trajectories from single cell data. In: *Nature Biotechnology* 34: 637–645.
- Spanjaard, B. et al. (2018): Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. In: *Nature Biotechnology* 36: 469–473.
- Ståhl, P. L. et al. (2016): Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. In: *Science* 353(6294): 78–82.

Stoeckius, M. et al. (2017): Simultaneous epitope and transcriptome measurement in single cells. In: *Nature Methods* 14: 865–868.

Su, T. et al. (2018): Single-cell analysis of early progenitor cells that build coronary arteries. In: *Nature* 559: 356–362.

Wagner, D. E. et al. (2018): Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. In: *Science* 360(6392): 981–987.



### **3. EINZELZELL-OMICS IN DER BIOMEDIZIN GESTERN, HEUTE UND MORGEN**

#### **3.1 NEUE WEGE DER EINZELZELL-OMICS<sup>1</sup> FÜR EINE NEUE WAHRNEHMUNG UND BEHANDLUNG VERBREITETER ERKRANKUNGEN**

Die zelluläre Heterogenität innerhalb von Geweben war immer ein großes Hindernis im Verständnis und in der Behandlung von Krankheiten wie Krebs, chronisch entzündlichen Erkrankungen, Autoimmunerkrankungen, Infektionen oder neurodegenerativen Erkrankungen. Frühere Genomikansätze waren auf Massenanalysen beschränkt, die nur gemittelte Ergebnisse über alle Probezellen lieferten. Die Untersuchung der zellulären Heterogenität in der Einzelzellauflösung ist erst seit einigen Jahren möglich, und kommt heute weltweit zum Einsatz, um die Grundmechanismen und somit auch die Pathogenese solcher Erkrankungen zu begreifen. In diesem Kapitel soll ein Überblick darüber gegeben werden, wie Einzelzell-Omics langsam unser Verständnis von Krebs und neurodegenerativen Erkrankungen revolutionieren, wie aktuell Grundlagenforschung im klinischen Bereich Anwendung findet, und mit welchen innovativen und experimentellen Ideen neue Wege eingeschlagen werden können, um Patienten zu helfen und Behandlungen individuell gestalten zu können.

#### **3.2 EINZELZELL-OMICS GESTERN**

##### **Krebs als fortgeschrittenstes Beispiel für die Anwendung von Einzelzell-Omics auf Krankheiten**

Die Mikroumgebung eines Tumors zeichnet sich nicht nur durch unterschiedliche Zusammensetzungen aus Krebszellen aus, sondern auch durch infiltrierende Immunzellen sowie Stromazellen, die den Tumor umgeben. Da die umfassende Analyse von Krebs im Einzelzellbereich ein lang verfolgtes Ziel ist, stellt die Tumorforschung eine treibende Kraft im Bereich der Einzelzell-Omics dar.

<sup>1</sup> Einzelzell-Omics sind ein Forschungsbereich, in dem es um die kollektive Charakterisierung und Quantifizierung einzelner Zellen geht, die sich in der Funktion und Dynamik von Geweben niederschlagen. Dazu gehören mehrere unterschiedliche Technologien (z. B. Transkriptomik, Proteomik oder Metabolomik), die zu vielen unterschiedlichen Ebenen von Informationen über Zellen führen (z. B. über die Transkripte, Proteine oder Metaboliten).

Um das Ökosystem eines Tumors zu begreifen, wurden bereits Einzelzellatlanten von Brustkrebs (Wagner et al., 2019), Kopf- und Hals- (Puram et al., 2017), Lungen- (Lavin et al., 2017; Zilionis et al., 2019) und Nierenkrebs (Chevrier et al., 2017) – um nur einige Beispiele zu nennen – erstellt, die uns wertvolle Einblicke in die Tumorbiologie geben, neuartige Biomarker beschreiben, die Rückschlüsse auf pathogene Prozesse zulassen, und somit neue attraktive Ziele für die therapeutische Intervention definieren.

Die Massenzytometrie ist eine Methode zur Beurteilung von bis zu 50 Proteinen einer einzelnen Zelle und ermöglicht bereits heute einen großen Durchsatz in der Einzelzellanalyse. Die Charakterisierung von 26 Millionen Zellen aus insgesamt 144 Brusttumoren mithilfe von Massenzytometrie hat bereits gezeigt, dass 18 % der Tumore Muster einer starken T-Zell-Erschöpfung in Verbindung mit der Expression des ko-suppressiven Moleküls PD-1 aufweisen (Wagner et al., 2019). T-Zellen erkennen und greifen normalerweise fremde bzw. nicht eigene Zellen (wie z.B. Krebszellen) im Körper an. Jedoch entkommen manche Tumoren diesem Schutzmechanismus unseres Immunsystems. Sie deaktivieren die zytotoxische Aktivität von T-Zellen durch die Expression von PD-L1, dem Bindungspartner für PD-1, und tragen somit zur T-Zell-Erschöpfung bei. In diesem Fall sprechen Patienten normalerweise nicht gut auf Standardtherapien an, oder entwickeln eine Therapieresistenz und Metastasen. Aber mit diesem besonderen Wissen über die Tumorerheterogenität und die Stratifizierung von Patienten könnten Ärzte die Behandlung so anpassen, dass manche Patienten von einer Anti-PD-1- und Anti-PD-L1-Therapie profitieren könnten, die man auch als „Checkpoint-Blockade“ bezeichnet – ein Immuntherapieansatz, der bereits für viele unterschiedliche Tumortypen erfolgreich eingeführt wurde.

Im Gegensatz zu der großen Anzahl der Zellen, deren Proteine mit Massenzytometrie analysiert werden können, sind Einzelzell-Sequenzierungsansätze, die das Transkriptom abbilden, aufgrund der relativ hohen Kosten eingeschränkt. Das führt dazu, dass die Analyse auf kleinere Zellmengen aus Tumorbiopsien beschränkt werden muss. Die Beurteilung einer begrenzten Zellanzahl stellt daher möglicherweise nicht das vollständige Gesamtbild dar, das für die korrekte Diagnose und Klassifizierung des Tumors erforderlich ist, was aber eine Grundvoraussetzung für die richtige Behandlung wäre. Auf diese Aspekte wird jedoch in mehreren Studien auf der ganzen Welt eingegangen, und die Bemühungen, diese Studien zu kombinieren, sind bereits in vollem Gange.

## Einzelzell-Omics am Gehirn und dessen Erkrankungen

Obwohl menschliche Gehirne nur schwer für Biopsien zugänglich sind, wurde inzwischen von der ersten Studie berichtet, die Einzelzell-RNA-Sequenzierung bei der Biopsie von postmortalen Gehirnen von Alzheimer-Patienten anwendet (Mathys et al., 2019). Die Alzheimer-Krankheit (AK) ist eine progressive neurodegenerative Erkrankung, die weltweit für die meisten Fälle altersbedingter Demenz verantwortlich ist. Trotz enormer Forschungsanstrengungen – vorwiegend in Tiermodellen – haben wir noch immer kein umfassendes Verständnis der AK. Auch klinische Studien, die auf Moleküle abzielten, die zuvor hauptsächlich in Tiermodellen identifiziert und charakterisiert worden waren, sind bisher gescheitert. Um diese Krankheit bekämpfen zu können, ist daher eine viel detailliertere Zellanalyse dringend erforderlich – am besten am Menschen. In einer neueren Studie an etwa 80.000 Einzelzell-Transkriptomen von 48 Patienten mit unterschiedlich ausgeprägter AK-Pathologie (Mathys et al., 2019) wurde als Schlüsselfaktor in der AK-Pathophysiologie eine veränderte Myelinisierung identifiziert, die mit einer schnelleren Wanderung von Nervenimpulsen einhergehen könnte. Außerdem wurde bei mehreren Zelltypen, einschließlich der Myelin-produzierenden Oligodendrozyten, eine geschlechtsabhängige molekulare Reaktion festgestellt. Diese Studie ist ein hervorragendes Beispiel dafür, wie Einzelzell-Omics unser Verständnis von verbreiteten Erkrankungen schärft, die für unsere Gesellschaft eine wichtige Rolle spielen. Nur die Einzelzellauflösung ermöglichte die Klärung dieser neuartigen pathophysiologischen Mechanismen, die jetzt mit ganz neuen therapeutischen Strategien angegangen werden können. Dies ist nur die Spitze des Eisbergs, weitere Erkenntnisse bei Erkrankungen des Gehirns und anderer wichtiger Organe sind zu erwarten.

### 3.3 EINZELZELL-OMICS HEUTE

#### Einzelzell-Omics entwickeln sich schnell und halten Einzug in die klinische Forschung

Die Grundlagenforschung profitiert im Moment stark von den Fortschritten mit neuartigen Einzelzelltechnologien. Dieses neue Auflösungs-niveau ermöglicht eine ungeahnte Sicht auf heterogene Zellpopulationen, die Gewebezusammensetzung sowie die Infiltration veränderter Immunzellen bei Krankheiten. Neu entwickelte Techniken können nun zudem auf mechanische oder enzymatische Zellvereinzelnung

verzichten und so zusätzlich räumliche Informationen innerhalb des Gewebes integrieren. Das prominenteste Beispiel aktueller Forschungsprojekte in diesem Bereich ist der Human Cell Atlas (HCA) – ein großer internationaler Zusammenschluss vieler Wissenschaftler zur systematischen, hochaufgelösten Kartierung aller Zellen des menschlichen Körpers als Basis für ein besseres Verständnis fundamentaler humanbiologischer Prozesse und schließlich als Referenzquelle zur Beleuchtung pathologischer Vorgänge (HCA, 2017; Regev et al., 2017). Neben dem HCA zur Charakterisierung unserer gesunden Zellen bilden sich weltweit große Konsortien zur Anwendung von Einzelzell-Omics auf krankheitsbezogene Fragen. Darunter wird die europäische LifeTime-Initiative (siehe hierzu Junker, Popp, Rajewsky, Kapitel 2) als eines der vielversprechendsten Expertennetzwerke wahrgenommen, die künftige Herausforderungen der Präzisionsmedizin mithilfe von Einzelzell-Omics-Technologien angehen wollen.

#### **Die Anwendung von Einzelzell-Omics in klinischen Studien ist zum Greifen nah**

Fortschritte in den Einzelzell-Omics haben es möglich gemacht, von reinen Machbarkeitsstudien zur Anwendung von Einzelzell-Omics in breiteren Zusammenhängen überzugehen. In der Grundlagenforschung sind die wichtigsten Ziele das Verständnis von Differenzierungsprozessen in der Entwicklung, die Beleuchtung von Immunzellheterogenität in klassisch definierten Zellpopulationen und die Untersuchung der Pathogenese in Modellsystemen. Neuere technische und rechnergestützte Fortschritte ermöglichen inzwischen, größere klinische Studien anzugehen, und eröffnen neue Wege sowohl zur Erforschung von Krankheiten als auch zur Entwicklung von Diagnostik, Therapien und Therapiemanagement (siehe Junker, Popp, Rajewsky, Kapitel 2).

Da Blut als humane Gewebeprobe am zugänglichsten ist, standen isolierte periphere Blutzellen lange im Fokus der Erforschung von Krankheiten oder dienen als Stellvertreter für Krankheiten in anderen Organen. Einzelzellansätze haben bereits zu Erkenntnissen zellulärer Heterogenität ausgebreiteter zirkulierender Immunzellpopulationen bei Leukämie geführt und Daten generiert, die mit klinischen Ergebnissen verknüpft werden können, beispielsweise zur Entwicklung von Signaturen zur Vorhersage der Überlebenschancen eines Patienten (Gawad, 2014; Levine, 2015). Humane Festgewebeproben sind schwieriger zu bekommen, aber unglaublich nützlich zur Untersuchung der Zellzusammensetzung und des funktionellen Zustandes der Zellen im erkrankten Gewebe.

### 3.4 EINZELZELL-OMICS MORGEN

#### Erweiterung der Anwendung von Einzelzell-Technologien auf größere Patientenkohorten

Die Einzelzell-Omics sind mittlerweile technisch so weit entwickelt, dass sie auch auf größere Patientenkohorten mit klinisch relevanten Erkrankungen angewendet werden können – ein Weg, der unser Verständnis von Krankheiten auf eine ganz neue Ebene heben wird. Umfassende molekulare Profilierung – unabhängig von bisherigen Annahmen oder Erkenntnissen – mithilfe von Einzelzell-Omics-Technologien wird es uns ermöglichen, die molekularen Pfade sowie die an pathophysiologischen Vorgängen beteiligten Moleküle innerhalb jeder einzelnen Zelle zu definieren. Das ist als Grundlage für die Präzisionsmedizin von entscheidender Wichtigkeit, da sie auf eine Verbindung zwischen molekularen Mechanismen mit Einzelzellauflösung und klinischen Phänotypen angewiesen ist. Wir erwarten, dass sich hieraus eine Veränderung der traditionellen Krankheitsklassifizierungen ergeben wird. Die Stratifizierung von Patienten wird sich anhand der durch Einzelzell-Omics gelieferten Informationen verbessern und präzisieren. Dadurch wird auch eine genauere Identifizierung der entsprechenden Biomarker, einschließlich derer zur Vorhersage und Überwachung von Krankheiten, ermöglicht. Zudem wird es möglich sein, neuartige und auch weniger gut definierte, seltene Krankheiten oder klinische Fälle mit unbestimmter Diagnose umfassend zu charakterisieren. Das gewonnene Wissen wird die Ansätze der Präzisionsmedizin wesentlich vorantreiben.

Da perspektivisch auch Vorhersagen des Ansprechens auf verfügbare Arzneimittel gemacht werden können, können durch das Screening von einzelnen Patienten fundierte Entscheidungen über geeignete therapeutische Wege getroffen werden, die molekulare Befunde auf Einzelzellebene berücksichtigen. Ein weiterer Anwendungsbereich werden zellbasierte Therapien sein, bei denen eine Einzelzellanalyse helfen kann, die verwendeten Zellpopulationen besser zu charakterisieren und zu verfeinern. Die Reinheit der Produkte zur CAR-T-Zell-Therapie<sup>2</sup> (CAR: chimärer Antigenrezeptor) könnte beispielsweise durch gezielte Einzelzell-Omics-Analysen vor der Produktanwendung am Patienten verbessert werden.

2 Bei dieser Therapie werden patienteneigene Immunzellen mit gentherapeutischen Methoden in die Lage versetzt, Tumorzellen zu erkennen und zu vernichten.

## Die klinische Anwendung bei immunvermittelten Erkrankungen ist äußerst vielversprechend

Die Humanimmunologie steht an der Spitze der Anwendung von Einzelzell-Omics-Technologien, da sich Immunzellen leicht aus peripherem Blut gewinnen lassen (Schultze/Aschenbrenner, 2019; Bassler et al., 2019). Außerdem ist bei Blut für nachgelagerte analytische Verfahren kein Gewebeaufschluss erforderlich, weil die Zellen bereits vereinzelt sind. Immunologen können ferner auf profundem Wissen aus molekularer Profilierung von Einzelzellen mithilfe von Durchflusszytometrie aufbauen – ein Messverfahren, das es ermöglicht, die Expression mehrerer Proteine in einer Zelle nachzuweisen, und seit fast 50 Jahren dazu genutzt wird, Immunzellen zu beschreiben und anhand von Zelloberflächenmarkern zu definieren. Deshalb kann davon ausgegangen werden, dass Krankheiten mit einer immunologischen Komponente, wie zum Beispiel chronische Infektionen, metabolische Syndrome, neurodegenerative Erkrankungen oder Krebsformen, für die Profilierung ganz oben auf der Liste stehen.

Solche Systemansätze, die an immer größeren Kohorten angewendet werden, werden auch die Variabilität zwischen einzelnen Patienten bezüglich ihrer Organfunktionen, inklusive der des Immunsystems, aufklären. Es hat sich deutlich abgezeichnet, dass die Kombination aus genetischer Anfälligkeit und Umweltfaktoren Einfluss auf die Gesundheit eines Menschen hat. Was das Immunsystem angeht, konnte kürzlich gezeigt werden, dass Immunseneszenz, das Altern des Immunsystems, stark von genetischen und umweltbedingten Einflussfaktoren abhängt und stärker mit dem Ausgang von Krankheiten und deren Entwicklung in Zusammenhang steht als das tatsächliche chronologische Altern (Alpert et al., 2019). Das große Potenzial der Beleuchtung von Variationen in Immunreaktionen bei gesunden Probanden wurde ebenfalls im Rahmen des Human Functional Genome Project gezeigt (HFGP; Ter Horst et al., 2016; Li et al., 2016). Wir gehen davon aus, dass solche großen Kohorten das Grundgerüst für künftige Einzelzell-Omics-Ansätze zur Beantwortung klinischer Fragen bilden werden. Diese Kohortenstudien liefern bereits heute wichtige Informationen zum Einfluss der Umwelt und zur genetischen Anfälligkeit von Immunfunktionen, die bei verbreiteten entzündlichen Erkrankungen (Krebs, Autoimmunerkrankungen, chronische Infektionen) gestört sind. Neben der Charakterisierung von Krankheitspathologie im Einzelzellbereich wird die weitere Abgrenzung der Faktoren, die sich auf Variationen im menschlichen Immunsystem auswirken, eine der

wichtigen Aufgaben für die nahe Zukunft sein, da dadurch die Entwicklung von Ansätzen der Präzisionsmedizin und Vorhersage von Erkrankungsrisiken gestützt werden. Die Häufigkeit einer bestimmten Untergruppe von Monozyten erwies sich beispielsweise als stark prädiktiv für den Behandlungsausgang mit einer Anti-PD-1-Immuntherapie, was zeigt, wie hochdimensionale Einzelzellanalysen das Ansprechen auf eine Checkpoint-Blockade vorhersagen kann (Krieg et al., 2018).

### Multidisziplinäre Teams werden klinische Fragen angehen

Die kooperative Arbeit von Klinikärzten, Biologen, Bioinformatikern, KI-Spezialisten, aber auch Entwicklern medizinischer Geräte und Ingenieuren wird ein besonders wichtiger Bereich zur Ableitung medizinisch relevanter Implikationen aus der Vielzahl von Informationen sein, die Ansätze der Einzelzell-Omics hervorbringen. Wir gehen davon aus, dass Einzelzellansätze krankheitsspezifische Signaturen liefern werden, die anstelle herkömmlicher analytischer Parameter, die oft nicht spezifisch genug sind (wie z.B. die Anzahl weißer Blutkörperchen), für die Diagnostik eingesetzt werden können. Mit der laufenden Verbesserung der Techniken und Erfahrungen im Bereich der Einzelzell-Omics werden die Kosten langsam sinken, sodass diese viel tiefergreifenden Ansätze klinisch anwendbar werden. Das gewonnene Wissen kann eine Umstellung auf Tests definierter Marker oder sogar zurück zur Sequenzierung von Zellgemischen (z.B. aus Blut) begünstigen, sodass Einzelzell-Omics auch aus klinischer Sicht noch attraktiver werden könnten.

### Ausblick – Künftige Entwicklungen und Erfordernisse

Einzelzell-Omics sind ein Gebiet, das sich extrem schnell weiterentwickelt. Die folgenden wichtigen Aspekte werden dieses Gebiet in den nächsten zehn Jahren vorantreiben. Zusätzlich zu Einzelzell-Omics von Zellen, die natürlicherweise (z. B. aus Blut) oder artifiziell (aus Gewebe) vereinzelt wurden, werden Technologien zum Erhalt von *räumlicher Information* über den Ursprung einzelner Zellen im klinischen Bereich zunehmend an Bedeutung gewinnen. Dies liegt an der Notwendigkeit, pathophysiologische Mechanismen im räumlichen Kontext zu verstehen. Zahlreiche unterschiedliche Technologien für die räumliche Einzelzell-Omics werden aktuell entwickelt. Dies ist ein Bereich mit großem Potenzial, insbesondere im Zusammenhang mit klinischen Anwendungen. Problematischer

wird dann aber die Beurteilung der zeitlichen Komponente einer Krankheit. Mit Ausnahme von Blut ist es schwierig, wiederholte Biopsien aus festen Organen zur Analyse mittels Einzelzell-Omics zu nehmen. Eine Strategie für dieses Problem könnte die Entwicklung blutbasierter Einzelzell-Omics als Surrogat für die Vorgänge im Gewebe sein. Das zweite Erfordernis ist die Anwendung von *KI-Methoden* auf alle analytischen Wege für Einzelzell-Omics-Daten. Diese Daten sind sowohl umfangreich als auch spärlich, was ganz besondere Herausforderungen mit sich bringt. Basierend auf ersten erfolgreichen Anwendungen von KI-Methoden (Eraslan et al., 2019) gehen wir davon aus, dass dieser Bereich für klinische Anwendungen von entscheidender Bedeutung sein wird (siehe auch Aliee, Sacher, Theis, Kapitel 4). Zur Steigerung des prädiktiven Werts von Modellsystemen werden *humane Organoide* zusammen mit Analysen auf Einzelzellebene unser Verständnis für Humanbiologie und verbreitete Erkrankungen weiter voranbringen (Velasco et al., 2019; Bolhaqueiro et al., 2019; Klaus et al., 2019; Xiang et al., 2019; Gehart et al., 2019; Roerink et al., 2018; siehe auch Junker, Popp, Rajewsky, Kapitel 2). Die Kombination aus Einzelzellanalyse, humanen Organoiden und KI wird außerdem die Entwicklung besserer Tiermodelle vorantreiben, die auch weiterhin erforderlich sein werden, um kausale Zusammenhänge zwischen molekularen Mechanismen zu bestimmen, die für verbreitete Erkrankungen verantwortlich sind. Die wahrscheinlich wichtigste Anforderung ist aber die Entwicklung von geeigneten Forschungsstrukturen, die alle genannten und für die Einzelzell-Omics relevanten unterschiedlichen Fachkompetenzen miteinander verbinden können (z. B. in internationalen Netzwerken). Für die klinische Entwicklung der Einzelzell-Omics ist es von entscheidender Bedeutung, dass diese Sektoren für eine nahtlose Zusammenarbeit effizient miteinander verbunden werden können.

### 3.5 LITERATUR

Alpert, A. et al. (2019): A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. In: *Nat. Med.* 25(3): 487–495.

Bassler, K. et al. (2019): The myeloid cell compartment-cell by cell. In: *Annu. Rev. Immunol.* 37: 269–239.

Bolhaqueiro, A. C. F. et al. (2019): Ongoing chromosomal instability and karyotype evolution in human colorectal cancer organoids. In: *Nat. Genet.* 51: 824–834.

Chevrier, S. et al. (2017): An immune atlas of clear cell renal cell carcinoma. In: *Cell* 169(4): 736–749.e18.



- Eraslan, G. et al. (2019): Single-cell RNA-seq denoising using a deep count autoencoder. In: *Nat. Commun.* 10(1): 390.
- Gawad, C. et al. (2014): Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. In: *Proc. Natl. Acad. Sci.* 111(50): 17947–17952.
- Gehart, H. et al. (2019): Identification of enteroendocrine regulators by real-time single-cell differentiation mapping. In: *Cell* 176(5): 1158–1173.e16.
- HCA (The Human Cell Atlas Consortium) (2017): The human cell atlas white paper, Online-Publikation 18.10.2017. Unter: [https://www.humancellatlas.org/files/HCA\\_WhitePaper\\_18Oct2017.pdf](https://www.humancellatlas.org/files/HCA_WhitePaper_18Oct2017.pdf) [25.06.2019].
- Klaus, J. et al. (2019): Altered neuronal migratory trajectories in human cerebral organoids derived from individuals with neuronal heterotopia. In: *Nat. Med.* 25(4): 561–568.
- Krieg, C. et al. (2018): High-dimensional single-cell analysis predicts response to anti-PD-1 immunotherapy. In: *Nat. Med.* 24(4): 144–153.
- Lavin, Y. et al. (2017): Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses. In: *Cell* 169(4): 750–765.e17.
- Levine, J. H. et al. (2015): Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. In: *Cell* 162(1): 184–197.
- Li, Y. et al. (2016): A functional genomics approach to understand variation in cytokine production in humans. In: *Cell* 167(4): 1099–1110.e14.
- Mathys, H. et al. (2019): Single-cell transcriptomic analysis of alzheimer's disease. In: *Nature*, Online-Publikation 01.06.2019. DOI: 10.1038/s41586-019-1195-2.
- Puram, S. V. et al. (2017): Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. In: *Cell* 171(7): 1611–1624.e24.
- Regev, A. et al. (2017): The human cell atlas. In: *eLife* 6, Online-Publikation 05.12.2017. DOI: 10.7554/eLife.27041.
- Roerink, S. F. et al. (2018): Intra-tumour diversification in colorectal cancer at the single-cell level. In: *Nature* 556(7702): 457–462.
- Schultze, J. L./Aschenbrenner, A. C. (2019): Systems immunology allows a new view on human dendritic cells. In: *Semin. Cell Dev. Biol.* 86: 15–23.
- Ter Horst, R. et al. (2016): Host and environmental factors influencing individual human cytokine responses. In: *Cell* 167(4): 1111–1124.e13.

Velasco, S. et al. (2019): Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. In: *Nature*, Online-Publikation 05.06.2019. DOI: 10.1038/s41586-019-1289-x.

Wagner, J. et al. (2019): A single-cell atlas of the tumor and immune ecosystem of human breast cancer. In: *Cell* 177(5): 1330-1345.e18.

Xiang, Y. et al. (2019): hESC-derived thalamic organoids form reciprocal projections when fused with cortical organoids. In: *Cell Stem Cell* 24(3): 487–497.e7.

Zilionis, R. et al. (2019): Single-cell transcriptomics of human and mouse lung cancers reveals conserved myeloid populations across individuals and species. In: *Immunity* 50(5): 1317–1334.e10.

## 4. ANALYSE VON EINZELZELLENOMIK-DATEN MIT METHODEN DES MASCHINELLEN LERNENS

### 4.1 EINFÜHRUNG

Sowohl in der Wissenschaft als auch in der Industrie werden Datensätze stetig größer, da immer mehr moderne Geräte und Technologien sogenannte *Big Data* generieren und sammeln. „Big Data“ ist ein Begriff für große und komplexe, oft unstrukturierte Datenmengen, die eine Bearbeitung mit rechnerbasierten Hilfsmitteln erforderlich machen, um bedeutsame Informationen daraus ziehen zu können. Die Analyse solcher Daten wird als „Data Science“ (Datenwissenschaft) bezeichnet. Sie eröffnet neue Möglichkeiten der Kombination von Daten aus unterschiedlichen Quellen und ermöglicht so tiefere Einblicke in ein spezifisches Problem für eine bessere Entscheidungsfindung.

Zur Extraktion von Werten aus Daten kommen oft Methoden des *maschinellen Lernens* (ML) zum Einsatz, eine der großen Triebfedern der Big-Data-Revolution. ML ist ein Unterbegriff der *künstlichen Intelligenz* (KI), mit der im Allgemeinen menschliche Intelligenz für bestimmte Aufgaben nachgeahmt werden soll. Genauer bezeichnet ML unterschiedliche Berechnungsalgorithmen für autonomes Lernen aus gekennzeichneten und ungekennzeichneten Daten<sup>1</sup>, um datengestützte Einblicke zu liefern und Entscheidungsfindungen sowie Vorhersagen zu unterstützen. Durch die immer weiter anwachsende Datenflut kann es allerdings sein, dass konventionelles ML nicht mehr leistungsstark genug ist, um die Komplexität innerhalb der Daten richtig zu erfassen. Aus diesem Grund entstand *Deep Learning* (DL) als neuer Bereich des ML. Deep Learning ist eine ML-Technik basierend auf künstlichen neuronalen Netzwerken, die einfache nicht-lineare Verarbeitungseinheiten<sup>2</sup> in mehreren Schichten verknüpfen. Die Architektur von Deep-Learning-Modellen kann komplizierte, hierarchische, statistische Muster innerhalb von Datensätzen überwacht („supervised“, beispielsweise zur

- 1 Ungekennzeichnete Daten („unlabeled data“) sind Daten ohne zugehörige Informationen, wohingegen gekennzeichnete Daten („labeled data“) zusätzliche Informationen über diese Daten umfassen.
- 2 Diese Einheiten, die man als *künstliche Neuronen* bezeichnet, bilden grob die Neuronen in einem biologischen Gehirn nach. Über eine Verbindung kann ein Signal von einem künstlichen Neuron an ein anderes weitergegeben werden. Das empfangende Neuron verarbeitet das Eingangssignal und leitet es an andere verbundene künstliche Neuronen weiter.

Klassifizierung) oder nicht überwacht („unsupervised“, z. B. für Gruppierungen) erfassen.<sup>3</sup> Der Hauptvorteil von DL-Algorithmen ist, dass sie schrittweise immer komplexere Merkmale aus Daten lernen können. Diese Merkmalsauswahl benötigt kein Expertenwissen für das vorliegende Problem, allerdings sind für gewöhnlich auch größere Datensätze erforderlich.

DL hat in den letzten Jahren viele Bereiche wie *Computervision* (zum maschinellen Verständnis digitaler Bilder und Videos) und *Natural Language Processing* (zum maschinellen Verständnis natürlicher Sprache) revolutioniert und breite Anwendungsgebiete u. a. in der Astronomie, Robotik, im Finanz- und Gesundheitswesen gefunden. In diesem Kapitel liegt der Fokus auf der Gesundheitsforschung, insbesondere auf der Genomik. Dieser Bereich hat in den letzten Jahren durch neue Fortschritte in biomedizinischen Techniken wie *Next-Generation-Sequencing* (NGS), die heute routinemäßig riesige Mengen genomischer Daten generieren, ein exponentielles Wachstum gezeigt. Bei NGS-basierten Technologien wie Genomik, Transkriptomik, Proteomik und Epigenomik<sup>4</sup> liegt der Fokus immer stärker auf der Profilerstellung einzelner Zellen. Im Gegensatz zu traditionellen Profilierungsmethoden zur Beurteilung großer Zellpopulationen werden bei Einzelzelltechnologien einzelne Zellen isoliert, um zellspezifische Sequenzierungsbibliotheken, eine Sammlung ähnlicher Moleküle, zu erstellen. Dabei wird jede Zelle individuell mit einem zellspezifischen molekularen Barcode markiert. Einzelzelltechnologien ermöglichen dann eine Profilierung der Informationen über Tausende bis Millionen einzelner Zellen in einem einzigen Experiment. Dadurch ist es inzwischen möglich geworden, die Heterogenität auch zwischen ähnlichen Zelltypen aufzudecken (siehe Aschenbrenner, Mass, Schultze, Kapitel 3) und potenziell komplexe und seltene Zellpopulationen, Zelldynamiken, regulatorische Beziehungen zwischen Genen und sog. Entwicklungstrajektorien<sup>5</sup> bestimmter Zelllinien zu erkennen (Hwang et al., 2018; siehe Junker, Popp, Rajewsky, Kapitel 2). Die Komplexität der Einzelzelldaten gepaart mit ihrem

3 „Supervised“ bedeutet, dass Funktionen auf der Basis bekannter Datensätze abgeleitet werden, die die Klassifizierung unbekannter Daten ermöglichen. „Unsupervised“ bedeutet, dass unbekannte Daten untersucht und Strukturen innerhalb dieser Daten identifiziert werden, was dann Gruppierungen ermöglicht.

4 In der Genomik geht es um die Untersuchung des gesamten Genoms, in der Transkriptomik um die Untersuchung aller Transkripte von Genen (Genexpressionsprodukte, RNA), in der Proteomik wird die Gesamtheit aller Proteine untersucht, und in der Epigenomik alle epigenetischen Daten innerhalb von Zellen. Siehe Walter/Gasparoni, Kapitel 1.

5 Eine Entwicklungstrajektorie bezeichnet den Verlauf einer Entwicklung über die Zeit hinweg. Eine solche Entwicklungstrajektorie kann aus Einzelzelldaten rekonstruiert werden, wenn man Zellen zu vielen einzelnen Zeitpunkten vermisst und sie dann entlang einer virtuellen Zeitachse anordnet (pseudotemporale Trajektorie).

gigantischen Umfang macht die Einzelzellanalyse zu einem paradigmatischen Fall von Big Data. Daraus ergibt sich die Notwendigkeit der Entwicklung von Analysemethoden, die große Datensätze über eine große Anzahl von Zellen verarbeiten können. Dieses Kapitel liefert einen Überblick über die Einzelzell-Transkriptomik als eine der populärsten Einzelzelltechnologien und deren Herausforderungen und Möglichkeiten aus der Perspektive der modernen Analytik basierend auf ML und DL.

## 4.2 MASCHINELLES LERNEN IN DER EINZELZELL-TRANSKRIPTOMIK

Die Einzelzell-RNA-Sequenzierung (Single-cell-RNA-Sequencing, scRNA-seq) umfasst die Profilierung aller Messenger-RNA (mRNA)<sup>6</sup> in einer einzelnen Zelle und liefert das Genexpressionsprofil von Hunderttausenden und sogar Millionen individueller Zellen. Damit erhebt scRNA-seq Big Data mit überlegener statistischer Aussagekraft, die neue Möglichkeiten der Anwendung von maschinellem Lernen und Deep Learning für die Einzelzelldatenanalyse eröffnen.

Da sich die Messungen einzelner Zellen im Femtoliterbereich bewegen, sind technische Störfaktoren erhöht und die erhaltenen Expressionswerte oft verrauscht.<sup>7</sup> Daraus ergeben sich mehrere Herausforderungen für die Berechnung und Statistik in der Erkennung von Mustern in der Genexpression, zum Beispiel von Zelltypen. Häufig wird eine zusätzliche Qualitätskontrolle durchgeführt, um fehlerhafte Messungen auszuschließen, verursacht zum Beispiel durch sogenannte Ausreißer („drop-out effect“) oder mögliche Dubletten,<sup>8</sup> gefolgt von einer *Normalisierung*,<sup>9</sup> bei der zellspezifische Unterschiede in der Sequenzierung und andere technische Störfaktoren korrigiert werden. Anschließend wird eine *Merkmalsauswahl*<sup>10</sup> und

6 Messenger-RNA werden vom Genom der Zelle abgeschrieben und am Ribosom in Proteine übersetzt. Sie stellen damit das aktive Genom dar.

7 „Rauschen“ bedeutet in diesem Zusammenhang, dass die Daten in Bezug auf die zu untersuchende Frage irrelevante oder zufällige Signale enthalten, die zunächst herausgefiltert werden müssen, um signifikante Signale identifizieren zu können.

8 Als „Ausreißer“ bezeichnet man Zellen, die vom durchschnittlichen Expressionsgrad ihres Zelltyps abweichen und somit die Identifizierung gewöhnlich exprimierter Gene erschweren. Dubletten sind Expressionsprofile, die versehentlich aus zwei und nicht aus einer Zelle generiert werden, oft infolge von Fehlern beim Sortieren oder Erfassen der Zellen. Sie können die korrekte Interpretation der Ergebnisse beeinträchtigen, z. B. indem sie auf das Vorhandensein von Zwischenpopulationen oder Übergangsstadien hinweisen, die es in Wahrheit gar nicht gibt.

9 Bei der Normalisierung werden Auszählungsdaten skaliert, um eine korrekte relative Fülle von Genexpressionen zwischen verschiedenen Zellen zu erhalten.

10 Bei der Merkmalsauswahl wird der Datensatz gefiltert, um nur Merkmale/Variablen (in diesem Zusammenhang: Gene) zu behalten, die etwas über die Variabilität innerhalb der Daten aussagen.

eine *Dimensionsreduktion*<sup>11</sup> durchgeführt, wobei die am meisten informativen Gene und die stärksten Signale aus dem Hintergrundrauschen herausgefiltert werden (Luecken/Theis, 2019). Die Qualitätskontrolle ist für die nachgeschaltete Analyse, zum Beispiel die Gruppierung von Zellen zur Erkennung von Teilpopulationen oder zur Ableitung künftiger Zellentwicklungen, erforderlich. Im Folgenden werden einige Anwendungsgebiete überwachter und nicht überwachter Lerntechniken in der Downstream-Analyse transkriptomischer Daten erörtert.

## Überwachtes Lernen in der Einzelzell-Transkriptomik

Das überwachte Lernen („supervised learning“) ist ein Bereich des maschinellen Lernens, der Training mit bekannten Daten erfordert, sodass aus diesen gekennzeichneten Daten eine Funktion abgeleitet werden kann, die sich zur Kartierung von ungekennzeichneten Daten und Ausgangsvariablen eignet. Ein überwachtes Lernmodell wird zunächst mit einem *Trainingssatz* aus Eingangszielpaaren trainiert, damit es die Modellparameter lernt. Um messen zu können, wie gut eine Funktion zum Trainingssatz passt, wird eine Verlustfunktion definiert, um Fehler in der Vorhersage zu bestimmen. Dadurch sollen die Modellparameter durch Minimierung der Vorhersagefehler optimiert werden. Außerdem wird das Modell mit einem bestimmten *Validierungssatz* verifiziert und anschließend die Leistung der abgeleiteten Funktion mit einem vom Trainingssatz separaten *Testsatz* beurteilt. Die Genauigkeit der Vorhersagen wird mit unterschiedlichen Beurteilungsmessgrößen wie zum Beispiel dem Pearson-Korrelationskoeffizienten gemessen. Die Hauptanwendungsbereiche von überwachtem Lernen sind Klassifizierung und Regression.<sup>12</sup>

Im Bereich der Einzelzell-Transkriptomik wird überwachtes Lernen vor allem zur *Zellannotation* verwendet. Hier werden unbekanntes Zellen Zelltypen anhand einer Reihe von Referenzdatensätzen mit gekennzeichneten Zelltypen zugeordnet. Dies entspricht im Bereich der Genomikdaten in etwa dem routinemäßigen Einsatz der Durchflusszytometrie<sup>13</sup> zur Diagnose von Erkrankungen wie Blutkrebs. Konventionell wurden Zellen basierend auf einer Reihe von Markern annotiert,

11 Dimensionsreduktion ist ein Verfahren zur Senkung der Anzahl von Zufallsvariablen durch Identifizierung einer gewissen Anzahl relevanter Variablen.

12 Regression ist eine Reihe statistischer Verfahren zur Einschätzung des Verhältnisses zwischen den Variablen.

13 Durchflusszytometrie ermöglicht das Bestimmen von molekularen und physikalischen Eigenschaften von Zellen mithilfe von Fluoreszenz-markierten Proben wie z. B. Antikörpern.

was arbeitsaufwendig ist und eine umfassende Literaturübersicht über zelltypspezifische Gene erforderlich macht. Außerdem variieren diese Gene oft zwischen einzelnen experimentellen Gegebenheiten, was den Vergleich der Ergebnisse erschwert (Pliner et al., 2019). Klassische Techniken des überwachten Lernens sind besser, da sie aus den gekennzeichneten Daten automatisch wichtige Merkmale (oder Gene) erfassen, was eine exaktere Zellannotation und weniger Diskrepanzen zwischen den Klassifizierungen verschiedener Experimente ermöglicht. In der rechnerbasierten Analyse werden zahlreiche Klassifizierungsmodelle wie zum Beispiel die logistische Regression, *Support Vector Machines* und *Random Forests* eingesetzt. Allerdings werden bei immer größerem Datenvolumen Deep-Learning-Modelle bei Zellannotationsaufgaben gegebenenfalls den klassischen Machine-Learning-Modellen vorgezogen.

### Unüberwachtes Lernen in der Einzelzell-Transkriptomik

Beim unüberwachten Lernen („unsupervised learning“) werden nützliche Strukturen oder Muster aus nicht gekennzeichneten Datensätzen abgeleitet. Klassisch kommen nicht überwachte Lernalgorithmen zur Gruppierung von Daten, für die Dimensionsreduktion und zur Visualisierung und Einbettung zum Einsatz. Künstliche neuronale Netzwerke können einige dieser Ansätze generalisieren. *Autoencoder* komprimieren die Daten beispielsweise in einen niedrigdimensionalen Code und dekomprimieren diesen dann zur Rekonstruktion der ursprünglichen Eingangsdaten. Ein Autoencoder ermöglicht aber nur die ungefähre Kopie der Eingangsdaten in den Output, was das Modell dazu zwingt, eine Dimensionsreduktion durchzuführen, indem es lernt, Rauschen zu ignorieren. Im Bereich der Einzelzell-Transkriptomik kommen Autoencoder für Entrauschen sowie zur Dimensionsreduktion zum Einsatz. Einbettungstechniken wie t-SNE können zur Kartierung der komprimierten Daten in einer 2D-Ebene angewendet werden. Spezifische Geräuschmerkmale der scRNA-seq-Daten wie zum Beispiel *Zero-Inflated Negative Binomial* (ZINB) können ebenfalls mit maßgeschneiderten Verlustfunktionen innerhalb des Autoencoder-Rahmens behandelt werden (Eraslan et al., 2019).

Eine weitere leistungsstarke Anwendung von nicht überwachtem Lernen ist die *Clusteranalyse* zur Definition von Zelltypen innerhalb der scRNA-seq-Daten. Das Ziel der Clusteranalyse ist die Gruppierung von Zellen basierend auf Ähnlichkeiten in deren Genexpressionsprofilen. Die Clusteranalyse ist die

Basis mehrerer Atlasprojekte, insbesondere des Human Cell Atlas.<sup>14</sup> Bei solchen Forschungsprojekten werden mehrere Einzelzell Datensätze in einen Atlas aufgenommen, der umfassende Referenzkarten aller menschlichen Zellen enthält. Damit ein Zellatlas einen bestimmten Zweck erfüllen kann, ist eine der wichtigsten rechnerischen Herausforderungen die Entwicklung zuverlässiger Methoden zur unüberwachten Gruppierung der Zellen (Kiselev et al., 2019).

Die Zelldiversität wird allerdings in einem einzelnen Klassifizierungssystem wie dem Clustering nicht ausreichend beschrieben. Immerhin sind die biologischen Prozesse, die zur Entwicklung der beobachteten Heterogenität führen, kontinuierliche Prozesse. Um also Übergänge zwischen Zelltypen, Prozesse der verzweigten Differenzierung oder graduelle, nicht synchronisierte Veränderungen in der biologischen Funktion erfassen zu können, benötigen wir dynamische Modelle der Genexpression. Diese Methodenklasse wird als *Trajectory Inference* bezeichnet. In der Trajektorienanalyse (siehe Junker, Popp, Rajewsky, Kapitel 2) werden die Daten als Schnappschuss eines dynamischen Prozesses bzw. auf einer virtuellen Zeitachse abgebildet. Die Zellen werden dann entlang einer solchen virtuellen Zeitachse geordnet und durch eine kontinuierliche Variable beschrieben, die man als *Pseudozeit* bezeichnet. Die Pseudozeitanalyse basiert oft auf der transkriptionellen Distanz von Zellen zu einer Ursprungszelle und beschreibt die Entwicklung als Übergang im transkriptomischen Zustand (d.h. Trajektorie), und nicht als Übergang in Echtzeit. Die pseudotemporale Ordnung von Zellen hilft dabei, zu verstehen, wie sich Zelltyphäufigkeiten als Reaktion auf Entwicklungs- oder Umweltsignale verändern, die physiologischen Mechanismen von Gesundheit und Krankheit unterliegen. Beispielsweise wird damit festgestellt, wie die Häufigkeit eines bestimmten Zelltyps im Laufe eines Prozesses abnimmt, indem die Sterblichkeitsrate steigt oder eine Differenzierung zu anderen Zelltypen stattfindet. Es ist dabei wichtig, die Ursache einer solchen Verschiebung zu verstehen, insbesondere wenn der Prozess im Zusammenhang mit einer Erkrankung steht.<sup>15</sup> Eine weitere interessante Frage, die eine Pseudozeitanalyse beantworten kann, ist die Frage, wie sich Stammzellen oder Vorläuferzellen differenzieren, um ein Organ aus unterschiedlichen Zelltypen auszubilden. Um die allgemeine Topologie der Daten zu verstehen und Trajektorien abzuleiten, kommen häufig nicht-lineare Methoden der Dimensionsreduktion (z. B. *Manifold Learning*) zum Einsatz (Wolf et al., 2019).

<sup>14</sup> Siehe: <https://www.humancellatlas.org/> [21.06.2019] und Walter/Gasparoni, Kapitel 1.

<sup>15</sup> Ein Beispiel hierfür wäre der Zusammenhang zwischen einem Rückgang in der Häufigkeit von Betazellen der Bauchspeicheldrüse und Diabetes.



### 4.3 AUSBLICK

Die Einzelzell-RNA-Sequenzierung ist eine leistungsstarke Methode zur Entdeckung interzellulärer Heterogenität. Dabei liegt der Fokus auf der Charakterisierung individueller Zellen. Dadurch können komplexe und seltene Zellpopulationen identifiziert werden, ebenso wie regulatorische Beziehungen zwischen Genen und die Trajektorien bestimmter Zelllinien in der Entwicklung. Mehrere Studien haben gezeigt, wie nützlich scRNA-seq Technologien insbesondere zur Untersuchung der frühen embryonalen Entwicklung sowie zur Erkennung der Komplexität von Krebs und anderen Erkrankungen sind. Sowohl die Komplexität als auch der gigantische Umfang der Einzelzelldaten stellen große Herausforderungen in der Datenanalyse dar. Dazu kommt, dass es sich bei der Einzelzellanalyse um ein neues Forschungsfeld handelt, für das standardisierte Analysemethoden erst noch entwickelt werden müssen.

### 4.4 LITERATUR

Eraslan, G. et al. (2019): Single-cell RNA-seq denoising using a deep count autoencoder. In: *Nat. Commun* 10(1): 390.

Hwang, B. et al. (2018): Single-cell RNA sequencing technologies and bioinformatics pipelines. In: *Exp. Mol. Med.* 50(8): 96.

Kiselev, V. Y. et al. (2019): Challenges in unsupervised clustering of single-cell RNA-seq data. In: *Nat. Rev. Genet* 20(5): 273–282.

Luecken, M. D./Theis, F. J. (2019): Current best practices in single-cell RNA-seq analysis: a tutorial. In: *Mol Syst Biol* 15: e8746.

Pliner, H. A. et al. (2019): Supervised classification enables rapid annotation of cell atlases. In: *bioRxiv*, Online-Publikation 04.02.2019. DOI: 10.1101/538652.

Wolf, F. A. et al. (2019): PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. In: *Genome Biol* 20(1): 59.

## 5. EINZELZELL-TRANSKRIPTOMANALYSE IN PFLANZEN

### 5.1 PFLANZLICHE EINZELZELL-TRANSKRIPTOME

In der Pflanzenforschung wird die Einzelzell-Transkriptomanalyse (single-cell RNA-seq, scRNA-seq, siehe Walter/Gasparoni, Kapitel 1) gerade erst etabliert, gewinnt aber schnell an Bedeutung. Im Unterschied zu tierischen Zellen besitzen Pflanzenzellen eine recht starre Zellwand. Diese besteht aus unterschiedlichen Kohlenhydratpolymeren mit einer variablen Zusammensetzung, je nach Zelltyp und Differenzierungszustand. Dazu gehören Cellulose, Hemicellulose und weitere Komponenten, unter anderem in die Zellwand integrierte Proteine. Die Zellen bilden einen stabilen Gewebsverbund und müssen vor einer typischen Einzelzell-Transkriptomanalyse zunächst einmal voneinander getrennt werden. Dies geschieht durch die Behandlung der pflanzlichen Gewebe mit unterschiedlichen Enzymen, die die Zellwand abbauen; dabei entstehen sogenannte Protoplasten (also Pflanzenzellen ohne Zellwand), die dann einer Einzelzell-Transkriptomanalyse unterzogen werden. Da die Herstellung der Protoplasten selbst bereits zu einer Veränderung der Transkriptionsmuster führen kann, müssen entsprechende Kontrollen durchgeführt werden, zum Beispiel Vergleiche mit bereits bekannten Genexpressionsmustern von unbehandelten Pflanzenzellen, um herauszufinden, ob sich diese Muster auch bei den Protoplasten erkennen lassen.

Die bisher publizierten Arbeiten zur Einzelzell-Transkriptomanalyse in Pflanzen wurden an Wurzeln der Pflanze *Arabidopsis thaliana* (Ackerschmalwand) durchgeführt (Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulze et al., 2019; Turco et al., 2019). Sie stellen ein seit Langem untersuchtes und daher inzwischen gut verstandenes Modellsystem zur Erforschung von Entwicklungsprozessen in Pflanzen dar. Zahlreiche Gene, die die Entwicklung der pflanzlichen Wurzel steuern, sowie ihre Reaktion auf Umwelteinflüsse sind bekannt. Diese Forschung hat auch zur Identifizierung von Markergenen geführt, die nur in bestimmten Zelltypen der Wurzel aktiv sind, zum Beispiel in Stammzellen oder in haarbildenden Zellen der Wurzelepidermis. Die mittels Einzelzell-Transkriptomanalyse erhaltenen Expressionsdaten konnten daher mit früher gewonnenen Genaktivitätskarten der Wurzel verglichen und damit für Gene, deren Expression in unterschiedlichen Zellen bereits bekannt war, validiert werden. In den bisher publizierten Studien gelang es, die Transkriptome von jeweils etwa 400 bis 12.000 einzelnen Zellen zu analysieren.

Die Einzelzell-Transkriptomanalyse in Pflanzen hat bisher unter anderem zu folgenden Erkenntnissen geführt:

- scRNA-seq erfasst raumzeitliche Informationen zur Genexpression mit hoher Präzision;
- scRNA-seq erlaubt die Identifizierung von neuen Regulatoren für Prozesse in einzelnen Zelltypen;
- mittels scRNA-seq lassen sich regulatorische Pfade der zellulären Entwicklung mit höherer Genauigkeit untersuchen, als dies bisher mit anderen Methoden möglich war,
- und es lassen sich bisher nicht bekannte Subtypen von Zellen aufgrund ihrer spezifischen Genexpressionsmuster identifizieren.

In den bisherigen scRNA-seq-Studien an Pflanzen wurden neben Wildtyp-Pflanzen auch bereits Mutanten bzw. transgene Pflanzen mit Defekten in der Wurzelentwicklung untersucht (Ryo et al., 2019; Turco et al., 2019) sowie die Effekte von externen Faktoren wie Verfügbarkeit von Saccharose im Wachstumsmedium (Shulse et al., 2019) oder Hitzestress (Jean-Baptiste et al., 2019) analysiert. Pflanzen – als ortsgebundene Organismen – können sich den Umweltbedingungen, denen sie ausgesetzt sind, nicht durch Mobilität entziehen. Jedoch besitzen sie eine ausgeprägte Entwicklungsplastizität, d.h. sie können sich in vielfältiger Weise durch Veränderung ihres Wachstums und ihrer Morphologie an unterschiedliche Umweltsituationen anpassen, und zwar ohne Veränderungen ihrer genetischen Konstitution (Sequenz der Erbinformation) (Bradshaw, 2006; Salazar-Henao et al., 2016). Diese unterschiedlichen Anpassungsszenarien und die ihnen zugrunde liegenden Gene haben sich im Laufe der Evolution etabliert und zur Anpassung der Pflanzen an bestimmte ökologische Nischen geführt. Die Einzelzell-Transkriptomanalyse wird es in Zukunft erlauben, die diesen komplexen – und variablen – Entwicklungsprozessen zugrunde liegenden molekularen und zellulären Mechanismen mit deutlich höherer Detailtreue als bisher zu analysieren.

## 5.2 TRANSKRIPTOMANALYSEN MITTELS ISOLIERTER ZELLKERNE

Wie oben erläutert, erfordert die Einzelzell-Transkriptomanalyse in Pflanzen bisher noch die Protoplastierung pflanzlicher Gewebe. Da die Zusammensetzung der Zellwand zwischen unterschiedlichen Zellen einer Pflanze bzw. Zellen in verschiedenen Pflanzen variiert, und darüber hinaus durch Umwelteinwirkungen

moduliert wird, müssen zunächst jeweils geeignete Protoplastierungsprotokolle entwickelt werden. Allein dies kann bereits viel Zeit beanspruchen, was den Zugang zu Einzelzell-Transkriptomen in nicht zu vernachlässigender Weise erschwert. Eine mögliche Alternative könnte darin bestehen, anstelle von Protoplasten als Untersuchungsobjekt direkt auf die Transkripte in den Zellkernen der untersuchten Pflanzen und ihren Geweben zuzugreifen. Dies hätte zum einen den Vorteil, dass nicht für jede Pflanze und jedes Gewebe genau adaptierte Protoplastierungsprotokolle entwickelt werden müssten. Zum anderen liegen bereits gut etablierte Protokolle für eine rasche und unkomplizierte Anreicherung von Zellkernen aus komplexen pflanzlichen Geweben oder Organen vor. Mithilfe solcher Methoden könnte es in naher Zukunft möglich werden, die Einzelzell-Transkriptome auch solcher Pflanzen zu analysieren, die bisher nicht zu den typischen Modellsystemen zählen, aber die in ökologischer Hinsicht oder in Bezug auf ihre speziellen physiologischen oder biotechnologischen Eigenschaften von besonderer Relevanz sind. Zum Beispiel ist es prinzipiell möglich, Zellkerne quasi „en bloc“ aus einem Organ (mit seinen unterschiedlichen Zelltypen) zu isolieren, um sie anschließend weiteren Analysen zu unterziehen. Bei dieser Methode kann gänzlich ohne transgene Pflanzen gearbeitet werden. Jedoch besteht auch die Möglichkeit, gezielt Zellkerne aus bestimmten Zelltypen zu isolieren. Dazu ist es notwendig, Zellkerne entsprechender Zelltypen zu markieren. Dies kann unter anderem dadurch erreicht werden, dass die Zellkerne mittels bestimmter Proteine, zum Beispiel dem grün fluoreszierenden Protein (GFP), versehen werden (dazu werden die Pflanzen mit geeigneten Genkonstruktionen gentechnisch modifiziert). Auf diese Weise gekennzeichnete Zellkerne können dann mit geeigneten biochemischen Methoden, zum Beispiel mittels Immunpräzipitation unter Verwendung von Antikörpern, die GFP erkennen, oder durch FACS („fluorescence-activated cell sorting“), isoliert und anschließend einer Transkriptomanalyse unterzogen werden. Allerdings werden dabei die Genexpressionsmuster in anderen Zellen, deren Zellkerne nicht spezifisch markiert sind, nicht erfasst, wodurch dem Experimentator möglicherweise wichtige Informationen für eine umfassendere Interpretation zellulärer Prozesse in Gewebeverbänden entgehen.

### 5.3 ZUKÜNFTIGE FORSCHUNG

Künftige wissenschaftliche Fragestellungen, die mithilfe der Einzelzell-Transkriptomanalyse in der Pflanzenforschung angegangen werden können, sind äußerst vielfältig und zum jetzigen Zeitpunkt nur ansatzweise abschätzbar. Beispiele sind, neben vielen anderen:

1. die Analyse der Genexpressionsmuster in Pflanzen, deren Wachstum und Entwicklung sich aufgrund von Umwelteinwirkungen verändern;
2. die Entschlüsselung von genregulatorischen Netzwerken mit hoher räumlicher und zeitlicher Präzision – davon wird in besonderem Maße auch die Bioinformatik profitieren.
3. Durch den Vergleich der Einzelzell-Genexpressionsmuster in diversen Pflanzen wird es möglich werden, die Evolution der Pflanzen und deren Diversifizierung und Anpassung an unterschiedliche ökologische Nischen auf molekularer Ebene besser zu verstehen.
4. Zunehmend gewinnen synthetisch-biologische Ansätze auch in Pflanzen an Bedeutung. Zweifelsohne werden Einzelzell-Transkriptomanalysen erheblich dazu beitragen, die Variabilität zwischen Zellen hierbei besser zu verstehen. Dadurch wird es möglich werden, die Robustheit synthetisch-biologischer Modifikationen in Pflanzen auf eine solidere Basis zu stellen.

#### 5.4 FAZIT

Im Bereich der Pflanzenforschung wurde über Einzelzell-Transkriptomanalysen bisher ausschließlich für die Modellpflanze *Arabidopsis thaliana* in wissenschaftlichen Publikationen berichtet, und hier lediglich in Hinblick auf Prozesse in der Wurzel. Somit stehen zum jetzigen Zeitpunkt noch keine umfangreichen Datensätze für eine weitergehende Analyse durch die wissenschaftliche Gemeinschaft zur Verfügung. Jedoch kann trotz der zurzeit noch geringen Anzahl an Publikationen in diesem Forschungsgebiet davon ausgegangen werden, dass die Einzelzell-Transkriptomanalyse einen wesentlichen Einfluss auf die zukünftige Pflanzenforschung haben wird. Dies betrifft in besonderer Weise auch die Forschung an für die menschliche und tierische Ernährung wichtigen Kulturpflanzen. Besonders interessant ist dabei, dass nicht nur „traditionelle“ Kulturpflanzen in die Analysen einbezogen werden können, sondern auch solche, die bisher in der Forschung eher unterrepräsentiert sind und noch einer weiteren züchterischen Optimierung in den kommenden Jahren bedürfen. Hierfür sollten hinreichend Forschungsmittel zur Verfügung gestellt werden.

## 5.5 LITERATUR

Bradshaw, A. D. (2006): Unravelling phenotypic plasticity – why should we bother? In: *New Phytol* 170(4): 644–648.

Denyer, T. et al. (2019): Spatiotemporal developmental trajectories in the arabidopsis root revealed using high-throughput single-cell RNA sequencing. In: *Dev Cell* 48(6): 840–852.e5.

Jean-Baptiste, K. et al. (2019): Dynamics of gene expression in single root cells of arabidopsis thaliana. In: *Plant Cell* 31(5): 993–1011.

Ryu, K. H. et al. (2019): Single-cell RNA sequencing resolves molecular relationships among individual plant cells. In: *Plant Physiol* 179(4): 1444–1456.

Salazar-Henao, J. E. et al. (2016): The regulation and plasticity of root hair patterning and morphogenesis. In: *Development* 143(11): 1848–1858.

Shulse, C. N. et al. (2019): High-throughput single-cell transcriptome profiling of plant cell types. In: *Cell Rep* 27(7): 2241–2247.e4.

Turco, G. M. et al. (2019): Molecular mechanisms driving switch behavior in xylem cell differentiation. In: *Cell Report* 28 (2): 342–351.e4.

## 6. EINZELZELLANALYSEN UND ÜBERLEGUNGEN ZUR ETHIK

### 6.1 EINFÜHRUNG

Das derzeit vorherrschende Biomedizinische Modell fußt erkenntnistheoretisch auf den Säulen des „Universalismus“, „Reduktionismus“ und der „Modellierung“ (Strasser, 2014). Schon lange spielt in diesem Dreiklang die Zelle als kleinste bzw. „grundlegendste funktionelle Einheit“ (Regev et al., 2017),<sup>1</sup> die ihre eigene Genexpression reguliert, eine herausragende Rolle als Ankerpunkt für jeden dieser Ansätze. Lange galt das Dogma, dass was für eine Zelle gelte, auch universell gültig sei, dass die Analyse des Lebendigen auf Zellebene reduziert betrachtet werden könne und die eine Zellart wiederum als Modell für eine andere Zellart im Forschungsprozess herangezogen werden könnte.

Das Forschungsfeld der Einzelzellanalyse setzt auf der einen Seite genau an diesem Punkt an, auf der anderen Seite differenziert es die genannten Säulen weiter aus, indem es von der Idee ausgeht, dass sich einzelne Zellen weit mehr voneinander unterscheiden als bisher angenommen. Die Besonderheiten der individuellen Zellen mit Blick auf ihr Genom, Epigenom, Transkriptom oder Proteom<sup>2</sup> sind in den letzten Jahren vermehrt in den Mittelpunkt der forschersischen Anstrengungen gerückt.

Zwar gilt immer noch, dass die Einzelzellanalyse als Forschungsgebiet in den Kinderschuhen steckt, wie Michael Speicher 2013 festhielt, doch die Potenziale, die er damals mit dem Ansatz verband, haben an Strahlkraft nicht verloren. So dienen die entsprechenden Methoden unter anderem dem besseren Verständnis davon, wie unterschiedliche Zellen sich differenzieren, altern oder auf Schadstoffe reagieren und welche unterschiedlichen Zelltypen sich charakterisieren lassen.<sup>3</sup>

- 1 Hier findet sich auch eine gute Übersicht zu den verschiedenen Methoden der Single Cell Genomics (S. 4).
- 2 Als Genom einer Zelle bezeichnet man die Summe aller Gene, als Epigenom alle epigenetischen Modifikationen, das Transkriptom bezeichnet die Summe aller in einer Zelle transkribierten (von DNA in RNA umgeschriebenen) Gene und das Proteom besteht aus der Gesamtzahl aller in der Zelle befindlichen Proteine (siehe hierzu auch Walter/Gasparoni, Kapitel 1). Die Verfahren, mit denen Daten über diese Eigenschaften erhoben werden können, werden als „Omics-Technologien“ bezeichnet.
- 3 Zu diesem Zweck werden verschiedene internationale Initiativen durchgeführt, etwa das Projekt Human Cell Atlas zur Charakterisierung der unterschiedlichen Zelltypen des Menschen oder die LifeTime Initiative, bei der große Datenmengen gesammelt und ausgewertet werden. Zum Human Cell Atlas siehe Regev et al., 2017. Zu LifeTime siehe unter: <https://www.mdc-berlin.de/de/lifetime> [14.05.2019]. Siehe hierzu auch Junker, Popp, Rajewsky, Kapitel 2.

In der medizinischen Anwendung wiederum kann dieses Wissen dazu beitragen, die Tumorentstehung besser zu verstehen oder metastasierende Zellen aufzufinden. Auch für die Präimplantationsdiagnostik und andere Formen der Krankheitsprädiktion werden Impulse erwartet (Speicher, 2013). Insgesamt – so die Hoffnung – sollen Einzelzellanalysen in der Medizin dazu beitragen, eine laborzentrierte prädiktive und so genannte „personalisierte“ oder „individualisierte“ Medizin Wirklichkeit werden zu lassen, indem sie einerseits helfen, zuverlässig vorherzusagen, wann eine Krankheit ausbricht und wie sie verläuft, sowie andererseits die Testung von Therapeutika auf Zellebene erlauben (Shalek/Benson, 2017).

Der permanente ausdrückliche Verweis auf mögliche medizinische Einsatzbereiche der Einzelzellanalyse lässt die Frage aufkommen, inwieweit dieses Arbeitsfeld eventuell eigene oder neue medizinethische Fragestellungen evoziert. Immerhin geht es bei diesem noch jungen Verfahren um ein Gebiet mit Hoffnung auf eine große medizinische Wirkung und damit um ein Gebiet, das Patienten und Patientinnen, Ärzte und Ärztinnen und andere Gesundheitsakteure und -akteurinnen persönlich, professionell und im Umgang miteinander unmittelbar betreffen könnte. Darüber hinaus stellt sich die Frage, ob die methodische Verknüpfung der Einzelzellanalyse mit Komplexen wie zum Beispiel der Genomeditierung (zur Modellbildung) die hier debattierten ethischen Dilemmata einfach reproduziert oder ihnen ein neues Gesicht gibt.

## 6.2 ETHISCHE THEMENFELDER

Um es vorwegzunehmen: Es erscheint nicht so, als müsse durch die Einzelzellanalyse ein gänzlich neues Subgebiet der Moraltheorie geschaffen werden. Vielmehr berührt die Methode eine Reihe von ethischen Fragen, die in der Bioethik schon lange im Umfeld anderer Biotechnologien diskutiert werden. Im Zentrum stehen Wertkonflikte, die im Zusammenhang mit neuen Methoden der (Massen-) Datenerhebung und -verarbeitung im Umfeld der Medizin immer wieder diskutiert werden: Dem Interesse an Datenschutz und Datensouveränität oder der Sorge vor Datenmissbrauch steht das Interesse der möglichst umfassenden Datenerhebung mit dem Ziel der Erkenntnisgewinnung gegenüber. Gerade die interdisziplinäre Anknüpfbarkeit, Nutzbarkeit und Anschlussfähigkeit der Einzelzellanalyse an vielfältigste Forschungs- und Anwendungsbereiche macht allerdings eine Sammlung der durch sie berührten ethischen Themen umfänglich und im Zweifel lückenhaft. Im Folgenden bemühen wir uns um eine Sortierung der für die Einzelzellanalyse



spezifischen ethischen Herausforderungen, bei denen auf der einen Seite immer der mögliche medizinische Wert und die Freiheit der Forschung stehen und auf der anderen Seite (meist gesellschaftliche oder individuelle) Werte, die das Potenzial haben, mit diesen in Konflikt zu geraten.<sup>4</sup>

### **Aussagekraft und Validität der Daten**

Die Möglichkeit, große Datenmengen zu erheben und zu interpretieren, ermöglicht erst die Durchführung von Einzelzellanalysen. Wie in anderen Feldern der Medizin besteht die Hoffnung, dass viele Daten auch zu einer besseren Diagnostik führen können. Wie in anderen Bereichen der Bioinformatik, medizinischen Diagnostik und datengetriebenen Forschung ist hier darauf zu achten, dass Standards der Statistik eingehalten, Probleme der Korrelation beachtet und Daten somit korrekt und einheitlich interpretiert werden. Andernfalls drohen Über- oder Unterschätzung von Zusammenhängen, Fehlinterpretationen und im Zweifel eine auf ungenauer Datenbasis arbeitende schlechte Medizin.<sup>5</sup> So wie die Einzelzellanalyse erkenntnistheoretisch im Blick behalten muss, dass die untersuchten Zellen aus dem Zellverband, ihrem System, herausgelöst betrachtet werden, was Effekte auf ihr Verhalten haben kann, so ist insgesamt selbstkritisch in Hinsicht auf mögliche Datenerhebungen im Blick zu behalten, dass möglicherweise die Einkommenssituation eines Menschen ein besserer Risikoindikator für bestimmte Krankheitsrisiken ist als Omics-Daten.

### **Datenschutz**

Wie immer, wenn biologische Daten für die Forschung erhoben und verarbeitet werden, stellen sich Datenschutz- und Dateneigentumsfragen. Es ist zu klären, ob mögliche an menschlichen Zellen erhobene Daten den Forschern und Forscherinnen oder den Probanden und Probandinnen bzw. Patienten und Patientinnen gehören. Ferner ist auf den Datenschutz und angemessenes Datenmanagement zu achten, da beispielsweise eine Anonymisierung der Daten schwierig sein könnte, wenn eine Verbindung zu Individuen zum Beispiel für weitere Forschungen gewünscht ist. Assoziiert ist das praktische Problem des Datenaustausches zwischen

4 Weitere Informationen zu den folgenden Oberthemen finden sich in: Lenk et al., 2014 und Düwell, 2011; eine Übersicht zum Themenbereich „Big Data“ in: Deutscher Ethikrat, 2018.

5 Zur Fehleranfälligkeit großer Datenmengen siehe etwa Bertram, 2019.

Forschenden gegebenenfalls auch über Grenzen hinweg bei internationalen Forschungsprojekten mit Partnern, deren Datenschutzbestimmungen nicht unbedingt denen in der Europäischen Union entsprechen. Da Datenschutz kein Zweck an sich ist, sondern dem Schutz von Menschen vor Stigmatisierung oder Missbrauch eines Wissens über sie dienen soll, sind vor allem auch die Domänen zu bestimmen, in denen die Einzelzellanalyse überhaupt entsprechende Daten hervorbringen kann.

### **Informed Consent**

Adressiert werden können Datenschutzfragen durch einen angemessenen Informed Consent, wenn betroffene Personen der Teilnahme an Forschung und der Lagerung bzw. Speicherung und Nutzung ihrer Biomaterialien zur Datengewinnung nach spezifischer und verständlicher Aufklärung zustimmen. Die Reichweite einer Zustimmung zwischen einer zweckgebundenen, spezifischen („narrow consent“) oder einer allumfassenden, offenen („broad consent“) Einwilligung ist dabei umstritten. Neben dem Problem der verständlichen Aufklärung über ungewisse Nutzungsmöglichkeiten, die sich angesichts der Offenheit von Forschung immer ergeben können, stellt sich die Frage, wie weit eine Zustimmung hier gehen kann und wie dynamisch Zustimmungen gehalten werden können, wenn Forschende nicht permanent neue oder aktualisierte Einwilligungen einholen wollen („dynamic consent“).

### **Zusatz- und Zufallsbefunde**

Eng mit der Frage des Informed Consent verknüpft ist die Frage des Umgangs mit Befunden und Nebenbefunden, die außerhalb einer vorher klar umrissenen Diagnostik erhoben werden. Dilemmata entstehen hier vor allem, wenn relevante Informationen über behandelbare oder nicht behandelbare zukünftige Erkrankungen erhoben werden und der vorige Informed Consent die Frage der Information über solche Befunde ausgeklammert hatte oder die entsprechende Entscheidung der betroffenen Person der moralischen Intuition des Forschers oder der Forscherin widerspricht (z. B. im Falle der Ablehnung einer Information über eine behandelbare, unbehandelt aber tödlich verlaufende Erkrankung).

## Gesellschaftliche Implikationen/Verteilungsgerechtigkeit

Wie bei jeder neuen Medizin- oder Biotechnologie stellen sich mit Blick auf die Gesellschaft vor allem Gerechtigkeits- und Priorisierungsfragen. Diese reichen von der Allokation von Forschungsgeldern auf übergeordneter Ebene bis hin zur Ebene der Patientenversorgung. Werden die Kosten für den Zugang zu den erhobenen Daten aus Einzelzellanalysen sehr hoch, könnten diese für nur wenige oder selbstzahlende Patienten und Patientinnen zur Behandlungsplanung zur Verfügung gestellt werden. Eine Stratifizierung in Patientenuntergruppen könnte dazu führen, dass relativ kleine Gruppen relativ teure Medikamente benötigen und/oder bestimmte Gruppen von der Versorgung wegen zu hoher Kosten ausgeschlossen werden. Patienten und Patientinnen mit seltenen Erkrankungen sind hier besonders gefährdet.

### 6.3 Schluss

Keiner der hier skizzierten ethischen Themenkomplexe ist neu oder spezifisch für die Einzelzellanalyse. Per se scheint die Methode relativ frei von durch sie allein hervorgerufenen moralischen Entscheidungssituationen zu sein. Gleichwohl weist sie eine Reihe von Anknüpfungspunkten für aus anderen Bereichen bekannte Probleme und Vorschläge zu ihrer Lösung auf. So bleibt insgesamt festzuhalten, dass der aktuelle Hype um die Einzelzellanalyse als von der Zeitschrift *Science* 2018 zum „Durchbruch des Jahres“ gekürtes Verfahren<sup>6</sup> nicht dazu führen darf, dass in anderen Feldern etablierte ethische Standards abgesenkt oder negiert werden. Wie immer steht die Forderung im Raum, die Standards guter wissenschaftlicher, guter klinischer und guter ethischer Praxis einzuhalten, damit mehr Daten auch zu mehr Wissen führen.

### 6.4 Literatur

Bertram, I. (2019): Big Data, Big Errors – Interview mit Gerd Antes. In: GID 248, Februar 2019. Unter: <http://www.gen-ethisches-netzwerk.de/biobanken-und-big-data/genestsund-genomsequenzierung/big-data-big-errors> [14.05.2019].

6 Siehe unter: <https://www.sciencemag.org/tags/2018-breakthrough-year> [14.05.2019].

Deutscher Ethikrat (2018): Big Data und Gesundheit. Datensouveränität als informationelle Freiheitsgestaltung: Stellungnahme. Deutscher Ethikrat, Berlin.

Düwell, M. et al. (Hrsg.) (2011): Handbuch Ethik. 3. aktualisierte Auflage. Metzler, Stuttgart, Weimar.

Lenk, C. et al. (Hrsg.) (2014): Handbuch Ethik und Recht der Forschung am Menschen. Springer, Berlin, Heidelberg.

Regev, A. et al. (2017): The Human Cell Atlas. In: eLife 6, Online-Publikation 05.12.2017. DOI: 10.7554/eLife.27041.

Shalek, A. K./Benson, M. (2017): Single-Cell Analysis to Tailor Treatments. In: Sci Transl Med, Online-Publikation 20.09.2017. DOI: 10.1126/scitranslmed.aan4730.

Speicher, M. R. (2013): Single-Cell Analysis. Toward the Clinic. In: Genome Medicine 5(8), Online-Publikation 27.08.2013. DOI: 10.1186/gm478.

Strasser, B. J. (2014): Biomedicine. Meanings, Assumptions, and Possible Futures. Report to the Swiss Science and Innovation Council (SSIC) 1/2014. Swiss Science and Innovation Council, Bern.

## 7. PROBLEMFELDER UND INDIKATOREN ZUM THEMA EINZELZELLANALYSE

### 7.1 EINFÜHRUNG: MOTIVATION UND ZIELSETZUNG

Die IAG *Gentechnologiebericht* an der Berlin-Brandenburgischen Akademie der Wissenschaften hat die Aufgabe, die verschiedenen Entwicklungen auf dem Gebiet der Gentechnologie in Deutschland langfristig zu beobachten und im Rahmen von Publikationen und Veranstaltungen für die interessierte Öffentlichkeit aufzubereiten. Ihre Ergebnisse sollen als allgemein verständliche Informationsquelle dienen und so einen wohlinformierten öffentlichen Diskurs über dynamische und zum Teil gesellschaftlich umstrittene Themenfelder fördern. Neben der qualitativen Auseinandersetzung mit verschiedenen (z. B. naturwissenschaftlichen, rechtlichen und ethischen) Aspekten der Gentechnologien hat es sich die IAG zur Aufgabe gemacht, das komplexe Feld der Gentechnologien in einer (öffentlich) zugänglichen und messbaren Form für die interessierte Öffentlichkeit zu erschließen (Diekämper/Hümpel, 2015: 16 ff., 2012: 51–60). Als zentrales Instrument wird hierfür die sozialwissenschaftlich motivierte Methode der Problemfeld- und Indikatorenanalyse genutzt: Ausgehend von einer qualitativen Erhebung (Problemfeldanalyse) werden quantitative Daten (Indikatoren) zusammengetragen.<sup>1</sup>

### 7.2 PROBLEMFELDER

In der Öffentlichkeit und insbesondere in den Medien werden viele gentechnologische Themen kontrovers diskutiert. Die IAG *Gentechnologiebericht* wendet die Methode der Problemfelderhebung an, um diese komplexen Diskussionen in Themenbereiche und Aspekte (Problemfelder) aufzuschlüsseln. Die Problemfeldanalyse hat somit das Ziel, die öffentliche Wahrnehmung der Gentechnologien übersichtlich abzubilden (Diekämper/Hümpel, 2015: 16). Im Rahmen der Analyse werden verschiedene Print- und Online-Medien ausgewertet. Im Anschluss an diese Auswertung werden die identifizierten Problemfelder innerhalb eines gewählten Koordinatensystems eingeordnet. Die Eckpunkte des

1 Die Problemfeld- und Indikatorenanalyse ist eine der zentralen Methoden der IAG. So sind einführende und allgemeine Überlegungen sowie Ausführungen zu diesem Ansatz im Wortlaut ähnlich bereits in vorherigen Veröffentlichungen der IAG beschrieben (siehe etwa: Marx-Stölting, 2017; Diekämper/Hümpel, 2012).

Koordinatensystems sind vier Leitdimensionen, die sich im Kontext von Gentechnologien primär abzeichnen: die ökonomische Dimension, die wissenschaftliche Dimension, die ethische Dimension sowie die soziale Dimension. Als letzter Schritt werden diesen Problemfeldern relevante Indikatoren zugeordnet.

Die Basis der Problemfelderhebung bildet ein Textkorpus, das qualitativ ausgewertet wird. Dieses Textkorpus wird mittels einer Stichwortrecherche in den Leit-Printmedien *Süddeutsche Zeitung* und *Frankfurter Allgemeine Zeitung*, in den wöchentlich erscheinenden Medien *Die Zeit* und *Der Spiegel* sowie anderen in Online-Suchmaschinen erhoben (Google und Metager).<sup>2</sup>

Abbildung 1 zeigt die identifizierten Problemfelder für den Themenbereich der Einzelzellanalyse und deren quantitative Gewichtung innerhalb des analysierten Textkorpus. Die Größe und Einfärbung bilden die quantitative Gewichtung der Problemfelder ab. Je öfter das Problemfeld im Textkorpus thematisiert wurde, desto größer und dunkler wird es dargestellt.

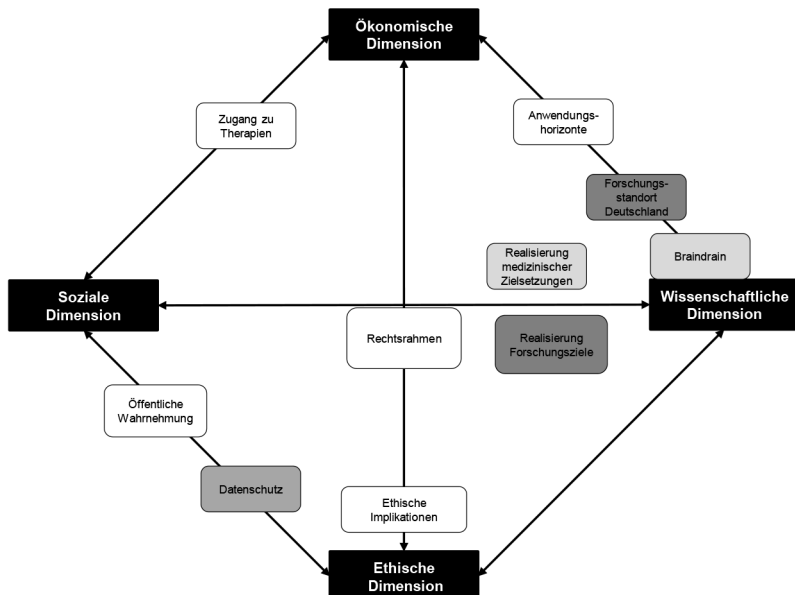
Folgende Problemfelder wurden mithilfe der qualitativen Auswertung des Textkorpus ermittelt:

*Anwendungshorizonte:* Anwendungshorizonte im Bereich der Einzelzellanalyse werden bereits laufend diskutiert. Sie schließen visionäre Ziele mit hohem Innovationspotenzial ein, deren Durchführbarkeit entsprechend ungewiss ist.

*Forschungsstandort Deutschland:* Zur internationalen Attraktivität eines Forschungsstandortes tragen eine Vielzahl an Faktoren bei: die vorhandene wissenschaftliche Infrastruktur, das Ausmaß und die Art an Fördermaßnahmen aber auch nationale rechtliche Regelungen, die die wissenschaftliche Praxis beeinflussen. Der internationale Ruf und die Vernetzung innerhalb der globalisierten Forschungslandschaft spielen ebenfalls eine Rolle.

2 Für die Printmedien wurde im Zeitraum vom 05. bis zum 12.03.2019 eine Suche mit den folgenden Stichwörtern durchgeführt: „Einzelzellsequenzierung“, „Einzelzellbiologie“, „Einzelzell-Transkriptomik“, „Einzelzell-Genomik“, „Einzelzelldiagnostik“ und „Einzelzellanalyse“. Üblicherweise wurde im Rahmen der Recherche für andere Themenbereiche der IAG lediglich mit deutschen Stichwörtern gesucht. Da nur sehr wenige Artikel gefunden wurden, wurden für dieses Thema auch englische Begriffe verwendet („single cell analysis“, „single cell biology“, „single cell sequencing“, „single cell genomics“, „single cell diagnostics“, „single cell transcriptomics“). Es wurden insgesamt 4 Artikel gefunden. Die Recherche mithilfe der Suchmaschinen Google und Metager wurde vom 12. bis zum 26.03.2019 durchgeführt. Es wurden sowohl erneut die bereits oben genannten deutschen Suchbegriffe genutzt als auch die Kombination aus diesen Suchbegriffen (mit Trunkierung) und „Stellungnahme“. Die ersten zehn Treffer aus den Suchmaschinen wurden zusammengeführt und abgeglichen. Die Treffer aus der Suchmaschinen- und der Printmedienrecherche bilden gemeinsam das Textkorpus, das anschließend im Hinblick auf die Problemfelder qualitativ ausgewertet wurde.

**Abbildung 1:** Erhobene Problemfelder zum Themenbereich Einzelzellanalyse



*Zugang zu Therapien:* Falls die Kosten für mögliche medizinische Anwendungen oder auch für den Zugang zu den erhobenen Daten aus Einzelzellanalysen sehr hoch sind, stellt sich die Frage nach der Kostenübernahme durch die gesetzlichen Krankenkassen. An dieser Stelle ist es notwendig, Fragen der Priorisierung und Verteilung zu diskutieren.

*Braindrain:* In einer dynamischen, globalisierten Forschungslandschaft mit ihrer Mobilitätsanforderung läuft Deutschland Gefahr, wissenschaftliche Talente zu verlieren, ohne dass im vergleichbaren Maß Wissenschaftler und Wissenschaftlerinnen gewonnen werden können. Für den Bereich Einzelzellanalyse kann dies bedeuten, dass hochqualifizierte Wissenschaftlerinnen und Wissenschaftler aus beruflichen, wirtschaftlichen oder rechtlichen Gründen das Land verlassen. Damit würde im globalen Standortwettbewerb und Forschungswettlauf wichtiges Know-how verloren gehen und ökonomisches Potenzial ungenutzt bleiben.

*Realisierung medizinischer Zielsetzungen:* Ein Ziel der Forschung auf dem Gebiet der Einzelzellanalyse ist, neue Erkenntnisse zu gewinnen, um unter anderem Entwicklungen im Bereich der personalisierten Medizin voranzutreiben. Probleme ergeben sich dann, wenn nicht alle Zielsetzungen umsetzbar sind oder sich diese als schwieriger oder zeitraubender herausstellen als zunächst angenommen.

*Realisierung Forschungsziele:* Wissenschaftliche Forschung will neue Erkenntnisse und Technologien generieren. Zu ihrem Wesen gehört eine begrenzte Planbarkeit und Ergebnisoffenheit. Gleichwohl beeinflussen die vorhandenen Rahmenbedingungen – wie die wissenschaftliche Infrastruktur, Förderungsmöglichkeiten oder geltendes Recht – die Realisierung von gesetzten Forschungszielen, die sich quantifizierbar zum Beispiel in Veröffentlichungen, Forschungspreisen oder akademischen Abschlüssen niederschlagen.

*Rechtsrahmen:* Der rechtliche Rahmen auf nationaler, europäischer und internationaler Ebene bestimmt über die Zulässigkeit von Forschung, insbesondere den Umgang mit Forschungsdaten. Der Rechtsrahmen definiert den Einsatz in der wissenschaftlichen Praxis bzw. formuliert dafür notwendige Rahmenbedingungen. Er hat eine Funktion bei der Vermittlung von einander widersprechenden Interessen und Schutzgütern. Diskutiert wird im Bereich der Einzelzellanalysen unter anderem der Datenaustausch über Grenzen hinweg im Rahmen von internationalen Forschungsprojekten.

*Ethische Implikationen:* Forschung – vor allem in den Biowissenschaften und verschärft im biomedizinischen Bereich – generiert Wissen und Anwendungen, die eine Auseinandersetzung mit etwaigen Konsequenzen für den Menschen, die Gesellschaft und die Umwelt verlangen. Dabei spielen soziale oder rechtliche Aspekte ebenso eine Rolle wie ethische Fragen, die es gesellschaftlich zu diskutieren gilt und die u. U. politischen Handlungsbedarf nach sich ziehen. Im Fall der Einzelzellanalyse birgt zum Beispiel der Umgang mit umfangreichen Datenmengen und den damit verbundenen Wertkonflikten ein großes Diskussionspotenzial. Die Frage nach dem Umgang mit Zufalls- und Zusatzbefunden spielt auf diesem Gebiet ebenfalls eine Rolle.

*Datenschutz:* Die Erhebung und Speicherung von Forschungsdaten im Bereich der Einzelzellanalyse ermöglicht prinzipiell eine weitergehende Nutzung, die individuelle Rechte tangieren kann. In diesem Zusammenhang wird auch das Recht auf informationelle Selbstbestimmung wie auch ein „Recht auf Nichtwissen“ diskutiert.



*Öffentliche Wahrnehmung:* Der Einsatz und die Etablierung neuer technologischer Verfahren hängen zentral von deren gesellschaftlicher Wahrnehmung ab. Anhand der Präsenz des Themas Einzelzellanalyse in den Printmedien und dem Internet sowie der Quantität öffentlicher Veranstaltungen und der Allgemeinheit zugänglichen Publikationen zeigt sich das öffentliche Interesse an der Thematik.

### 7.3 INDIKATOREN

Ausgehend von der qualitativen Problemfeldanalyse werden Indikatoren<sup>3</sup> (quantitative Daten) zusammengetragen, um aktuelle Entwicklungen zu illustrieren. Eine Auswahl von Indikatoren wird im Folgenden vorgestellt und ausgewertet. Mithilfe dieser Daten können erste Hinweise auf den aktuellen Stand wie auch die Entwicklungen im Bereich Einzelzellanalyse gegeben werden.

Im Bereich der Einzelzellanalyse erscheinen zurzeit – vor allem vor dem Hintergrund, dass es aktuell viele neue Entwicklungen auf diesem Gebiet gibt – die Indikatoren „Anzahl internationaler Fachartikel“, „Online-Suchanfragen“ und „Neuerscheinungen“ als geeignet, ausgewählte Problemfelder auszuleuchten.

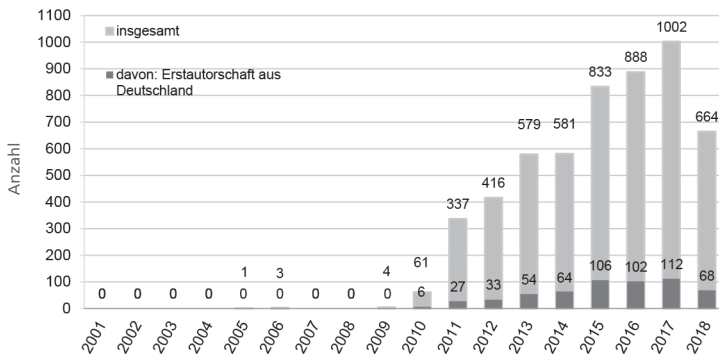
Der Indikator „Anzahl internationaler Fachartikel“ zur Einzelzellanalyse wurde den Problemfeldern „Realisierung Forschungsziele“ und „Forschungsstandort Deutschland“ zugeordnet. Gesucht wurde in der kostenlosen und öffentlich zugänglichen Online-Zitationsdatenbank PubMed des US-amerikanischen National Center for Biotechnology Information (NCBI) (Zugriff: März 2019, Stand: 2018). Nach eigenen Angaben führt die Datenbank gegenwärtig ca. 24 Millionen Zitationen für biomedizinische Literatur aus MEDLINE (= Medical Literature Analysis and Retrieval System Online), einschlägigen Fachzeitschriften und E-Büchern. Generell sind Fachartikel ab 1946 berücksichtigt, zum Teil auch ältere. Der Schwerpunkt liegt auf englischsprachiger Literatur. Für die Recherche können zum einen frei gewählte Stichwörter verwendet werden, zum anderen kann der Katalog der „Medical Subject Headings“ (MeSH) genutzt werden, der für die Indizierung der PubMed-Zitationen verwendet wird und kontinuierlich von der US-amerikanischen National Library of Medicine (NLM) gepflegt und erweitert wird (vgl. [www.nlm.nih.gov/mesh](http://www.nlm.nih.gov/mesh) [03.04.2019]).<sup>4</sup> Für die Recherche wurde ein einschlägiger MeSH-Term

3 Die Indikatoren werden im Rahmen bisheriger Publikationen der IAG mittels standardisierter Indikatorenblätter aufbereitet. Zuletzt wurden sie im „Vierten Gentechnologiebericht“ veröffentlicht (Marx-Stöltzing et al., 2018: 299–340).

4 Zusätzlich wurde eine Recherche in der Datenband Web of Science durchgeführt (Zugriff: 05.04.2019). Die Entwicklung der Publikationszahlen ist vergleichbar mit denen aus PubMed. Es besteht seit Anfang des Erhebungszeitraums ein stetiger Anstieg. Die Daten sind allerdings nur bedingt vergleichbar, da in Web of Science keine MeSH-Terme genutzt werden.

(„single cell analysis“) aus dem aktuellen MeSH-Katalog verwendet. Zusätzlich wurden außerdem Erstautorschaften aus Deutschland identifiziert. Die Daten sind hier ab 2001, dem Jahr, in dem die IAG *Gentechnologiebericht* ihre Arbeit aufgenommen hat, bis 2018 dargestellt. Der Indikator spiegelt die weltweiten Forschungsaktivitäten auf dem Gebiet der „single cell analysis“ wider. Anhand des Umfangs der veröffentlichten Publikationen kann beobachtet werden, wie intensiv ein Themenbereich über die Jahre beforscht wird und welche Länder dabei jeweils eine Vorrangstellung im internationalen Forschungswettbewerb einnehmen. Dabei ist zu berücksichtigen, dass trotz des großen Umfangs der Datenbank keine vollständige Erfassung der Zitationen erwartet werden kann: Relevante Veröffentlichungen sind unter Umständen von vornherein nicht in der Datenbank enthalten oder nicht unter den verwendeten MeSH-Kategorien verschlagwortet. Ebenfalls muss beachtet werden, dass eine Veröffentlichung eine gleichwertige Kollaboration von Autoren und Autorinnen mehrerer Länder darstellen kann, wobei die MEDLINE-Datenbank hier nur die Landeszugehörigkeit von Erstautoren und -autorinnen standardmäßig erfasst.

**Abbildung 2:** Anzahl internationaler Fachartikel zur Einzelzellanalyse (gesamt und mit deutscher Erstautorschaft [2001–2018])



Jahr	2001	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18
Insgesamt	0	0	0	0	1	3	0	0	4	61	337	416	579	581	833	888	1002	664
davon: Erstautorschaft aus Deutschland	0	0	0	0	0	0	0	0	0	6	27	33	54	64	106	102	112	68

Die Darstellung für 2018 ist möglicherweise unvollständig, da eventuell noch nicht alle Veröffentlichungen in die Datenbank aufgenommen wurden.

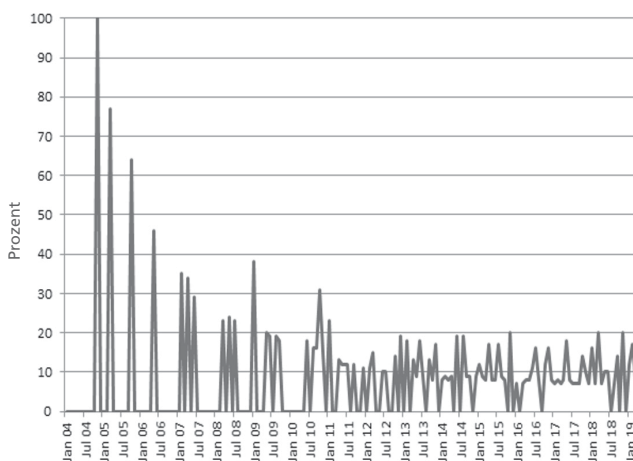
Der Indikator „Online-Suchanfragen“ wurde mithilfe des kostenlosen Tools „Google Trends“ erhoben (siehe unter: [www.google.com/trends](http://www.google.com/trends) [26.03.2019], Stand: März 2019). Zugeordnet wurde der Indikator dem Problemfeld „Öffentliche Wahrnehmung“. Das Online-Tool analysiert einen prozentualen Anteil der Sucheingaben in die Google-Suche. Allerdings sind der verwendete Analyse-Algorithmus und die absoluten Zahlen zu den Suchanfragen nicht öffentlich einsehbar. Die Daten spiegeln die Nachfrage eines bestimmten Suchbegriffs in Relation zum gesamten Suchaufkommen in Google innerhalb einer ausgewählten Zeitspanne wider. Aktuell nutzt ein großer Anteil der Bevölkerung in Deutschland regelmäßig privat das Internet (87% in 2018; siehe unter: [destatis.de](http://destatis.de) [03.04.2019]). Zur Informationssuche werden vorwiegend Internet-Suchmaschinen genutzt: An erster Stelle steht Google (siehe unter: <http://de.statista.com> [03.04.2019]). So können Online-Suchanfragen als Indikator für das Interesse der Öffentlichkeit an verschiedenen Themen gesehen werden.

In den bisherigen Publikationen der IAG *Gentechnologiebericht* wurden üblicherweise deutsche Suchbegriffe verwendet. Begonnen wurde mit der Recherche der Stichwörter „Einzelzellsequenzierung“, „Einzelzellbiologie“, „Einzelzell-Transkriptomik“, „Einzelzell-Genomik“, „Einzelzellanalyse“, „Einzelzelldiagnostik“ (Trunkierungen wie „einzelzell\*“ sind bei Google Trends nicht möglich). Für diese Suchbegriffe lagen allerdings keine ausreichenden Daten vor („Suchvolumen ist zu gering“ = 0). Dies zeigt, dass das Thema noch sehr jung ist und in der Öffentlichkeit noch nicht breit diskutiert wird. In einem nächsten Schritt wurden die englischen Begriffe „single cell analysis“, „single cell biology“, „single cell sequencing“, „single cell genomics“, „single cell diagnostics“ und „single cell transcriptomics“ abgefragt. Lediglich für die englischen Stichwörter „single cell analysis“ und „single cell sequencing“ wurden Treffer erzielt. Es besteht die Möglichkeit, Suchergebnisse nach Regionen (Länder, Städte) und festgelegten Sachkategorien zu filtern.<sup>5</sup> Zudem können mehrere Stichwörter gleichzeitig abgefragt werden. Die Daten sind ab 2004 öffentlich einsehbar. So wurde im Zeitraum Januar 2004 bis März 2019 in Deutschland nach Einträgen gesucht (Erhebungsdatum: 26.03.2019, Stand: März 2019).

5 Zur Eingrenzung der Aussagefähigkeit: Ein höheres Suchvolumen ist nicht gleichzusetzen mit einer Zunahme der Suchanfragen, die Berechnungen basieren auf Stichproben, Mehrdeutigkeiten der Suchbegriffe könnten eine Rolle spielen und der Anlass der Informationssuche ist nicht nachvollziehbar.

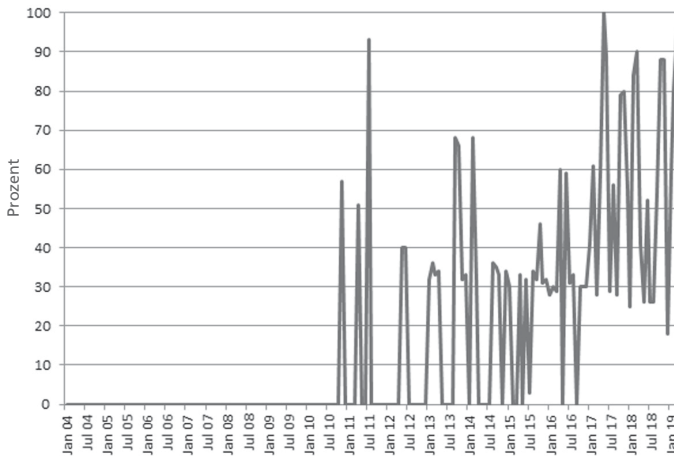
Die relative Nachfrage nach dem Stichwort „single cell analysis“ hat Peaks in den Jahren 2004, 2005 und 2006. Danach liegt die Nachfrage bis heute auf geringem Niveau. Das Stichwort „single cell sequencing“ wird vermehrt erst ab dem Jahr 2010 abgefragt, in den nachfolgenden Jahren sanken die Nachfragen auf ein mittleres Niveau ab mit Peaks in den Jahren 2013 und 2014. Ab 2017 bis 2019 steigt die Nachfrage wieder in etwa auf das Niveau aus dem Jahr 2010.

**Abbildung 3:** Relative Nachfragen nach dem Stichwort „single cell analysis“ in Google Trends für Deutschland (2004–2019)



Der Indikator „Neuerscheinungen“ zeigt die Publikationsdichte von Büchern in Deutschland. Zugeordnet wurde der Indikator den Problemfeldern „Öffentliche Wahrnehmung“ und „Forschungsstandort Deutschland“. Zur Erhebung dieses Indikators erfolgte eine Stichwortrecherche in der Datenbank der Deutschen Nationalbibliothek (Erhebungsdatum: 26.03.2019, Stand: März 2019). Die Nationalbibliothek (DNB) ist eine bundesunmittelbare Anstalt des öffentlichen Rechts. Ihre Aufgabe ist die Archivierung und bibliografische Erfassung in Deutschland veröffentlichter Medienwerke (Monografien, Zeitungen, Zeitschriften, Loseblattwerke, Karten, Musikalien, Tonträger, elektronische Publikationen). Darüber hinaus werden auch im Ausland veröffentlichte deutschsprachige Medienwerke, im Ausland veröffentlichte Übersetzungen deutschsprachiger Medienwerke, fremdsprachige Medienwerke über Deutschland sowie

**Abbildung 4:** Relative Nachfragen nach dem Stichwort „single cell sequencing“ in Google Trends für Deutschland (2004–2019)



Exilpublikationen deutschsprachiger Emigranten und Emigrantinnen zwischen 1933 und 1950 erfasst. Seit 2006 werden zusätzlich Online-Publikationen systematisch berücksichtigt. Der Katalog der DNB erlaubt eine kostenlose Recherche innerhalb der umfassenden Bibliotheksbestände seit 1913. Nach Anbieterangaben werden eingegangene Publikationen mit einer Bearbeitungszeit von ca. einem Monat in den Katalog und in die DNB aufgenommen. Für die Recherche relevanter Titel wurden die Suchbegriffe „Einzelzellsequenz\*“, „Einzelzellbiolog\*“, „Einzelzell-Transkriptom\*“, „Einzelzell-Genom\*“, „Einzelzellanaly\*“, „Einzelzelldiagnost\*“ im Modus „Expertensuche“ im gesamten Bestand des Katalogs der DNB ab 2001 (Beginn der Arbeit der IAG *Gentechnologiebericht*) gesucht. Da es sich um eine Suche nach speziellen Begriffen handelte, wurden über die Titelfelder hinausgehende Suchfunktionen (Index = woe) verwendet. Bei früheren Indikatorenerhebungen wurden die im Bestand vermerkten Hochschulschriften ausgeklammert, da sie für den interessierten Laien schwer zugänglich sind.<sup>6</sup> Schaut man sich die recherchierten Publikationen an, zeichnet sich allerdings folgendes Bild ab: Im Erhebungszeitraum der deutschsprachigen Veröffentlichungen wurden lediglich 17 deutschsprachige Hochschulschriften gefunden.

6 Ausgeschlossen wurden: Periodika, Normdaten für einzelne Personen, Organisationen, Veranstaltungen, Geografika, Sachbegriffe und Werktitel, Doppelnennungen (physische und Online-Publikationen). Englischsprachige Medien wurden händisch aussortiert. Es fand keine weiterführende qualitative Filterung der Suchergebnisse statt.

Publikationen, die in der DNB gelistet werden und für die Öffentlichkeit einsehbar sind, sind ein möglicher Indikator und ein Gradmesser für die öffentliche Wahrnehmung eines Themengebietes.

#### 7.4 ZUSAMMENFASSUNG

Zusammenfassend kann man folgende Punkte festhalten:

- Die Einzelzellanalyse ist weltweit ein Forschungsbereich mit zunehmender Relevanz. So steigen die Publikationszahlen seit 2009 stetig an. Auch die Anzahl der Artikel mit deutscher Erstautorschaft spiegelt diese Entwicklung wider.
- Auch wenn der Indikator „Online-Suchanfragen“ lediglich die relative Suchhäufigkeit anzeigt, ist es interessant, dass bei der Suche mit deutschen Stichwörter in „Google Trends“ das Suchvolumen zu gering war und nur die englischen Begriffe „single cell analysis“ und „single cell sequencing“ – wenn auch vergleichbar geringe – relative Suchhäufigkeiten erzielten.
- Bei den Neuerscheinungen, die in der Deutschen Nationalbibliothek erfasst wurden, gab es nur wenige deutschsprachige Publikationen. Zudem waren lediglich Hochschulschriften verzeichnet. Das zeigt, dass das Wissen auf dem Gebiet der Einzelzellanalyse zwar in der Fachcommunity wissenschaftlich aufbereitet und publiziert wird, aber noch nicht in der Öffentlichkeit angekommen ist bzw. diskutiert wird.

Der stete und starke Anstieg der internationalen Fachpublikationen auf dem Gebiet der Einzelzellanalyse, die zum Teil geringen relativen Suchhäufigkeiten der Online-Suchanfragen und die geringe Anzahl an Neuerscheinungen sind Indikatoren für die Neuartigkeit dieser Methode. Zudem können steigende Publikationszahlen ein Hinweis für die Zunahme der Forschungsaktivitäten auf internationaler und nationaler Ebene sein. Wie bereits erwähnt, ist die Etablierung der Methode noch relativ neu, was ein Grund dafür sein könnte, dass das Themenfeld der Einzelzellanalyse nur wenig in der Öffentlichkeit diskutiert wird beziehungsweise in den Medien nur wenig sichtbar ist. Perspektivisch und mit Blick auf die zunehmende Datenmenge, die mit der Entwicklung und Etablierung dieser Methode verbunden ist, können ethische Implikationen und rechtliche Aspekte wie zum

Beispiel Fragen des Datenschutzes, des Informed Consents oder auch verschiedene gesellschaftliche Implikationen eine Rolle im öffentlichen Diskurs spielen (siehe Fangerau, Marx-Stölting, Osterheider, Kapitel 6). Hierfür geben die qualitative Auswertung des Textkorpus und die Erhebung der Problemfelder erste Hinweise.

## 7.5 LITERATUR

Diekämper, J./Hümpel, A. (2012): Synthetische Biologie in Deutschland. Eine methodische Einführung. In: Köchy, K./Hümpel, A. et al. (Hrsg.): Synthetische Biologie. Entwicklung einer neuen Ingenieurbiologie. Forum W, Dornburg: 51–60.

Diekämper, J./Hümpel, A. (2015): Einleitung: Gentechnologien in Deutschland im Langzeit-Monitoring. In: Müller-Röber, B. et al. (Hrsg.): Dritter Gentechnologiebericht. Analyse einer Hochtechnologie. Nomos, Baden-Baden: 13–23.

Marx-Stölting, L. et al. (2018): Ausgewählte Indikatoren zu den unterschiedlichen Gentechnologien. In: Hucho, F. et al. (Hrsg.): Vierter Gentechnologiebericht. Nomos, Baden-Baden: 299–340.

Interdisziplinäre Arbeitsgruppe *Gentechnologiebericht*:

Jörn Walter, Heiner Fangerau, Boris Fehse, Jürgen Hampel, Ferdinand Hucho, Martin Korte, Bernd Müller-Röber, Jens Reich, Jochen Taupitz, Martin Zenke

## **8. KERNAUSSAGEN UND HANDLUNGSEMPFEHLUNGEN ZUR EINZELZELLANALYTIK**

### **8.1 KERNAUSSAGEN ZUR EINZELZELLANALYTIK**

#### **Die Bedeutung von Einzelzellanalysen für die Biomedizin**

Höhere Lebewesen bestehen aus einer Vielzahl unterschiedlicher Zellen – der erwachsene Mensch beispielsweise aus ca. 38 Billionen Zellen. Die Zusammensetzung von Zellen und ihre Funktion verändern sich im Verlauf des Lebens, der Entwicklung, der Regeneration, des Alterns und im Falle von Erkrankung. Mit der modernen Einzelzellanalytik entsteht ein Forschungsgebiet, in dem grundlegend neue biologische Daten gewonnen werden, die tiefe Einblicke in die molekulare Funktionsweise von Zellen eröffnen. Die Einzelzellanalytik erschließt damit neue Ebenen kontext- und personenbezogener Interpretation biologischer Zusammenhänge mit zentraler Bedeutung für die Lebenswissenschaften, Biotechnologie, Medizin und Pharmaforschung. Bislang basierten Interpretationen zumeist auf der Analyse von Gruppen von Zellen oder ganzen Geweben und Organen, spiegelten also „Durchschnittswerte“ wider. Die Funktion und Variationsbreite einzelner Zellen war nur bedingt oder gar nicht erfassbar. Die neuen Methoden und Anwendungen der Einzelzellanalytik bieten hier bislang unerreichte, tiefe Einblicke und werden die biologische Forschung und die Medizin nachhaltig beeinflussen. So zeigen Ergebnisse der Einzelzellanalyse von „Zelltypen“, die mittels Oberflächenproteinen sortiert wurden, zum Beispiel, dass Gruppen von vermeintlich gleichen Zellen mit ähnlichen, aber nicht mit gleichen Programmen ausgestattet sind. Auch die Erforschung grundlegender biologischer Prinzipien und Mechanismen der Krankheitsgenese und -ursachen erfährt neue Möglichkeiten. Einzelne Zellen von Patienten und Patientinnen oder zum Beispiel auch von Organoiden, die aus Patientenzellen gewonnen wurden, können präzise untersucht werden, um unter anderem zu ermitteln, inwieweit sie die „normalen“ Zellen des Organs darstellen oder davon abweichen. So können tiefe Rückschlüsse über individuelle Krankheits Hintergründe erlangt werden oder getestet werden,



wie bestimmte Zellen im Körper auf Medikamente ansprechen. Damit ermöglicht die Einzelzellanalyse einen wichtigen Schritt hin zu einer personalisierten Medizin.

### Einzelzellanalytik durch Next-Generation-Sequencing und andere Omics-Technologien

Nach der Entschlüsselung des Humangenoms um die Jahrtausendwende wurde deutlich, dass die Sequenz von Genomen allein keine abschließenden Antworten liefert, sondern erst interpretiert und ihre Funktion in Zellen validiert werden muss. Genomsequenzen müssen in RNA und Proteine übersetzt werden, beides Moleküle, die die Zellfunktionen steuern. Daneben gibt es andere Vorgänge in Zellen, die für deren Funktionsweise wichtig sind, wie etwa ihr Stoffwechsel (Metabolismus). Die Untersuchung und Erfassung zellspezifischer molekularer Programme erfolgt auf multiplen Ebenen: auf der Ebene der Genome („genomics“), der RNA-Transkripte („transcriptomics“), der Proteine („proteomics“), der Stoffwechselprodukte („metabolomics“), der Lipide („lipidomics“) und epigenetischer Zustände („epigenomics“), um nur die wesentlichen zu nennen. Im Mittelpunkt der Forschung steht damit nicht mehr nur die Analyse einzelner DNA-Bausteine, sondern die Untersuchung komplexer Vorgänge in der Zelle.

Parallel zu den „klassischen“ Anwendungen dieser Omics-Methoden auf viele Zellen wurden Einzelzell-Omics-Technologien auf der Basis des sogenannten Next-Generation-Sequencing (NGS) weiterentwickelt. Die Entwicklung neuer NGS-Verfahren ermöglicht heutzutage eine schnelle und effiziente Sequenzierung von Milliarden von DNA-Molekülen in kurzer Zeit. Ein wesentlicher Schritt für die Entwicklung der Einzelzellanalytik war die Nutzung von Mikrofluidik-Methoden, um RNA- und DNA-Moleküle vieler einzelner Zellen separat zu vermehren und mit Hochdurchsatz-Sequenzierungsprotokollen zu verknüpfen. Durch die Parallelisierbarkeit von Sequenzierungen kann man mittlerweile umfassende molekulare Einzelzell-Signaturen wie zum Beispiel Transkriptome (mRNA), Epigenome (DNA-Methylierung) oder Chromatinveränderungen (Chromatinzugänglichkeit) von Tausenden bis Millionen von Zellen in einem Sequenzierlauf erfassen. Diese NGS-basierten Ansätze werden in jüngster Zeit durch Verfahren ergänzt, die Proteom- und Metabolom-Daten aus einzelnen Zellen erheben. Darüber hinaus wurden erste Einzelzell-Multi-Omics-Ansätze entwickelt, die gleichzeitig Daten des Transkriptoms, des Chromatins und der DNA-Methylierung erzeugen und

dadurch erstmalig ermöglichen, eine direkte Beziehung der Genaktivierung und genregulatorischen Veränderung in einer Zelle zu verstehen.

### **Anwendungsbreite der Einzelzellanalytik in der Biologie, Biotechnologie und Medizin**

Die moderne NGS-basierte Einzelzellanalytik erlaubt es, die molekularen Signaturen nicht nur hunderter, sondern mittlerweile mehrerer Millionen einzelner Zellen zu erfassen. Hierdurch ergeben sich vollkommen neue Perspektiven für die Biologie. Komplexe Prozesse wie zum Beispiel die Strukturbildung in Fliegenlarven oder die Entwicklung von Organen können zellgenau erfasst werden. Ergänzt durch neue hochauflösende und dynamische bildgebende Verfahren ist es möglich, auch die räumliche Zuordnung und entwicklungsbiologische Dynamik einzelner Zellen im Organ oder Gewebe zu modellieren. Hieraus ergeben sich aktuell und perspektivisch tief greifende neue Einblicke in Prozesse der Entwicklung und Erkrankung. Für den Menschen deutet sich unmittelbar ein breites Spektrum neuer direkter medizinischer Anwendungsmöglichkeiten an. Diese reichen von der exakten Bestimmung der Zusammensetzung und Verteilung von Zellpopulationen (z. B. Stammzellen, Immunzellen) über die Erfassung zellulärer Veränderungen bei chronischen Erkrankungen und die Bestimmung der Auswirkungen genetischer Erkrankungen auf einzelne Zelltypen bis hin zur hochauflösenden Analyse individueller Tumore für eine auf das Individuum bezogene Therapie (personalisierte Medizin). Einzelzellanalysen werden darüber hinaus eine wichtige Rolle für das sich schnell entwickelnde Feld der Organoiden darstellen.

Tief gehende Einzelzellanalytik lässt sich jedoch nicht nur auf Mensch und Tier, sondern auch auf Mikroorganismen und Pflanzen anwenden. Bei Bakterien wird zum Beispiel untersucht, wie sich einzelne Zellen einer Bakterienkolonie unterscheiden und ob diese Unterschiede Auswirkungen auf die Pathogenität haben. Die Untersuchung von Pflanzen ist aufgrund der festen Zellwände von Pflanzenzellen schwieriger. Zu den Problemstellungen, die im Kontext der Pflanzenzüchtung untersucht werden, gehören Fragen nach der Reaktion von Zellen auf einen Angriff von Pathogenen und nach Mechanismen der Resistenz gegen Krankheitserreger, nach dem Einfluss von Variationen der Umweltbedingungen auf zelluläre Prozesse und Entwicklungsabläufe und nach der Rolle von genetischen Netzwerken. Neu gewonnene Erkenntnisse könnten zur gezielteren Züchtung und Verbesserung von Eigenschaften von Nutzpflanzen führen.

## Datenanalyse und Infrastruktur

Die Einzelzellanalyse wird in Deutschland bereits an vielen spezialisierten Zentren durchgeführt. Diese Zentren haben neben der experimentellen Infrastruktur in der Regel auch Methoden zur Datenerfassung, Datenspeicherung und Dateninterpretation erarbeitet. Die Deutsche Forschungsgemeinschaft (DFG) hat kürzlich vier neue DNA-Sequenzierungszentren mit neuester Infrastruktur ausgestattet, die im Hochdurchsatz Daten auch für Einzelzellanalysen erzeugen können. Die der Einzelzellsequenzierung nachfolgende bioinformatische (statistische und modellierende) Datenanalyse stellt die Biologie und Medizin vor große Herausforderungen, auf die die Bioinformatik und Dateninfrastruktur noch nicht flächendeckend eingestellt sind. Einzelzellanalysen benötigen daher neue, komplexe bioinformatisch-methodische Verfahren der Datenerfassung und Nutzung. Zudem müssen in diesem Bereich neue Standards und Referenzdaten generiert werden, um vergleichbare Interpretationen zu ermöglichen. Um die wachsenden Datenmengen effizient verarbeiten und für die Forschung zugänglich und nutzbar machen zu können, werden für die Analysen, insbesondere für komplexe Prozessmodellierungen, zunehmend Methoden der künstlichen Intelligenz und des maschinellen Lernens (z. B. sogenannte Deep-Learning-Methoden) genutzt. Für die Verwendung von Einzelzellanalysen in der klinischen Diagnostik wird es notwendig sein, die komplexen Einzelzellanalysen zu abstrahieren und in klinisch nutzbare Kernaussagen zu übersetzen.

## Fachspezifische Implikationen

Die Einzelzellanalytik entwickelt sich mit einer rasanten Geschwindigkeit. Für die Wissenschaft bedeutet dies, dass einerseits technische Anpassungen schnell erfolgen müssen, wenn einzelne Standorte hier kompetitiv bleiben wollen, und dass andererseits der Fortbildung im Umgang mit den Technologien ein breiter Raum eingeräumt werden muss. Auch das Wissen um die Anwendungsmöglichkeiten und ihre Grenzen muss dabei bezogen auf die jeweiligen Forschungsrichtungen noch geschärft werden. Ein weiterer wichtiger Aspekt, der durch die Einführung der Einzelzellanalytik in die Lebenswissenschaften induziert wird, ist der deutlich zunehmende Einfluss anderer Fachdisziplinen, insbesondere der Mathematik, der Bioinformatik und der Computerwissenschaften. Hier wird es besonders wichtig sein, die Ausbildung und den konstruktiven kritischen Dialog über Fachgrenzen hinweg stärker auszubauen. Neben dieser fachspezifischen Ausbildung müssen

geeignete experimentelle und bioinformatische Rahmenbedingungen vorhanden sein, um Einzelzelldaten nachhaltig nutzen zu können.

Die Anwendungs- und Nutzungsbedingungen der Einzelzellanalytik in den Lebenswissenschaften, der Biotechnologie und Medizin werden breit gestreut sein. Während in den Lebenswissenschaften die Entwicklung und breite Anwendung unterschiedlicher NGS-Technologien für die Grundlagenforschung im Vordergrund stehen werden, bilden die Entwicklung und Anwendung standardisierter Verfahren sehr wahrscheinlich den Schwerpunkt in biotechnologischen Anwendungen und der Medizin. In der Biologie können auf Basis der Einzelzellanalytik Prinzipien der funktionalen Gemeinsamkeit und Diversität zwischen Organismen auf einem neuen molekularen Niveau erfasst werden. Aspekte der Biodiversität, individueller und ökologischer Adaptation sind so viel genauer zu bestimmen. Auf der medizinischen Seite werden Einzelzelldaten neue Möglichkeiten für eine personenbezogene molekulare Diagnostik (z. B. von Krebserkrankungen) ermöglichen und in der Forschung und Anwendung von zellbasierten Verfahren (Stammzellen, Regeneration, Organoide) unverzichtbar sein. Auch im Bereich von zellbasierten Test- und Produktionsverfahren innerhalb der Pharmakologie wird die Einzelzellanalyse eine wichtige Rolle für die Qualitätssicherung spielen.

### Technologiefolgenabschätzung

Wie bei allen neuen biotechnologischen Verfahren wird es wichtig sein, den Nutzen, das Anwendungsspektrum, aber auch die Unschärfen und Grenzen der neuen Technologie kritisch zu analysieren und mit einer breiten Öffentlichkeit zu diskutieren. Die Einzelzellanalyse berührt eine Reihe ethischer Fragen, die auch im Kontext von anderen Biotechnologien erörtert werden und von großer gesellschaftlicher Relevanz sind, allen voran der Umgang mit sensiblen medizinischen Daten. Es ist zu prüfen, inwieweit bestehende Regeln zum verantwortungsvollen Umgang und für eine ausreichende Datensicherheit und Datensouveränität an die neuen Möglichkeiten angepasst werden müssen. Im Bereich der Forschung muss die kritische Auseinandersetzung mit den Daten gestärkt werden, um Fehlinterpretationen und Fehleinschätzungen zu vermeiden.

Die Zusammenführung und gemeinsame Analyse von genetischen (Genomdaten) und Einzelzelldaten wird das Interpretationsspektrum vertiefen und neue Dimensionen der individuellen Bestimmtheit eröffnen. Diese Möglichkeiten

müssen im Hinblick auf ihre ethischen und sozialpolitischen Implikationen diskutiert werden. Die Datensouveränität von möglichen Probanden und Probandinnen sowie Patienten und Patientinnen ist dabei unbedingt zu wahren. Die Kenntnis von individualisierten Einzelzelldaten hebt die Frage nach Individualität bzw. individueller Ausprägung der genetischen Basis zudem auf eine andere Ebene. Der Zusammenhang zwischen Genotyp und Phänotyp könnte in einem Maße erkennbar werden, der Vorhersagen des Phänotyps anhand zellulärer Merkmale erlaubt. Dann könnten etwa anhand einer Biopsie noch genauere Vorhersagen über zukünftige Erkrankungen oder Krankheitsverläufe als bisher möglich werden. Dies würde gegenüber herkömmlichen Gentests einen enormen Erkenntniszuwachs bedeuten. Es ist dabei wichtig, Befunde so zu kommunizieren, dass sie von den Betroffenen verstanden und in ihrer Bedeutung eingeschätzt werden können. Welcher Umfang neuer Erkenntnisse bezüglich Genfunktion und deren zellulärer Ausprägung sich hieraus ergeben wird, ist jedoch noch unklar.

## **8.2 HANDLUNGSEMPFEHLUNGEN FÜR DEN UMGANG MIT DER EINZELZELLTECHNOLOGIE UND EINZELZELLDATEN**

- Die Einzelzellanalytik ist eine Zukunfts- und Schlüsseltechnologie für die Biologie und die Medizin. Ihre Bedeutung wird in den kommenden Jahren stark zunehmen. Die Technologie sollte daher in den Agenden der Forschungsförderung einen prominenten Platz einnehmen.
- Deutschland hat sehr gute erste Ansätze (Zentren) in der Einzelzelltechnologie und ausgewiesene Kompetenz in der bioinformatischen Bearbeitung von Einzelzelldaten. Es gilt, diese Stärken zu bewahren und weiter auszubauen, zum Beispiel durch Forschungsinitiativen wie LifeTime oder Single Cell Omics Germany (SCOG) und Strukturen wie die DFG-Sequenzierungszentren.
- Für die klinische Nutzung von Einzelzelldaten müssen standardisierte Abläufe geschaffen werden. Die „Medizininformatik-Initiative“ kann hierfür geeignete Rahmenbedingungen schaffen. Es müssen Standards entwickelt werden, um komplexe Einzelzelldaten für die klinische Applikation aufzubereiten, die für die Entwicklung neuer Diagnose- und Therapieverfahren benutzt werden können. Die im klinischen Zusammenhang erhobenen Einzelzelldaten müssen in einem geschützten Raum verbleiben und vor dem Zugriff Unbefugter effektiv bewahrt werden.

- Für eine breite Nutzung der generierten Daten (auch zur Erhebung von Referenzen, mit denen etwa Proben von Patienten und Patientinnen verglichen werden können) sollten geeignete Datenstrukturen aufgebaut werden, die auf einheitlichen Dokumentationsstandards aufbauen, um eine größtmögliche Kompatibilität zu erreichen. Geeignete Rahmenbedingungen für die Datensicherung und Datensicherheit sind dabei zu schaffen (analog zu Genomdaten). Dies sollte sich auch in der nationalen Forschungsdateninfrastruktur (NFDI) widerspiegeln.
- Es wird wichtig sein, Verfahren für eine Aufklärung und informierte Zustimmung sowohl für die Anwendung einzelzelldiagnostischer Verfahren für Patienten und Patientinnen als auch für die Forschung mit personalisierten Daten zu etablieren.
- Im Bereich personenbezogener Daten wirft die Einzelzellbiologie keine prinzipiell neuen rechtlichen und ethischen Fragen auf. Jedoch kann es durch die Analyse der individuellen Ausprägung des Genoms in Einzelzellen zu neuen Erkenntnissen, aber auch zu noch unsicheren Interpretationsspielräumen kommen, die beide gleichermaßen Aspekte persönlicher Stigmatisierung bzw. Diskriminierung betreffen. Nicht nur bezogen auf Zufallsbefunde sind das Recht auf Nichtwissen und der Schutz der Persönlichkeit in diesem Zusammenhang erneut und verstärkt zu diskutieren und gegebenenfalls rechtlich abzusichern.
- Grundsätzlich darf der aktuell beginnende Hype um die Einzelzellanalyse keinesfalls dazu führen, dass in anderen Feldern bereits etablierte ethische Standards abgesenkt oder negiert werden. Wie immer müssen die Standards guter wissenschaftlicher, guter klinischer und guter ethischer Praxis eingehalten werden, damit mehr Daten auch zu mehr Wissen führen – und zwar zum Wohl der Individuen wie auch der Gesellschaft.

## 9. ANHANG

### 9.1 AUTORINNEN UND AUTOREN

**Dr. Hananeh Aliee**

Postdoc am Institute of Computational Biology, Helmholtz Zentrum München.

**Dr. Anna C. Aschenbrenner**

Postdoc am Life & Medical Sciences-Institut, Universität Bonn.

**Prof. Dr. Heiner Fangerau**

Lehrstuhlinhaber und Direktor des Instituts für Geschichte, Theorie und Ethik der Medizin, Centre for Health and Society, Medizinische Fakultät, Universität Düsseldorf; Mitglied der IAG *Gentechnologiebericht*.

**Prof. Dr. Boris Fehse**

Leiter der Forschungsabteilung Zell- und Gentherapie an der Klinik für Stammzelltransplantation, Universitätsklinikum Hamburg-Eppendorf; Präsident der Deutschen Gesellschaft für Gentherapie; Sprecher der IAG *Gentechnologiebericht*.

**Dr. Nina Gasparoni**

Postdoc in der Arbeitsgruppe EpiGenetik, Universität des Saarlandes.

**Dr. Jürgen Hampel**

Assistent am Lehrstuhl für Technik- und Umweltsoziologie am Institut für Sozialwissenschaften, Universität Stuttgart; Mitglied der IAG *Gentechnologiebericht*.

**Prof. Dr. Ferdinand Hucho**

Emeritierter Professor für Biochemie an der Freien Universität Berlin, Institut für Chemie und Biochemie; Mitglied der Berlin-Brandenburgischen Akademie der Wissenschaften; stellvertretender Sprecher der IAG *Gentechnologiebericht*.

**Dr. J. Philipp Junker**

Leiter der AG Junker zum Thema Quantitative Entwicklungsbiologie am Berlin Institute of Medical Systems Biology des Max-Delbrück-Centrums für Molekulare Medizin.

**Prof. Dr. Martin Korte**

Leiter der Abteilung für Zelluläre Neurobiologie, Molekulare Zellbiologie, Zoologisches Institut, TU Braunschweig; Mitglied der Berlin-Brandenburgischen Akademie der Wissenschaften; Mitglied der IAG *Gentechnologiebericht*.

**Yaroslav Koshelev**

Studentische Hilfskraft der Geschäftsstelle der IAG *Gentechnologiebericht*.

**Dr. Lilian Marx-Stölting**

Wissenschaftliche Mitarbeiterin der Geschäftsstelle der IAG *Gentechnologiebericht*.

**Dr. Elvira Mass**

Arbeitsgruppenleiterin zur Entwicklungsbiologie des angeborenen Immunsystems am Life & Medical Sciences-Institut, Universität Bonn.

**Prof. Dr. Bernd Müller-Röber**

Professor für Molekularbiologie am Institut für Biochemie und Biologie, Universität Potsdam; Mitglied der Berlin-Brandenburgischen Akademie der Wissenschaften und bis 2019 der IAG *Gentechnologiebericht*.

**Angela Osterheider**

Wissenschaftliche Mitarbeiterin der Geschäftsstelle der IAG *Gentechnologiebericht* und Doktorandin am Institut für Publizistik- und Kommunikationswissenschaft, Freie Universität Berlin.

**Dr. Christian Popp**

Abteilung Programm-Management BIMSB am Max-Delbrück-Centrum für Molekulare Medizin, Berlin.

**Prof. Dr. Nikolaus Rajewsky**

Wissenschaftlicher Direktor am Berlin Institute for Medical Systems Biology, Max-Delbrück-Centrum für Molekulare Medizin, Berlin.

**Prof. Dr. Jens Reich**

Emeritierter Professor für Molekularbiologie am Max-Delbrück-Centrum für Molekulare Medizin und an der Humboldt-Universität zu Berlin; Mitglied der Berlin-Brandenburgischen Akademie der Wissenschaften; Mitglied der IAG *Gentechnologiebericht*.



**Marlen Reinschke**

Studentische Hilfskraft der Geschäftsstelle der IAG *Gentechnologiebericht*.

**Dr. Anna Sacher**

Head of Operations & Science Management am Helmholtz Zentrum München, Institute of Computational Biology.

**Hannah Schickl**

Wissenschaftliche Mitarbeiterin und Koordinatorin der Geschäftsstelle der IAG *Gentechnologiebericht* und wissenschaftliche Mitarbeiterin am Lehrstuhl für Systematische Theologie II (Ethik), Friedrich-Alexander-Universität Erlangen-Nürnberg.

**Prof. Dr. Joachim L. Schultze**

Leiter der Abteilung für Genomik & Immunoregulation am Life & Medical Sciences-Institut der Universität Bonn. Direktor der PRECISE Platform for Single Cell Genomics and Epigenomics am Deutschen Zentrum für Neurodegenerative Erkrankungen und der Universität Bonn.

**Prof. Dr. Jochen Taupitz**

Geschäftsführender Direktor des Instituts für Deutsches, Europäisches und Internationales Medizinrecht, Gesundheitsrecht und Bioethik der Universitäten Heidelberg und Mannheim (IMGB); Inhaber des Lehrstuhls für Bürgerliches Recht, Zivilprozessrecht, internationales Privatrecht und Rechtsvergleichung der Universität Mannheim; Mitglied der IAG *Gentechnologiebericht*.

**Prof. Dr. Fabian J. Theis**

Direktor des Institute of Computational Biology am Helmholtz Zentrum München.

**Prof. Dr. Jörn Walter**

Professor für Genetik/Epigenetik, Universität des Saarlandes, Mitglied der IAG *Gentechnologiebericht*.

**Prof. Dr. Martin Zenke**

Direktor des Instituts für Biomedizinische Technik – Zellbiologie, Universitätsklinikum der RWTH Aachen und Helmholtz-Institut für Biomedizinische Technik, RWTH Aachen; Mitglied der IAG *Gentechnologiebericht*.

## 9.2 MITGLIEDER DER INTERDISZIPLINÄREN ARBEITSGRUPPE *GENTECHNOLOGIEBERICHT*

**Prof. Dr. Boris Fehse** (Sprecher der IAG)

Universitätsklinikum Hamburg-Eppendorf, Klinik für Stammzelltransplantation

**Prof. Dr. Ferdinand Hucho\*** (Stellvertretender Sprecher der IAG)

Freie Universität Berlin, Institut für Chemie und Biochemie

**Prof. Dr. Heiner Fangerau**

Universität Düsseldorf, Institut für Geschichte, Theorie und Ethik der Medizin

**Dr. Jürgen Hampel**

Universität Stuttgart, Institut für Sozialwissenschaften

**Prof. Dr. Martin Korte\***

TU Braunschweig, Institut für Zelluläre Neurobiologie

**Prof. Dr. Stefan Mundlos\***

Charité Berlin, Institut für Medizinische Genetik und Humangenetik

**Prof. Dr. Bernd Müller-Röber\***

Universität Potsdam, Institut für Biochemie und Biologie

**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. Jochen Taupitz**

Universität Mannheim, Fakultät für Rechtswissenschaft  
und Volkswirtschaftslehre

**Prof. Dr. Jörn Walter**

Universität des Saarlandes, Institut für Biowissenschaften

**Prof. Dr. Martin Zenke**

RWTH Aachen, Universitätsklinikum, Institute for Biomedical Engineering

\* Mitglied der Berlin-Brandenburgischen Akademie der Wissenschaften

### 9.3 PUBLIKATIONEN DER INTERDISZIPLINÄREN ARBEITSGRUPPE *GENTECHNOLOGIEBERICHT*

#### Bücher

Hucho, F. et al. (Hrsg.) (2018): **Vierter Gentechnologiebericht**. Bilanzierung einer Hochtechnologie. Nomos, Baden-Baden.

Zenke, M. et al. (Hrsg.) (2018): **Stammzellforschung**. Aktuelle wissenschaftliche und gesellschaftliche Entwicklungen. Nomos, Baden-Baden. Unter: <https://www.nomos-elibrary.de/10.5771/9783845287720.pdf> [21.08.2019].

Walter, J./Hümpel, A. (Hrsg.) (2017): **Epigenetik**. Implikationen für die Lebens- und Geisteswissenschaften. Nomos, Baden-Baden.

Müller-Röber, B. et al. (Hrsg.) (2015): **Dritter Gentechnologiebericht**. Analyse einer Hochtechnologie. Nomos, Baden-Baden.

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 neuer 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**

IAG Gentechnologiebericht (Hrsg.) (2018): **Vierter Gentechnologiebericht**. Bilanzierung einer Hochtechnologie. Kurzfassung. BBAW, Berlin.

IAG Gentechnologiebericht (Hrsg.) (2018): **Stammzellforschung**. Aktuelle wissenschaftliche und gesellschaftliche Entwicklungen. Kurzfassung. BBAW, Berlin.

IAG Gentechnologiebericht (Hrsg.) (2017): **Epigenetik**. Implikationen für die Lebens- und Geisteswissenschaften. Kurzfassung. BBAW, Berlin.

IAG Gentechnologiebericht (Hrsg.) (2015): **Dritter Gentechnologiebericht**. Analyse einer Hochtechnologie. Kurzfassung. BBAW, Berlin.

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.

IAG Gentechnologiebericht (Hrsg.) (2013): **Grüne Gentechnologie**. Aktuelle wissenschaftliche, wirtschaftliche und gesellschaftliche Entwicklungen. Kurzfassung. BBAW, Berlin.

IAG Gentechnologiebericht (Hrsg.) (2012): **Synthetische Biologie**. Entwicklung einer neuen Ingenieurbiologie? Kurzfassung. BBAW, Berlin.

IAG Gentechnologiebericht (Hrsg.) (2011): **Gentherapie in Deutschland**. Eine interdisziplinäre Bestandsaufnahme. Kurzfassung. BBAW, Berlin.

IAG Gentechnologiebericht (Hrsg.) (2009): **Zweiter Gentechnologiebericht**. Analyse einer Hochtechnologie in Deutschland. Kurzfassung. BBAW, Berlin.

Beier, H. et al. (2009): **Neue Wege der Stammzellforschung**. Reprogrammierung von differenzierten Körperzellen. Hg v. Berlin-Brandenburgische Akademie der Wissenschaften, Berlin. Unter: [http://www.bbaw.de/service/publikationen-bestellen/manifeste-und-leitlinien/BBAW\\_Stammzellforschung.pdf](http://www.bbaw.de/service/publikationen-bestellen/manifeste-und-leitlinien/BBAW_Stammzellforschung.pdf) [21.08.2019].

IAG Gentechnologiebericht (Hrsg.) (2008): **Gentherapie in Deutschland**. Eine interdisziplinäre Bestandsaufnahme. Zusammenfassung. BBAW, Berlin.

IAG Gentechnologiebericht (Hrsg.) (2007): **Gendiagnostik in Deutschland**. Status quo und Problemerkundung. Zusammenfassung. BBAW, Berlin.

IAG Gentechnologiebericht (Hrsg.) (2005): **Gentechnologiebericht**. Analyse einer Hochtechnologie in Deutschland. Kurzfassung. BBAW, Berlin.

### Sonstige Publikationen

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: <http://www.bbaw.de/publikationen/neuerscheinungen/pdf/Debatte-Heft-19> [21.08.2019].

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> [21.08.2019].

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: <http://www.bbaw.de/publikationen/stellungnahmen-empfehlungen/Stellungnahmen-Gendiagnostik.pdf> [21.08.2019].

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.



Diese Stellungnahme ist in deutscher und englischer Sprache online abrufbar unter:  
[www.gentechnologiebericht.de/publikationen/einzelzellanalyse-in-forschung-und-medizin-2019/](http://www.gentechnologiebericht.de/publikationen/einzelzellanalyse-in-forschung-und-medizin-2019/)

ISBN: 978-3-939818-84-7