

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

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

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



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

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

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

In Kooperation mit dem Deutschen Humangenom-Phänomarchiv

In cooperation with The German Human Genome-Phenome Archive



Berlin Institute of Health at Charité (BIH)

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

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

Layout/Graphic: Polygraph Design

Druck/Printing: Druckerei Rüss

Übersetzung/Translation: Baker & Company

Auflage/Circulation: 500

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

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

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



ISBN: 978-3-00-078676-1

DOI: <http://dx.doi.org/10.17169/refubium-43070>

Berlin, April 2024

Inhaltsverzeichnis

Jörn Walter und Boris Fehse

Vorwort

6

Olaf Rieß und Julia Schulze-Hentrich

1.	Die Bedeutung und klinische Nutzung moderner Genomsequenzierungstechnologien – eine Einführung am Beispiel seltener Erkrankungen	8
1.1	Einleitung	8
1.2	Darstellung aktueller Fortschritte in der klinischen Genomsequenzierung	9
1.3	Herausforderungen und Weiterentwicklungen hin zur genomischen Medizin	12
1.4	Fazit	13
1.5	Literaturverzeichnis	14

Thorsten Schlomm und Nadine Biernath

2.	Die Relevanz und der praktische Nutzen von Genomdaten im Bereich der Krebsdiagnostik	15
2.1	Einleitung	15
2.2	Kurze Historie der molekularen Therapie	16
2.3	Moderne Ansätze in der onkologischen Genomik: Von Standardsequenzierung bis zum Whole-Genome-Sequencing	16
2.4	Von der Analyse zur Anwendung: Fortschritte in der Genomik und deren Nutzbarmachung in der Onkologie	17
2.5	Ausblick: Mit „Schwarmintelligenz“ die kollektive Erfahrung der medizinischen Gemeinschaft nutzen	19
2.6	Fazit	20
2.7	Literaturverzeichnis	20

Jörn Walter, Nina Gasparoni und Oliver Kohlbacher

3.	Die sichere Archivierung komplexer Genomdaten und ihre effiziente Nutzung – eine Herausforderung	21
3.1	Einführung	21
3.2	Genomdaten in Forschung und Medizin	21
3.3	Herausforderungen einer zentralisierten Datennutzung: Datenschutz und informierte Einwilligung	25
3.4	Strukturen für eine Archivierung und den kontrollierten Zugang	26
3.5	Fazit	29
3.6	Literaturverzeichnis	30

	<i>Martin Jungkunz, Christoph Schickhardt und Eva C. Winkler</i>	
4.	Der ethische Rahmen für die Nutzung von Genomdaten in der Forschung	31
4.1	Einleitung	31
4.2	Ethische Rahmenbedingungen für die Forschungsnutzung genomischer Daten	32
4.3	Ausblick	37
4.4	Interessenkonflikt	38
4.5	Literaturverzeichnis	38
	<i>Ralf Müller-Terpitz und Hannah Lilly Straub</i>	
5.	Der rechtliche Rahmen für die Nutzung von Patient*innen- und Genomdaten – Schutz und Nutzen aus medizinrechtlicher Sicht	39
5.1	Schutz der Patient*innen und ihrer Daten durch das GenDG	39
5.2	Schutz der Genomdaten durch datenschutzrechtliche Vorschriften	41
5.3	Fazit: Das Potenzial von Genomdaten im Rahmen rechtlicher Vorgaben	44
5.4	Literaturverzeichnis	45
	<i>Jörn Walter</i>	
	Ausblick: Der Weg in eine zukunftsorientierte humane Genomdiagnostik und Genomforschung	46
	Autor*innen	48
	Mitglieder der Arbeitsgruppe Gentechnologiebericht	49
	Publikationen der Arbeitsgruppe Gentechnologiebericht	50

Vorwort

Die Entschlüsselung des menschlichen Genoms im Jahr 2000 setzte einen Meilenstein der modernen Medizin mit großen und weitreichenden Auswirkungen für die genetische Diagnostik und die Erforschung menschlicher Entwicklung, Gesundheit und Erkrankungen. Mit der Nutzung höchst effizienter Genomsequenzierungstechnologien, ergänzt durch weitere Omics-Technologien und -Daten wie z. B. aus der Epigenomik, Proteomik, Transkriptomik (einschließlich der hochaktuellen Einzelzellsequenzierung), stehen wir mittlerweile vor einer Zeitenwende in der genetischen Diagnostik und bewegen uns mit raschen Schritten in die Richtung einer datengestützten, funktionellen Genominterpretation, die ganz neue Perspektiven eröffnen wird. Sie wird ein grundlegend neues Verständnis für die Beurteilung, die individuelle Diagnose und Behandlung von Krankheiten mit sich bringen. Die rasanten technischen Entwicklungen der vergangenen Jahre eröffnen darauf aufbauend die praktische Möglichkeit, die Entschlüsselung eines Patientengenoms in die gesundheitliche Routineversorgung einzubeziehen.

In dieser Broschüre erörtern wir die Möglichkeiten und den gegenwärtigen Stand der personenbezogenen Genomsequenzierung, ihren Nutzen für die Erforschung von seltenen Erkrankungen und Krebserkrankungen und diskutieren einige wichtige sich aus der Genomforschung ergebende ethische, rechtliche und soziale Fragen.

Den Anlass für die aktuelle Beschäftigung mit den verschiedenen Aspekten und Fragen der Genomsequenzierung bietet das am 1. April 2024 beginnende Modellvorhaben § 64e SGB, in dem zum ersten Mal Genomanalysen in die klinische Routineversorgung mit einbezogen werden. Dieses Konzept wird zunächst modellhaft für ausgewählte Bereiche genetisch bedingter Erkrankungen durchgeführt. Es soll dabei überprüft werden, inwieweit eine individuelle Genomanalyse in der medizinischen Routinediagnostik genutzt werden kann und welchen Nutzen die systematische Genomanalyse für die aktuelle Forschung bietet.

Die Voraussetzungen für den Erfolg dieses Modellvorhabens sind zweifellos ausgezeichnet. Die technische Machbarkeit einer breit angelegten und bezahlbaren Genomdiagnostik ist gegeben. Bereits laufende, breit angelegte vergleichende Genomstudien in anderen Ländern bezeugen, welche große und breite Bedeutung nationale und internationale Genomprogramme für die Forschung und das Gesundheitsmanagement haben. Neben dem unmittelbaren gesundheitlichen Nutzen für den Einzelnen im Bereich Diagnostik ergeben sich aus Genomprogrammen weitreichende neue Erkenntnisse für die medizinische und pharmazeutische Forschung. Aufbauend auf immer besseren datengestützten Analysen und der Nutzung computergestützter Methoden des maschinellen Lernens (Künstliche Intelligenz, KI) wird dieses Wissen zunehmend genutzt, um die Bedeutung genetischer Variation für die allgemeine und individuelle Gesundheit zu erschließen. Die Kenntnis und vor allem der Vergleich von vielen individuellen Genomen werden neben einer immer besseren

personenbezogenen Diagnostik auch essenzielle Daten für breit angelegte Gesundheitsstudien und das allgemeine Gesundheitsmanagement in Deutschland liefern.

Diese Broschüre bietet in fünf Beiträgen und einem kurzen Ausblick Einblicke in den Stand der aktuellen Genomsequenzierung in Deutschland und ihre Anwendung im Bereich der Humandiagnostik. Thematisch widmen sich die Kapitel der Entschlüsselung seltener Erkrankungen (Kapitel 1), der Diagnostik von Krebserkrankungen (Kapitel 2), den zentralen Aspekten der Datensicherung, Datenarchivierung und der Datennutzung (Kapitel 3) sowie des Datenschutzes (Kapitel 3, 4 und 5). In Kapitel 4 setzen sich die Autor*innen mit der Frage der ethischen Bewertung im individuellen und gesellschaftsbezogenen Umgang mit Genomdaten auseinander. Abschließend wird der rechtliche Rahmen der Genomdatenutzung für die zuvor genannten Aspekte beleuchtet (Kapitel 5) und ein kurzer Ausblick auf mögliche folgende Schritte gegeben.

Die als Mitherausgeber der Broschüre fungierende Arbeitsgruppe „Gentechnologiebericht – Monitoring und interdisziplinärer Dialog“ ist interdisziplinär zusammengesetzt aus Natur-, Geistes- und Sozialwissenschaftler*innen. Unsere AG widmet ihre Arbeit einem breiten Spektrum von Schwerpunkten im Bereich Gentechnologien und behält deren Entwicklungen langfristig im Blick. Unsere regelmäßig erscheinenden Veröffentlichungen richten sich überwiegend an eine breite Öffentlichkeit. Im Rahmen unseres „Im Fokus“-Formates reagieren wir auf aktuelle Themen und Entwicklungen.

Für diesen Band konnten wir eine Reihe auf dem Gebiet der Genomdiagnostik und Genomforschung aktiver Wissenschaftler*innen gewinnen. Zusätzlich zu der Expertise innerhalb der AG mit den Mitgliedern Eva C. Winkler, Ralf Müller-Terpitz und Jörn Walter konnten wir externe Expert*innen gewinnen wie Olaf Rieß, Julia Schulze-Hentrich, Oliver Kohlbacher und Thorsten Schlomm, die gemeinsam mit ihren Mitarbeiter*innen Nadine Biernath, Nina Gasparoni, Martin Jungkunz, Christoph Schickhardt und Hannah Lilly Straub ein Spektrum von Themen wissenschaftlich fundiert und allgemein verständlich ausgearbeitet haben. Hervorzuheben ist dabei, dass die Autor*innen Kohlbacher, Rieß, Schlomm, Walter, Winkler, Gasparoni und Schickhardt nicht nur Expert*innen in den jeweiligen Gebieten sind, sondern auch aktiv in den Forschungsinitiativen GenomDE und dem Deutschen Humangenom-Phänomarchiv (GHGA) tätig sind und so auch in anderen Kontexten die deutsche Genomforschung wie auch das Modellvorhaben begleiten. Wir danken dem GHGA sehr herzlich für die Kooperation bei dieser Publikation wie auch den Autor*innen für ihre Mitwirkung an unserer Broschüre. Zu großem Dank verpflichtet sind wir dem Berlin Institute of Health (BIH) in der Charité für die Aufnahme und Unterstützung der AG.

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

Jörn Walter und Boris Fehse

Stellvertretender Sprecher und Sprecher der AG *Gentechnologiebericht* am BIH
Berlin/Hamburg, im März 2024

1. Die Bedeutung und klinische Nutzung moderner Genomsequenzierungstechnologien – eine Einführung am Beispiel seltener Erkrankungen

Olaf Rieß und Julia Schulze-Hentrich

1.1 Einführung

Zahlreiche Erkrankungen sind durch Veränderungen im Erbmateriale, der DNA, eines Menschen bedingt. Diese genetischen Veränderungen liegen entweder bereits in der Keimbahn vor und sind damit vererbbar oder sie entstehen im Laufe des Lebens in einem Teil der Zellen wie beispielsweise bei Tumorerkrankungen und werden nur in den Tumorzellen, aber nicht an Nachkommen vererbt. Um diese Veränderungen zu erkennen und zu verorten, ist die Sequenzierung des Genoms¹ und dessen Vergleich mit Referenzsequenzen aus „gesunden“ Proband*innen notwendig.

In den letzten Jahren gab es beeindruckende technische Fortschritte bei der Genomsequenzierung, sowohl in Bezug auf Schnelligkeit und Kosten als auch bei den immer effizienteren Möglichkeiten einer bioinformatischen Auswertung und der medizinischen Bewertung. Diese Entwicklung stellt einen Quantensprung in der Genomik dar, da sie einen umfänglichen Nutzen der Technologie und der Genomdaten in der Klinik in greifbare Nähe rückt. Genomdaten werden dazu beitragen, Krankheitsursachen individuell aufzudecken, personalisierte Therapieansätze vorzuschlagen, aber auch, individuelle (genetische) Krankheitsrisiken zu bestimmen (bevor Krankheitszeichen auftreten).

Insbesondere die Genomforschung von seltenen Erkrankungen und von Tumorerkrankungen hat neue Erkenntnisse zur Genotyp-Phänotyp-Korrelation ermöglicht und gezeigt, dass es bei vielen Erkrankungen einen klaren Zusammenhang zwischen Krankheitsrisiko und der Genomsequenz einer Person gibt. Daher hat man über die letzten Jahre weltweit bereits mit einer breiten Einführung der genomischen Diagnostik in die Medizin begonnen. In Deutschland wird diese Entwicklung momentan in einem Modellvorhaben (siehe Abschn. 1.2.1 und 1.2.2) auf dem Gebiet der seltenen Erkrankungen, bei denen in ca. 80 % der Patient*innen eine genetische Ursache angenommen wird, bei familiären Tumorerkrankungen sowie für zielgerichtete Tumorthapien vollzogen. Darüber hinaus hat die Methodik ein

¹ Das Genom, auch Erbgut eines Lebewesens, ist die Gesamtheit der vererbaren Information (Gene) eines Individuums. Bei der Genomsequenzierung wird die Sequenz des Genoms, also die Abfolge der Basen der DNA, bestimmt.

hohes Potenzial, künftig auch individuelle epigenetische Veränderungen² im Rahmen der Diagnostik und für die Therapie zu bestimmen.

1.2 Darstellung aktueller Fortschritte in der klinischen Genomsequenzierung

1.2.1 Die Idee des Modellvorhabens

Aktuelle Entwicklungen in der Sequenzierung und damit der genomischen Diagnostik bergen das Potenzial, die Diagnose und Behandlung sowohl seltener als auch häufiger Krankheiten erheblich zu verbessern. Die breite Nutzung der Genomsequenzierung in der routinemäßigen Gesundheitsversorgung wurde daher von vielen Expert*innen als Notwendigkeit erkannt, wurde jedoch bisher in Deutschland durch finanzielle und infrastrukturelle Herausforderungen erschwert. In einem ersten wichtigen Schritt soll in Deutschland nun im Juli 2024 im Rahmen des Modellvorhabens nach § 64e SGB V³ für ausgewählte Personen mit seltenen Erkrankungen und erblichen Tumorerkrankungen sowie für Patient*innen im Rahmen einer individuellen zielgerichteten Tumorthherapie eine Genomsequenzierung ermöglicht und durch die Krankenkassen finanziert werden. Hierfür erarbeitet eine große Arbeitsgruppe an Ärzt*innen, Humangenetiker*innen, Wissenschaftler*innen, und Bioinformatiker*innen im Rahmen des Konzeptes genomDE⁴ Kriterien für den Zugang und die Durchführung der Diagnostik sowie für den Umgang mit den erhobenen Daten.⁵ Das Modellvorhaben nach § 64e SGB V ermöglicht die Umsetzung dieses Konzeptes an universitären Zentren und hat damit das Ziel, langfristig Strukturen zu schaffen, um die genomische Diagnostik in die Gesundheitsversorgung in Deutschland zu integrieren.

1.2.2 Fokus auf seltene Erkrankungen

Das Modellvorhaben fokussiert sich zunächst in einem Arbeitszweig auf seltene Erkrankungen, in denen bei etwa 80 % der Patient*innen eine lokalisierbare genetische Ursache angenommen wird. Eine Erkrankung gilt als selten, wenn weniger als 5 von 10.000 Menschen an ihr erkranken. Da es mehr als 6.000 abgegrenzte seltene Erkrankungen gibt und damit allein in Deutschland mehrere Millionen Betroffene, ist die Gesamtzahl der von dem Modellvorhaben profitierenden Menschen potenziell sehr hoch.

Seltene Erkrankungen sind im Erscheinungsbild häufig individuell sehr heterogen, dennoch haben sie gemeinsam, dass eine genetische Veränderung in einem krankheitsrelevanten Gen zugrunde liegt (Amberger et al., 2009; Amberger et al., 2015). Trotz enormer Anstrengungen und Fortschritte in der Sequenzierungstechnologie in den letzten Jahren, kann bisher nur in etwa 40 % der Fälle die genetische Ursache genau geklärt werden (Wright et al., 2018). Ein Grund dafür findet sich in der Herausforderung, die genetische Verände-

² Epigenetische Veränderungen umfassen chemische Veränderungen des sogenannten Chromatins, der Verpackungsstruktur der DNA, oder z. B. auch Methylierung der DNA selbst.

³ Für weitere Informationen zu dem Modellvorhaben siehe Walter et al., Kap. 3, und Müller-Terpitz/Straub, Kap. 5.

⁴ genomDE ist eine deutsche Genominitiative gefördert vom Bundesgesundheitsministerium. Das Projekt startete im Oktober 2021 und endet im Dezember 2024. genomDE ist Bestandteil der nationalen Strategie für Genommedizin und soll zukünftig einen wichtigen Baustein in der Dateninfrastruktur im deutschen Gesundheitswesen bilden. Siehe unter: <https://genom.de/de> [09.02.2024].

⁵ Siehe unter: <https://www.bundesgesundheitsministerium.de/themen/gesundheitswesen/personalisierte-medizin/genomde-de.html> [09.02.2024].

rung im Patienten/in der Patientin zu finden, die schwerwiegende funktionelle Konsequenzen hat und damit krankheitsrelevant ist. Genetische Veränderungen sind nämlich häufig, jedoch in den meisten Fällen nicht funktionell relevant. Um krankheitsverursachende Veränderungen zu identifizieren bzw. abgrenzen zu können, ist es bedeutsam die allgemeine Häufigkeit genetischer Variation in der Bevölkerung genau zu kennen. Eine seltene Veränderung, die bei einer Person (oder mehreren Personen) mit einer spezifischen Erkrankung gefunden wird, ist mit größerer Wahrscheinlichkeit relevant, sofern sie sich von häufigen Varianten bei gesunden Personen im gleichen Gen abgrenzt. Um dies zu bestimmen, sind daher umfangreiche Vergleichsdaten, d.h. Datenbanken mit ausreichenden „Referenzgenomen“, notwendig.⁶

Die am häufigsten beobachtete Veränderung sind Punktmutationen⁷, die ein einziges oder wenige Basenpaare (bp) betreffen. Allerdings gibt es auch seltene Erkrankungen, in denen mehr als eine Base vom Referenzgenom abweicht und zu krankheitsverursachenden Veränderungen führt. Dazu gehören kleine Insertionen und Deletionen (InDels), strukturelle Varianten (SVs), die mehr als 50 Basenpaare betreffen, Umlagerungen wie größere Inversionen, Copy-Number-Varianten⁸ und Trinukleotidrepeat-Expansionen⁹. Diese können mit den neusten Sequenzieretechnologien deutlich besser erfasst und damit interpretiert werden (siehe Abbildung 1 sowie Schlomm/Biernath, Kap. 2). Bisher wurden Veränderungen vor allem im Exom¹⁰ bestimmt, das alle protein-codierenden Sequenzen umfasst. Bei der Genomsequenzierung wird die komplette DNA eines Organismus sequenziert. Dies schließt alle regulatorischen Bereiche innerhalb und außerhalb bekannter Gene mit ein und damit auch funktionell bislang noch nicht zugeordnete Bereiche. Es mehren sich Befunde, dass genetische Veränderungen in diesen regulatorischen Regionen ebenfalls krankheitsverursachend sein können.

Das Modellvorhaben fokussiert sich auf die Genomsequenzierung, um all diese möglichen genetischen Ursachen in seltenen und Tumorerkrankungen bestmöglich zu identifizieren, und schließt damit die bisher bestehenden diagnostischen Lücken.

1.2.3 Übersicht aktueller Sequenzieretechnologien

Sequenzierungstechnologien haben sich in den letzten drei Jahrzehnten rasant weiterentwickelt. Die ersten Versuche, DNA zu sequenzieren, nutzten den chemischen Abbau oder die enzymatische Spaltung der DNA, um Fragmente zu erzeugen, die einzeln analysiert werden mussten (First-Generation-Sequencing, siehe Abbildung 1). Obwohl diese Techniken bereits wertvolles Wissen im Bereich der Molekulargenetik lieferten und bis vor wenigen Jahren der Goldstandard in der Humangenetik waren, waren sie extrem zeitintensiv

⁶ Siehe für weitere Informationen zu Referenzgenomen und Datenbanken Walter et al., Kap. 3.

⁷ Man unterscheidet Substitution (Basenaustausch; mit Abstand häufigste Punktmutation), Insertion (Einfügung einer Base) und Deletion (Verlust einer Base).

⁸ Copy Number Variants (CNV) sind eine Art struktureller Variation des Erbguts („structural variation“, SV), die Abweichungen der Anzahl der Kopien eines bestimmten DNA-Abschnittes erzeugt.

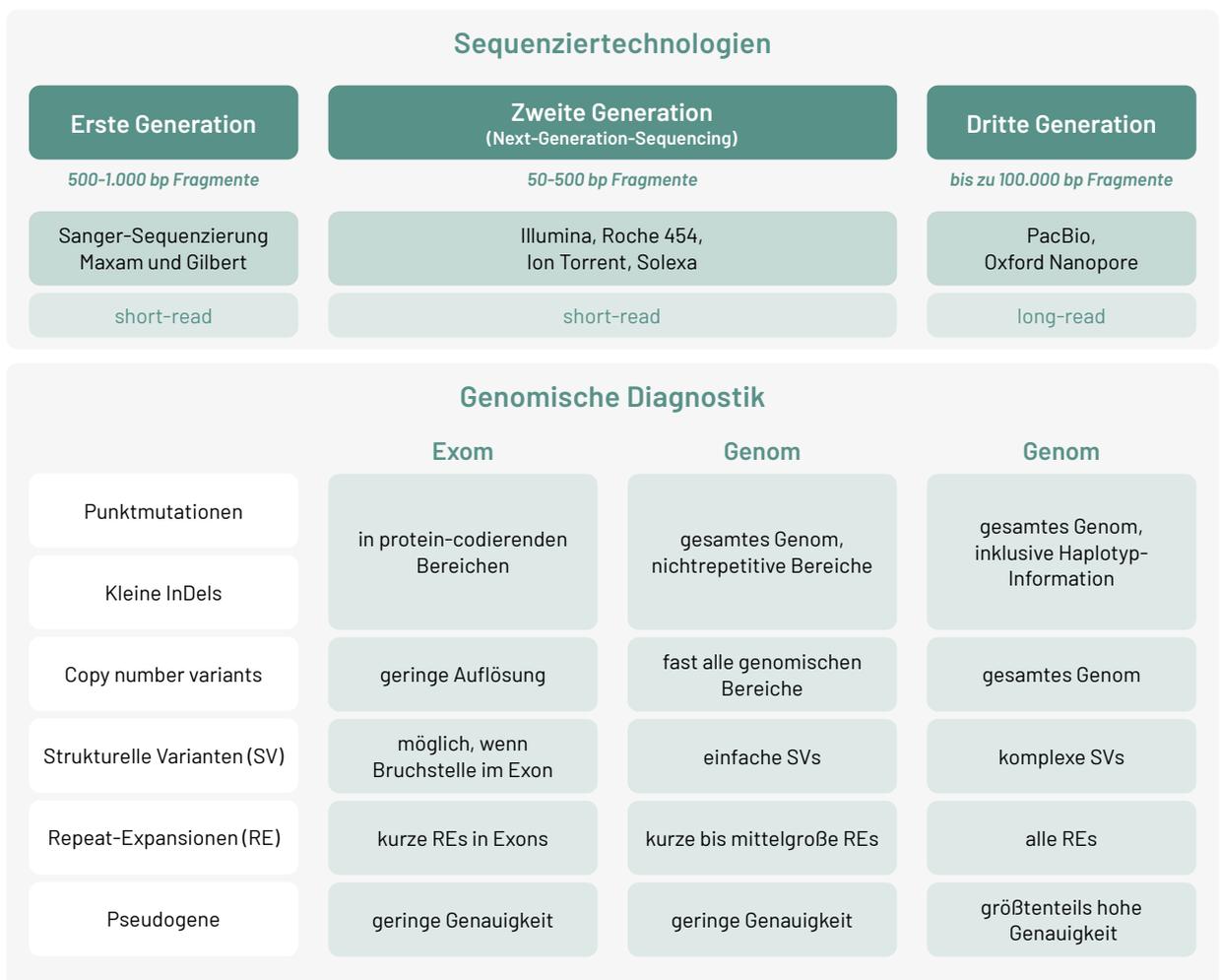
⁹ Trinukleotid-Expansionen sind DNA-Abschnitte, in denen eine Nukleotid-Dreierfolge wiederholt auftritt (z. B. CAG-CAG-CAG-CAG). Übersteigt die Repeatanzahl einen bestimmten Wert, kann dies eine Ursache für eine Erbkrankheit sein, die sogenannten Trinukleotiderkrankungen.

¹⁰ Beim Exom handelt sich um die DNA-Sequenz, die die etwa 20.000 Proteine codiert. Sie umfasst etwa 1-2% des gesamten humanen Erbgutes, also etwa 30 Millionen Basenpaare. Im Unterschied dazu umfasst das gesamte Genom auch nicht codierende Sequenzen inklusive Introns, regulatorische Sequenzen und verschiedene sich wiederholende Sequenzen.

und arbeitsaufwendig. Erst die zweite Generation der Sequenzierung (Next-Generation-Sequencing, NGS) brachte den Quantensprung in der Genomik und bot eine leistungsstarke Technologie für die Genomforschung. Diese sogenannten „short read Sequencing“-Plattformen erlauben die Sequenzierung des gesamten Genoms in kurzer Zeit, indem Millionen kleiner DNA-Abschnitte auf einmal sequenziert werden und dann wie in einem Puzzle wieder mithilfe einer Blaupause (Referenz) durch bioinformatische Tools zusammengesetzt werden. Die jüngsten Fortschritte in dieser Technologie konzentrieren sich vornehmlich auf schnellere, kostensparende und genauere Sequenzierung, sowie eine effizientere und verbesserte Datenanalyse. Verglichen mit den Kosten für die Entschlüsselung des „ersten“ menschlichen Genoms im Jahre 2001 von ca. 3 Mrd. Dollar, könnte man heute 3 Millionen Genome generieren.

Im Gegensatz zu der heute am meisten angewandten Methode der Short-Read-Sequenzierung sind Sequenziertechnologien der „dritten Generation“ darauf ausgerichtet, sehr große, d. h. lange DNA-Fragmente (Long-Read-Sequenzierung) von mehr als 100.000 Basenpaaren, auf einmal auszulesen (siehe Abbildung 1).

Abbildung 1: Übersicht Sequenziertechnologien in der genomischen Diagnostik



Die neuen Sequenzieretechnologien ermöglichen es, i) komplexe Strukturvarianten des Genoms wie große Insertionen, Deletionen, Inversionen und Duplikationen zu identifizieren, ii) nachzuweisen und festzustellen, auf welchem elterlichen Allel eine Veränderung liegt, iii) Pseudogene¹¹ von aktiven Genen zu unterscheiden, iv) epigenetische Modifikationen in der DNA zusammen mit dem Genom auszulesen. Die große Bedeutung dieser neuen Technologien zeigt sich u. a. darin, dass sie die Grundlage boten für „ein“ völlig neues Gesamtverständnis unseres humanen Genoms, des Telomer-zu-Telomer-Genoms (T2T Genom)¹². Nur mithilfe der neuen Technologien ließen sich bislang „verborgene“ DNA-Abschnitte mit komplexen Strukturen entschlüsseln und zuordnen.

Zusätzlich zu Genomanalysen nutzt man moderne Sequenzierungstechnologien für die umfassende Analyse abgelesener RNAs (Transkriptom), für epigenomische Analysen und für die Sequenzanalyse des menschlichen Mikrobioms. In diesem Bereich der funktionellen Genomik haben Methoden zur Einzelzellsequenzierung sich als ein besonders erkenntnisgewinnendes Feld in den letzten Jahren hervorgetan (Walter/Schickl, 2019).

In Bezug auf seltene Erkrankungen bleibt abzuwarten, ob tatsächlich 80 % aller Fälle allein durch Genomsequenzierung aufgeklärt werden können. Vor allem für die funktionelle Bewertung und Interpretation von genetischen Veränderungen in regulatorischen Bereichen wird es essenziell sein, zusätzliche molekulare Profile wie das Epigenom¹³ und Transkriptom¹⁴ zu erstellen und diese vielschichtigen Sequenzierungsdaten zu integrieren.

1.3 Herausforderungen und Weiterentwicklungen hin zur genomischen Medizin

1.3.1 Notwendigkeit der Schulung von qualifiziertem Personal

Mit dem neuen Arbeitsfeld der genomischen Diagnostik bedarf es ebenfalls einer neu durchdachten Ausbildung und Zusammenarbeit von unterschiedlich qualifiziertem Personal im Bereich der klinischen Genetiker*innen, Dokumentar*innen und Genomanalytiker*innen. Da der gesamte Prozess mehrstufig ist, müssen alle an der Genomdiagnostik beteiligten Personen für ihre jeweilige Rolle (Präanalytik, Probenentnahme, Sequenzierung im Labor, Bioinformatik, Diagnostik, Validierung, medizinische Bewertung) qualifiziert sein und an der entsprechenden Schnittstelle effizient kommunizieren. In diesem Zusammenhang ist neben den gesetzlichen Vorgaben des Gendiagnostikgesetzes¹⁵ auch die S2k-Leitlinie der Deutschen Gesellschaft für Humangenetik (GfH)¹⁶ zu beachten. Eine erweiterte

¹¹ Ein Pseudogen ist ein DNA-Abschnitt, der ähnlich wie ein Gen aufgebaut ist, aber für kein funktionales Protein codiert. Man nimmt an, dass Pseudogene einst funktionierende Gene waren und durch genetische Veränderungen nicht mehr nutzbar gemacht wurden.

¹² Telomere sind Strukturelemente der DNA, die an den Enden der Chromosomen vorkommen und diese stabilisieren. Die DNA-Bereiche sind besonders schwer zu sequenzieren, da sie viele repetitive Bereiche enthalten, die damit nur schwer genau zugeordnet werden können.

¹³ Unter dem Epigenom versteht man die Gesamtheit aller chemischen Veränderungen der DNA sowie deren Verpackungsstruktur, dem sogenannten Chromatin. Epigenetische Veränderungen beeinflussen die Zugänglichkeit von Genen und regulieren damit unter anderem die Genaktivität (siehe auch Walter/Hümpel, 2017 und Walter/Gasparoni, 2021).

¹⁴ Das Transkriptom ist die Gesamtheit aller in einer Zelle hergestellten RNA-Moleküle (siehe auch Fehse et al., 2022).

¹⁵ Das Gendiagnostikgesetz regelt die Voraussetzungen für genetische Untersuchungen und für die im Rahmen dieser Untersuchungen durchgeführten genetischen Analysen sowie die Verwendung genetischer Proben und Daten in Deutschland (siehe unter: <https://de.wikipedia.org/wiki/Gendiagnostikgesetz> [14.02.2024]). Zum rechtlichen Rahmen siehe Müller-Terpitz/Straub, Kap. 5.

¹⁶ Siehe unter: <https://www.degruyter.com/document/doi/10.1007/s11825-018-0223-1/html> [09.02.2024].

Facharztausbildung wäre hier wünschenswert, da neben einer ausreichenden personellen Ausstattung für die Durchführung der Sequenzierungsprozesse, Dateninterpretation und Dokumentation die dringende Notwendigkeit der Einbindung entsprechend ausgebildeter Wissenschaftler*innen (z. B. Humangenetiker*innen) und Ärzt*innen besteht. Verantwortliche für die Datenauswertung und -dokumentation sollten neben der fachlichen Qualifikation über mehrjährige diagnostische Erfahrung in der molekulargenetischen Analyse von Sequenzveränderungen verfügen.

1.3.2 Klinische Krankenversorgung mit dem Ziel der Prävention

Das große Potenzial der Genomsequenzierung liegt nicht nur in der Diagnosestellung von genetisch bedingten Erkrankungen, bzw. zur Planung einer zielgerichteten individuellen Therapie, sondern auch in der Prävention von Erkrankungen. Prävention im Kontext klinischer Krankenversorgung umfasst das Ziel, Erkrankungen zu vermeiden, zu verzögern oder weniger wahrscheinlich zu machen. Dies gilt sowohl für seltene als auch häufige Erkrankungen („from rare to common diseases“) und bietet großes Potenzial für die genomische Diagnostik. Etwa 5 % unserer Bevölkerung tragen monogene pathogene/wahrscheinlich pathogene Varianten in Genen, die zu unmittelbaren Vorsorgeanpassungen führen, in den sogenannten „actionable genes“ (Rieß et al., 2024). Die Auswertung der „actionable genes“ wird in manchen deutschen humangenetischen Einrichtungen mit sehr gutem Erfolg bereits seit Jahren angewandt. Entsprechende Arbeitsgruppen wurden eingesetzt und Stellungnahmen erarbeitet.¹⁷ Die Anwendung der Diagnostik für häufige Erkrankungen, die i. d. R. durch zahlreiche „niedrigpathogene“ Varianten in Kombination (polygene Risikoscores, PRS) und Umwelteinwirkungen bedingt werden (multifaktorielle Erkrankungen), ist bisher bis auf Ausnahmen u. a. aufgrund fehlender Abrechenbarkeit der Leistungen noch nicht im Gesundheitswesen etabliert. Eigene diesbezügliche Genomanalysen an gesunden Personen für 10 häufige Erkrankungsgruppen zeigen, dass ca. 16 % der gesunden Personen ein hohes Risiko für diese Erkrankungen tragen (Rieß et al., 2024). Im Rahmen des deutschen Netzwerkes für familiären Brust- und Eierstockkrebs wird ein PRS (Mavaddat et al., 2019) in einigen Zentren bereits angewandt (Witt et al., eingereicht). Es ist anzunehmen, dass die Implementierung und Anwendung der Diagnostik von „actionable genes“ sowie von PRS im Rahmen des Modellvorhabens 64e/genomDE in den kommenden Jahren eine größere Verbreitung finden wird, da dadurch Krankheiten nicht nur besser diagnostiziert und behandelt werden, sondern auch individuell angepasste Präventionsmaßnahmen über die initiale diagnostische Fragestellung hinaus angegangen werden können. Es ist dabei selbstverständlich, und auch im Rahmen des deutschen Gendiagnostikgesetzes fixiert, dass der Patient/die Patientin vorher über die Ausweitung der Diagnostik beraten wird, sowie sein/ihr Einverständnis schriftlich belegt sein muss.

1.4 Fazit

Die Möglichkeiten der Humangenomforschung mit einer umfassenden Sequenzierung humaner Genome haben sich in den letzten Jahren rasant weiterentwickelt. Die Analyse des einzelnen menschlichen Genoms birgt daher auch für medizinische Versorgungskonzepte

¹⁷ Siehe unter: <https://doi.org/10.1515/medgen-2023-2026> [09.02.2024].

in der Krankenversorgung ein immer bedeutenderes Potenzial. Aktuell ist mit dem Modellvorhaben die klinische Nutzung der Genomsequenzierung auf seltene Erkrankungen und einige Krebserkrankungen (siehe hierzu Schlomm/Biernath, Kap. 2) beschränkt. Zukünftig muss die genomische Diagnostik jedoch für eine Vielzahl von Erkrankungen gedacht und genutzt werden, denn sie verspricht einen wesentlichen Beitrag zur Schaffung eines modernen Gesundheitssystems in Deutschland und weltweit.

1.5 Literaturverzeichnis

- Amberger, J. et al. (2009):** McKusick's Online Mendelian Inheritance in Man (OMIM). In: *Nucleic Acids Res* 37: D793–796. DOI: 10.1093/nar/gkn665.
- Amberger, J. S. et al. (2015):** OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. In: *Nucleic Acids Res* 43: D789–798. DOI: 10.1093/nar/gku1205.
- Fehse, B. et al. (Hrsg.) (2022):** Im Fokus: RNA. Eine Bestandsaufnahme der Arbeitsgruppe Gentechnologiebericht. BIH, Berlin. DOI: 10.17169/refubium-36831.
- Mavaddat, N. et al. (2019):** Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. In: *Am J Hum Genet.* 104(1): 21–34. DOI: 10.1016/j.ajhg.2018.11.002.
- Rieß, O. et al. (2024):** Genomes in clinical care. In: *Npj Genomic Medicine* 9(1): 20. DOI: 10.1038/s41525-024-00402-2.
- Walter, J./Hümpel, A. (Hrsg.) (2017):** Epigenetik. Implikationen für die Lebens- und Geisteswissenschaften. Nomos, Baden-Baden. Unter: <https://www.nomos-elibrary.de/10.5771/9783845270838/epigenetik> [19.02.2024].
- Walter, J./Gasparoni, N. (2021):** Themenbereich Epigenetik: von Zellidentitäten bis hin zu Krankheiten und Therapien. In: Fehse, B. et al. (Hrsg.): Fünfter Gentechnologiebericht. Sachstand und Perspektiven für Forschung und Anwendung. Nomos, Berlin: 93–113. Unter: <https://www.nomos-elibrary.de/10.5771/9783748927242-93/3-themenbereich-epigenetik-von-zellidentitaeten-bis-hin-zu-krankheiten-und-therapien?page=1> [19.02.2024].
- Walter, J./Schickl, H. (Hrsg.) (2019):** Einzelzellanalyse in Forschung und Medizin. Eine Stellungnahme der interdisziplinären Arbeitsgruppe Gentechnologiebericht. BBAW, Berlin. Unter: https://edoc.bbaw.de/files/3278/BBAW_Einzelzellanalyse_Walter_Schickl.pdf [19.02.2024].
- Witt, D. et al.:** Clinical genome sequencing in patients with hereditary breast and ovarian cancer: Concept, implementation and benefits. Eingereicht.
- Wright, C. F. et al. (2018):** Paediatric genomics: diagnosing rare disease in children. In: *Nat Rev Genet* 19(5): 253–268. DOI: 10.1038/nrg.2017.116.

2. Die Relevanz und der praktische Nutzen von Genomdaten im Bereich der Krebsdiagnostik

Thorsten Schlomm und Nadine Biernath

2.1 Einleitung

Genomdaten spielen eine immer wichtigere Rolle in der Krebsdiagnostik sowie Therapieplanung und eröffnen neue, personalisierte Behandlungsansätze. Durch Fortschritte im Next-Generation-Sequencing (NGS) können jetzt zahlreiche Gene oder das gesamte Genom eines Individuums effizient und genau analysiert werden, um ein breites Spektrum an krebsrelevanten genetischen Faktoren zu erfassen. Die Whole-Genome-Sequencing(WGS)-Technologie erlaubt zudem die Analyse struktureller DNA-Veränderungen¹ und epigenetischer Prozesse² des gesamten Genoms, was einen umfassenderen Einblick in Tumorgenome ermöglicht. Die Herausforderung liegt jedoch in der Interpretation dieser Daten, wobei die Unterscheidung zwischen klinisch relevanten und irrelevanten genetischen Veränderungen fortgeschrittene bioinformatische Werkzeuge und Expertise erfordert. Zudem sind die Verfügbarkeit und Nutzung genetischer Daten für Forschung und Real-World-Evidenz (RWE)³-Synthese ein wichtiges Thema. Initiativen wie GenomDE, das Nationale-Centrum-für-Tumorerkrankungen(NCT)-Netzwerk, das German Human Genome-Phenome Archive (GHGA), das Deutsche Netzwerk für Präzisionsmedizin (DNPM) und das Modelvorhaben Gensequenzierung (Details siehe Abschn. 2.4.3) haben eine standardisierte Datenverwendung zum Ziel, um das Potenzial der Genomik in der Krebsforschung und -behandlung zu maximieren.

Dieses Kapitel fokussiert auf die Rolle der Genomdaten in der Krebsdiagnostik, die Fortschritte und Herausforderungen der Gensequenzierungstechnologien sowie die Bedeutung der Datenzugänglichkeit für die Forschung und Entwicklung neuer Therapien, um die personalisierte Medizin und Krebsbehandlung voranzutreiben.

1 Strukturelle Genomveränderungen beziehen sich auf größere Veränderungen in der DNA-Struktur, wie Deletionen (Entfernung von DNA-Abschnitten), Amplifikationen (Vervielfältigung bestimmter DNA-Bereiche) oder Translokationen (Verlagerung von DNA-Abschnitten an andere Stellen). Diese Veränderungen können die Funktion von Genen beeinflussen.

2 Epigenetische Veränderungen sind Anpassungen in der Genaktivität, die nicht durch Änderungen der DNA-Sequenz selbst verursacht werden, sondern durch chemische Modifikationen an der DNA oder den Histonen (kleine basische Chromatin-Proteine im Zellkern), um die herum die DNA gewickelt ist. Im Zusammenhang mit Krebs können solche Veränderungen dazu führen, dass Gene, die normalerweise das Zellwachstum kontrollieren, abnorm aktiviert oder deaktiviert werden, was zu unkontrolliertem Zellwachstum und Tumorentwicklung führen kann. Für weitere Informationen zur Epigenetik siehe Walter/Hümpel, 2017 und Walter/Gasparoni, 2021.

3 Real-World-Evidenz (RWE) bezieht sich auf medizinische Informationen, die aus der tatsächlichen Anwendung von Behandlungen in der alltäglichen Praxis gewonnen werden, im Gegensatz zu Daten aus streng kontrollierten klinischen Studien. RWE umfasst Daten aus verschiedenen Quellen wie elektronischen Gesundheitsakten, Patient*innenregistern und Versicherungsdatenbanken. Sie bietet Einblicke in die Wirksamkeit, Sicherheit und Kosten von Therapien unter realen Bedingungen und hilft dabei, die Patient*innenversorgung und medizinische Entscheidungsfindung zu verbessern.

2.2 Kurze Historie der molekularen Therapie

Die Therapie von Krebserkrankungen hat seit der Zulassung der ersten Präzisionsonkologiemedikamente wesentliche Fortschritte gemacht. Rituximab, ein monoklonaler Antikörper gegen das CD20-Antigen von B-Zellen, wurde 1997 für die Behandlung von Non-Hodgkin-Lymphomen zugelassen und war anfänglich ausschließlich für diese Tumorentität vorgesehen (Grillo-López et al., 2000). Trastuzumab, ein weiteres zielgerichtetes Medikament gegen den HER2-Rezeptor, folgte 1998 mit einer Indikation spezifisch für HER2-positiven Brustkrebs (Cobleigh et al., 1999; Maadi et al., 2021).

In jüngerer Zeit hat sich das Konzept der zielgerichteten Therapien zu einem tumoragnostischen Ansatz⁴ weiterentwickelt, wie die Zulassungen der amerikanischen Arzneimittelbehörde FDA (Food and Drug Administration) für die Medikamente Pembrolizumab für Tumoren mit hoher Mutationslast⁵ oder Mikrosatelliteninstabilität⁶ und Larotrectinib für Krebsarten mit NTRK-Genfusionen zeigen. Beide Therapien werden unabhängig von der Tumorentität eingesetzt (Chakravarty et al., 2022; Farago/Demetri, 2020). Diese Medikamente sind Beispiele für eine Behandlung, die nicht mehr nur auf den Ursprungsort des Tumors abzielt, sondern auf spezifische genetische Aberrationen (Veränderungen).

Die American Society of Clinical Oncology (ASCO) erkennt die Bedeutung dieser Entwicklung an und empfiehlt seit 2022 eine breite genetische Testung mittels Multipanel-NGS (gleichzeitige Analyse mehrerer spezifischer Gene) für alle Patient*innen mit fortgeschrittenen Tumorerkrankungen, wenn mehrere präzisionsonkologische Behandlungsoptionen zur Verfügung stehen (Chakravarty et al., 2022).

2.3 Moderne Ansätze in der onkologischen Genomik: von Standardsequenzierung bis zum Whole-Genome-Sequencing

In der klinischen Onkologie unterscheiden wir zwei Hauptmethoden, um das Erbgut von Tumoren zu untersuchen: einfache und umfassende Ansätze. Die einfacheren Methoden, wie das Single-Gene-Sequencing oder Panel-Gene-Sequencing, konzentrieren sich auf spezifische, bereits bekannte Genveränderungen, die für die Behandlung von Krebs wichtig sind. Diese Methoden sind kostengünstiger und einfacher in der Auswertung.

Die umfassenderen Methoden, wie das Whole-Genome-Sequencing (WGS), untersuchen das gesamte Erbgut des Tumors. Dies gibt ein vollständigeres Bild und kann neue Erkenntnisse über den Krebs liefern. Eine weitere Methode ist das Whole-Exome-Sequencing (WES), das alle proteincodierenden Bereiche des Genoms untersucht. WES ist hilfreich, um genetische Veränderungen zu identifizieren, die Proteine beeinflussen und somit direkten Einfluss auf die Krebsentwicklung haben können.

4 Tumoragnostische Therapien sind nicht auf die spezifischen Gewebeeigenschaften, das Ursprungsorgan oder die spezifische Krebsart abgestimmt. Die Behandlung basiert stattdessen auf einem Nachweis bestimmter molekulargenetischer Veränderungen im Tumor.

5 Eine hohe Tumormutationslast ist definiert als ≥ 10 Mutationen/Megabase.

6 Eine Mikrosatelliteninstabilität liegt häufig bei Mutationen in einem bestimmten DNA-Reparaturmechanismus – den „Mismatch-Repair“-Genen – vor.

WGS und WES sind zwar teurer und komplizierter, aber sie werden zunehmend wichtiger, da sie deutlich umfassendere Daten liefern. Dies könnte in der Zukunft entscheidend sein, um bessere Behandlungen zu entwickeln. Zudem gibt es kombinierte Ansätze, die sowohl breite Veränderungen im Erbgut als auch spezielle, für die Behandlung wichtige Veränderungen erfassen, was ein detailliertes Bild des Tumors bietet und gleichzeitig praktischer und kostengünstiger ist als das komplette Sequenzieren des ganzen Genoms.

Die Herausforderungen bei der Interpretation und Nutzbarmachung von breiten genomischen Daten wie WES- oder WGS-Daten in der klinischen Praxis sind jedoch vielschichtig. Die Komplexität und der Umfang der Daten erfordern den Einsatz fortschrittlicher bioinformatischer Werkzeuge und Algorithmen. Die Unterscheidung zwischen pathogenen und benignen (gutartigen) Varianten, die Integration und Korrelation von Daten auf unterschiedlichen Ebenen, das Fehlen umfassender Referenzdaten und Wissenslücken, die Notwendigkeit multidisziplinärer Expertise und der erhebliche Zeit- und Ressourcenaufwand sind nur einige der zu meisternenden Hürden. Diese Faktoren verdeutlichen die Notwendigkeit kontinuierlicher Fortschritte in der Bioinformatik, einer verbesserten Datenintegration, Daten-Nutzbarmachung und eines tieferen Verständnisses der Krebsgenomik. Nur so kann das volle Potenzial der WES und WGS in der personalisierten Krebsmedizin realisiert und die Behandlung von Krebserkrankungen zukünftig weiter verbessert werden. Erste Programme zum breiteren Einsatz der WES und WGS in der Onkologie werden derzeit in großen nationalen Netzwerken initiiert (siehe Abschn. 2.4.3).

2.4 Von der Analyse zur Anwendung: Fortschritte in der Genomik und deren Nutzbarmachung in der Onkologie

2.4.1 Die transformative Rolle molekularer Tumorboards in der genetischen Datenanalyse und Behandlungsplanung

Molekulare Tumorboards spielen eine entscheidende Rolle in der modernen Krebsbehandlung, indem sie genetische Informationen von Tumoren nutzen, um personalisierte Behandlungspläne zu entwickeln. Ein wesentlicher Aspekt dabei ist die Bewertung der biologischen Wertigkeit von molekularen Veränderungen. Dies geschieht durch umfassende Datenbankrecherchen, in denen die gefundenen Aberrationen mit bestehenden Variantendatenbanken abgeglichen werden. Diese Bewertung ist besonders wichtig bei breiten Sequenzierungsansätzen wie der Multipanel-NGS-Diagnostik, dem WES und dem WGS.

Zusätzlich wird die genetische Information in den Kontext der gesamten Tumorphistorie des/der Patient*in gesetzt. Dies umfasst die Betrachtung früherer Behandlungen, das Ansprechen auf diese Therapien und die Entwicklung des Tumors über die Zeit. Diese ganzheitliche Sichtweise ermöglicht es den Fachleuten im molekularen Tumorboard, einen umfassenden Behandlungsplan zu entwickeln, der nicht nur auf den aktuellen genetischen Status des Tumors, sondern auch auf die gesamte Behandlungsgeschichte des/der Patient*in abgestimmt ist.

Die Herausforderungen in diesem Prozess umfassen nicht nur die technischen Aspekte der genetischen Analyse, sondern auch die Interpretation und klinische Anwendung dieser

komplexen Daten. Es erfordert ein hohes Maß an Fachwissen und Erfahrung, um die genetischen Daten korrekt zu interpretieren und sie in effektive Behandlungsstrategien umzusetzen. Darüber hinaus müssen Datenschutz und ethische Überlegungen berücksichtigt werden,⁷ insbesondere wenn es um den Umgang mit potenziell vererbbaaren genetischen Keimbahninformationen geht.

Trotz dieser Herausforderungen bieten molekulare Tumorboards einen unschätzbaren Nutzen für die personalisierte Krebsbehandlung, indem sie die genetischen Details eines Tumors entschlüsseln und so eine maßgeschneiderte Therapie ermöglichen, die auf die spezifischen Bedürfnisse jedes/jeder einzelnen Patient*in zugeschnitten ist.

2.4.2 Automatisierte molekulare Auswertungsalgorithmen (Clinical Decision Support System; CDSS) in der personalisierten Onkologie

CDSS sind fortschrittliche Softwarelösungen, die Tumorsequenzierungsbefunde automatisiert analysieren und standardisierte Berichte für die klinische Entscheidungsfindung erstellen. Sie verarbeiten effizient komplexe genetische Daten aus Sequenzierungstechniken wie Multipanel-NGS, WES oder WGS. Mit Algorithmen identifizieren und klassifizieren diese Lösungen genetische Mutationen und Aberrationen und übersetzen die Daten in klare, interpretierbare Berichte, die klinisch relevante Informationen und potenzielle Therapien und Studienmöglichkeiten hervorheben.

Diese Systeme sparen Zeit und erhöhen die Präzision in der Dateninterpretation, indem sie menschliche Fehler reduzieren. Herausforderungen bestehen jedoch in der Abhängigkeit von der Datenqualität, der Integration in klinische Prozesse und dem Umgang mit sensiblen Patient*innendaten. Insgesamt tragen CDSS zur Weiterentwicklung der personalisierten Medizin und zur Optimierung der Patient*innenversorgung bei und bieten Potenzial für innovative Entwicklungen wie die Integration von Schwarmintelligenz und Echtzeitdaten (siehe Abschn. 2.5).

2.4.3 Synergie in der Genomik: Vernetzung deutscher Initiativen zur Förderung der personalisierten Onkologie

Für eine flächendeckende Implementierung breiter genomischer Ansätze, wie z. B. dem WGS in der deutschen Onkologie, einschließlich einer effektiven und breiten Nutzbarmachung der Daten, ist ein Zusammenschluss bereits bestehender Initiativen und Netzwerke essenziell. Ein solch integrativer Ansatz würde die Stärken und Ressourcen verschiedener Akteur*innen bündeln, um eine umfassende und kohärente Strategie in der genombasierenden Onkologie zu entwickeln und umzusetzen. Durch die Vernetzung bestehender Initiati-

⁷ Zu ethischen bzw. rechtlichen Fragen siehe Jungkunz et al., Kap. 4, und Müller-Terpitz/Straub, Kap. 5.

ven wie der GenomDE-Initiative⁸, des GHGA⁹, des DNPM¹⁰, des NCT¹¹-Netzwerkes und die Integration des Modellvorhabens § 64e SGB V¹² zur umfassenden Diagnostik und Therapiefindung mittels Genomsequenzierung kann eine umfassende Infrastruktur für die Genommedizin in Deutschland geschaffen werden.

Diese Kooperation würde die Standardisierung der genetischen Sequenzierung und den Datenaustausch erleichtern, die Forschung in der personalisierten Onkologie fördern und eine einheitliche Datenquelle für Forschung und klinische Anwendung, inkl. RWE-Generierung schaffen. Sie würde auch die Ausbildung des Fachpersonals in Genomik und Onkologie verbessern, was für die effektive Integration genombasierter Ansätze in die klinische Praxis entscheidend ist. Zudem würden gemeinsame Richtlinien und Standards für Datenschutz und Datensicherheit etabliert, was die Genommedizin sicherer machen und das öffentliche Vertrauen stärken würde. Insgesamt würde ein solcher Zusammenschluss die genombasierte Onkologie in Deutschland effizienter und wirksamer machen, die Patient*innenversorgung optimieren und die Forschung in diesem Bereich voranbringen.

2.5 Ausblick: Mit „Schwarmintelligenz“ die kollektive Erfahrung der medizinischen Gemeinschaft nutzen

Die Integration von Echtzeitdaten und Schwarmintelligenz in die Nutzung von Genomdaten könnte die Krebsdiagnostik wesentlich voranbringen. Durch das Zusammenführen von Patient*innenhistorien und genetischen Befunden könnten CDSS, die bereits in der personalisierten Onkologie eingesetzt werden, ähnlich wie Google Maps für Verkehrsinformationen, personalisierte Behandlungsempfehlungen auf Basis der Daten vieler Patient*innen geben. Ein Projekt an der Charité – Universitätsmedizin Berlin in Kooperation mit dem Deutschen Netzwerk für angewandte Präzisionsmedizin (DNA-Med) arbeitet daran, heterogene Patient*innendaten in eine einheitliche, strukturierte Form zu bringen. Dies ist entscheidend, um Muster zu erkennen und individuelle Behandlungspläne zu erstellen. Die Herausforderung liegt in der Strukturierung und Standardisierung der Patient*innendaten sowie in der Interoperabilität zwischen verschiedenen Datenbanken.¹³

Dieses Vorhaben könnte eine Wende in der medizinischen Versorgung herbeiführen, bei der Entscheidungen auf der Grundlage eines breiten Spektrums von Daten getroffen werden,

8 Diese Initiative zielt darauf ab, die genetische Diagnostik und personalisierte Medizin in Deutschland zu stärken. Sie soll den Zugang zu genetischer Beratung und Testung verbessern und die Integration genetischer Informationen in die klinische Praxis fördern. In Kombination mit dem Modellvorhaben § 64e würde dies eine verbesserte Infrastruktur und Ressourcen für genetische Untersuchungen und personalisierte Behandlungen bedeuten.

9 Das GHGA dient der Speicherung und dem Austausch von genomischen und phänotypischen Daten, um die Forschung und medizinische Innovation in Deutschland zu fördern. Die Integration des GHGA in das Modellvorhaben und die GenomDE-Initiative würde den Austausch von Daten und Erkenntnissen erleichtern, was die Forschung und Entwicklung in der personalisierten Medizin beschleunigen könnte.

10 Schrittweise Verbesserung, qualitätsgesicherte Implementierung und fortlaufende Datenerhebung im Sinne eines selbstlernenden Systems sind die Ziele dieses durch das Comprehensive Cancer Center getragenen Verbundes, der derzeit durch den Innovationsfond der Krankenversicherungen unterstützt wird. Zusätzlich wurde durch die Deutsche Krebsgesellschaft ein System der Zertifizierung im Rahmen des Zertifizierungssystems der Organkrebszentren und onkologischen Zentren entwickelt.

11 Die Nationalen Centren für Tumorerkrankungen sind auf die Krebsbehandlung und -forschung in frühen klinischen Studien spezialisiert. Ihre Einbindung in das Modellvorhaben § 64e und die Verknüpfung mit der GenomDE-Initiative und dem GHGA könnte die Krebsforschung und -behandlung beschleunigen. Durch den Austausch von genetischen Daten und Erkenntnissen könnten effektivere und personalisierte Behandlungsstrategien für Krebspatient*innen entwickelt werden.

12 Dieses Vorhaben ermöglicht es, innovative Versorgungsformen in der gesetzlichen Krankenversicherung zu erproben. Der Fokus liegt auf der Verbesserung der Patient*innenversorgung durch neue Ansätze wie WGS, die über die herkömmliche medizinische Versorgung hinausgehen. Im Kontext der Genomik und personalisierten Medizin würde dies die Einbindung fortschrittlicher genetischer Diagnostik und individualisierter Therapieansätze beinhalten.

13 Für weitere Informationen zu Genomdatenbanken siehe Walter et al., Kap. 3.

die die kollektive Erfahrung der medizinischen Gemeinschaft widerspiegeln. Die Herausforderungen, die sich dabei stellen – von technischen Aspekten über KI-Entwicklung bis hin zu Datenschutz und ethischen Fragen – sind bedeutend, aber das Ziel ist eine verbesserte, datengestützte und auf den Einzelnen zugeschnittene Patient*innenversorgung in der Onkologie.

2.6 Fazit

Die rasante technologische Entwicklung in der Genomsequenzierung ermöglicht eine immer umfassendere Anwendung in der Krebsdiagnostik. Es bestehen dabei aber noch große Herausforderungen, die Daten effektiv zu interpretieren und in den gesamtmedizinischen Kontext der/des Patient*in zu stellen. Zudem müssen die Daten für eine breitere Forschung geschützt zugänglich gemacht werden, um mithilfe KI gesteuerter Verfahren neue und erweiterte Erkenntnisse zu erhalten. Die Bewältigung dieser Herausforderungen wird entscheidend sein, um die personalisierte Medizin weiter voranzutreiben und die Behandlung von Krebserkrankungen zu revolutionieren.

2.7 Literaturverzeichnis

- Chakravarty, D. et al. (2022):** Somatic genomic testing in patients with metastatic or advanced cancer: ASCO provisional clinical opinion. In: *J Clin Oncol* 40(11): 1231–1258. DOI: 10.1200/JCO.21.02767.
- Cobleigh, M. A. et al. (1999):** Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease. In: *Journal of Clinical Oncology* 17(9): 2639–2648. DOI: 10.1200/JCO.1999.17.9.2639.
- Farago, A. F./Demetri, G. D. (2020):** Larotrectinib, a selective tropomyosin receptor kinase inhibitor for adult and pediatric tropomyosin receptor kinase fusion cancers. In: *Future Oncol* 16(9): 417–425. DOI: 10.2217/fon-2019-0647.
- Grillo-López, A. J. et al. (2000):** Rituximab: the first monoclonal antibody approved for the treatment of lymphoma. In: *Curr Pharm Biotechnol* 1(1): 1–9. DOI: 10.2174/1389201003379059.
- Maadi, H. et al. (2021):** Trastuzumab mechanism of action; 20 years of research to unravel a dilemma. In: *Cancers (Basel)* 13(14): 3540. DOI: 10.3390/cancers13143540.
- Walter, J./Gasparoni, N. (2021):** Themenbereich Epigenetik: von Zellidentitäten bis hin zu Krankheiten und Therapien. In: Fehse, B. et al. (Hrsg.): *Fünfter Gentechnologiebericht. Sachstand und Perspektiven für Forschung und Anwendung*. Nomos, Berlin: 93–113. Unter: <https://www.nomos-elibrary.de/10.5771/9783748927242-93/3-themenbereich-epigenetik-von-zellidentitaeten-bis-hin-zu-krankheiten-und-therapien?page=1> [19.02.2024].
- Walter, J./Hümpel, A. (Hrsg.) (2017):** Epigenetik. Implikationen für die Lebens- und Geisteswissenschaften. Nomos, Baden-Baden. Unter <https://www.nomos-elibrary.de/10.5771/9783845270838/epigenetik> [19.02.2024].

3. Die sichere Archivierung komplexer Genomdaten und ihre effiziente Nutzung – eine Herausforderung

Jörn Walter, Nina Gasparoni und Oliver Kohlbacher

3.1 Einführung

Menschliche Genomdaten und verwandte Omics-Daten¹ spielen in der humanmedizinischen Forschung eine immer bedeutendere Rolle. Aus zahlreichen internationalen Genomsequenzierungsprojekten heraus entsteht eine schnell anwachsende Zahl von Genomdaten, die bedeutende neue Erkenntnisse über die genetischen Grundlagen von Gesundheit und Erkrankungen liefern. Deutschland beschreitet einen eigenen Weg und generiert erstmalig Genomdaten im Rahmen medizinisch abrechenbarer klinischer Leistungen im Rahmen des deutschen Modellvorhabens „Genomsequenzierung“ (§ 64e SGB V, siehe Abschn. 3.2.2). Die primär zu diagnostischen Zwecken sequenzierten Genome sollen aber sekundär auch für die Forschung nutzbar sein.

Eine kombinierte klinische und forschungsorientierte Nutzung der Genomdaten stellt uns vor mehrere Herausforderungen: Genomdaten müssen einerseits sicher und geschützt archiviert werden, andererseits aber auch einer breiten forschungsorientierten Nutzung zur Verfügung stehen. Komplexe Genomdaten können mithilfe computergestützter Verfahren nur dann optimal und nachvollziehbar interpretiert werden, wenn sie mit möglichst umfangreichen anderen Datensätzen verglichen werden können. Mit einfließen muss dabei auch die Berücksichtigung personenbezogener Lebens- und Gesundheitsdaten. Eine sichere und gleichzeitig flexibel such- und nutzbare Zusammenführung von Genomdaten und personenbezogenen (medizinischen) Daten in einem Datenarchiv ist daher eine essenzielle Voraussetzung. Nur diese Verknüpfung ermöglicht eine optimale Interpretation der Genominformationen für die Diagnose und Behandlung von Erkrankungen.

3.2 Genomdaten in Forschung und Medizin

3.2.1 Die wachsende Komplexität von Genomdaten

Die Genomforschung erlebt seit einigen Jahren einen tiefgreifenden Wandel. Angetrieben von atemberaubend schnellen Technologieverbesserungen eröffnen sich vollkommen neue Dimensionen (Menge und Tiefe) der Genomanalyse, die Genomdaten tief ins Zentrum

¹ Omics-Daten sind umfassende Datensätze, die auf der Analyse des gesamten zellulären Inhalts der untersuchten Moleküle beruhen, z. B. die Gesamtheit der genetischen Information (Genom), der Gentranskripte (Transkriptom), epigenetischer Modifikationen (Epigenom), die Gesamtheit der Proteine und ihrer Modifikationen (Proteom) oder aller Stoffwechselprodukte (Metabolom).

des medizinischen Interesses rücken. Die Entschlüsselung des menschlichen Genoms liefert (immer noch) bahnbrechende Beiträge für die Biomedizin. Der Informationsgewinn individueller Genomdaten für die personenbezogene Diagnose und die Erforschung der Ursachen von Krankheiten wie auch die Entwicklung und Anwendung von Wirkstoffen ist unschätzbar groß und wird in nahezu jedem Bereich der medizinischen Versorgung und Forschung Nutzen und Anwendungen finden.

Im Gegensatz zu dem immensen Aufwand, den die Entschlüsselung des ersten menschlichen Referenzgenoms² (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001) benötigte, können heutzutage Genome mithilfe verschiedenster NGS-Technologien³ für einen Bruchteil der Kosten⁴ und in kurzer Zeit sequenziert werden.

Die Herausforderungen liegen daher nicht mehr vornehmlich in der Herstellung (Sequenzierung), sondern in der geordneten, verknüpften Ablage sowie einer standardisierten Auswertung und Interpretation von Genomdaten. Diese Prozesse erfordern um ein Vielfaches mehr Zeit sowie intensiver computergestützter (Bioinformatik, KI) und personengestützter Arbeit (Bewertung) als die reine Herstellung der Daten. Schätzungen gehen davon aus, dass allein in den kommenden zwei Jahren 100 Millionen Genomsequenzen weltweit erstellt werden.⁵ Für die Verarbeitung dieser großen Datenmengen müssen entsprechende Dateninfrastrukturen aufgebaut und dauerhaft vorgehalten werden.

Weltweite Genomforschungsprogramme der vergangenen Jahre haben verdeutlicht, dass das individuelle menschliche Genom viel komplexer ist als ursprünglich angenommen. So hat das Pangenom-Projekt⁶ (Liao et al., 2023) eindrucksvoll gezeigt, dass der Bezug zu einem einzigen Referenzgenom allein zu kurz greift und neue Wege beschritten werden müssen, Genomveränderungen im Populationskontext zu analysieren und zu bewerten. Diese populationsbezogene Einordnung individueller Genome ist bedeutsam, da sie hilft, individuelle krankheitsrelevante Genomveränderungen von bereits vorhandenen Veränderungen in der Bevölkerungsgruppe abzugrenzen und besser zu bewerten. Neueste Sequenzieretechnologien (siehe hierzu Rieß/Schulze-Hentrich, Kap. 1), die die Sequenzierung längerer zusammenhängender Sequenzabschnitte ermöglichen, schaffen hierfür zudem eine immer genauere Datenausgangslage.

Neben den reinen Genomdaten (DNA-Sequenz) gewinnen ergänzende, flankierende Omics-Daten zunehmend an Bedeutung. Dazu gehören sowohl die dem Genom „aufgedruckten“

2 Im Rahmen des internationalen Forschungsprojekts Humangenomprojekt und einer privatwirtschaftlichen Initiative der Firma Celera arbeiteten von 1990 bis 2003 Tausende von Wissenschaftler*innen weltweit an der Entschlüsselung des menschlichen Genoms. Durch Sequenzierung wurde dabei die nahezu vollständige DNA-Sequenz des Menschen identifiziert. Zur Veröffentlichung der vollständigen Version in 2003 (erste lückenhafte Version in 2001) siehe unter: <https://www.genome.gov/25520492/online-education-kit-2003-human-genome-project-completed> [23.10.2023].

3 NGS ist die Abkürzung für Next-Generation-Sequencing, Hochdurchsatz-Sequenzierung (siehe hierzu Schlomm/Biernath, Kap. 2).

4 Siehe unter: <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost> [23.10.2023].

5 Siehe unter: <https://www.labiotech.eu/in-depth/cloud-genomics-big-data-problem> [08.11.2023].

6 Im Rahmen des internationalen Pangenom-Projekts arbeiten Wissenschaftler*innen an der Sammlung von 350 Referenzgenomen, wobei die Diversität der Bevölkerung abgebildet werden soll. Siehe unter: <https://humanpangenome.org/index.html> [23.10.2023].

epigenetischen Modifikationen (das Epigenom)⁷ als auch Informationen zur Genomnutzung in Form von RNA (das Transkriptom)⁸. Beide Omics-Datenarten werden ebenfalls durch Sequenzierung abgelesen und im Kontext des Genoms interpretiert. Diese ergänzenden sogenannten funktionellen Genomdaten werden wesentlich dazu beitragen, die Steuerung des Genoms im Zellkontext personen- und krankheitsspezifisch zu analysieren und zu interpretieren (Walter/Gasparoni, 2021; Fehse et al., 2022). Besonders hervorzuheben sind hier Sequenzierdaten von Einzelzellen (Transkriptom und Chromatinzugänglichkeit)⁹, die mit einer bisher noch nie dagewesenen Granularität zur Krankheitsursachenforschung beitragen (Walter/Schickl, 2019). Genomdaten können zudem mit weiteren Omics-Daten für Biomoleküle wie Proteine (Proteom) oder Stoffwechselprodukte (Metabolom) sowie mit phänotypischen und (para-)klinischen Daten, wie z. B. Gewicht, klinischen Diagnosen, Laborergebnissen, bildgebenden Verfahren und pathologischen Befunden aus Gewebeproben, verknüpft werden. Moderne Gesamtgenomdaten werden entsprechend zunehmend „multimodal“ und dies bedingt, dass die verschiedenen Datensätze zell- und fallbezogen miteinander verknüpfbar sein müssen. Länder wie Estland, Dänemark und Finnland sind in ihren Genomprogrammen schon sehr weit fortgeschritten, verschiedene Datenformate und Datenbanken miteinander zu verknüpfen, um komplexe diagnostische Fragen adressieren zu können (Smetana/Brož, 2022).

Insgesamt ist festzustellen, dass die Interpretation von Genomdaten in Verbindung mit erweiterten Referenzgenomen (pangenomische Ansätze liefern hier eine bedeutende Erweiterung), funktionellen Genomdaten sowie weiteren bildgebenden und diagnostischen Daten immer tiefere, aber auch neue Einblicke in die Bedeutung und Folgen individueller genomischer Veränderungen ermöglichen (werden). Die genomgestützte Diagnose und die Vorhersage von Krankheitsursachen und -verläufen wird sich so noch deutlich verbessern. In der Forschung werden kontinuierlich neue komplexe integrative Analyseansätze erarbeitet, die in absehbarer Zeit auch in der Diagnostik zum Einsatz kommen und der Genommedizin ein differenziertes fallbezogenes Anwendungsspektrum eröffnen werden.

Um dies in Zukunft zu ermöglichen, sind zwei wesentliche Schritte erforderlich. Erstens ist es wichtig, die Forschung in diesem Bereich zu stärken und zweitens sind Genomdatenbanken perspektivisch so anzulegen, dass sie die integrierte Nutzung (sprich Zusammenführung) erweiterter funktioneller Daten ermöglichen, da nicht allein die Anzahl genomischer Daten, sondern auch die zu erwartende Zunahme der Informationsdichte eine Herausforderung für Datenbanken und Analysemethoden darstellt. Datenarchive müssen daher leistungsstark ausgestattet und mit sicheren Rechenressourcen verbunden werden, um eine Verknüpfung, Auffindbarkeit und Analyse der Daten zu ermöglichen (siehe auch Abschn. 3.4.1). Hierzu bedarf es flankierender und standardisierter bioinformatischer

7 Epigenetische Modifikationen sind molekulare Veränderungen an der DNA und den sie umgebenden Proteinen, die die Aktivität von Genen beeinflussen, ohne dabei die eigentliche DNA-Sequenz zu verändern. Ein Beispiel ist die DNA-Methylierung, bei der Methylgruppen an bestimmte Bausteine der DNA in zellspezifischen Mustern gebunden sind und die Bindung von weiteren Faktoren zum Ablesen der Gene regulieren.

8 Die DNA wird in einem Prozess namens Transkription in RNA umgeschrieben. Die Gesamtheit aller RNA-Moleküle, die zu einem bestimmten Zeitpunkt in einer Zelle vorhanden sind (Transkriptom) liefert Informationen zur Genaktivität.

9 Die DNA liegt im Zellkern kompakt mit Proteinen in einer Struktur namens Chromatin verpackt vor. Der Verpackungsgrad variiert an verschiedenen Stellen und beeinflusst die Zugänglichkeit der DNA für andere Faktoren, die beispielsweise Gene ablesen können. Die Untersuchung der Chromatinzugänglichkeit liefert wertvolle Einblicke in die Regulation der Genaktivität.

Arbeitsabläufe und Auswertungsverfahren (siehe auch Abschn. 3.4.2), aber auch erheblicher Maßnahmen bei Datenschutz und Informationssicherheit, um die besonders schützenswerten Gesundheitsdaten vor Missbrauch zu schützen.¹⁰

3.2.2 Nutzung klinischer Genomdaten: Verknüpfung von Patient*innenversorgung mit Forschung

Das Erstellen von Gesamtgenomdaten zur Diagnose oder personalisierten Therapie war in Deutschland bislang nicht Teil der medizinischen Regelversorgung. Teilgenomdaten wurden vor allem im Kontext humangenetischer Fragestellungen und Indikationen (Vererbung von Anlagen) oder der personalisierten Onkologie im Rahmen krankheitsfokussierter Studien und Netzwerke personalisierter Medizin erhoben, wie z. B. nNGM¹¹, DNPM¹² und MASTER¹³. In der Humangenetik waren Analysen zum Auffinden von Mutationen in Kandidatengenen, Gen-Panels oder Exom-Analysen¹⁴ abrechenbar. Die Komplexität von Gesamtgenomdaten stellt jedoch ganz neue und weitreichende Herausforderungen an die Datengenerierung, -analyse und -interpretation.

Das innovative Modellvorhaben des Bundesministeriums für Gesundheit (BMG) nach § 64e SGB V¹⁵ eröffnet einzigartige Möglichkeiten, moderne Strukturen für eine breite Anwendung der Gesamtgenomsequenzierung in der klinischen Gesundheitsversorgung aufzubauen. Die Maßnahme konzentriert sich zunächst auf die Genomsequenzierung bestimmter Erkrankungen und wird nach einem Zeitraum von vier Jahren evaluiert. Für den Erfolg der Maßnahme wird es wichtig sein, eine bundesweit koordinierte Zusammenarbeit der über Deutschland verteilten Genomzentren aufzuzeigen und diese Zentren durch eine gemeinsame Dateninfrastruktur eng zu verbinden. Die Ausbildung und Schulung von Expert*innen für Datenanalyse im humangenetischen Bereich der Genomzentren, gemeinsam mit den behandelnden Ärzt*innen und/oder Kliniker*innen, wird dabei von großer Bedeutung sein, um eine optimale Nutzung von Genomdaten zu erreichen.

Die durch das Modellprojekt generierten und fachlich ausgewerteten Omics-Daten sind einerseits wertvolle Quellen für die individuelle Gesundheitsversorgung, sie eröffnen aber auch neue Möglichkeiten zur Erforschung der genetischen Ursachen von Krankheiten auf molekularer Ebene. Das Ziel der Maßnahme ist, wie es die nationale Genominitiative genomDE¹⁶ formuliert, „die Nutzung der Genomsequenzierung zum Wohle der Patientinnen und Patienten als ei[n] innovative[r] Bestandteil der Regelversorgung in Deutschland [...]. Gleichzeitig dienen diese Vorhaben der Stärkung der nationalen und internationalen genommedizinischen Forschung im Sinne einer wissensgenerierenden Versorgung“.¹⁷

¹⁰ Zum Datenschutz aus ethischer bzw. rechtlicher Perspektive siehe Jungkunz et al., Kap. 4, und Müller-Terpitz/Straub, Kap. 5.

¹¹ Nationales Netzwerk Genomische Medizin Lungenkrebs; siehe unter: <https://nngm.de/> [15.12.2023].

¹² Deutsches Netzwerk für Personalisierte Medizin; siehe unter: <https://dnpm.de/> [15.12.2023].

¹³ Molecularly Aided Stratification for Tumor Eradication; siehe unter: <https://www.nct-heidelberg.de/forschung/molecular-stratification/master.html> [15.12.2023].

¹⁴ Dies bezieht sich auf die Untersuchung der proteincodierenden Abschnitte der Gene (Exons). Viele genregulatorische Bereiche liegen in den nicht codierenden Bereichen des Genoms, die dabei nicht erfasst werden.

¹⁵ Siehe unter: <https://www.sozialgesetzbuch-sgb.de/sgbv/64e.html> [23.10.2023]. Zum rechtlichen Rahmen des Modellvorhabens siehe Müller-Terpitz/Straub, Kap. 5.

¹⁶ genomDE ist eine vom Bundesministerium für Gesundheit geförderte Initiative. Sie dient der Entwicklung von inhaltlichen und strukturellen Konzepten zur Vorbereitung der Umsetzung des Pilotprojekts Genomsequenzierung nach § 64e SGB V. Siehe unter: <https://www.genom.de/de> [16.11.2023].

¹⁷ Siehe unter: <https://www.genom.de/de/ss-64e-sgb-v-modellvorhaben-regelt-die-rechtlichen-grundlagen> [16.11.2023].

Um dieses Ziel zu erreichen, bedarf es einer standardisierten Zusammenführung der individuellen klinisch erhobenen Genomdaten aus dem Modellprojekt mit Daten anderer Zentren und bereits existierenden Daten aus nationalen und internationalen Genom- und Forschungsprojekten. Nur so können die benötigten Vergleichskohorten und Referenzen entstehen, die eine Klassifizierung ursächlicher Veränderungen/Varianten im Krebs oder bei Patient*innen mit seltenen Erkrankungen verbessern und diese von nicht kausalen genomischen Veränderungen abgrenzen.¹⁸ Die klinisch erstellten Genomdaten sollen – bei Vorliegen einer informierten Einwilligung der Probenspender*innen bzw. Patient*innen¹⁹ – in zentralen Datenbanken abgelegt werden, und dadurch für die Erforschung von Krankheitsursachen wie auch für die Entwicklung präventiver Maßnahmen nutzbar sein. Hierzu müssen auch die klinisch erhobenen Genomdaten von Beginn an standardisiert prozessiert, zentral archiviert und mit den zugehörigen klinischen Daten zusammengeführt werden. Die Archivierung ist so zu gestalten, dass der Schutz sensibler Daten gewährleistet ist und gleichzeitig pseudonymisierte, datensatzrelevante klinische Parameter individueller Proben für die Forschung nutzbar sind.

3.3 Herausforderungen einer zentralisierten Datennutzung: Datenschutz und informierte Einwilligung

Die einheitliche Ablage und Verfügbarkeit von Omics-Daten erfordert umfangreiche rechtliche, technische und organisatorische Maßnahmen. Daten müssen rechtssicher und vor Missbrauch geschützt gespeichert werden und gleichzeitig effizient suchbar und nachhaltig nutzbar sein. Hierzu müssen Genomdaten mit zusätzlichen (Meta-)Daten des/der Spender*in (Patient*innen oder Forschungsteilnehmende) verknüpft werden, wie beispielsweise Alter, biologisches Geschlecht und Informationen zu Krankheiten. Aufgrund der Sensitivität dieser personenbezogenen Daten – Gesundheitsdaten und Genomdaten zählen zu den ‚besonderen Kategorien‘ personenbezogener Daten nach Art. 9 Abs. 1 der Datenschutz-Grundverordnung (DSGVO) – sind diese Daten besonders schützenswert. In der Regel werden zu Forschungszwecken erhobene Genomdatensätze pseudonymisiert²⁰, sind also nicht auf einfache Art einer bestimmten Person zuzuordnen. Einzelne Genomabschnitte können u. U. dennoch zu einer Re-Identifizierung führen, insbesondere durch Verknüpfung mit anderen Datenquellen, wie z. B. öffentlich zugänglichen Ahnenforschungsdatenbanken (Gymrek et al., 2013). Daher werden pseudonymisierte Daten auch Forschenden nur unter strengen Zugangsbeschränkungen und Sicherheitsvorkehrungen bereitgestellt. Menschliche Genomforschungsdaten sind geschützt zu archivieren, wie z. B. im Europäischen Genomarchiv (European Genome-phenome Archive, EGA)²¹ (siehe Abschnitt 3.4.3). Die dort abgelegten humanen Sequenzdaten sind nicht öffentlich zugänglich und werden stets nur verschlüsselt gespeichert und übertragen. Einen kontrollierten Zugang

¹⁸ Siehe zu der Bedeutung von Genomdaten bei Krebserkrankungen Schlomm/Biernath, Kap. 2, und zu der Bedeutung bei seltenen Erkrankungen Rieß/Schulze-Hentrich, Kap. 1.

¹⁹ Siehe hierzu Jungkunz et al., Kap. 4.

²⁰ Bei der Pseudonymisierung werden direkte persönliche Informationen wie Namen oder Geburtsdaten durch (meist zufällig generierte) Zahlen- und Buchstabenfolgen ersetzt. Dies verhindert eine direkte Zuordnung der Daten zu einzelnen Personen. Informationen zur Rückverfolgbarkeit, um Betroffene von Erkenntnissen der Genomanalyse zu informieren, verbleiben separat und werden nicht mit anderen geteilt.

²¹ Siehe unter: <https://ega-archive.org/> [23.10.2023].

zu solchen Daten erhalten nur autorisierte Forschende mit einem legitimen Forschungsvorhaben (Shabani et al., 2016). Hierfür muss ein Zugangsantrag bei der für die Daten verantwortlichen Person oder Institution gestellt werden. Ein Datenzugangskomitee oder eine vergleichbare Instanz prüft den Antrag und entscheidet, ob der Zugang gewährt wird. Nur durch solche vertraglich geregelten, kontrollierten Abläufe können Genom- und Metadaten eingesehen und genutzt werden, um alle durch die DSGVO vorgegebenen Auflagen zu erfüllen.

Die vorangehende informierte Einwilligung der Datenspende*r*innen schließt eine klare und transparente Aufklärung über die Ziele, Methoden, aber auch Risiken einer Studie sowie einer Datenfreigabe mit ein. Diese Einwilligung bildet in der Regel die Rechtsgrundlage für den Zugriff auf die Daten. Sie stellt auch sicher, dass die Patient*innen über die Verwendung ihrer genetischen Informationen informiert sind, freiwillig zustimmen und ihre Zustimmung jederzeit widerrufen können. Die umfangreichen Aspekte der rechtlichen und ethischen Fragen von Genomdaten werden in den kommenden Kapiteln diskutiert.

3.4 Strukturen für eine Archivierung und den kontrollierten Zugang

3.4.1 Herausforderungen und Lösungen bei der Etablierung einer sicheren Genomdateninfrastruktur

Wie bereits mehrfach angesprochen, bedarf eine sichere Archivierung diagnostisch erhobener Genomdaten mit kontrolliertem Zugang einer sehr gut geplanten Dateninfrastruktur. Zur ersten Aufgabe einer Genomdateninfrastruktur gehört es, Daten aus verschiedenen Quellen, Institutionen und Studien empfangen zu können (Data-Submission), diese geordnet einzuholen (mit Metadaten), zu archivieren und sicher zusammenzuführen (zu „verlinken“), um sie auf einer kontrolliert zugänglichen Plattform suchbar und nutzbar zu machen. Mit einem Netz von umfassenden IT-Sicherheitsmaßnahmen (Cybersicherheit) muss dabei sichergestellt sein, dass die Daten vor unbefugtem Zugriff und Datenverlust geschützt sind. Gleichzeitig muss berechtigten Kliniker*innen und Forschenden in einem DSGVO-konformen rechtlichen Rahmen der kontrollierte Zugang zu den für sie (auf Antrag) freigegebenen Daten ermöglicht werden. Die Dateninfrastruktur muss nachhaltig operieren, skalierbar und an veränderte Bedarfe (z. B. wachsende Datenvolumina, ergänzende (Meta-)Daten, alternative Datenformate, Analysemethoden) anpassbar sein. Insbesondere private Cloud-Infrastrukturen basierend auf skalierbaren Objektspeichern (z. B. betrieben von öffentlichen Einrichtungen) stellen diese Flexibilität auf technischer Seite zur Verfügung und sind einfach mit leistungsfähigen Rechnerinfrastrukturen (CPUs, GPUs)²² kombinierbar. Während die Daten in dieser Infrastruktur physisch verteilt gehalten werden, ist es wichtig, die Metadaten und die Verantwortung für den Betrieb auf nationaler Ebene zentral zu bündeln.

Mithin müsste eine solche Dateninfrastruktur die folgenden Anforderungen erfüllen:

- Sicherstellung der Einhaltung strenger Richtlinien und Verfahren zur Identifizierung und Autorisierung von Nutzer*innen und Nutzer*innenberechtigungen

22 CPU steht für „central processing unit“ und bezeichnet die zentrale Rechen- und Steuereinheit eines Computers. GPU steht für „graphics processing unit“ und bezeichnet den Grafikprozessor eines Gerätes.

- Standardisierte Abläufe für die Verwaltung von Zugangsanfragen, sicherer Zugang über verschlüsselte Downloads, DSGVO-konforme Datenverarbeitung
- Etablierung einer zentralen Zugangsstelle, die Datennutzungsanfragen schnell bescheiden kann
- Unterstützung von Anwendenden, um die notwendigen Schritte für die Daten-Submission, die Suche nach Datensätzen, die Beantragung des Zugriffs und den Zugriff auf Datensätze so einfach wie möglich zu gestalten
- Implementierung der FAIR-Prinzipien²³ (Wilkinson et al., 2016): Leichte Auffindbarkeit, Zugänglichkeit, Interoperabilität und Wiederverwendbarkeit für eine bessere Sekundärnutzung von Daten

Diese Anforderungen werden vom im Aufbau befindlichen Deutschen Humangenom-Phänomarchiv (GHGA)²⁴ (The GHGA Consortium, 2023) (siehe Abschnitt 3.4.3) in Deutschland erfüllt werden. GHGA übernimmt auch die Rolle eines nationalen Datenknotens im Rahmen des Föderierten Europäischen Genom-Phänom-Archivs (fEGA) (siehe Abschnitt 3.4.3). Im Modellvorhaben „Genomsequenzierung“ sind darüber hinaus unter Federführung des BfArM²⁵ und RKI²⁶ weitere Strukturen in Planung. Lösungen für eine integrierte und FAIRe Nutzung von Genom- und Omics-Daten aus Forschung und Patient*innenversorgung werden derzeit diskutiert und implementiert. Die nationale Genominitiative genomDE hat hierzu eine Reihe praktischer Vorschläge vorgelegt, die Rahmenbedingungen für notwendige Infrastrukturen definieren, u. a. wie eine logistische und rechtssichere Verknüpfung von Sequenzierzentren mit (einem) Genomdatenzentrum aussehen könnte.

3.4.2 FAIRe Datennutzung und standardisierte Workflows

Mit dem Aufbau einer Genomdateninfrastruktur müssen (Meta-)Datenmodelle und Datenmanagementprozesse implementiert werden, die Daten in den Infrastrukturen auffindbar machen und eine standardisierte Analyse unterschiedlicher Datensätze ermöglichen. Für eine FAIRe Datennutzung ist die Verknüpfung mit ontologisch klar definierten technischen und individuellen Metadaten entscheidend, um Omics-Daten nicht nur auffindbar, sondern auch interpretierbar und integrierbar zu machen. Metadaten liefern dabei nicht nur Informationen über die Eigenschaften der Datensätze. Sie beschreiben idealerweise auch die genutzten experimentellen Protokolle und Instrumente, die Software für die Prozessierung, aber auch den Ursprung der Proben (Zell- oder Gewebetyp), den Krankheitsstatus und deidentifizierte Details über Datenspender*innen wie Alter oder biologisches Geschlecht. Die Harmonisierung der Beschreibung von Daten ist von großer Bedeutung und sollte unter Verwendung etablierter und weit verbreiteter Ontologien und Vokabulare erfolgen. Die Implementierung klar definierter einheitlicher Metadatenstandards erleichtert die gemeinsame Nutzung von Daten und ermöglicht den interoperablen Datenaustausch

²³ Bei den FAIR-Prinzipien (Findable, Accessible, Interoperable, Reusable) handelt es sich um einen Leitfaden zur Verbesserung der Sekundärnutzung von Daten. Findable (auffindbar): Daten sollten mithilfe umfassender Beschreibungen und eindeutiger Bezeichnungen leicht auffindbar sein. Accessible (zugänglich): Daten sollten über klare und kontrollierte Zugangswege verfügbar sein. Interoperable (interoperabel): Daten sollten so strukturiert sein, dass sie sich problemlos mit anderen Datenquellen kombinieren und integrieren lassen. Reusable (wiederverwendbar): Durch detaillierte Metadaten sollten Daten optimal für Sekundärnutzung und Wiederverwendung aufbereitet sein.

²⁴ Siehe unter: <https://www.ghga.de/de/> [16.11.2023].

²⁵ Bundesinstitut für Arzneimittel und Medizinprodukte.

²⁶ Robert Koch-Institut.

über nationale Archive hinweg. Dies erleichtert den klinischen Nutzer*innen und Forschenden ihre primäre Aufgabe, nicht nur ihre eigenen, sondern auch Daten Dritter besser einordnen und somit gewinnbringend nutzen zu können.

Für die Prozessierung der Genomdaten und ihre nachfolgende Interpretation sollten idealerweise standardisierte Analyse-Workflows genutzt werden. Diese sollten sich zudem – auch im Hinblick auf Vergleichbarkeit – soweit möglich an optimierten Analysemethoden und globalen Standards orientieren und modular anpassbar sein, um reproduzierbare und effiziente Forschung und Diagnostik zu ermöglichen. Insbesondere für studienübergreifende Vergleiche und die gemeinsame Analyse mehrerer Kohorten ist eine einheitliche Verarbeitung der Datensätze notwendig.

3.4.3 Datenarchive: Gegenwärtiger Stand und Perspektiven

Beispiele, wie ein zentrales nationales Datenarchiv effizient organisiert sein könnte, kann man aus dem Bereich der europäischen Forschungsdateninfrastruktur ableiten. Es gibt bereits einige Bemühungen, Omics-Daten offen und FAIR – aber unter Einhaltung der notwendigen technischen und organisatorischen Sicherheitsvorkehrungen – verfügbar zu machen, auch über nationale Grenzen hinweg. Zu nennen ist hier das Europäische Genom-Phänom-Archiv (EGA), das gemeinsam vom European Bioinformatics Institute (EBI; Hinxton, England) und dem Centre for Genomic Regulation (CRG; Barcelona, Spanien) betrieben wird. Dieses europäische Humangenomarchiv speichert personenbezogene genetische und phänotypische Daten aus Forschungsprojekten in Europa und ist ein Vermittler für den kontrollierten Zugang. Um der unterschiedlichen Auslegung und Umsetzung der DSGVO in den EU-Mitgliedsstaaten gerecht zu werden und die nationalen Datenschutzbestimmungen zu berücksichtigen, wurde beschlossen, die EGA-Infrastruktur im Jahr 2022 durch eine international föderierte Infrastruktur²⁷ zu ergänzen – das Föderierte Europäische Genom-Phänom-Archiv (fEGA). Nationale Knotenpunkte innerhalb von fEGA archivieren und verwalten Daten im Einklang mit den jeweiligen nationalen Datenschutzbestimmungen und ermöglichen gleichzeitig einen kontrollierten Datenaustausch innerhalb der föderierten Struktur. Dabei werden ausschließlich Metadaten zu Studien und Datensätzen an das zentrale Metadatenarchiv von EGA übermittelt; die Sequenzdaten und personenbezogene Metadaten verbleiben in den jeweiligen nationalen Datenknoten. Dies ermöglicht die länderübergreifende Auffindbarkeit der Daten über alle Knoten hinweg.

In Deutschland ist das Deutsche Genom-Phänomarchiv (GHGA)²⁸ der nationale Datenknoten in fEGA. GHGA wird im Rahmen der NFDI-Initiative (Nationale Forschungsdateninfrastruktur)²⁹ der Deutschen Forschungsgemeinschaft (DFG) gefördert. Die Arbeiten an einer Genomdateninfrastruktur sind bereits weit vorangeschritten. In Zukunft werden/können die in der Forschung generierten und nach den FAIR-Prinzipien veröffentlichten menschlichen Genomdaten aus Deutschland archiviert und verwaltet werden – unter Einhaltung der nationalen Datenschutzbestimmungen (The GHGA Consortium, 2023). In den Teilen, in denen eine

²⁷ Siehe unter: <https://blog.ega-archive.org/the-federated-ega-network> [24.04.2024].

²⁸ German Human Genome-Phenome Archive; siehe unter: <https://www.ghga.de/de/> [24.10.2023].

²⁹ Siehe unter: <https://www.nfdi.de/> [24.10.2023].

Einwilligung zur Forschungsnutzung vorliegt, sind in GHGA auch die Genomdaten des Modellvorhabens „Genomsequenzierung“ verwaltbar. GHGA plant, in späteren Phasen neben der Archivierung auch zusätzliche Dienste wie die GHGA Cloud (für Analysen) und den GHGA Atlas (für eine tiefergehende Interpretation von Genomdaten) anzubieten.

Als Knotenpunkt innerhalb von fEGA bietet GHGA zudem Anknüpfungen und Austauschmöglichkeiten mit anderen europäischen Initiativen, an denen Deutschland beteiligt ist, wie z. B. die 1+ Million Genomes Initiative (1+MG)³⁰ und das Projekt Genomic Data Infrastructure (GDI)³¹. Zudem wäre z. B. über Daten-Analyse-Netzwerke wie ELIXIR³² eine Zusammenarbeit mit anderen Netzwerken für personalisierte beziehungsweise Genommedizin möglich, wie die NHS Genomic Strategy³³ (England), die France Genomic Medicine 2025 Initiative³⁴ (Frankreich), Genomic Medicine Sweden³⁵ (Schweden), das Danish National Genome Center³⁶ (Dänemark) und die Precision Medicine Infrastructure associated with Science and Technology³⁷ (Spanien).

3.5 Fazit

Die umfassende Sequenzierung humaner Genome wird die gesundheitliche Versorgung in der Medizin und die medizinische Forschung nachhaltig verändern. Das Modellvorhaben „Genomsequenzierung“ nach § 64e SGB V trägt dieser zu erwartenden Entwicklung Rechnung und wird die Gesundheitsversorgung wie auch die wissenschaftliche Rolle Deutschlands auf dem medizinisch wichtigen Gebiet der humanen Genomforschung stärken. Für diese Aufgabe ist die Etablierung einer nachhaltigen, breit nutzbaren, sicheren und von hohen Standards geleiteten Genomdateninfrastruktur essenziell. Die wesentlichen Bausteine dieser Dateninfrastruktur sind bereits vorhanden und sollten schnell und mit viel Energie ausgebaut, zusammengeführt und verstetigt werden. In diesem Prozess darf nicht aus den Augen verloren werden, dass die Etablierung hoher Datenstandards und das zentrale Archivieren von Genomdaten unverzichtbar sind, sowohl für eine optimale Nutzung und Interpretation der Daten für die medizinische Grundversorgung als auch für eine weiterführende Forschung. Genome sind nicht selbsterklärend und ihre Nutzung und Interpretation bedürfen einer hohen fachlichen Expertise. Die Initiative genomDE bildet mit ihren Partnern in Gesundheitszentren einen sehr guten Rahmen, diese Verbindung für eine verbesserte gesundheitliche Versorgung und Weiterentwicklung durch Forschung zu generieren. Operativ klare Strukturen sind dabei wichtige Voraussetzungen – auch um die Verbindung mit internationalen Genomdateninfrastrukturen zu gewährleisten und einen sichtbaren Beitrag zu Programmen wie 1+MG und GDI darzustellen.

30 Siehe unter: <https://digital-strategy.ec.europa.eu/en/policies/1-million-genomes> [24.04.2024].

31 Siehe unter: <https://gdi.onemilliongenomes.eu/> [24.10.2023].

32 Siehe unter: <https://elixir-europe.org/> [16.11.2023].

33 Siehe unter: <https://www.england.nhs.uk/genomics/genomics-strategy/> [16.11.2023].

34 Siehe unter: <https://pfgm2025.aviesan.fr/> [24.04.2024].

35 Siehe unter: <https://genomicmedicine.se/en/> [16.11.2023].

36 Siehe unter: <https://eng.ngc.dk/> [16.11.2023].

37 Siehe unter: <https://www.isciii.es/QueHacemos/Financiacion/IMPACT/Paginas/default.aspx> [16.11.2023].

3.6 Literaturverzeichnis

Fehse, B. et al. (Hrsg.) (2022): Im Fokus: RNA. Eine aktuelle Bestandsaufnahme der Arbeitsgruppe Gentechnologiebericht. Berlin Institute of Health at Charité, Berlin. DOI: 10.17169/refubium-36831.

Gymrek, M. et al. (2013): Identifying personal genomes by surname inference. In: *Science* 339(6117): 321–324. DOI: 10.1126/science.1229566.

International Human Genome Sequencing Consortium (2001): Initial sequencing and analysis of the human genome. In: *Nature* 409(6822): 860–921. DOI: 10.1038/35057062.

Liao, W.-W. et al. (2023): A draft human pangenome reference. In: *Nature* 617(7960): 312–324. DOI: 10.1038/s41586-023-05896-x.

Shabani, M. et al. (2016): Controlled access under review: Improving the governance of genomic data access. In: *PLoS biology* 13(12): e1002339. DOI: 10.1371/journal.pbio.1002339.

Smetana, J./Brož, P. (2022): National genome initiatives in Europe and the United Kingdom in the era of whole-genome sequencing: a comprehensive review. In: *Genes* 13(3): 556. DOI: 10.3390/genes13030556.

The GHGA Consortium (2023): The German Human Genome-Phenome Archive – GHGA Brochure. Zenodo. DOI: 10.5281/zenodo.8359328.

Venter, J. C. et al. (2001): The sequence of the human genome. In: *Science* 291(5507): 1304–1351. DOI: 10.1126/science.1058040.

Walter, J./Gasparoni, N. (2021): Themenbereich Epigenetik: von Zellidentitäten bis hin zu Krankheiten und Therapien. In: Fehse, B. et al. (Hrsg.): Fünfter Gentechnologiebericht. Sachstand und Perspektiven für Forschung und Anwendung. Nomos, Berlin: 93–113. Unter: <https://www.nomos-elibrary.de/10.5771/9783748927242-93/3-themenbereich-epigenetik-von-zellidentitaeten-bis-hin-zu-krankheiten-und-therapien?page=1> [19.02.2024].

Walter, J./Schickl, H. (Hrsg.) (2019): Einzelzellanalyse in Forschung und Medizin – Eine Stellungnahme der interdisziplinären Arbeitsgruppe Gentechnologiebericht. Berlin Brandenburgische Akademie der Wissenschaften, Berlin. Unter: https://edoc.bbaw.de/files/3278/BBAW_Einzelzellanalyse_Walter_Schickl.pdf [19.02.2024].

Wilkinson, M. D. et al. (2016): The FAIR guiding principles for scientific data management and stewardship. In: *Scientific data* 3(1): 1–9. DOI: 10.1038/sdata.2016.18.

4. Der ethische Rahmen für die Nutzung von Genomdaten in der Forschung

Martin Jungkunz, Christoph Schickhardt und Eva C. Winkler

4.1 Einleitung

Die Bedeutung moderner Genomsequenzierungsverfahren ist in den letzten Jahrzehnten stark gestiegen.¹ In der biomedizinischen Forschung ist der Einsatz genomischer Analyseverfahren nicht mehr wegzudenken – sowohl in der Grundlagenforschung als auch in der versorgungsnahen Forschung. In der klinischen Versorgung werden Genomanalysen aktuell ebenfalls schon verwendet, bspw. in der humangenetischen Diagnostik bei entwicklungsverzögerten Kindern und in der Krebsmedizin,² um die genetischen Ursachen einer Erkrankung herauszufinden und über Behandlungsstrategien zu informieren. Dadurch, dass die Kosten für Genomsequenzierung und -analyse deutlich gesunken sind, und wegen des bereits heute ersichtlichen Nutzens genomischer Forschung und Diagnostik ist zu erwarten, dass genomische Analyseverfahren noch an Bedeutung gewinnen werden. An Vorhaben wie dem Modellvorhaben genomDE³ zeigt sich, dass eine strukturierte Ausweitung genomischer Analyseverfahren in der Regelversorgung in der Zukunft absehbar ist. In diesem Modellvorhaben ist geplant, Genomanalysen unter definierten Standardbedingungen bei seltenen und bei onkologischen Erkrankungen für Diagnostik und Therapiefindung einzusetzen.

Die Daten, die bei Genomanalysen entstehen, genomische Daten, sind für die biomedizinische Forschung äußerst wertvoll – gewöhnlich in Verbindung mit anderen klinischen Daten. Da man große Datensätze genomischer Daten benötigt, um auch seltene Zusammenhänge zu entdecken, gibt es zahlreiche Initiativen, die sich das Teilen und Vernetzen von Genomdatensätzen zu Forschungszwecken zum Ziel gemacht haben.⁴ Einige dieser Initiativen zielen darauf ab, eine Infrastruktur aufzubauen und bereitzustellen, in der Daten archiviert und/oder vernetzt werden und die Forscher*innen den Zugang zu großen Datenmengen ermöglichen soll. Stand jetzt bleiben die Daten dabei i. d. R. im Einflussbereich der Datenerzeuger*innen (Forscher*innen, Kliniken), die auch über den Datenzugriff Dritter mitentscheiden. Neben der

¹ Siehe für weitere Informationen zu modernen Genomsequenzierungstechnologien Rieß/Schulze-Hentrich, Kap. 1.

² Siehe zu Genomdaten in der Krebsdiagnostik Schlomm/Biernath, Kap. 2.

³ Siehe unter: <https://www.genom.de/> [15.12.2023].

⁴ Siehe bspw. das „German Human Genome Phenome Archive“ (GHGA), aber auch die Euroäische „1+ Million Genomes“-Initiative und die „Global Alliance for Genomics and Health“ (GA4GH). Zu ethischen und rechtlichen Rahmenbedingungen für Infrastrukturprojekte siehe Schickhardt et al., 2023.

Vernetzung von Daten mithilfe großer Infrastrukturprojekte findet der Austausch von Daten aber auch in und zwischen Forschungskonsortien statt. In beiden Fällen werden gewöhnlich genomische Daten zusammen mit klinischen Daten geteilt, die ursprünglich aus der Versorgung stammen oder direkt in Forschungsprojekten erhoben wurden.

So groß der Nutzen ist, so wichtig ist es, dass Forschung mit genomischen Daten in einem Rahmen stattfindet, der ethische Aspekte berücksichtigt und eine vertrauenswürdige Governance garantiert.⁵ Ein solcher Rahmen soll, ausgehend von den ethischen Herausforderungen, die bei der Forschungsnutzung genomischer Daten entstehen, in diesem Kapitel skizziert werden.

4.2 Ethische Rahmenbedingungen für die Forschungsnutzung genomischer Daten

Die ethischen Herausforderungen bei der Forschungsnutzung genomischer Daten hängen eng mit einigen grundsätzlichen spezifischen Eigenschaften dieser Art von Daten zusammen. Mit „genomischen Daten“ sind hier diejenigen umfangreichen Daten gemeint, die aus einer Ganzgenom- oder Ganzexomsequenzierung stammen, sowie Daten, die im Rahmen sehr umfangreicher Genpanels erhoben wurden.

Genomische Daten zeichnen sich zum einen dadurch aus, dass sie für jede Person spezifisch sind (mit Ausnahme von eineiigen Zwillingen, die das gleiche Erbgut haben). Eine Reidentifizierung der Person anhand ihrer einzigartigen Genomdaten ist also grundsätzlich möglich (Gymrek et al., 2013), wenngleich hierfür einiges vorausgesetzt werden muss. Eine Person kann nur anhand ihres genomischen Datensatzes identifiziert werden, wenn Dritte noch Zugang zu anderen persönlichen Daten wie Klarnamen oder Adresse zusammen mit genetischen Daten dieser Person haben. Eine Anonymisierung genomischer Daten ist nicht möglich. Der Personenbezug bzw. die potenzielle Personenbeziehbarkeit ist den Daten inhärent. Möglich und in der Forschung Standard ist jedoch eine Entkopplung der Daten von direkt identifizierenden Daten (bspw. dadurch, dass Name, Geburtsdatum oder Adresse entfernt oder durch Codes ersetzt werden). Eine weitere spezifische Eigenschaft genomischer Daten ist ihre prädiktive Aussagekraft. Dies bedeutet, dass genomische Daten nicht nur Informationen über bereits manifestierte Merkmale des Datensubjekts (wie bspw. die Haar- oder Augenfarbe) oder über eine bereits diagnostizierte Erkrankung enthalten. Genomische Daten können auch Wahrscheinlichkeitsaussagen über das mögliche Auftreten von Krankheiten in der Zukunft ermöglichen. Eine dritte spezifische Eigenschaft genomischer Daten ist, dass sie auch Aussagen über Verwandte der Datensubjekte erlauben, sowohl in Bezug auf manifeste Eigenschaften als auch in Bezug auf mögliche Krankheitsprädispositionen. Weiterhin können genomische Daten dazu dienen, Verwandte zu identifizieren.

⁵ Ein solcher Governance-Rahmen wird bspw. von GA4GH angestrebt (Rehm et al., 2021).

4.2.1 Beste Nutzbarkeit und Schutz der Patient*innen durch vertrauenswürdige Governance

Betrachten wir die drei genannten Eigenschaften genomischer Daten (Re-Identifizierbarkeit, prädiktive Eigenschaften, Relevanz für Verwandte), so wird klar, dass diese Daten als sehr sensibel einzustufen sind. Ihr Schutz ist somit von großer Wichtigkeit. Wie bei allen Prozessen, in denen persönliche Daten verarbeitet werden, muss das Recht auf informationelle Selbstbestimmung der Datensubjekte, in unserem Fall der Patient*innen oder Proband*innen, von denen die Daten stammen, berücksichtigt werden.⁶ Damit ist das Recht des Einzelnen gemeint, „grundsätzlich selbst über die Preisgabe und Verwendung seiner persönlichen Daten zu bestimmen“ (BVerfGE 65, 1, 43). Bei der Forschung mit Genomdaten (bzw. mit personenbezogenen Daten im Allgemeinen) bedeutet dieses Recht zum einen, dass die Daten eigentlich nur genutzt werden dürfen, wenn die Datensubjekte über die jeweilige Nutzung aufgeklärt wurden und ihre Einwilligung erteilt haben. Zum anderen beinhaltet das Recht auf informationelle Selbstbestimmung weitere (Mikro-)Rechte, die nach Beginn der Datennutzung bestehen bleiben und ausgeübt werden können, wie z. B. das Recht auf Erhalt einer Kopie der Daten, auf Widerruf der Einwilligung oder auch das Recht auf Löschung der Daten.

4.2.1.1 Schutz des Rechts auf informationelle Selbstbestimmung

Der Goldstandard für die Teilnahme an Forschung liegt in einer Aufklärung der Proband*innen durch die jeweiligen Forscher*innen und einer spezifischen informierten Einwilligung. Ein solches Vorgehen wäre bei der genomischen Forschung wie auch bei anderen Arten der datengetriebenen medizinischen Forschung extrem aufwendig, wenn Daten nicht nur für ein einziges, sondern für weitere Forschungsprojekte genutzt werden sollen, die mitunter zum Zeitpunkt der Datenerhebung noch gar nicht bekannt sind. Eine bisher vorwiegend theoretisch diskutierte Methode, Datensubjekten auch bei datenintensiver Forschung die Möglichkeit zu geben, jeder Datennutzung für jedes einzelne Forschungsprojekt zuzustimmen, ist das Modell der sogenannten „dynamischen Einwilligung“. Hier erhalten die Datensubjekte vor jeder Nutzung ihrer Daten durch ein bestimmtes Projekt elektronisch über eine App spezifische Informationen zum Projekt und können der Nutzung zustimmen oder sie verweigern. Die dynamische Einwilligung ist jedoch sehr aufwendig in der Implementierung, erschwert die Nutzbarkeit der Daten und bringt technische Datenschutzrisiken mit sich.⁷

Um den Willen der Datensubjekte und somit deren Recht auf informationelle Selbstbestimmung zu berücksichtigen, ohne die Forschung übermäßig zu behindern, wird deshalb teilweise auf die sogenannte breite Einwilligung (Broad Consent) zurückgegriffen.⁸ Hier werden die Datensubjekte im Vorfeld der genomischen Untersuchung über die Forschungsnutzung ihrer Daten im Allgemeinen aufgeklärt und können in diese Nutzung einwilligen (vgl. z. B. EURAT, 2015). Es erfolgt keine spezifische Aufklärung zu einzelnen Projekten.

⁶ Zum rechtlichen Rahmen siehe Müller-Terpitz/Straub, Kap. 5.

⁷ Zu unterschiedlichen Einwilligungsmodellen bei der Sekundärnutzung von Daten siehe Wiertz/Boldt, 2022 sowie zu ihrer Bewertung Jungkunz et al., 2022.

⁸ In einer repräsentativen Befragung unter Krebspatient*innen zur Nutzung ihrer klinischen Daten in Baden-Württemberg hat die breite Einwilligung unter verschiedenen Einwilligungsmodellen die größte Zustimmung erhalten (Königeter et al., 2022).

Die Verwendung der breiten Einwilligung gilt als forschungsfreundlich, bedeutet aber auch, dass die Datensubjekte ihre informationelle Selbstbestimmung nicht vollständig informiert ausüben können. Zur Kompensation dieser Einschränkung ist es ethisch geboten, begleitend einen Governance-Rahmen zu etablieren, der die organisatorischen, technischen, rechtlichen, ethischen und prozeduralen Rahmenbedingungen des Datenzugriffs und die Aufsicht so gestaltet, dass die Forschung mit den Genomdaten berechtigtes Vertrauen genießt. Im Prozess der Aufklärung und Einwilligung sollten neben den Nutzenpotenzialen für die Forschung auch die Risiken und Implikationen der Nutzung genetischer Daten für die betroffenen Personen angemessen angesprochen und erklärt werden. Textbeispiele für die Aufklärung für die sekundäre Forschungsnutzung genomischer Daten finden sich bspw. bei den Aufklärungsmaterialien des Deutschen Genomarchivs GHGA (Bruns et al., 2022). Dieses kann durch ein zusätzliches (Online-)Informationsangebot ergänzt werden, wie es die Autor*innen für die Medizininformatik-Initiative (MII) entwickelt haben.⁹

Die Einwilligung ist nur ein Teil des Rechts auf informationelle Selbstbestimmung. Um die genannten Mikrorechte bzw. deren mögliche Ausübung zu gewährleisten (Recht auf Widerspruch, Datenlöschung, Erhalt einer Kopie der Daten etc.), müssen zusätzlich technisch-organisatorische Maßnahmen etabliert werden. Hiermit sind konkrete technische Strukturen und die Festlegung von Verantwortlichkeiten gemeint, die die Verwirklichung genannter Rechte technisch und organisatorisch ermöglichen. Es muss z. B. ein klar definiertes Verfahren existieren, Daten auf Wunsch der Datensubjekte zu löschen oder herauszugeben.¹⁰

4.2.1.2 Forschungsfreundlicher Datenzugang bei gleichzeitigem Schutz der Privatsphäre

Der Governance-Rahmen für die Forschungsnutzung genomischer Daten sollte zum einen, wie erwähnt, berechtigtes Vertrauen in die Forschung fördern, zum anderen aber den Forschungsnutzen der Daten so groß wie möglich halten. Hierfür ist es wichtig, die Daten möglichst vielen Forscher*innen mit einfachen Regeln zugänglich zu machen, dabei aber trotzdem einen möglichst hohen Schutz der Daten zu gewährleisten. Als Orientierung für die Gewährleistung eines möglichst hohen Forschungsnutzens von Daten gelten die FAIR-Prinzipien: Daten sollen „findable“, „accessible“, „interoperable“ und „reusable“ sein. Ein erster Schritt, um die Daten auffindbar zu machen, ist der Zugang zu den Daten über eine zentrale Stelle, an die Forscher*innen sich wenden können, wenn sie mit genomischen Daten forschen wollen. Der zweite Schritt ist die Bereitstellung qualitativ hochwertiger Metadaten. Dabei handelt es sich um eine Beschreibung der Daten selbst, d. h. um Informationen zu Fragen wie z. B., um welche Daten es sich handelt, woher sie kommen, von wem sie erhoben und für welche Nutzung sie freigegeben wurden. Gute, d. h. umfangreiche Metadaten sind zwingend notwendig, um Forscher*innen überhaupt zu ermöglichen, Daten für ihre Forschungsprojekte zu identifizieren. Diejenigen Metadaten, die für diesen Zweck notwendig sind, können bspw. über ein Online-Portal zur Verfügung gestellt werden.

⁹ Vgl. hierzu <https://www.medizininformatik-initiative.de/de/genetische-daten> [15.12.2023].

¹⁰ Zur Herausgabe genomischer Rohdaten vgl. EURAT, 2019.

Nachdem Forscher*innen die Daten identifiziert haben, mit denen sie forschen wollen, müssen sie Zugriff auf die eigentlichen Daten beantragen. Für die Berechtigung auf Datenzugriff müssen vorab klare Kriterien kommuniziert werden, die bei der Bearbeitung eines Antrags auf Datenzugang und -nutzung geprüft werden. Beispiele für solche Kriterien sind das Vorliegen eines Ethikvotums und einer rechtlichen Grundlage für die Datennutzung. Auch muss plausibel sein, dass die angeforderten Daten entsprechend dem geplanten Studiendesign benötigt werden. Für das Verfahren der Antragsprüfung und -bewilligung sind unterschiedliche Konstellationen denkbar. Im Rahmen von GHGA werden bspw. die Anträge auf Datenzugriff von den jeweiligen datengebenden Stellen selbst bearbeitet. Prinzipiell ist aber auch denkbar, dass die Anträge an einer zentralen Stelle geprüft werden. In beiden Konstellationen ist für die Bearbeitung der Anträge die Einsetzung eines Datenzugangskomitees denkbar und oft schon Praxis, wie es bspw. die Medizininformatik-Initiative (MII)¹¹ – eine Infrastruktur für die Sekundärnutzung von Daten aus Kliniken – vorsieht. Das Datenzugangskomitee sollte aus Vertreter*innen unterschiedlicher relevanter Wissenschaftsdisziplinen (z. B. Genomik, Biometrie, Ethik, Recht) sowie aus Datenschutzbeauftragten und Vertreter*innen der dem Komitee jeweils übergeordneten Institution bestehen (z. B. der entsprechenden Klinik).

4.2.2 Der Umgang mit genetischen Zusatzbefunden

Ein weiterer wichtiger Aspekt bei der Forschungsnutzung genomischer Daten ist die mögliche Entdeckung sogenannter Zusatzbefunde. Hierbei handelt es sich um krankheitsrelevante Befunde (oder Befunde, die relevant für die Reproduktion sind), die im Rahmen genetischer oder genomischer Analysen entdeckt werden, die jedoch nicht mit der intendierten Fragestellung in Verbindung stehen und nach denen nicht aktiv gesucht wurde (EURAT, 2023). Dem Begriff nach gelten bei Genomdaten, die im reinen Forschungskontext erhoben wurden, diejenigen Befunde als Zusatzbefunde, die nicht mit der ursprünglichen Forschungsfrage in Verbindung stehen. Im Falle der sekundären Forschungsnutzung von Genomdaten aus der Humangenetik können Zusatzbefunde dadurch entstehen, dass die vorhandenen Daten zu Forschungszwecken mit einer anderen Fragestellung oder Methodik analysiert werden als bei ihrer Erhebung im Kontext der ursprünglichen Patient*innenversorgung. Zusatzbefunde sind auch deshalb ethisch herausfordernd, weil im Vorfeld einer genomischen Analyse nur ganz allgemein über verschiedene Kategorien von Zusatzbefunden aufgeklärt werden kann, da prinzipiell eine große Anzahl möglicher gesundheitsrelevanter Varianten entdeckt werden kann, über die eine Aufklärung im Vorfeld schlicht zu aufwendig wäre. Diese generische Aufklärung widerspricht dem humangenetischen Prinzip, nur Befunde zurückzumelden, über deren Eigenschaften und Konsequenzen im Vorfeld spezifisch informiert wurde, um der betroffenen Person zu ermöglichen, sich möglichst informiert für oder gegen eine Rückmeldung entscheiden zu können.

In der Bioethik wird seit Längerem diskutiert, ob Zusatzbefunde, die im Forschungskontext auftreten, zurückgemeldet werden sollen oder nicht, und wenn ja, welche und wie genau. Weitgehende Einigkeit besteht darin, dass Proband*innen vor der genomischen Untersuchung klar kommuniziert werden muss, ob Zusatzbefunde zurückgemeldet werden, und

¹¹ Siehe unter: <https://www.medizininformatik-initiative.de/de/start> [15.12.2023].

wenn ja, welche Kategorien von Befunden. Empfehlungen, wie mit Zusatzbefunden umgegangen werden soll, unterscheiden sich jedoch gerade im internationalen Vergleich mitunter stark (Miller et al., 2022; GfH, 2023).

Die tatsächliche Rückmeldung von Zusatzbefunden sollte über die jeweils behandelnden Ärzt*innen stattfinden, da diese bereits im Kontakt mit den Patient*innen stehen und wissen, ob der jeweilige Befund ggf. sogar schon bekannt ist. Zudem müssen sie sicherstellen, dass der Befund validiert wird und können eine genetische Beratung anbieten, was bei Rückmeldung genetischer Befunde dringend erforderlich und gesetzlich vorgeschrieben ist.

Je nach Infrastruktur, in der die Genomforschung stattfindet, kann die Rückmeldepraxis von Zusatzbefunden sehr unterschiedlich ausfallen. Findet die Forschung bspw. versorgungsnah statt, d. h. besteht eine Möglichkeit für die Forscher*innen, einen Zusatzbefund mit zumutbarem Aufwand an den/die behandelnde/n Arzt/Ärztin zu übermitteln, so ist es geboten, dass Zusatzbefunde, die eine Prädisposition für eine behandelbare Erkrankung anzeigen, zur Rückmeldung angeboten werden. Bei großen internationalen Forschungskonsortien besteht diese Möglichkeit des Kontaktes mit dem/der behandelnden Arzt/Ärztin eventuell nicht, was eine Rückmeldung sehr schwierig und somit für Forscher*innen ggf. unzumutbar macht.

Ist aber eine Rückmeldung von Zusatzbefunden in der jeweiligen Forschungskonstellation möglich, so sollte sich die Rückmeldepraxis am Patient*innenwillen orientieren (Winkler/Schickhardt, 2014; EURAT, 2015) bzw. bei minderjährigen Patient*innen am Kindeswohl (EURAT, 2023). Die Entscheidung über die Rückmeldung, die im Aufklärungs- und Einwilligungsprozess getroffen wird, sollte den Patient*innen (im Falle von Minderjährigen i. d. R. den Eltern) obliegen. Es sollten ihnen keine Zusatzbefunde „aufgezwungen“ oder gegen ihren Willen vorenthalten werden. Zum einen ist diese Orientierung am Patient*innenwillen geboten, da diese nur so ihr Recht auf Wissen und Nichtwissen ausüben können. Zum anderen geht es bei der Beantwortung der Frage, ob ein Zusatzbefund zurückgemeldet werden soll oder nicht, stets auch um eine Abschätzung des Nutzens der Rückmeldung im Verhältnis zu einer möglichen Belastung. Auch wenn empirische Daten nahelegen, dass die psychische Belastung durch genetisches Wissen (Crozier et al., 2015) bzw. Zusatzbefunde (Nambot et al., 2021; Schoot et al., 2021) in aller Regel weniger groß ist, als dies intuitiv erscheinen mag, so ist doch anzunehmen, dass diese mögliche Belastung individuell sehr unterschiedlich sein kann und (erwachsene) Patient*innen die Abwägung der möglichen Belastungen gegen den Nutzen am besten selbst vornehmen können.

4.2.3 Verantwortung von Forscher*innen in der Genomforschung

Es muss die Frage gestellt werden, welche Verantwortung Forscher*innen im Umgang mit genomischen Daten zukommt. Genomforscher*innen übernehmen insofern eine wichtige Rolle, als sie durch die Analyse genomischer Daten Zugriff auf Wissen haben, das für die betroffenen Patient*innen gesundheitsrelevant sein kann. Forscher*innen haben also in gewisser Weise ärztliches Wissen, sind aber nicht notwendigerweise einem ärztlichen Berufsethos verpflichtet. Im Gegenteil orientiert sich die Forschungsethik an anderen Prinzipien als die ärztliche Ethik. Ärzt*innen sind dem einzelnen Patienten/der einzelnen Patien-

tin verpflichtet und müssen diese/n nach dem Stand der ärztlichen Heilkunst behandeln (gute klinische Praxis). Forscher*innen hingegen sind der guten wissenschaftlichen Praxis verpflichtet, die das Streben nach überindividueller Erkenntnis in den Mittelpunkt stellt, nicht die Verantwortung für einzelne Patient*innen. Es stellt sich also die Frage, inwieweit Forscher*innen, die mit genomischen Daten arbeiten und somit potenziell über besagtes Gesundheitswissen verfügen, neben dem wissenschaftlichen Berufsethos auch Aspekten des ärztlichen Ethos gerecht werden müssen.

Es ist sinnvoll, die Verantwortlichkeiten von Genomforscher*innen im Rahmen eines Verhaltenskodex festzuhalten. Dies dient nicht nur dem Zweck, die Verhaltensregeln für alle in schriftlicher und nachvollziehbarer Form verfügbar zu machen. Je nachdem, wie ein solcher Verhaltenskodex in einer Institution implementiert wird, kann er normativ-rechtliches Gewicht gewinnen, bspw. dadurch, dass ein Forschungsinstitut den Kodex als allgemeine Dienstanweisung verankert, an die sich Forscher*innen halten müssen. Die Heidelberger EURAT-Gruppe hat hierzu einen Vorschlag erarbeitet (EURAT, 2015). In diesem Vorschlag werden die wichtigsten ethischen Grundsätze festgehalten, die dem Handeln der Forscher*innen in Bezug auf Patient*innen (Datensubjekte) zugrunde liegen sollten. Hierzu gehören:

- die Achtung der Selbstbestimmung
- Schadensvermeidung
- Sorgfalt
- Nichtdiskriminierung
- Schutz der Privatsphäre
- Unentgeltlichkeit
- gute wissenschaftliche Praxis
- Streben nach gesellschaftlichem Nutzen der eigenen Forschung
- Schutz zukünftiger Generationen

Neben diesen allgemeinen Grundsätzen sollten sich Forscher*innen auf einige wichtige Verfahren verpflichten. Hierzu gehören:

- die Risikoeinschätzung und -mitteilung an die Datensubjekte,
- das Einholen der informierten Einwilligung und eines Ethik-Votums,
- die Pseudonymisierung der Daten und Proben,
- die Etablierung von Zugriffsbeschränkungen für die Daten und
- eine Pflicht, ethische Richtlinien auch von etwaigen Projektpartner*innen zu fordern.

4.3 Ausblick

Vor dem Hintergrund der benannten wichtigsten ethischen Rahmenbedingungen für die vertrauenswürdige Forschungsnutzung genomischer Daten ist es besonders interessant, die aktuellen Gesetzesvorhaben in den Blick zu nehmen. Das Gesundheitsdatennutzungsgesetz (GDNG) und das Digital-Gesetz (DigiG) zeigen die Relevanz des Themas Sekundärnutzung von Patient*innendaten. Ziel des GDNG ist es, unterschiedliche Patient*innendaten, darunter auch genomische Daten zusammenzuführen und der Forschung zur

Verfügung zu stellen. Das GDNG sieht hierfür kein Einwilligungsmodell, sondern eine Widerspruchslösung vor. Damit verlagert sich ein großer Teil der Verantwortung auf einen vertrauenswürdigen Governance-Rahmen und macht diesen umso relevanter.

4.4 Interessenkonflikt

Die beiden Autor*innen C. Schickhardt und E. C. Winkler sind seit Jahren in der Medizin-informatik-Initiative (MII) und dem German Human Genome-Phenome Archive (GHGA) tätig. M. Jungkunz gibt an, dass kein Interessenkonflikt besteht.

4.5 Literaturverzeichnis

- Bruns, A. et al. (2022):** Consent modules for data sharing via the German Human Genome-Phenome Archive (GHGA). Unter: <https://www.ghga.de/resources/consent-modules> [15.12.2023].
- Crozier, S. et al. (2015):** The psychological impact of predictive genetic testing for Huntington's disease: a systematic review of the literature. In: *J Genet Couns* 24(1): 29–39. DOI: 10.1007/s10897-014-9755-y.
- EURAT (2015):** Eckpunkte für eine Heidelberger Praxis der Ganzgenomsequenzierung. Forum Marsilius Kolleg, Heidelberg.
- EURAT (2019):** Zur Herausgabe genomischer Rohdaten an Patient_innen und Studienteilnehmende. Forum Marsilius Kolleg, Heidelberg.
- EURAT (2023):** Stellungnahme zur Rückmeldung genetischer Zusatzbefunde Minderjähriger. Forum Marsilius Kolleg, Heidelberg.
- GfH (2023):** Stellungnahme der Deutschen Gesellschaft für Humangenetik zu Zusatz- und Zufallsbefunden in der genetischen Diagnostik. In: *Medizinische Genetik* 35(4): 313–321. DOI: 10.1515/medgen-2023-2060.
- Gymrek, M. et al. (2013):** Identifying personal genomes by surname inference. In: *Science* 339(6117): 321–324. DOI: 10.1126/science.1229566.
- Jungkunz, M. et al. (2022):** Stellungnahme zur Etablierung der sekundären Forschungsnutzung von Behandlungsdaten in Deutschland. Ergebnisse des Verbundprojekts LinCDat: „Learning from clinical data. Ethical, social and legal aspects“. Forum Marsilius Kolleg, Heidelberg.
- Köngeter, A. et al. (2022):** Patients' willingness to provide their clinical data for research purposes and acceptance of different consent models: Findings from a representative survey of cancer patients in Germany. In: *J Med Internet Res* 24(8): e37665. DOI: 10.2196/37665.
- Miller, D. T. et al. (2022):** ACMG SF v3.1 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). In: *Genet Med* 24(7): 1407–1414. DOI: 10.1016/j.gim.2022.04.006.
- Nambot, S. et al. (2021):** Incidental findings in a series of 2500 gene panel tests for a genetic predisposition to cancer: Results and impact on patients. In: *Eur J Med Genet* 64(5): 104196. DOI: 10.1016/j.ejmg.2021.104196.
- Rehm, H. L. et al. (2021):** GA4GH: International policies and standards for data sharing across genomic research and healthcare. In: *Cell Genom* 1(2). DOI: 10.1016/j.xgen.2021.100029.
- Schickhardt, C. et al. (2023):** Dateninfrastrukturen für die Gesundheitsforschung. Ethische Rahmenbedingungen und rechtliche Umsetzung. In: *Bundesgesundheitsbl* 66(2): 160–167. DOI: 10.1007/s00103-022-03648-2.
- Schoot, V. V. et al. (2021):** The impact of unsolicited findings in clinical exome sequencing, a qualitative interview study. In: *Eur J Hum Genet* 29(6): 930–939. DOI: 10.1038/s41431-021-00834-9.
- Wiertz, S./Boldt, J. (2022):** Evaluating models of consent in changing health research environments. In: *Med Health Care Philos* 25(2): 269–280. DOI: 10.1007/s11019-022-10074-3.
- Winkler, E. C./Schickhardt, C. (2014):** Ethische Herausforderungen der Genomsequenzierung in der translationalen Forschung und Antworten aus dem EURAT-Projekt. In: *LaboratoriumsMedizin* 38(4): 211–220. DOI: 10.1515/lab-med-2014-0027.

5. Der rechtliche Rahmen für die Nutzung von Patient*innen- und Genomdaten – Schutz und Nutzen aus medizinrechtlicher Sicht

Ralf Müller-Terpitz und Hannah Lilly Straub

5.1 Schutz der Patient*innen und ihrer Daten durch das GenDG

Im Juli 2021 wurde das Fünfte Sozialgesetzbuch (SGB V) um die rechtliche Grundlage für ein Modellvorhaben zur umfassenden Diagnostik und Therapiefindung mittels Genomsequenzierung bei seltenen onkologischen Krankheiten¹ ergänzt. Der neue § 64e Abs. 2 S. 2, Abs. 15 SGB V sieht nunmehr vor, dass bei Genomsequenzierungen die Bestimmungen des Gendiagnostikgesetzes (GenDG)² zu beachten sind. Das GenDG dient insbesondere der Erfüllung staatlicher Schutzpflichten in Bezug auf die Menschenwürdegarantie (Art. 1 Abs. 1 GG) und das Recht auf informationelle Selbstbestimmung (Art. 2 Abs. 1 i. V. m. Art. 1 Abs. 1 GG).³ Dazu knüpft es die Durchführung genetischer Untersuchungen an strenge Voraussetzungen.

5.1.1 Informierte Einwilligung durch Aufklärung

Zum Schutz der Patient*innenautonomie und ihres informationellen Selbstbestimmungsrechts stehen die Durchführung genetischer Untersuchungen und die Mitteilung der daraus gewonnenen Ergebnisse unter einem Einwilligungsvorbehalt (§ 8 Abs. 1 S. 1 und 2 GenDG) seitens der betroffenen Personen. Voraussetzung für die Abgabe einer wirksamen Einwilligung ist die vorherige Aufklärung der Patient*innen nach § 9 GenDG.⁴ Sie soll die Patient*innen in die Lage versetzen, die Untersuchung und potenzielle Folgewirkungen zu verstehen, um Risiko und Nutzen für sich abwägen zu können.⁵ § 9 Abs. 1 und 2 Nr. 1 GenDG setzt daher voraus, dass zum Inhalt der Aufklärung das Wesen, die Bedeutung und die Tragweite der Untersuchung sowie insbesondere ihr Zweck, ihre Art, ihr Umfang, ihre Aussagekraft und ihre erzielbaren Ergebnisse gehören. Einbezogen werden sollen auch mögliche Befunde außerhalb des Untersuchungszwecks (sog. Zufallsbefunde).⁶ Die Aufklärung im Vorfeld einer Genomsequenzierung stellt deshalb eine Herausforderung dar. Eine Genomsequenzierung fördert eine solche Vielzahl an Informationen zu Tage, dass es sich als

¹ Zur Genomsequenzierung im Kontext von Krebserkrankungen siehe Schlomm/Biernath, Kap. 2.

² Gendiagnostikgesetz vom 31.07.2009, BGBl. I 2529, zuletzt geändert durch Art. 15 Abs. 4 des Gesetzes v. 04.05.2021, BGBl. I 882.

³ Zum Recht auf informationelle Selbstbestimmung aus ethischer Perspektive siehe Jungkunz et al., Kap. 4.

⁴ Stockter, in: Prütting, MedR, § 8 GenDG Rn. 12; Kern, in: Kern, GenDG, § 9 Rn. 1.

⁵ Richtlinie der GEKO für die Anforderungen an die Inhalte der Aufklärung, 963 (964 f.).

⁶ Richtlinie der GEKO für die Anforderungen an die Inhalte der Aufklärung, 963 (965 f.).

schwierig gestaltet, umfassend über die gesamte denkbare Tragweite der Untersuchung oder gar jedes einzelne erzielbare Untersuchungsergebnis zu informieren (Molnár-Gábor, 2017: 168 ff.; EURAT, 2015: 70). Mögliche Zufallsbefunde sind kaum abschätzbar und reichen von der Anlageträgerschaft für bestimmte Erkrankungen, über Prädispositionen bis hin zu präzisen prädiktiven Aussagen zu späteren Krankheitsausbrüchen und -verläufen.

Es gilt daher, eine praktikable Lösung zu finden, die garantiert, dass die Patient*innen trotz der Fülle an genetischen Informationen eine Vorstellung von der Untersuchung und ihren Folgen erhalten und hierauf eine informierte Einwilligung stützen können. Die Aufklärung muss selbstverständlich mehr Aspekte umfassen, als es bei einer gezielten genetischen Untersuchung der Fall wäre. So sollte darauf aufmerksam gemacht werden, dass um ein Vielfaches mehr Informationen entstehen, als zu Diagnosezwecken benötigt werden. Es dürfte aber ausreichen, die Fülle dieser Informationen exemplarisch durch eine Kategorisierung möglicher Ergebnisse zu konkretisieren (z. B. Krankheiten, die therapierbar/nicht therapierbar sind) (Molnár-Gábor et al., 2014: 81/92; Fleischer, 2018: 86 f.; Duttge et al., 2019: 129/139). Auf diese Weise kann in die Einwilligung der Patient*innen auch einfließen, ob und über welche Art von Zufallsbefunden sie informiert werden möchten. Hervorzuheben ist jedoch, dass eine spezifische Aufklärung über sämtliche erzielbare Ergebnisse einer Genomsequenzierung weder praktisch möglich noch rechtlich notwendig ist (Rüping et al., 2022: 663/665 ff.; GEKO, 2013: 17). Nach der Rechtsprechung des Bundesgerichtshofs (BGH) geht es lediglich darum, dass die Patient*innen „im Großen und Ganzen“ verstanden haben, wozu sie ihre Einwilligung erteilen.⁷ Optional kann der Untersuchungszweck kraft Einwilligung auch auf die Feststellung weiterer genetischer Eigenschaften als die, die für die originäre Diagnose von Relevanz sind, ausgeweitet werden. In diesem Fall würden einige zusätzliche Gensequenzen analysiert, die sonst mangels Relevanz für die originäre Diagnostik nicht beachtet worden wären. Dann läge allerdings eine eigene genetische Untersuchung vor, für die alle Anforderungen an eine Aufklärung, Einwilligung und Beratung erfüllt sein müssten. Handelte es sich um eine prädiktive genetische Untersuchung, wäre auch das Angebot einer genetischen Beratung (§ 10 GenDG) verpflichtend (Rüping et al., 2022: 663/665).

5.1.2 Ergebnisse der Untersuchung

Die Genomsequenzierung zielt auf Erkenntnisse für den angestrebten Diagnosezweck. Als originäres Ziel und Anlass für die Untersuchung sind sie in die Aufklärung und Einwilligung der Patient*innen eingeflossen. Liegt diese Einwilligung wirksam vor und wurde nicht widerrufen, können die Ergebnisse den Patient*innen gemäß § 11 Abs. 1 und 4 i. V. m. § 8 Abs. 1 und 2 GenDG mitgeteilt werden.

Wie bereits ausgeführt (5.1.1), lassen sich aus einer Genomsequenzierung indes auch zahlreiche andere diagnostische Ergebnisse deduzieren, die für die Patient*innen von Interesse sein könnten. Nach den zitierten Bestimmungen des GenDG dürfen diese jedoch nur mitgeteilt werden, wenn diese Mitteilung von einer wirksamen Einwilligung umfasst ist. Durch die einwilligungsbedingte Begrenzung der Untersuchung müssen deshalb alle ande-

⁷ BGH, NJW 1977, 337; 1971, 1887; 1984, 1397 (1398); 2000, 1784 (1786).

ren überschüssigen Informationen ausgeblendet bleiben (Rüping et al., 2022: 663/666). Zum Schutz des sog. Rechts auf Nichtwissen, das aus dem informationellen Selbstbestimmungsrecht fließt, dürfen den Patient*innen mithin keine genetischen Informationen aufgedrängt werden. Eine Offenbarung von Zufallsbefunden trotz ursprünglich verweigerter Einwilligung in deren Mitteilung kann – wenn überhaupt – nur ausnahmsweise bei schweren, aber behandelbaren oder die Allgemeinheit gefährdenden Krankheiten in Betracht kommen.⁸ Hat die ursprüngliche Aufklärung das Thema Zufallsbefunde pflichtwidrig ausgespart, kann unter diesen engen Voraussetzungen (schwere behandelbare oder die Allgemeinheit gefährdende Krankheit) möglicherweise von einer hypothetischen Einwilligung in die Mitteilung des Zufallsbefunds ausgegangen werden.⁹

5.1.3 Aufbewahrung der Ergebnisse genetischer Untersuchungen

Gemäß § 12 Abs. 1 GenDG sind die Ergebnisse genetischer Untersuchung für zehn Jahre aufzubewahren. Zu Recht wird kritisiert, dass dieser Aufbewahrungszeitraum wegen der Aussagekraft genetischer Daten sehr kurz bemessen ist (Vossenkuhl, 2013: 140). Die Ergebnisse können noch viele Jahre später für die Patient*innen selbst oder ihre Angehörigen zu Diagnose- oder Therapiezwecken von Nutzen sein. Sie sind zudem schnell verfügbar und können dazu beitragen, erneute Sequenzierungskosten zu verhindern. Eine längere Aufbewahrung der Untersuchungsergebnisse ist gemäß § 12 Abs. 1 S. 3 GenDG immerhin möglich, wenn andernfalls schutzwürdige Interessen der Patient*innen beeinträchtigt würden oder sie eine solche verlangen. Besteht Hoffnung, dass die Patient*innen von künftigen Forschungsergebnissen profitieren könnten, wäre es vor diesem Hintergrund am einfachsten, ihre Einwilligung für eine längere Aufbewahrung der Ergebnisse einzuholen. Andernfalls muss die verantwortliche ärztliche Person zu der Bewertung kommen, dass mit der Vernichtung der Untersuchungsergebnisse schutzwürdige Interessen der Patient*innen beeinträchtigt werden und deshalb von ihrer Vernichtung abzusehen ist.

5.2 Schutz der Genomdaten durch datenschutzrechtliche Vorschriften

Laut § 64e Abs. 2 S. 2 SGB V sind neben den Vorschriften des GenDG auch die datenschutzrechtlichen Vorschriften zu beachten. Diese Vorschriften sind vor allem für die Verarbeitung von Genomdaten zu Forschungszwecken relevant, da die Verarbeitung zu medizinischen Zwecken umfassend durch das GenDG geregelt wird. Bei den Daten aus einer Genomsequenzierung handelt es sich um genetische Daten i. S. d. Art. 4 Nr. 13 der Datenschutzgrundverordnung (DSGVO)¹⁰ und deshalb um solche, die unter die besondere Kategorie personenbezogener Daten i. S. d. Art. 9 Abs. 1 DSGVO fallen.

5.2.1 Verarbeitung zu Forschungszwecken

Die Verarbeitung genomischer Daten zu Forschungszwecken kann einerseits durch eine

⁸ So die – freilich nicht unumstrittene – Auffassung von Duttge et al., 2019: 129/139; Siehe ferner Empfehlungen der BMBF-Projektgruppe „Recht auf Nichtwissen“ (Duttge/Lenk, 2016: 402).

⁹ EURAT, 2023: 45 für den Fall einer Gesundheitsgefährdung für Leib und Leben bei nicht erfolgter Aufklärung der Eltern von minderjährigen Kindern.

¹⁰ Verordnung (EU) 2016/679 des europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG, ABLEU Nr L 119 v. 04.05.2016.

ausdrückliche Einwilligung (Art. 9 Abs. 2 lit. a DSGVO) der betroffenen Person legitimiert sein; andererseits existieren Normen, die eine Verarbeitung zu Forschungszwecken auch ohne Einwilligung des Betroffenen erlauben (sog. gesetzliche Verarbeitungstatbestände – Art. 9 Abs. 2 lit. j DSGVO).

Bei forschungsbezogenen Datenverarbeitungen auf der Grundlage einer Einwilligung stellt sich die Frage, ob sich diese auf mehrere noch nicht im einzelnen konkretisierte Forschungsvorhaben beziehen kann (Broad Consent). Denn nach Art. 6 Abs. 1 lit. a DSGVO muss sich die Einwilligung auf „einen oder mehrere bestimmte Zwecke“ beziehen (Zweckbindungsgrundsatz); und auch Art. 9 Abs. 2 lit. a DSGVO sieht vor, dass die betroffene Person in die Verarbeitung ihrer genomischen Daten „für einen oder mehrere festgelegte Zwecke“ eingewilligt hat. Da der Forschungszweck zum Zeitpunkt der Datenerhebung oftmals noch nicht vollständig determiniert werden kann, sieht DSGVO-Erwägungsgrund Nr. 33, der zur Auslegung der einschlägigen Normen herangezogen werden kann, eine Lockerung dieses Zweckbindungsgrundsatzes dergestalt vor, dass eine Einwilligung für „bestimmte Forschungsbereiche“ möglich sein soll. In Teilen der rechtswissenschaftlichen Literatur wird diese Möglichkeit eines „Broad Consent“ allerdings kritisiert, da er mangels hinreichender Informiertheit keine effektive Ausübung des Rechts auf informationelle Selbstbestimmung erlaube. Diese Vorgehensweise müsse deshalb auf eng umrissene Ausnahmefälle beschränkt bleiben (Fröhlich/Spieker, 2022: 346/352 f.; Bronner/Wiedemann, 2023: 77/81).¹¹ Richtigerweise sollte die Einwilligung in die Verarbeitung personenbezogener Daten für näher bestimmte Forschungsbereiche indes grundsätzlich genügen (Schreiber, 2019: 111 f. m.w.N.).¹² Die grundrechtlich geschützte Dispositionsbefugnis über eigene Daten muss auch das Recht beinhalten, sie für weiter gefasste Zwecke freizugeben, solange dies in Kenntnis von Informations- bzw. Detaillierungslücken erfolgt (Fleischer, 2018: 287 m.w.N.) und die Einwilligung im Übrigen jederzeit widerrufen werden kann.

Auf der Grundlage des Art. 9 Abs. 2 lit. j DSGVO haben die nationalen Gesetzgeber zudem gesetzliche Verarbeitungstatbestände geschaffen. Sowohl auf nationaler als auch auf unionaler Ebene genießt die Forschungsfreiheit grundrechtlichen Schutz (Art. 5 Abs. 3 GG, Art. 13 GRCh). Ihr gegenüber steht das ebenfalls mitgliedstaatlich- wie unionsrechtlich geschützte Recht auf informationelle Selbstbestimmung (Art. 2 Abs. 1 i. V. m. Art. 1 Abs. 1 GG, Art. 8 GRCh) der betroffenen Personen. Bei Rückgriff auf eine gesetzliche Forschungsklausel erfolgt die Legitimation der Verarbeitung personenbezogener Daten zu wissenschaftlichen Zwecken deshalb durch eine Abwägungsentscheidung, die unter bestimmten Voraussetzungen zugunsten des Forschungsinteresses auszufallen vermag. Nach Art. 89 Abs. 1 DSGVO sind in diesem Fall jedoch stets „geeignete Garantien für die Rechte und Freiheiten der betroffenen Personen“ zu gewährleisten, beispielsweise durch eine Pseudonymisierung der verwendeten Daten.¹³

¹¹ Kritisch zur Möglichkeit eines Broad Consent auch Jungkunz et al., 2022: 92.

¹² Die durch das BMBF geförderte Medizin-Informatik-Initiative hat insoweit 2020 eine Mustereinwilligung erarbeitet (siehe unter: <https://www.medizininformatik-initiative.de/de/mustertext-zur-patienteneinwilligung> [28.11.2023]). Diese wurde auch von der Konferenz der unabhängigen Datenschutzbeauftragten des Bundes und der Länder akzeptiert (siehe Pressemitteilung der DSK v. 24.04.2020).

¹³ Ähnlich die Formulierung in Art. 9 Abs. 2 lit. j DSGVO.

5.2.2 „Flickenteppich“ der datenschutzrechtlichen Forschungsklauseln

Dieser datenschutzrechtliche Status quo wird aufgrund der Anwendbarkeit unterschiedlicher Regelungssysteme zu Recht als unübersichtlich kritisiert. Die DSGVO gibt einen europäischen Rahmen vor, der an vielen Stellen durch nationale Bestimmungen konkretisiert wird. In Deutschland kommt es für die Frage, welche ausgestaltenden nationalen Regelungen anwendbar sind, auf die Frage an, wer die Daten verarbeitet und wo diese Verarbeitung erfolgt. Je nachdem müssen die Vorschriften des Bundesdatenschutzgesetzes (BDSG) oder die jeweiligen Vorschriften der 16 Landesdatenschutzgesetze beachtet werden. In diesen Rechtsgrundlagen sind die vorstehend erwähnten Forschungsklauseln nicht einheitlich geregelt, sondern unterscheiden sich in ihren Voraussetzungen. Während § 27 Abs. 1 S. 1 BDSG fordert, dass die Forschungsinteressen die Interessen der betroffenen Person „erheblich“ überwiegen müssen, verzichtet § 13 Abs. 1 LDSG BW¹⁴ auf dieses Erheblichkeitserfordernis. § 17 Abs. 1 DSG NRW¹⁵ hinwiederum sieht vor, dass schutzwürdige Belange der betroffenen Personen nicht überwiegen dürfen und § 13 NSDG¹⁶ verlangt, dass schutzwürdige Interessen nicht entgegenstehen oder das öffentliche Interesse an der Forschung überwiegt. Diese Rechtszersplitterung sowie die Tatsache, dass die Voraussetzungen der Abwägung zwischen der Forschung und dem Schutz personenbezogener Daten durch die Gesetzgeber nicht weiter konturiert wurde, führen zu Rechtsunsicherheit bei der Normanwendung (Weichert, 2020: 18 ff.; Specht-Riemenschneider/Radbruch, 2021: 1358/1359; Buchner, 2022: 555/559 ff.; Jungkuntz et al., 2022: 95 ff.). Das hat zur Folge, dass Forscher*innen eher selten darauf zurückgreifen (Graf von Kielmansegg, 2022: 341/342; AKEK, 2023: 1 f.), obschon diese Zurückhaltung in nicht wenigen Fällen als unbegründet erscheint.

Vor diesem Hintergrund arbeitet die EU an einer Verbesserung der rechtlichen Situation: So hat die EU-Kommission im Mai 2022 einen Entwurf für einen Europäischen Gesundheitsdatenraum (European Health Data Space-Entwurf – EHDS-E) veröffentlicht. Ziel dieser Initiative ist u. a. die einfachere Primär- und Sekundärnutzung von Gesundheitsdaten zu Forschungszwecken.¹⁷ Akteure des Gesundheitssystems sollen verpflichtet werden, bestimmte Kategorien elektronischer Gesundheitsdaten, darunter auch genetische und genomische Daten,¹⁸ zur Verfügung zu stellen. Die Genehmigung zur Nutzung der Daten soll im Weiteren durch eine Zugangsstelle innerhalb einer kurzen Frist erfolgen.¹⁹ Als möglicher Verarbeitungszweck, den Antragssteller*innen verfolgen können, wird auch die wissenschaftliche Forschung im Bereich des Gesundheits- oder Pflegesektors genannt.²⁰ Die geplante Verordnung soll die Anforderungen der DSGVO nicht aufheben, sondern vielmehr eine Rechtsgrundlage zur Datenverarbeitung i. S. d. Art. 9 Abs. 2 lit. j DSGVO schaffen. Auch sollen alle anderen Voraussetzungen, die die DSGVO an die Verarbeitung personenbezogener

¹⁴ Landesdatenschutzgesetz Baden-Württemberg.

¹⁵ Datenschutzgesetz Nordrhein-Westfalen.

¹⁶ Niedersächsisches Datenschutzgesetz.

¹⁷ Vgl. Erwägungsgrund Nr. 1 EHDS-E.

¹⁸ Art. 33 Abs. 1 lit. e EHDS-E.

¹⁹ Art. 46 Abs. 1 und 3 EHDS-E.

²⁰ Art. 34 Abs. 1 lit. e EHDS-E.

ner (unter Einschluss genomischer) Daten stellt, unberührt bleiben.²¹ Das Gesetzesvorhaben befindet sich noch im Anfangsstadium, weshalb abzuwarten bleibt, wann und wie es Gestalt annehmen wird. Abseits von Einwilligungserklärungen eröffnet der Entwurf zwar das Potenzial für eine einfachere und einheitlichere Nutzung von Genomdaten zu Forschungszwecken (Tschammler/Jecker, 2022: 01265; AKEK, 2023: 2), dennoch könnten Unsicherheiten, etwa in Gestalt des Verhältnisses zur DSGVO (Raji, 2023: 3/6 f.), fortbestehen. Anknüpfend an den EHDS-E existiert auf nationaler Ebene seit März 2023 ein erster Entwurf des Bundesministeriums für Gesundheit (BMG) für ein Gesundheitsdatennutzungsgesetz (GDNG).

5.3 Fazit: Das Potenzial von Genomdaten im Rahmen rechtlicher Vorgaben

Genomische Untersuchungen können mit Einwilligung der Patient*innen für Diagnostik- und Therapiezwecke eingesetzt werden. Hierfür müssen die strengen Anforderungen des GenDG beachtet werden, wobei insbesondere die ordnungsgemäße Aufklärung eine Herausforderung darstellt. Über zusätzliche Befunde, die neben den ursprünglich intendierten Ergebnissen zu Tage treten, dürfen Patient*innen nur informiert werden, wenn sie zuvor zumindest in die Mitteilung der einschlägigen Befundkategorie eingewilligt haben. Daneben besteht die Möglichkeit, den Untersuchungszweck von Anfang an auf weitere Merkmale jenseits des originären Diagnosezwecks zu erweitern.

Die Aufbewahrungszeit für die Untersuchungsergebnisse beträgt grundsätzlich zehn Jahre. Mit Einwilligung der Patient*innen oder wenn die Aussicht besteht, dass Patient*innen von neuen Forschungserkenntnissen profitieren könnten, ist auch eine längere Aufbewahrung möglich.

Die Forschung mit Genomdaten kann einerseits auf der Grundlage einer Einwilligung der Patient*innen erfolgen, andererseits auf der Grundlage gesetzlicher Forschungsklauseln. Für die Einwilligung kann auf das Konzept des „Broad Consent“ zurückgegriffen werden, weshalb sich die Einwilligung nicht auf konkrete Forschungsvorhaben beziehen muss, sondern die Daten breiter für eingegrenzte Forschungsbereiche genutzt werden können. Stützt sich die Verarbeitung genomischer Daten auf gesetzliche Forschungsklauseln, bestehen Rechtsunsicherheiten, die u. a. in der Rechtszersplitterung der datenschutzrechtlichen Regelungen ihre Ursache finden.

Die Möglichkeit zur einfacheren Nutzung von Gesundheitsdaten im Rahmen medizinischer Forschung steht derzeit auf der politischen Agenda. Es bleibt abzuwarten, ob die insoweit geplanten Gesetze auf europäischer und nationaler Ebene eine Forschung auf diesem Gebiet erleichtern werden.

²¹ Erwägungsgrund Nr. 37 EHDS-E.

5.4 Literaturverzeichnis

- AKEK = Arbeitskreis Medizinischer Ethikkommissionen (2023):** Stellungnahme zur Sekundärnutzung von Patientinnen- und Patienten-Daten zu Forschungszwecken in Deutschland, 09.08.2023. Unter: https://www.akek.de/wp-content/uploads/Stellungnahme-EHDS_V7.1.pdf [28.11.2023].
- Bronner, P./Wiedemann, F. (2023):** Rechtmäßigkeit der Datenverarbeitung bei wissenschaftlicher Forschung an staatlichen Hochschulen. In: ZD 2: 77-85.
- Buchner, B. (2022):** Forschungsdaten effektiver nutzen. In: DuD 9: 555-560. DOI: 10.1007/s11623-022-1658-8.
- DSK = Datenschutzkonferenz (2020):** Pressemitteilung der Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder vom 27.04.2020. Unter: https://www.datenschutzkonferenz-online.de/media/dskb/20200427_Beschluss_MII.pdf [14.12.2023].
- Duttge, G. et al. (2019):** Die genomische Hochdurchsatzsequenzierung. In: Duttge, G. et al. (Hrsg.): Next Generation Medicine. Mohr Siebeck, Tübingen: 129-146.
- Duttge, G./Lenk, C. (2019):** Empfehlungen zum anwendungspraktischen Umgang mit dem „Recht auf Nichtwissen“. Ergebnisse einer rechtsethischen Grundlagenanalyse auf erfahrungswissenschaftlicher Basis in den Anwendungsfeldern von Humangenetik und Psychiatrie. In: Duttge, G./Lenk, C. (Hrsg.): Das sogenannte Recht auf Nichtwissen. Brill, Leiden: 253-269. DOI: 10.30965/9783957437792_018.
- EURAT = Projektgruppe EURAT (2015):** „Ethische und Rechtliche Aspekte der Ganzgenomsequenzierung des menschlichen Genoms“: Stellungnahme Eckpunkte für eine Heidelberger Praxis der Ganzgenomsequenzierung. 2. Auflage, Heidelberg. Unter: https://www.uni-heidelberg.de/md/totalsequenzierung/stellungnahme_2.pdf [14.12.2023].
- EURAT = Projektgruppe EURAT (2023):** „Ethische und rechtliche Aspekte der Translationalen Medizin“: Stellungnahme zur Rückmeldung genetischer Zusatzbefunde Minderjähriger. Heidelberg. Unter: <https://journals.ub.uni-heidelberg.de/index.php/forum-mk/issue/view/5962> [14.12.2023].
- Fleischer, H. (2018):** Rechtliche Aspekte der Systemmedizin. LIT Verlag, Berlin.
- Fröhlich, W./Spiecker, I. (2022):** Die breite Einwilligung (Broad Consent) in die Datenverarbeitung zu medizinischen Forschungszwecken – der aktuelle Irrweg der MII. In: GesR 6: 346-353. DOI: 10.9785/gesr-2022-210605.
- GEKO = Gendiagnostik-Kommission (2013):** Erster Tätigkeitsbericht der Gendiagnostik-Kommission. Unter: https://www.rki.de/DE/Content/Kommissionen/GendiagnostikKommission/Taetigkeitsbericht/Taetigkeitsbericht_01.pdf?__blob=publicationFile [14.12.2023].
- Graf von Kielmansegg, S. (2022):** § 27 BDSG und die medizinische Forschung. In: GesR 6: 341-346. DOI: 10.9785/gesr-2022-210604.
- Jungkunz, M. et al. (2022):** Stellungnahme der Etablierung der sekundären Forschungsnutzung von Behandlungsdaten in Deutschland. Ergebnisse des Verbundprojekts LinCDat, Heidelberg. Unter: <https://journals.ub.uni-heidelberg.de/index.php/forum-mk/article/view/91697> [26.10.2023].
- Kern, B.-R. (2021):** Gendiagnostikgesetz Kommentar. C.H.Beck, München.
- Medizin-Informatik-Initiative (2020):** Mustereinwilligung (16.04.2020). Unter: <https://www.medizininformatik-initiative.de/de/mustertext-zur-patienteneinwilligung> [28.11.2023].
- Molnár-Gábor, F. et al. (2014):** Das Ende der Zufallsbefunde in der genetischen Diagnostik: Die Ganzgenomsequenzierung und die Erwartbarkeit zusätzlicher Befunde. In: Zeitschrift für Medizin-Ethik-Recht 2: 81-105.
- Molnár-Gábor, F. (2017):** Die internationale Steuerung der Biotechnologie am Beispiel des Umgangs mit neuen genetischen Analysen. Duncker & Humblot, Berlin.
- Prütting, D. (2022):** Medizinrecht Kommentar. 6. Auflage. Luchterhand, Köln.
- Raji, B. (2023):** Datenräume in der Europäischen Datenstrategie am Beispiel des European Health Data Space. In: ZD 1: 3-8.
- Rüping, U. et al. (2022):** Ärztliche Aufklärungspflichten über die Möglichkeit von Nebenbefunden in der genomischen Medizin. In: MedR 40: 663-667. DOI: 10.1007/s00350-022-6273-z.
- Schreiber, M. (2019):** Die medizinische Forschung mit abgetrennten Körpersubstanzen Minderjähriger. LIT Verlag, Münster.
- Specht-Riemenschneider, L./Radbruch, A. (2021):** Forschung braucht Daten. In: Deutsches Ärzteblatt 27-28: A 1358-1361.
- Tschammler, D./Uecker, P. (2022):** Verordnungsentwurf zur Schaffung eines Europäischen Gesundheitsdatenraums – erster Überblick mit Fokus auf die Nutzung von Gesundheitsdaten für Forschungs- und Entwicklungsvorhaben. In: ZD-Aktuell: 01265.
- Vossenkuhl, C. (2013):** Der Schutz genetischer Daten. Springer, Heidelberg.
- Weichert, T. (2020):** Die Forschungsprivilegierung in der DS-GVO. In: ZD 1: 18-24.

Ausblick: Der Weg in eine zukunftsorientierte humane Genomdiagnostik und Genomforschung

Jörn Walter

Das Modellvorhaben 2024 zur Genomsequenzierung markiert eine Zeitenwende in der humanen Genomforschung in Deutschland.¹ Es besteht die große Chance, dass Deutschland hier erstmalig eine Vorreiterrolle in der Genomforschung einnehmen kann, indem einerseits eine direkte Verknüpfung von Genomsequenzierungen mit der Krankheitsversorgung erfolgt und andererseits diese umfassenden Genomdaten (zustimmungsabhängig) für die Erforschung und Aufklärung komplexer Erkrankungen wie auch für die Entwicklung neuer Medikamente und Diagnoseverfahren nutzbar gemacht werden. Hieraus ergeben sich Synergien und weitreichende Möglichkeiten für eine moderne, zukunftsorientierte Genomdiagnostik und funktionelle medizinische Genomforschung.

Diese Entwicklung erfolgt zum richtigen Zeitpunkt. Durch den Einsatz neuester Technologien (vor allem der Sequenziergeräte der 3. Generation) können Genome präziser und umfassender als bislang analysiert werden. Dabei können genetische wie auch epigenetische Veränderungen gleichzeitig erfasst werden. Dies erweitert die Möglichkeiten der Diagnostik und der Forschung um zwei wesentliche Parameter. Darüber hinaus werden vernetzte Strukturen für eine Dateninfrastruktur aufgebaut, die eine koordinierte deutschlandweite sichere Archivierung und Datennutzung ermöglichen und so die Möglichkeiten der Genomforschung in Deutschland erweitern.²

Die koordinierte Archivierung von Genomen ermöglicht erstmalig einen umfassenden bevölkerungsbezogenen Abgleich und verspricht so eine deutliche Verbesserung der Identifikation krankheitsverursachender bzw. -fördernder Genvarianten. Die geplante und dringend notwendige Vernetzung eines deutschen Archivs mit anderen nationalen Archiven auf europäischer Ebene wird diese Möglichkeiten noch erheblich erweitern.³

Auf dem Weg zu einer qualifizierten Bewertung genetischer Veränderungen wird es essenziell sein, weitere funktionelle Genomdaten z. B. aus der Einzelzellforschung, der Epigenomik, der Proteomik und Transkriptomik in die Analysen einzubeziehen. Ein mit solch zusätzlichen Daten ausgestattetes Datenarchiv böte die Basis für einen nationalen humanen Genomatlas, d. h. ein die Interpretation unterstützendes Kompendium bevölkerungsbezogener Genom-

1 Siehe unter: <https://www.bundesgesundheitsministerium.de/en/en/international/european-health-policy/genomde-en> [18.03.2024]; <https://www.aerzteblatt.de/nachrichten/144447/Modellvorhaben-zur-Genomsequenzierung-soll-2024-starten> [18.03.2024]; <https://www.genom.de> [18.03.2024].

2 Siehe unter: <https://www.bundesgesundheitsministerium.de/en/en/international/european-health-policy/genomde-en> [18.03.2024]; <https://www.ghga.de/about-us/mission> [18.03.2024].

3 Siehe unter: <https://www.youtube.com/watch?v=wYDJXEkSjw> [18.03.2024].

daten. In diesen Atlas werden idealerweise auch für die Forschung relevante klinische „Phänom“-Daten bei Einwilligung und anonym nutzbar mit eingepflegt. Dieser Atlas wird so tiefer reichende Analysemöglichkeiten für die klinische Diagnostik einzelner Patient*innen ermöglichen. Die Daten des Atlas können dann in einem datensicheren Raum, z. B. einer geschützten Cloud, genutzt werden, um hier z. B. mithilfe KI-gestützter Verfahren gesamtgenomische Aspekte für eine personenbezogene Diagnose und Behandlung zu entschlüsseln.

Die Grundlage für all diese Optionen ist, dass personenbezogene Genomdaten freiwillig und auf der Basis einer Aufklärung (Informed Consent) gewonnen werden und in kontrollierten, nicht kommerziellen Datenarchiven sicher gelagert sind. Die geschützten Netzwerke müssen es ermöglichen, die Daten standardisiert und nachhaltig zu lagern. Die Sicherheit wird durch verschiedene Ebenen der Verschlüsselung und eine abgeschirmte Netzwerkarchitektur gewährleistet. Gleichzeitig sollte die Nutzbarkeit der Daten nicht eingeschränkt sein, um ihren Mehrwert für die biomedizinische Forschung abschöpfen zu können. Hierzu bedarf es klarer Regeln wie z. B. kontrollierte Zugangs- und Nutzungsbedingungen, aber auch Regeln und Kontrollen für den Umgang mit den Daten. Dies erfordert die begleitende Ausarbeitung datenrechtlicher und ethischer Rahmenbedingungen.

Für die Herstellung, Bearbeitung und Bewertung der komplexen genetischen Daten wird neues interdisziplinär ausgebildetes Fachpersonal benötigt. Die Ausbildung entsprechender genomischer Fachkompetenzen, insbesondere im Bereich der fachärztlichen Ausbildung (Humangenetik, Onkologie u. a.) und der medizinischen Bioinformatik (Datenbankmanagement, KI-gestützte Analysen), muss schnell und entschieden ausgebaut werden.

Es ist an der Zeit, die Notwendigkeit für eine moderne, verantwortungsvolle und zukunftsweisende Genomdiagnostik und -forschung breiter als bisher zu vermitteln und zu diskutieren. Die Erkenntnisse der Genomforschung der letzten 20 Jahre haben das Gesamtbild des Genoms und seine Bedeutung für komplexe physiologische, biochemische und medizinische Vorgänge grundlegend erweitert. Dieses Wissen sollte für die Diagnose und Behandlung von genetisch bedingten Erkrankungen, aber auch für die allgemeine Gesundheitsvorsorge genutzt und eingesetzt werden. Dabei ist es von großer Wichtigkeit, einen interdisziplinären Diskurs von Patient*innen und Patient*innenvertretungen, Kliniker*innen, Krankenversorger*innen, Wissenschaftler*innen und behördlichen Aufsichtsgremien auf verschiedenen Ebenen zu fördern, um so inhaltliche, rechtliche und ethische Aspekte prozessbegleitend zu diskutieren und ggf. Anpassungen vorzunehmen.

Auf dem Weg in eine moderne Genomdiagnostik braucht es zudem den kontinuierlichen und begleitenden Austausch mit der Öffentlichkeit/Gesellschaft. Die AG *Gentechnologiebericht* möchte mit dieser Broschüre hierzu einen Beitrag leisten. Eine erfolgreiche Umsetzung und Akzeptanz des Modellvorhabens wird davon abhängen, dass eine transparente Aufklärungs- und Öffentlichkeitsarbeit die gesellschaftliche Akzeptanz für die Genomdiagnostik und Genomforschung steigern wird und (noch) vorhandene Ängste vor den Gefahren eines „gläsernen Menschen“ genommen werden. Eine zentrale Voraussetzung für das Vertrauen von Patient*innen ist dabei die langfristige Gewährleistung von Datensicherheit und Datenschutz in technischer, rechtlicher und ethischer Hinsicht.

Autor*innen/Authors

Dr. Nadine Biernath

Funktionsoberärztin/Senior Physician interdisziplinäre Uro-Onkologie, Leiterin/Head klinische Tumorgenetik, Klinik für Urologie, Charité – Universitätsmedizin Berlin

Prof. Dr. Boris Fehse

Leiter/Head der Forschungsabteilung Zell- und Gentherapie, Laborleiter, Klinik für Stammzelltransplantation, Universitätsklinikum Hamburg-Eppendorf

Dr. Nina Gasparoni

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

Dr. Martin Jungkuntz

Wissenschaftlicher Mitarbeiter/Research Associate, Sektion für Translationale Medizinethik, Nationales Centrum für Tumorerkrankungen (NCT), Deutsches Krebsforschungszentrum (DKFZ), Heidelberg

Prof. Dr. Oliver Kohlbacher

Direktor/Director des Instituts für Biomedizinische Informatik, Universität Tübingen; Professor Angewandte Bioinformatik/Professor Applied Bioinformatics, University of Tübingen; Direktor/Director des Instituts für Translationale Bioinformatik, Universitätsklinikum Tübingen

Prof. Dr. Ralf Müller-Terpitz

Direktor/Director des Instituts für Deutsches, Europäisches und Internationales Medizinrecht, Gesundheitsrecht und Bioethik, Universität Mannheim

Prof. Dr. Olaf Rieß

Direktor/Director des Instituts für Medizinische Genetik und Angewandte Genomik, Universitätsklinikum Tübingen

Dr. Christoph Schickhardt

Wissenschaftlicher Mitarbeiter/Research Associate, Sektion für Translationale Medizinethik, Nationales Centrum für Tumorerkrankungen (NCT), Deutsches Krebsforschungszentrum (DKFZ), Heidelberg

Prof. Dr. Thorsten Schlomm

Direktor/Director der Klinik für Urologie, Charité – Universitätsmedizin Berlin

Prof. Dr. Julia Schulze-Hentrich

Professorin/Professor für Genetik, Universität des Saarlandes

Hannah Lilly Straub

Wissenschaftliche Mitarbeiterin/Research Associate, Universität Mannheim

Prof. Dr. Jörn Walter

Professor für Genetik/Professor of Genetics, Universität des Saarlandes

Prof. Dr. Dr. Eva C. Winkler

Professorin für Translationale Medizinethik/Professor for Translational Medical Ethics, Medizinische Fakultät der Universität Heidelberg; Direktorin/Director des Nationalen Centriums für Tumorerkrankungen (NCT), Heidelberg

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

Prof. Dr. Boris Fehse

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

Prof. Dr. Jörn Walter

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

Prof. Dr. Sina Bartfeld

Technische Universität Berlin, Medizinische Biotechnologie, Institut für Biotechnologie/
Der Simulierte Mensch

Prof. Dr. Stephan Clemens

Universität Bayreuth, Lehrstuhl Pflanzenphysiologie

Prof. Dr. Tobias J. Erb

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

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

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

Dr. Jürgen Hampel

Universität Stuttgart, Institut für Sozialwissenschaften und Zentrum für interdisziplinäre
Risiko- und Innovationsforschung (ZIRIUS)

Prof. Dr. Martin Korte

Technische Universität Braunschweig, Zoologisches Institut und Institut für Zelluläre
Neurobiologie; Helmholtz-Institut für Infektionsforschung Braunschweig

Prof. Dr. Ralf Müller-Terpitz

Universität Mannheim, Institut für Deutsches, Europäisches und Internationales Medizin-
recht, Gesundheitsrecht und Bioethik

Prof. Dr. Stefan Mundlos

Charité Berlin, Institut für Medizinische Genetik und Humangenetik; Max-Planck-Institut
für molekulare Genetik, Berlin

Prof. em. Dr. Jens Reich

Max-Delbrück-Centrum für Molekulare Medizin und Humboldt-Universität, Berlin

Prof. Dr. Silke Schicktanz

Universitätsmedizin Göttingen, Institut für Ethik und Geschichte der Medizin

Prof. Dr. Dr. Eva C. Winkler

Ruprecht-Karls-Universität Heidelberg, Translationale Medizinethik; Universitätsklinikum
Heidelberg

Univ.-Prof. Dr. Martin Zenke

Universitätsklinikum RWTH Aachen, Medizinische Klinik IV

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

BÜCHER/BOOKS

- Fehse, B. et al. (Hrsg.) (2023):** Gen- und Zelltherapie 2023. Forschung, klinische Anwendung und Gesellschaft. Springer, Berlin. Unter: <https://link.springer.com/book/10.1007/978-3-662-67908-1> [08.02.2024].
- Fehse, B. et al. (Hrsg.) (2021):** Fünfter Gentechnologiebericht. Sachstand und Perspektiven für Forschung und Anwendung. Nomos, Baden-Baden. Unter: <https://www.nomos-elibrary.de/10.5771/9783748927242/fuenfter-gentechnologiebericht> [08.02.2024].
- Bartfeld, S. et al. (Hrsg.) (2020):** Organoide. Ihre Bedeutung für Forschung, Medizin und Gesellschaft. Nomos, Baden-Baden. Unter: <https://www.nomos-elibrary.de/10.5771/9783748908326/organoide> [08.02.2024].
- Hucho, F. et al. (Hrsg.) (2018):** Vierter Gentechnologiebericht. Bilanzierung einer Hochtechnologie. Nomos, Baden-Baden. Unter: <https://www.nomos-elibrary.de/10.5771/9783845293790/vierter-gentechnologiebericht> [08.02.2024].
- Zenke, M. et al. (Hrsg.) (2018):** Stammzellforschung. Aktuelle wissenschaftliche und gesellschaftliche Entwicklungen. Nomos, Baden-Baden. Unter: <https://www.nomos-elibrary.de/10.5771/9783845287720/stammzellforschung> [08.02.2024].
- Walter, J./Hümpel, A. (Hrsg.) (2017):** Epigenetik. Implikationen für die Lebens- und Geisteswissenschaften. Nomos, Baden-Baden. Unter: <https://www.nomos-elibrary.de/10.5771/9783845270838/epigenetik> [08.02.2024].
- Müller-Röber, B. et al. (Hrsg.) (2015):** Dritter Gentechnologiebericht. Analyse einer Hochtechnologie. Nomos, Baden-Baden. Unter: <https://www.nomos-elibrary.de/10.5771/9783845246956/dritter-gentechnologiebericht> [08.02.2024].
- Müller-Röber, B. et al. (Hrsg.) (2013):** Grüne Gentechnologie. Aktuelle wissenschaftliche, wirtschaftliche und gesellschaftliche Entwicklungen. 3. neubearb. u. erg. Aufl. Forum W, Limburg.
- Köchy, K./Hümpel, A. (Hrsg.) (2012):** Synthetische Biologie. Entwicklung einer neuen Ingenieurbiologie? Forum W, Dornburg.
- Fehse, B./Domasch, S. (Hrsg.) (2011):** Gentherapie in Deutschland. Eine interdisziplinäre Bestandsaufnahme. 2. akt. u. erw. Aufl. Forum W, Dornburg.
- Müller-Röber, B. et al. (Hrsg.) (2009):** Zweiter Gentechnologiebericht. Analyse einer Hochtechnologie in Deutschland. Forum W, Dornburg.
- Engelhard, M. et al. (2009):** Genetic Engineering in Livestock. Springer, Berlin, Heidelberg.
- Hucho, F. et al. (2008):** Gentherapie in Deutschland. Eine interdisziplinäre Bestandsaufnahme. Forum W, Dornburg.
- Schmidtke, J. et al. (Hrsg.) (2007):** Gendiagnostik in Deutschland. Status quo und Problemerkundung. Supplement zum Gentechnologiebericht. Forum W, Limburg.
- Müller-Röber, B. et al. (Hrsg.) (2007):** Grüne Gentechnologie. Aktuelle Entwicklungen in Wissenschaft und Wirtschaft. Spektrum, München.
- Wobus, A. M. et al. (Hrsg.) (2006):** Stammzellforschung und Zelltherapie. Stand des Wissens und der Rahmenbedingungen in Deutschland. Supplement zum Gentechnologiebericht. Spektrum, München.
- Hucho, F. et al. (Hrsg.) (2005):** Gentechnologiebericht. Analyse einer Hochtechnologie in Deutschland. Spektrum, München.
- Hucho, F./Köchy, K. (2003):** Materialien für einen Gentechnologiebericht. Grundlagenforschung, Medizinische Anwendung, ökonomische Bedeutung. Spektrum, Heidelberg.
- Köchy, K. et al. (Hrsg.) (2002):** Gentechnologie als Wirtschaftsfaktor. Spektrum, Heidelberg, Berlin.

BROSCHÜREN/BOOKLETS

- Fehse, B. et al. (Hrsg.) (2022):** Im Fokus: RNA. Eine aktuelle Bestandsaufnahme der Arbeitsgruppe Gentechnologiebericht. DOI: 10.17169/refubium-36831.
- IAG Gentechnologiebericht (Hrsg.) (2021):** Fünfter Gentechnologiebericht. Sachstand und Perspektiven für Forschung und Anwendung. Kurzfassung. BBAW, Berlin. Unter: https://edoc.bbaw.de/files/3609/BBAW_Gentechnologiebericht_V_Kurzfassung.pdf [24.04.2024].
- IAG Gentechnologiebericht/German Stem Cell Network (Hrsg.) (2020):** Organoide – von der Stammzelle zur zukunftsweisenden Technologie/Organoids – from stem cells to future technologies. White Paper. Berlin. Unter: https://edoc.bbaw.de/files/3436/BBAW_Whitepaper_GSCN_11_2020.pdf [24.04.2024].
- Walter, J./Schickl, H. (Hrsg.) (2019):** Einzelzellanalyse in Forschung und Medizin. Eine Stellungnahme der interdisziplinären Arbeitsgruppe Gentechnologiebericht. BBAW, Berlin. Unter: https://edoc.bbaw.de/files/3278/BBAW_Einzelzellanalyse_Walter_Schickl.pdf [24.04.2024].

IAG Gentechnologiebericht (Hrsg.) (2018): Vierter Gentechnologiebericht. Bilanzierung einer Hochtechnologie. Kurzfassung. BBAW, Berlin. Unter: https://edoc.bbaw.de/files/2991/BBAW_Broschuere_Gentechnologiebericht_4_Kurzfassung.pdf [24.04.2024].

IAG Gentechnologiebericht (Hrsg.) (2018): Stammzellforschung. Aktuelle wissenschaftliche und gesellschaftliche Entwicklungen. Kurzfassung. BBAW, Berlin. Unter: https://edoc.bbaw.de/files/2990/BBAW_Broschuere_Stammzellforschung_Kurzfassung.pdf [24.04.2024].

IAG Gentechnologiebericht (Hrsg.) (2017): Epigenetik. Implikationen für die Lebens- und Geisteswissenschaften. Kurzfassung. BBAW, Berlin. Unter: https://edoc.bbaw.de/files/2885/2017_BBAW_Epigenetik_Kurzfassung.pdf [24.04.2024].

IAG Gentechnologiebericht (Hrsg.) (2015): Dritter Gentechnologiebericht. Analyse einer Hochtechnologie. Kurzfassung. BBAW, Berlin. Unter: https://edoc.bbaw.de/files/2329/BBAW_DritterGentechnologiebericht_KF_2015.pdf [24.04.2024].

Reich, J. et al. (Hrsg.) (2015): Genomchirurgie beim Menschen. Zur verantwortlichen Bewertung einer neuen Technologie. Analyse der Interdisziplinären Arbeitsgruppe Gentechnologiebericht der Berlin-Brandenburgischen Akademie der Wissenschaften. BBAW, Berlin. Unter: https://edoc.bbaw.de/files/2483/2015_Analyse_GenomchirurgieBeimMenschen.pdf [24.04.2024].

IAG Gentechnologiebericht (Hrsg.) (2013): Grüne Gentechnologie. Aktuelle wissenschaftliche, wirtschaftliche und gesellschaftliche Entwicklungen. Kurzfassung. BBAW, Berlin. Unter: https://www.gentechnologiebericht.de/fileadmin/Gentechnologiebericht/Gruene_Gentechnologie_2013.pdf [24.04.2024].

IAG Gentechnologiebericht (Hrsg.) (2012): Synthetische Biologie. Entwicklung einer neuen Ingenieurbiologie? Kurzfassung. BBAW, Berlin. Unter: https://www.gentechnologiebericht.de/fileadmin/Gentechnologiebericht/Publikationen/Synthetische_Biologie__2012_/Synthetische_Biologie_2012.pdf [24.04.2024].

IAG Gentechnologiebericht (Hrsg.) (2011): Gentherapie in Deutschland. Eine interdisziplinäre Bestandsaufnahme. Kurzfassung. BBAW, Berlin. Unter: www.gentechnologiebericht.de/fileadmin/Gentechnologiebericht/Gentherapie_in_Deutschland_2011.pdf [24.04.2024].

IAG Gentechnologiebericht (Hrsg.) (2009): Zweiter Gentechnologiebericht. Analyse einer Hochtechnologie in Deutschland. Kurzfassung. BBAW, Berlin. Unter: www.gentechnologiebericht.de/fileadmin/Gentechnologiebericht/Zweiter_Gentechnologiebericht_2009.pdf [24.04.2024].

IAG Gentechnologiebericht (Hrsg.) (2007): Gendiagnostik in Deutschland. Status quo und Problemerkundung. Zusammenfassung. BBAW, Berlin. Unter: www.gentechnologiebericht.de/fileadmin/Gentechnologiebericht/Publikationen/Gendiagnostik_in_Deutschland__2007_/Gendiagnostik_in_Deutschland_2007.pdf [24.04.2024].

IAG Gentechnologiebericht (Hrsg.) (2005): Gentechnologiebericht. Analyse einer Hochtechnologie in Deutschland. Kurzfassung. BBAW, Berlin. Unter: www.gentechnologiebericht.de/fileadmin/Gentechnologiebericht/Gentechnologiebericht_2005.pdf [24.04.2024].

SONSTIGE PUBLIKATIONEN/OTHER PUBLICATIONS

Bartfeld, S. et al. (2020): Special Issue: 3D Organoids. In: Journal of Molecular Medicine 99(4). Unter: <https://link.springer.com/journal/109/volumes-and-issues/99-4> [24.04.2024].

Fehse, B. et al. (2018): Debatte 19 – Die Gentechnologie in der Gesellschaft: Von großen Versprechungen, hohen Erwartungen und Missverständnissen. Streitgespräche in den Wissenschaftlichen Sitzungen der Versammlung der Berlin-Brandenburgischen Akademie der Wissenschaften am 01. Dezember 2017. Hg. v. Grötschel, M., Berlin. Unter: https://edoc.bbaw.de/files/2926/BBAW_Debatte_19.pdf [24.04.2024].

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> [24.04.2024].

Ropers, H. H. et al. (2013): Stellungnahme zu den neuen Sequenzierungstechniken und ihren Konsequenzen für die genetische Krankenversorgung. Hg. v. Berlin-Brandenburgische Akademie der Wissenschaften, Berlin. Unter: <https://www.gentechnologiebericht.de/fileadmin/Gentechnologiebericht/Stellungnahmen-Gendiagnostik-1-2.pdf> [24.04.2024].

Beier, H. et al. (2009): Neue Wege der Stammzellforschung. Reprogrammierung von differenzierten Körperzellen. BBAW, Berlin. Unter: https://www.gentechnologiebericht.de/fileadmin/Gentechnologiebericht/Publikationen/Stellungnahme_Neue_Wege_zur_Stammzellforschung_2009/Stellungnahme_Neue_Wege_zur_Stammzellforschung.pdf [24.04.2024].

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

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

Contents

<i>Jörn Walter and Boris Fehse</i>	
Foreword	54
<i>Olaf Rieß and Julia Schulze-Hentrich</i>	
1. Significance and clinical use of modern genome sequencing technologies – an introduction focusing on rare diseases as an example	56
1.1 Introduction	56
1.2 Description of current progress in clinical genome sequencing	57
1.3 Challenges and further developments towards genomic medicine	60
1.4 Conclusion	61
1.5 References	61
<i>Thorsten Schlomm and Nadine Biernath</i>	
2. Relevance and practical use of genome data in the field of cancer diagnostics	62
2.1 Introduction	62
2.2 Short history of molecular therapy	62
2.3 Modern approaches in oncological genomics: from standard sequencing to whole-genome sequencing	63
2.4 From analysis to application: Progress in genomics and their practical application in oncology	64
2.5 Outlook: Using the collective experience of the medical community by way of “swarm intelligence”	66
2.6 Conclusion	66
2.7 References	66
<i>Jörn Walter, Nina Gasparoni and Oliver Kohlbacher</i>	
3. Secure archiving of complex genome data and their efficient use – a challenge	68
3.1 Introduction	68
3.2 Genome data in research and medicine	68
3.3 Challenges of centralized data use: data protection and informed consent	72
3.4 Structures for archiving and controlled access	73
3.5 Conclusion	75
3.6 References	76

	<i>Martin Jungkunz, Christoph Schickhardt and Eva C. Winkler</i>	
4.	The ethical framework for the use of genomic data in research	77
4.1	Introduction	77
4.2	Ethical framework conditions for genomic data used in research	78
4.3	Outlook	83
4.4	Conflict of interest	83
4.5	References	83
	<i>Ralf Müller-Terpitz and Hannah Lilly Straub</i>	
5.	The legal framework for the use of patient details and genome data – protection and benefits from a legal medical perspective	85
5.1	Protection of patients and their data by the GenDG (German Genetic Diagnostics Act)	85
5.2	Protection of genome data through legal data protection regulations	87
5.3	Conclusion: the potential of genome data within the framework of legal requirements	89
5.4	References	90
	<i>Jörn Walter</i>	
	Outlook: Charting the course to forward-looking human genome diagnostics and genome research	92
	Authors	48
	Members of the Working Group Gene Technology Report	49
	Publications of the Working Group Gene Technology Report	50

Foreword

The decoding of the human genome in 2000 marked a milestone in modern medicine, generating major, far-reaching impacts on genetic diagnostics and research and reaching into human development, health, and disease. Harnessing highly efficient genome sequencing technologies, supplemented by further omics technologies and data such as epigenomics, proteomics, and transcriptomics (including the highly topical single-cell sequencing), we are now facing a turning point in genetic diagnostics and moving rapidly toward data-assisted, functional genome interpretation which will open up wholly new vistas. It will open up a fundamentally new understanding of the assessment, individual diagnosis, and treatment of diseases. The rapid technical developments of the last few years offer a practical opportunity to include the decoding of a patient genome in routine healthcare.

In this brochure, we discuss the opportunities and current status of personal genome sequencing, its benefits for research into rare diseases and cancers, and we reflect on some important ethical, legal, and social issues resulting from genome research.

An occasion for highlighting the various aspects and issues revolving around genome sequencing is currently offered by the § 64e SGB pilot project starting on 1 April 2024 in which genome analyses will be included in routine clinical care for the first time. This concept will initially be conducted as a model for selected areas of genetic diseases. In the process, a study will be initiated to establish to what extent an individual genome analysis can be used in routine medical diagnostics and what benefit systematic genome analysis offers for current research.

The conditions for the success of this pilot project are undoubtedly excellent. The technical feasibility of broad-based, affordable genome diagnostics has been proven. Broad-based comparative genome studies already ongoing in other countries testify to the scale and breadth of the significance national and international genome programs for research and health management are enjoying. Besides the direct health benefits for individuals in the field of diagnostics, genome programs give rise to far-reaching, new findings for medical and pharmaceutical research. Building on steadily improving, data-assisted analyses and the use of computer-aided methods of machine learning (Artificial Intelligence, AI), this knowledge is being increasingly harnessed to reveal the significance of genetic variation for general and individual health. The knowledge and above all the comparison of many individual genomes will supply essential data for wide-ranging health studies and general health management in Germany, in addition to ever-improving personal diagnostics.

In five articles and a short outlook, this brochure offers insights into the state of current genome sequencing in Germany and its application in the field of human diagnostics. In terms of their subject-matter, the articles are devoted to the decoding of rare diseases

(Chapter 1), the diagnosis of cancer (Chapter 2), the key aspects of data backups, data archiving, and data use (Chapter 3) as well as data protection (Chapters 3, 4 and 5). In Chapter 4, the authors tackle the issue of ethical evaluation in our individual and society-based handling of genome data. Finally, the legal framework for the use of genome data is examined with respect to the aforementioned aspects (Chapter 5), and a short outlook is provided for possible next steps.

The Working Group “Gene Technology Report – Monitoring and Interdisciplinary Dialog”, which acts as the co-publisher of the brochure, has an interdisciplinary composition made up of natural scientists, humanists, and social scientists. Our Working Group focuses its activities on a broad range of priorities in the field of gene technologies while keeping an eye on their development over the long term. Our regular publications are mainly aimed at the broader public. As part of our “In Focus” strand, we respond to current issues and developments.

For this volume, we have been able to enlist the help of a series of scientists working in the field of genome diagnostics and genome research. In addition to the expertise within the Working Group provided by its members, Eva C. Winkler, Ralf Müller-Terpitz and Jörn Walter, we were able to attract external experts such as Olaf Rieß, Julia Schulze-Hentrich, Oliver Kohlbacher and Thorsten Schlomm who, together with their staff members Nadine Bier-nath, Nina Gasparoni, Martin Jungkunz, Christoph Schickhardt, and Hannah Lilly Straub have prepared a range of subjects on an academically sound and easily accessible basis. It should be emphasized that the authors, Kohlbacher, Rieß, Schlomm, Walter, Winkler, Gasparoni, and Schickhardt are not only experts in their respective fields, but also active in the GenomDE and German Human Genome-Phenome Archive (GHGA) initiatives and consequently also support German genome research in other contexts such as the pilot project. We would like to express our sincere thanks to the GHGA for its cooperation in this publication, as well as to the authors for their assistance with our brochure. We also owe a debt of gratitude to the Berlin Institute of Health (BIH) at Charité for hosting and supporting the Working Group.

Articles identified by name do not necessarily reflect the opinion of the publishers, the Working Group, or the BIH – but the latter fully endorse the quality of the work performed.

Jörn Walter and Boris Fehse

Deputy spokesman and spokesman of the Working Group *Gene Technology Report* at the BIH
Berlin/Hamburg, March 2024

1. Significance and clinical use of modern genome sequencing technologies – an introduction focusing on rare diseases as an example

Olaf Rieß and Julia Schulze-Hentrich

1.1 Introduction

Many diseases are caused by alterations in a person's genetic material, their DNA. These genetic changes are either already present in the germline of the parents and are thereby heritable, they may occur de novo in the oocyte/sperm cell or immediately after fertilization, or they occur in the course of life in somatic cells, as is the case with tumor cells, for example, and are only passed on in dividing tumor cells but usually not to subsequent generations. In order to detect and locate these changes, the genome has to be sequenced¹ and compared with reference sequences from "healthy" subjects.

Over the last years impressive technical progress has been made in genome sequencing, resulting in faster and more accurate sequencing, reduced costs, and improved data analysis. This development represents a quantum leap in genomics as it brings the extensive benefits of the technology and genome data for the patients into the clinic. Genome data will help to reveal the individual causes of diseases, and propose personalized therapeutic approaches while also determining individual (genetic) risks of certain diseases (before symptoms appear).

Genome research of rare diseases and cancer in particular, has enabled new findings in genotype-phenotype correlation, and has shown that with many diseases, there is a clear connection between a person's risk of developing a disease and their genome sequence. Consequently, genomic diagnostics has already been introduced in many areas of medicine on a global scale in recent years. In Germany, this development is currently taking place in a pilot scheme (see Sections 1.2.1 and 1.2.2) in the field of rare diseases in which a genetic cause is assumed in around 80 % of patients, in hereditary cancers running in families as well as for targeted tumor therapies. In addition, latest sequencing technologies hold the high potential of even determining individual epigenetic changes² in the future as part of diagnostics and target for therapy.

¹ The genome or genetic material of an organism comprises an individual's heritable information (genes). In genome sequencing, the sequence of the genome, in other words the sequence of the DNA bases, is determined.

² Epigenetic changes comprise changes of chromatin, the packaging structure for DNA, or, for example, methylation of the DNA itself.

1.2 Description of current progress in clinical genome sequencing

1.2.1 The idea behind the pilot project

Current developments in sequencing, and thus in genome-based diagnostics, have the potential to significantly improve the specificity and sensitivity of diagnosis and treatment of both, rare and more common diseases. The widespread use of genome sequencing in routine health provision has therefore been recognized by many experts but, to date, has been hampered by financial and infrastructure challenges in Germany. A first important step has been taken by facilitating genome sequencing in Germany as part of a pilot project in accordance with § 64e SGB V (German Social Code Book)³ for selected persons with rare diseases and hereditary tumor diseases, as well as for patients as part of an individual, targeted tumor therapy, funded by the health insurance providers starting in July 2024. To this end, a large working group of medical doctors, human geneticists, scientists and bioinformaticians are establishing criteria for accessing and carrying out diagnostics, as well as for handling the data gathered as part of genomDE^{4,5}. In accordance with § 64e SGB V, the pilot project enables and facilitates the implementation of this concept at university centers, with the aim of creating long-term structures to integrate genomic diagnostics with health provision in Germany.

1.2.2 Focus on rare diseases (RD)

Part of the pilot project will focus initially on rare diseases for which a genetic cause is assumed for around 80 % of patients. A disease is considered rare if fewer than 5 out of 10,000 people are affected. As there are more than 6,000 distinct rare diseases and thus several million people affected by them in Germany alone, the total number of people benefiting from the pilot project is likely very high.

While rare diseases are heterogenous in terms of their phenotype, the underlying cause is typically a genetic change in a disease-associated gene (Amberger et al., 2009; Amberger et al., 2015). Despite enormous efforts and progress in sequencing technologies in the last few years, the genetic cause can only be precisely identified in about 40 % of all RD patients so far (Wright et al., 2018). A potential reason might be based on the challenge to pinpoint the causal genetic change that leads to serious functional consequences and is therefore relevant for the disease in the individual patient. After all, genetic mutations are frequent, but in most cases not functionally relevant. In order to identify and pinpoint the mutations causing disease and distinguishing them for the ones with unknown clinical significance, it is important to develop a comprehensive understanding of general frequencies of genetic variation in a population. The likelihood that a rare change is causally linked to a specific disease in one or several patients is greater than a base pair change in frequently occurring variants in healthy individuals. To distinguish between rare and common variants, extensive comparative data are therefore required, i.e., databases with sufficient “reference genomes”.⁶

³ For further information on the pilot project, see Walter et al., Chapter 3 and Müller-Terpitz/Straub, Chapter 5.

⁴ genomDE is a German genome initiative financed by the Federal Ministry of Health. The project launched in October 2021 and ends in December 2024. genomDE is part of the national strategy for genome medicine and is intended to establish an important building block in data infrastructure in the German health system in the future. See: <https://genom.de/de> [09.02.2024].

⁵ See: <https://www.bundesgesundheitsministerium.de/themen/gesundheitswesen/personalisierte-medizin/genomde-de.html> [09.02.2024].

⁶ For further information on reference genomes and databases, see Walter et al., Chapter 3.

Point mutations⁷ relating to a single or few base pairs (bp) represent the most frequently observed change. However, there are also cases in which larger genomic regions associated with a disease differ from the reference genome. This includes small insertions and deletions (InDels), structural variants (SVs) relating to more than 50 base pairs, rearrangements such as larger inversions, copy number variants⁸ and repeat expansions.⁹ These can be much more accurately identified and interpreted using the latest sequencing technologies (see Figure 1, as well as Schlomm/Biernath, Chapter 2). To date, changes have primarily been determined in the exome¹⁰ comprising all protein-coding sequences. With genome sequencing, the (nearly) entire genome of an organism can be sequenced. This includes not only the coding sequence but all regulatory regions of genes likely not yet assigned to specific functions. There is increasing evidence that genetic changes in these regulatory regions can also be associated with diseases.

The pilot project focuses on genome sequencing in order to achieve the most comprehensive identification of all these possible genetic causes in rare diseases and tumors, thereby closing the diagnostic gaps that have existed so far.

1.2.3 Overview of current sequencing technologies

Sequencing technologies have developed at a rapid pace in the last decades. The first attempts to sequence DNA used chemical degradation or enzymatic digestion of DNA in order to produce fragments to be individually analyzed (first-generation sequencing, see Figure 1). Although these techniques delivered valuable knowledge in the field of molecular genetics and were the gold standard in human genetics up to a few years ago, they are time-consuming and laborious. It was not until second-generation sequencing was established (next-generation sequencing, NGS) that genomics made a quantum leap, offering a powerful technology for genome research. The so-called “short-read sequencing” platforms permit the almost entire genome to be quickly sequenced by simultaneously sequencing millions of small DNA segments and then reassembling them with bioinformatic tools like a jigsaw puzzle with the aid of a blueprint (reference). The latest progress in this technology focuses primarily on faster, more economical and more precise sequencing as well as more efficient, enhanced data analysis. With the large amount of money (around 3 billion dollars) required to sequence the “first” human genome in 2001, you can generate more than 3 million genomes today.

In contrast to short-read sequencing, “third-generation” sequencing technologies concentrate on reading very large, i.e. long DNA fragments (long-read sequencing) consisting of more than 100,000 base pairs at the same time (see Figure 1).

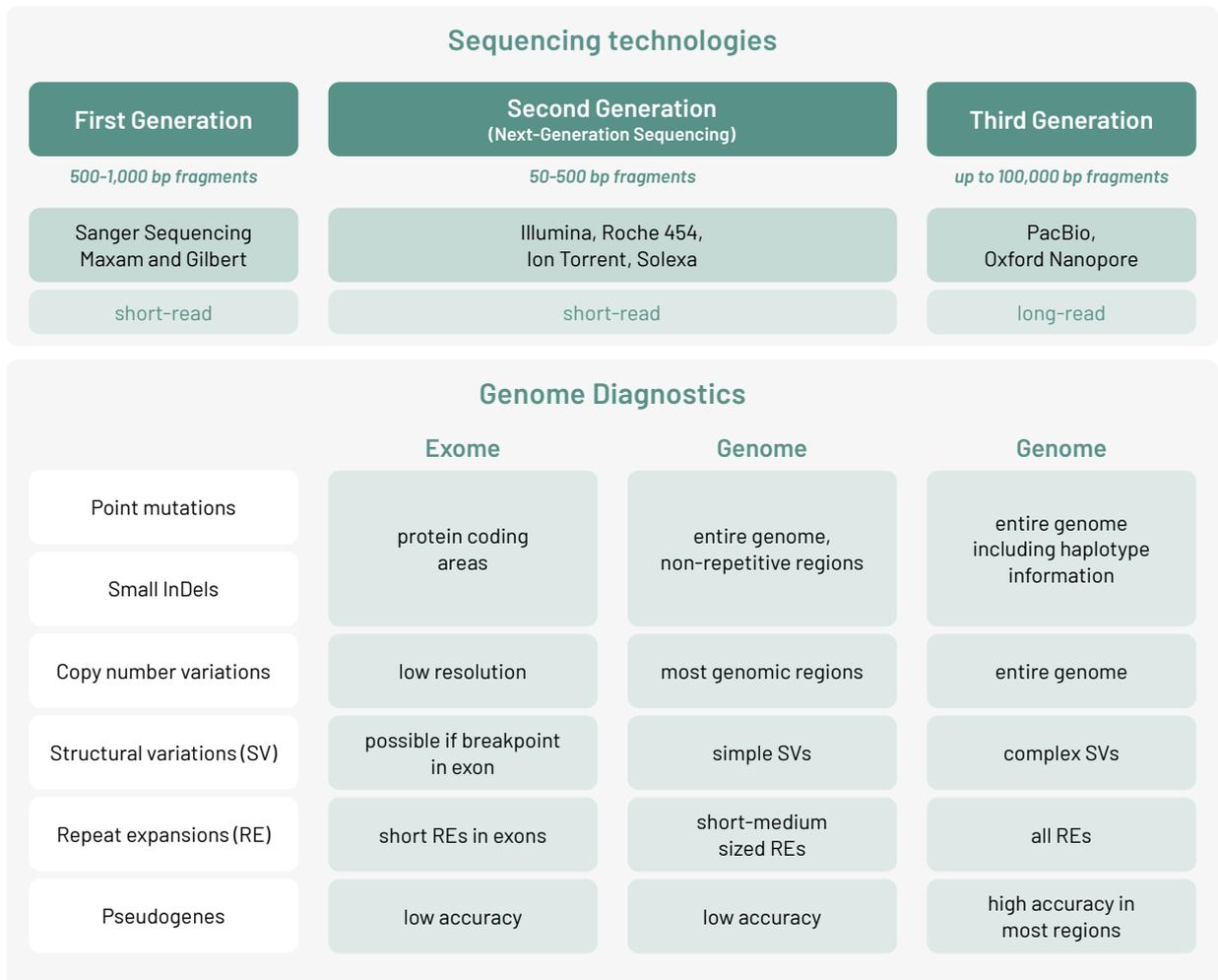
7 We distinguish between substitution (base exchanges; the most frequent point mutation by some margin), insertion (insertion of a base) and deletion (loss of a base).

8 Copy Number Variants (CNV) are a kind of structural variation of the genetic material (structural variation, SV), which produces variations in the number of copies of a certain DNA segment.

9 Trinucleotide expansions are DNA segments in which a three-letter nucleotide sequence occurs repeatedly (e.g. CAG-CAG-CAG-CAG). If the number of repeats exceeds a certain value, this can be the cause of a hereditary disease, so-called trinucleotide diseases.

10 The exome is the DNA sequence that codes for around 20,000 proteins. It comprises around 1-2 % of the entire human genome, i.e. around 30 million base pairs. By contrast, the entire genome also comprises non-coding sequences including introns, regulatory sequences and various repetitive sequences.

Figure 1: Overview of sequencing technologies in genome diagnostics



These long-read sequencing techniques make it possible to i) identify complex structural variants in the genome such as large insertions, deletions, inversions and duplications, ii) reveal and locate on which parental allele a change occurs, iii) distinguish pseudogenes¹¹ from active genes, and iv) chart selected epigenetic DNA modifications in parallel to the genome sequence. The major significance of these new techniques is reflected by a currently published understanding of our human genome, the Telomere to Telomere Genome (T2T Genome)¹². It was only possible to decode and assign previously “hidden” DNA segments with complex structures with the aid of these new techniques.

In addition to genome analyses, modern sequencing techniques are used for the comprehensive analysis of transcribed RNAs (transcriptome), epigenetic modifications (epigenomes) and the human microbiome. In this area of functional genomics, single-cell

¹¹ A pseudogene is a DNA segment which has a typical gene structure but does not code for a functional protein. We assume that pseudogenes used to be functioning genes and that they became non-functional as a result of genetic mutations.

¹² Telomeres are structural elements of DNA which occur at the ends of chromosomes to stabilize them. These DNA regions are particularly hard to sequence as they contain many repetitive elements which consequently are hard to assign precisely.

sequencing methods have proved themselves as an especially valuable technology in recent years (Walter/Schickl, 2019).

With regard to rare diseases, current approaches will clarify whether 80 % of all cases can in fact be solved by genome sequencing alone. With respect to functional evaluation and interpretation of genetic changes in regulatory areas above all, it will be essential to profile additional molecular layers such as the epigenome¹³ and transcriptome¹⁴ and to integrate these multimodal sequencing data.

1.3 Challenges and further developments towards genomic medicine

1.3.1 The need to train qualified staff

The new field of genome diagnostics also requires new training concepts and the collaboration of differently qualified staff in the area of clinical geneticists, documentalists, genome analysts, and clinical bioinformatician. As the entire process entails multiple steps, all those involved in genome diagnostics must be qualified for their particular role (preanalytics, sampling, lab sequencing, bioinformatics, diagnostics, validation, and medical evaluation) and be able to communicate efficiently at the corresponding interface. In this context, the S2k guideline from the German Society of Human Genetics (GfH) has to be complied with¹⁵ as well as statutory requirements contained in the Genetic Diagnostics Act.¹⁶ Specialization of medical doctors would benefit from new concepts and be desirable here – as besides sufficient staff for carrying out sequencing processes, data interpretation and documentation – there is also an urgent need to include trained scientists (e.g., human geneticists) and medical doctors.

1.3.2 Clinical healthcare provision with the aim of prevention

The great potential of genome sequencing lies not only in diagnosis of genetic diseases or in targeted, personalized therapy, but also in disease prevention. Prevention in the context of clinical healthcare aims at preventing or delaying disease onset as well as progression. This applies to rare as well as common diseases and offers great potential for genome diagnostics. Around 5 % of our population carry monogenic pathogenic/likely pathogenic variants in genes, which lead to direct screening adjustments, in so-called “actionable genes” (Rieß et al., 2024). The evaluation of actionable genes has already been used very successfully in many German human genetic institutions for many years. Corresponding working groups have been deployed and opinions drawn up.¹⁷ The use of diagnostics for common diseases which are usually caused by a combination of numerous “low pathogenic” variants (polygenic risk scores, PRS) and environmental effects (multifactorial diseases), has not yet been established in the health care system with some exceptions as, among other things, such ser-

¹³ The epigenome comprises chemical changes of DNA and its packaging structure called chromatin. Epigenetic changes influence the accessibility of genes, thus regulating gene activity (see also Walter/Hümpel, 2017 and Walter/Gasparoni, 2021).

¹⁴ The transcriptome refers to all RNA molecules produced in a cell (see also Fehse et al., 2022).

¹⁵ The Genetic Diagnostics Act defines the prerequisites for genetic testing and genetic analyses as well as the use of genetic samples and data in Germany (see: <https://de.wikipedia.org/wiki/Gendiagnostikgesetz> [14.02.2024]). For the legal framework, see Müller-Terpitz/Straub, Chapter 5.

¹⁶ See: <https://www.degruyter.com/document/doi/10.1007/s11825-018-0223-1/html> [09.02.2024].

¹⁷ See: <https://doi.org/10.1515/medgen-2023-2026> [09.02.2024].

vices are currently not covered by the health system. Genome analyses of healthy individuals for 10 common disease groups showed that around 16 % of healthy individuals carry a high risk of developing these diseases (Rieß et al., 2024). As part of the German network for familial breast and ovarian cancer, a PRS (Mavaddat et al., 2019) is already in use in some centers (Witt et al., submitted). It can be assumed that the implementation and application of diagnostics for actionable genes as well as of PRS as part of the 64e/genomDE pilot project will become more widespread in the coming years. It goes without saying here, and indeed it is also stipulated as part of the Genetic Diagnostics Act, that the patient must be advised beforehand on the extension of diagnostics, and their consent obtained in writing.

1.4 Conclusion

The opportunities of human genome research with the comprehensive sequencing of the genome have grown at an incredibly rapid pace in the last few years. Therefore, the analysis of individual human genomes offers ever more significant potential for medical health-care concepts. With the current pilot project, clinical use of genome sequencing is limited to rare diseases and some forms of cancer (see Schlomm/Biernath, Chapter 2). In the future, however, genome diagnostics must be considered and employed for a multiplicity of diseases as they promise to yield significant contributions towards a modern health care system in Germany and across the world.

1.5 References

- Amberger, J. et al. (2009):** McKusick's Online Mendelian Inheritance in Man (OMIM). In: *Nucleic Acids Res* 37: D793–796. DOI: 10.1093/nar/gkn665.
- Amberger, J. S. et al. (2015):** OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. In: *Nucleic Acids Res* 43: D789–798. DOI: 10.1093/nar/gku1205.
- Fehse, B. et al. (Ed.) (2022):** Im Fokus: RNA. Eine Bestandsaufnahme der Arbeitsgruppe Gentechnologiebericht. BIH, Berlin. DOI: 10.17169/refubium-36831.
- Mavaddat, N. et al. (2019):** Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. In: *Am J Hum Genet.* 104(1): 21–34. DOI: 10.1016/j.ajhg.2018.11.002.
- Rieß, O. et al. (2024):** Genomes in clinical care. In: *Npj Genomic Medicine* 9(1): 20. DOI: 10.1038/s41525-024-00402-2.
- Walter, J./Hümpel, A. (Ed.) (2017):** Epigenetik. Implikationen für die Lebens- und Geisteswissenschaften. Nomos, Baden-Baden. Unter: <https://www.nomos-elibrary.de/10.5771/9783845270838/epigenetik> [19.02.2024].
- Walter, J./Gasparoni, N. (2021):** Themenbereich Epigenetik: von Zellidentitäten bis hin zu Krankheiten und Therapien. In: Fehse, B. et al. (Hrsg.): Fünfter Gentechnologiebericht. Sachstand und Perspektiven für Forschung und Anwendung. Nomos, Berlin: 93–113. Unter: <https://www.nomos-elibrary.de/10.5771/9783748927242-93/3-themenbereich-epigenetik-von-zellidentitaeten-bis-hin-zu-krankheiten-und-therapien?page=1> [19.02.2024].
- Walter, J./Schickl, H. (Ed.) (2019):** Einzelzellanalyse in Forschung und Medizin. Eine Stellungnahme der interdisziplinären Arbeitsgruppe Gentechnologiebericht. BBAW, Berlin. Unter: https://edoc.bbaw.de/files/3278/BBAW_Einzelzellanalyse_Walter_Schickl.pdf [19.02.2024].
- Witt, D. et al.:** Clinical genome sequencing in patients with hereditary breast and ovarian cancer: Concept, implementation and benefits. Submitted.
- Wright, C. F. et al. (2018):** Paediatric genomics: diagnosing rare disease in children. In: *Nat Rev Genet* 19(5): 253–268. DOI: 10.1038/nrg.2017.116.

2. Relevance and practical use of genome data in the field of cancer diagnostics

Thorsten Schlomm and Nadine Biernath

2.1 Introduction

Genome data are playing an ever more important role in cancer diagnostics as well as in the planning of therapy, and they are opening up new, personalized approaches to treatment. Thanks to progress made in next-generation sequencing (NGS), numerous genes or the entire genome of an individual can now be efficiently and precisely analyzed in order to capture a broad range of genetic factors of relevance to cancer. Whole-genome sequencing (WGS) also allows structural DNA changes¹ and epigenetic processes² in the entire genome to be analyzed, facilitating more comprehensive insight into tumor genomes. The challenge, however, lies in the interpretation of these data, as advanced bioinformatic tools and expertise are required to differentiate between clinically relevant and irrelevant genetic mutations. In addition, the availability and use of genetic data for research and Real World Evidence (RWE)³ synthesis represent an important topic. Initiatives such as GenomDE, the National Center for Tumor Diseases (NCT) network, the German Human Genome-Phenome Archive (GHGA), the German Network for Personalized Medicine (DNPM), and the gene sequencing pilot project (for details see Section 2.4.3) aim for standardized data use in order to maximize the potential of genomics in cancer research and cancer treatment.

This chapter focuses on the role of genome data in cancer diagnostics, the progress and challenges of gene sequencing technologies as well as the importance of data accessibility for the research and development of new therapies in order to drive personalized medicine and cancer treatment.

2.2 Short history of molecular therapy

Since the certification of the first precision oncology drugs, cancer therapy has made significant progress. Rituximab, a monoclonal antibody against the CD20 antigen of B cells,

¹ Structural genome changes relate to larger changes in the DNA structure such as deletions (removal of DNA segments), amplifications (reproduction of certain DNA areas), or translocations (movement of DNA segments to other positions). These changes can influence the function of genes.

² Epigenetic changes are modifications in gene activity not caused by changes to the DNA sequence itself but by chemical modifications to the DNA or the histones (small basic chromatin proteins in the cell nucleus) around which the DNA is wound. In connection with cancer, such changes can lead to the abnormal activation or deactivation of genes that normally control cell growth which can result in uncontrolled cell growth and tumor development. For further information on epigenetics, see Walter/Hümpel, 2017 and Walter/Gasparoni, 2021.

³ Real-world evidence (RWE) relates to medical information gained from the actual application of treatments in daily practice by contrast with data from strictly controlled clinical trials. RWE comprises data from various sources such as electronic health files, patient registers, and insurance databases. This evidence offers insights into the efficacy, safety, and cost of therapies under real conditions, helping to improve patient care and medical decision-making.

was approved for the treatment of non-Hodgkin's lymphomas in 1997 and was initially only intended for this tumor entity (Grillo-López et al., 2000). Trastuzumab, a further drug targeted against the HER2 receptor, followed in 1998 with an indication specifically for HER2-positive breast cancer (Cobleigh et al., 1999; Maadi et al., 2021).

The concept of targeted therapies has recently progressed to a tumor-agnostic approach⁴ as shown by the approvals issued by the American drug regulatory authority, FDA (Food and Drug Administration), for the drugs Pembrolizumab for tumors with high mutational burden⁵ or microsatellite instability⁶ and Larotrectinib for types of cancer with NTRK gene fusions. Both therapies are used regardless of the tumor entity (Chakravarty et al., 2022; Farago/Demetri, 2020). These drugs are examples of treatment no longer aimed at the original site of the tumor but at specific genetic aberrations (changes).

The American Society of Clinical Oncology (ASCO) recognizes the importance of this development and since 2022, it has been recommending broad genetic testing using multi-panel NGS (simultaneous analysis of several specific genes) for all patients with advanced tumor diseases, if several precision oncology treatment options are available (Chakravarty et al., 2022).

2.3 Modern approaches in oncological genomics: from standard sequencing to whole-genome sequencing

In clinical oncology, we distinguish between two main methods of examining the genomes of tumors: simple and comprehensive approaches. The simpler methods such as single gene sequencing or panel gene sequencing focus on specific, already known gene changes that are important for the treatment of cancer. These methods are less expensive and simpler to evaluate.

The more comprehensive methods such as whole-genome sequencing (WGS) examine the entire genome of the tumor. This yields a more complete picture and can deliver new insights into the cancer. Whole-exome sequencing (WES) is a further method that examines all protein-coding areas of the genome. WES helps identify genetic changes that affect proteins and can thereby have a direct impact on the development of the cancer.

Although WGS and WES are more expensive and more complicated, they are becoming increasingly important as they supply significantly more comprehensive data. This could be crucial in the future in developing better treatments. There are also combined approaches that register both broad changes in the genome and specific changes of importance for the treatment which offer a detailed picture of the tumor, while at the same time being more practical and less expensive than complete sequencing of the entire genome.

⁴ Tumor-agnostic therapies are not tailored to the specific tissue properties, the original organ, or the specific type of cancer. Instead, treatment is based on an indication of certain molecular genetic changes in the tumor.

⁵ A high tumor mutational burden is defined as ≥ 10 mutations/megabases.

⁶ Microsatellite instability frequently applies in the case of mutations in a certain DNA repair mechanism - the mismatch repair genes.

However, the challenges involved in the interpretation and practical application of broad genomic data such as WES or WGS data in clinical practice, are multilayered. The complexity and scope of the data require the deployment of advanced bioinformatic tools and algorithms. The distinction between pathogenic and benign variants, the integration and correlation of data on different levels, the lack of comprehensive reference data and gaps in our knowledge, the need for multidisciplinary expertise, and the considerable investment of time and resources are only some of the hurdles to be overcome. These factors illustrate the need for continuous progress in bioinformatics, enhanced data integration, data exploitation, and a deeper understanding of cancer genomics. Only by such means can the full potential of WES and WGS be realized in personalized cancer medicine and the treatment of cancers further enhanced in the future. The first programs for the broader deployment of WES and WGS in oncology are currently being launched in large nationwide networks (see Section 2.4.3).

2.4 From analysis to application: Progress in genomics and their practical application in oncology

2.4.1 The transformative role of molecular tumor boards in genetic data analysis and treatment plans

Molecular tumor boards are playing a crucial role in modern cancer treatment by using genetic information from tumors to develop personalized treatment plans. One key aspect here is to assess the biological value of molecular changes. This is achieved by using comprehensive database research in which the aberrations detected are compared with existing variant databases. This evaluation is particularly important in broad sequencing approaches such as multi-panel NGS diagnostics, WES, and WGS.

In addition, the genetic information is viewed in the context of the entire tumor history of the patient. This includes looking at earlier treatments, the response to such therapies, and the growth of the tumor over time. This holistic perspective enables specialists on the molecular tumor board to develop a comprehensive treatment plan that is not only attuned to the current genetic status of the tumor but also to the patient's entire treatment history. The challenges in this process comprise not only the technical aspects of the genetic analysis but also the interpretation and clinical application of these complex data. A high degree of specialist knowledge and experience is required to interpret genetic data correctly and translate them into effective treatment strategies. In addition, data protection and ethical considerations must be taken into account,⁷ particularly when the treatment of potentially heritable genetic germline information is involved.

Despite these challenges, molecular tumor boards offer an invaluable benefit for personalized cancer treatment by decoding the genetic details of a tumor and thereby facilitating bespoke therapy tailored to the specific needs of each individual patient.

⁷ For ethical or legal issues, see Jungkunz et al., Chapter 4, and Müller-Terpitz/Straub, Chapter 5.

2.4.2 Automated molecular evaluation algorithms (Clinical Decision Support System; CDSS) in personalized oncology

CDSS represent advanced software solutions that analyze tumor sequencing findings on an automated basis and generate standardized reports for clinical decision-making. They efficiently process complex genetic data from sequencing techniques such as multi-panel NGS, WES, or WGS. Harnessing algorithms, these solutions identify and classify genetic mutations and aberrations and translate the data into clear, interpretable reports that highlight the clinically relevant information and potential therapies and opportunities for study.

These systems save time and increase the precision of data interpretation by reducing human error. However, challenges remain in the dependence on the quality of the data, integration with clinical processes, and the handling of sensitive patient data. Overall, CDSS contribute to the further development of personalized medicine and the optimization of patient care, offering potential for innovative developments such as the integration of swarm intelligence and real-time data (see Section 2.5).

2.4.3 Synergy in genomics: Networking German initiatives promoting personalized oncology

It is essential that already existing initiatives and networks join forces in order to implement genomic approaches such as WGS across the board in German oncology, including broad and effective utilization of the data. An integrative approach of this nature would pool the strengths and resources of various players in order to develop and implement a comprehensive, coherent strategy in genome-based oncology. The networking of existing initiatives such as the GenomDE initiative⁸, the GHGA⁹, the DNPM¹⁰, the NCT¹¹ network and integration of the pilot project pursuant to § 64e SGB V¹² for comprehensive diagnostics and therapeutics using genome sequencing can create a comprehensive infrastructure for genome medicine in Germany.

Such a collaboration would facilitate the standardization of genetic sequencing and data exchange, promote research in personalized oncology, and create a uniform source of data for research and clinical application incl. RWE generation. It would also improve the training of specialists in genomics and oncology which is crucial for the effective integration of genome-based approaches with clinical practice. In addition, joint guidelines and standards for data protection and data security would be established, making genome medi-

8 This initiative aims to bolster and reinforce genetic diagnostics and personalized medicine in Germany. The intention is to improve access to genetic consultations and testing and promote the integration of genetic information with clinical practice. Combined with the § 64e pilot project, this would mean enhanced infrastructure and resources for genetic examinations and personalized treatments.

9 The GHGA serves to store and exchange genomic and phenotype data in order to promote research and medical innovation in Germany. The integration of the GHGA with the pilot project and the GenomDE initiative would facilitate the exchange of data and insights which might accelerate research and development in personalized medicine.

10 Gradual improvement, quality-assured implementation, and continuous data collection in terms of a self-learning system – these are the goals of the alliance under the lead of the Comprehensive Cancer Center which is currently supported by the health insurance companies' innovation fund. In addition, a certification system has been developed by the German Cancer Society as part of the certification system of the Organ Cancer Centers and Oncology Centers.

11 The National Centers for Tumour Diseases specialize in cancer treatment and research in early clinical trials. Involving them in the § 64e pilot project and links to the GenomDE initiative and the GHGA could speed up cancer research and treatment. Through the exchange of genetic data and findings, more effective personalized treatment strategies could be developed for cancer patients.

12 This project enables innovative forms of care to be trialed within the statutory health insurance system. The focus lies on improving patient care through new approaches such as WGS which extend beyond traditional medical care. In the context of genomics and personalized medicine, this would include the integration of advanced genetic diagnostics and individualized therapeutic approaches.

cine safer while strengthening public trust. Overall, joining forces in this way would make genome-based oncology in Germany more efficient and effective, optimize patient care, and advance research in this field.

2.5 Outlook: Using the collective experience of the medical community by way of “swarm intelligence”

The integration of real-time data and swarm intelligence with the use of genome data could lead to major advances in cancer diagnostics. By combining patient histories and genetic findings, CDSS which are already deployed in personalized oncology, could give personalized recommendations for treatment based on the data from many patients in the same way as Google Maps is used for traffic information.

A project at the Charité – Universitätsmedizin Berlin in collaboration with the German Network for Applied Precision Medicine (DNA-Med) is working to convert a varied mix of patient data to a uniform, structured format. This is crucial for detecting patterns and preparing individual treatment plans. The challenge lies in structuring and standardizing the patient data as well as ensuring interoperability between different databases.¹³

This project could lead to a turning point in medical care where decisions are made based on a broad range of data reflecting the collective experience of the medical community. The challenges posed in the process – from technical aspects, and AI development to data protection and ethical questions – are considerable but the key objective is enabling enhanced, data-supported patient care in oncology tailored to the individual.

2.6 Conclusion

The extremely rapid pace of technological developments in genome sequencing allows their ever more extensive application in cancer diagnostics. However, there remain major challenges in interpreting the data effectively and placing them in the overall medical context of the patient. In addition, the data must be protected and made accessible for broader research in order to obtain new, extended findings with the aid of AI-driven techniques. Mastering these challenges will be decisive in further advancing the boundaries of personalized medicine and revolutionizing the treatment of cancers.

2.7 References

Chakravarty, D. et al. (2022): Somatic genomic testing in patients with metastatic or advanced cancer: ASCO provisional clinical opinion. In: J Clin Oncol 40(11): 1231-1258. DOI: 10.1200/JCO.21.02767.

Cobleigh, M. A. et al. (1999): Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease. In: Journal of Clinical Oncology 17(9): 2639-2648. DOI: 10.1200/JCO.1999.17.9.2639.

Farago, A. F./Demetri, G. D. (2020): Larotrectinib, a selective tropomyosin receptor kinase inhibitor for adult and pediatric tropomyosin receptor kinase fusion cancers. In: Future Oncol 16(9): 417-425. DOI: 10.2217/fon-2019-0647.

¹³ For further information on genome databases, see Walter et al., Chapter 3

Grillo-López, A. J. et al. (2000): Rituximab: the first monoclonal antibody approved for the treatment of lymphoma. In: *Curr Pharm Biotechnol* 1(1): 1-9. DOI: 10.2174/1389201003379059.

Maadi, H. et al. (2021): Trastuzumab mechanism of action; 20 years of research to unravel a dilemma. In: *Cancers (Basel)* 13(14): 3540. DOI: 10.3390/cancers13143540.

Walter, J./Gasparoni, N. (2021): Themenbereich Epigenetik: von Zellidentitäten bis hin zu Krankheiten und Therapien. In: Fehse, B. et al. (Ed.): *Fünfter Gentechnologiebericht. Sachstand und Perspektiven für Forschung und Anwendung*. Nomos, Berlin: 93-113. Unter: <https://www.nomos-elibrary.de/10.5771/9783748927242-93/3-themenbereich-epigenetik-von-zellidentitaeten-bis-hin-zu-krankheiten-und-therapien?page=1> [19.02.2024].

Walter, J./Hümpel, A. (Ed.) (2017): Epigenetik. Implikationen für die Lebens- und Geisteswissenschaften. Nomos, Baden-Baden. Unter <https://www.nomos-elibrary.de/10.5771/9783845270838/epigenetik> [19.02.2024].

3. Secure archiving of complex genome data and their efficient use – a challenge

Jörn Walter, Nina Gasparoni and Oliver Kohlbacher

3.1 Introduction

Human genome data and related omics data¹ are playing an ever more significant role in human medical research. Numerous international genome sequencing projects are generating a rapidly growing volume of genome data, supplying important, new findings on the genetic foundations of health and diseases. Germany is taking its own path and for the first time is generating genome data as part of medically billable clinical services within the German genome sequencing pilot project (§ 64e SGB V, see Section 3.2.2). These genomes are primarily sequenced for diagnostic purposes, but on a secondary level, they are also to be usable for research.

The combined clinical and research-driven use of genome data presents us with several challenges. On the one hand, genome data must be archived securely and protected, while on the other, they must be available for broad, research-related use. Complex genome data can only be optimally and transparently interpreted with the help of computer-aided techniques if they can be compared with other datasets which should be as extensive as possible. The personal details and health data of the person concerned must also be taken into account. The secure combining of genome data with personal (medical) data in a data archive that can be searched and used, is therefore an essential requirement. Only this link can facilitate the optimal interpretation of genome information for the diagnosis and treatment of diseases.

3.2 Genome data in research and medicine

3.2.1 The growing complexity of genome data

Genome research has been undergoing a profound transformation for some years. Driven by breathtakingly rapid technological improvements, entirely new dimensions (volume and depth) of genome analysis are opening up which are shifting genome data to the focus of medical interest. The decoding of the human genome is (still) making trailblazing contribu-

¹ Omics data are comprehensive datasets based on the analysis of the entire cellular content of the molecules examined, e.g., the entirety of the genetic information (genome), the gene transcripts (transcriptome), epigenetic modifications (epigenome), the entirety of proteins and their modifications (proteome) or all metabolic products (metabolome).

tions to biomedicine. The additional information to be derived from individual genome data for personalized diagnosis, researching the causes of diseases, and for the development and application of active agents is incalculably valuable and will find uses and applications in nearly every field of medical care and research.

In contrast to the huge investment in time and money required to decode the first human reference genome² (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001), genomes can nowadays be sequenced with the aid of a wide range of NGS techniques³ for a fraction of the cost⁴ and in rapid order.

The challenges therefore no longer lie mainly in the generation (sequencing) of genome data but in their orderly, linked storage as well as in their standardized evaluation and interpretation. These latter processes require far more time of staff-based intensive work on computer-aided analyses than the mere generation of the data. Estimates assume that in the next two years alone, 100 million genome sequences will be generated worldwide.⁵ Corresponding data infrastructures will have to be built and permanently maintained for the processing of such vast quantities of data.

Worldwide genome research programs in the last few years have made it clear that the individual human genome is much more complex than originally assumed. For example, the Pangenome project⁶ (Liao et al., 2023) showed impressively that reference to one single reference genome falls short of the mark and new avenues have to be explored in order to detect and include population specific genome changes. This population-related classification of individual genomes is significant as it helps to differentiate individual disease-related genome changes from already existing changes in the population and to evaluate them more effectively. The latest sequencing technologies (for this, see Rieß/Schulze-Hentrich, Chapter 1) which enable the sequencing of longer, contiguous segments of the genome, will create an ever more precise data basis for a precise genome interpretation.

Besides the pure genome data (DNA sequence), supplementary, supporting omics data are becoming increasingly important. This includes epigenetic modifications “imprinted” on the genome (the epigenome)⁷ as well as information on the use of the genome in form of RNA readouts (such as the transcriptome)⁸. Both types of omics data are also read through sequencing and interpreted in the context of the genome. These supplementary, so-called

2 As part of the international Human Genome Project and a private business initiative by Celera, thousands of scientists around the world worked on decoding the human genome from 1990 until 2003. Nearly the entire human DNA sequence has been identified by means of sequencing. For the publication of the full version in 2003 (first patchy version in 2001), see: <https://www.genome.gov/25520492/online-education-kit-2003-human-genome-project-completed> [23.10.2023].

3 NGS stands for next-generation sequencing, high-throughput sequencing (see Schlomm/Biernath, Chapter 2).

4 See: <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost> [23.10.2023].

5 See: <https://www.labiotech.eu/in-depth/cloud-genomics-big-data-problem> [08.11.2023].

6 As part of the international Pangenome project, scientists are working on collecting 350 reference genomes with the intention of reflecting the diversity of the population. See: <https://humanpangenome.org/index.html> [23.10.2023].

7 Epigenetic modifications are molecular changes to the DNA and the surrounding proteins that affect the activity of genes without changing the actual DNA sequence in the process. One example is DNA methylation in which methyl groups are bound to certain building blocks of DNA in cell-specific patterns and regulate the binding of further factors for reading genes.

8 The DNA is transcribed into RNA in a process by the name of transcription. The totality of all RNA molecules that are present in a cell at a certain moment (transcriptome) supply information on gene activity.

functional genome data will make a major contribution to analyzing and interpreting the control of the genome in a cellular context on a personal and disease-specific basis (Walter/Gasparoni, 2021; Fehse et al., 2022). A particular rich resource for the interpretation are sequencing data of individual cells (transcriptome and chromatin accessibility)⁹ which contribute to research into the causes of disease with so far unsurpassed granularity (Walter/Schickl, 2019). Genome data can also be linked to further omics data for biomolecules such as proteins (proteome) or metabolic products (metabolome) as well as to phenotype and clinical meta data, e.g., weight, clinical diagnoses, lab results, imaging techniques, and pathological findings from tissue samples. Modern whole-genome data are becoming increasingly “multimodal” and as a result, it must be possible to link the various datasets on a cellular and case-by-case basis. Countries such as Estonia, Denmark, and Finland are already very far advanced in their genome programs concerning linking various data formats and databases, enabling them to address complex diagnostic questions (Smetana/Brož, 2022).

Overall, it can be stated that the interpretation of genome data in conjunction with extended reference genomes (pan-genomic approaches deliver a significant extension here), functional genome data as well as further, imaging and diagnostic data (will) enable increasingly profound but also new insights into the significance and consequences of individual genomic changes. Genome-based diagnosis and the prediction of the causes and course of diseases will thus improve significantly. In research, new, complex integrative analytical approaches are constantly being developed which will also be used in diagnostics within the foreseeable future and will open up a differentiated, case-based range of applications for genome medicine.

Two essential steps are required in order to make this possible in the future. Firstly, it is important to bolster research in this area and secondly, genome databases must be set up in the future in such a way that they enable integrated use (i.e. the collation) of extended functional data, as not only the volume of genomic data but also the expected increase in information density represent a challenge for databases and analytical methods. Data archives must therefore be designed with sufficient power and connected to secure computing resources to enable the data to be linked, retrieved, and analyzed (see also Section 3.4.1). Supporting, standardized bioinformatic work processes and evaluation procedures are required for this (see also Section 3.4.2) as well as substantial measures to ensure data protection and information security in order to protect particularly sensitive health data from any misuse.¹⁰

3.2.2 Use of clinical genome data: linking patient care with research

The generation of whole-genome data for diagnosis or personalized therapy has so far not been part of standard medical care in Germany. Partial genome data have been collected primarily in the context of human genetic issues and indications (heritable traits) or personalized oncology as part of disease-focused studies and networks of personalized med-

⁹ The DNA lies compactly in the cell nucleus, packaged with proteins in a structure called chromatin. The degree of packaging varies at different sites, affecting the accessibility of the DNA for other factors that can read genes, for example. Examination of chromatin accessibility delivers valuable insights into the regulation of gene activity.

¹⁰ For data protection from an ethical or legal perspective, see Jungkunz et al., Chapter 4, and Müller-Terpitz/Straub, Chapter 5.

icine, such as nNGM¹¹, DNPM¹² and MASTER¹³. In human genetics, analyses for detecting mutations in candidate genes, gene panels, or exome analyses¹⁴ have been billable services. However, the complexity of whole-genome data poses completely new, far-reaching challenges to data generation, analysis, and interpretation.

The innovative pilot project from the Federal Ministry of Health (BMG) under § 64e SGB V¹⁵ opens up unique opportunities to build modern structures for the broad application of whole-genome sequencing in clinical healthcare. This venture is focusing initially on the genome sequencing of certain diseases and it will be evaluated after a period of four years. It will be important for the success of the project to highlight coordinated collaboration between the genome centers distributed across the whole of Germany and to connect these centers closely through a common data infrastructure. The training and education of experts in data analysis in the human genetics field of the genome centers, together with the consulting doctors and/or clinicians, will be of great importance in order to achieve optimum use of genome data.

The omics data generated by the pilot project and professionally evaluated are on the one hand valuable sources for individual healthcare, while on the other hand, they are also opening up new opportunities for researching the genetic causes of diseases on a molecular level. The aim of the project, as the national genome initiative genomDE¹⁶ puts it, is “to use genome sequencing for the benefit of patients as an innovative component of standard care in Germany [...]. At the same time, these projects serve to strengthen national and international medical genome research in terms of knowledge-generating care”.¹⁷

Achieving this goal requires the standardized combining of individual, clinically gathered genome data from the pilot project with data from other centers and already existing data from national and international genome and research projects. Only in this way can the necessary comparative cohorts and references be created which will improve the classification of causal changes/variants in cancers or patients with rare diseases and differentiate them from non-causal genomic changes.¹⁸ The clinically generated genome data are to be filed in central databases¹⁹ – provided there is informed consent from the sample donors or patients – thereby making them usable for researching the causes of diseases and the development of preventive measures. To this end, the clinically gathered genome data must be processed by standardized means from the outset, centrally archived, and com-

11 National Network Genomic Medicine Lung Cancer; see: <https://nngm.de/> [15.12.2023].

12 German Network for Personalized Medicine; see: <https://dnpm.de/> [15.12.2023].

13 Molecularly Aided Stratification for Tumor Eradication; see: <https://www.nct-heidelberg.de/forschung/molecular-stratification/master.html> [15.12.2023].

14 This relates to the examination of protein-coding segments of the genes (exons). Many gene regulatory regions lie in the non-coding areas of the genome that are not recorded in the process.

15 See: <https://www.sozialgesetzbuch-sgb.de/sgbv/64e.html> [23.10.2023]. For the legal framework of the pilot project, see Müller-Terpitz/Straub, Chapter 5.

16 genomDE is an initiative sponsored by the Federal Ministry of Health. It serves the development of substantive and structural concepts to prepare the implementation of the genome sequencing pilot project under § 64e SGB V. See: <https://www.genom.de/de> [16.11.2023].

17 See: <https://www.genom.de/de/ss-64e-sgb-v-modellvorhaben-regelt-die-rechtlichen-grundlagen> [16.11.2023].

18 For the significance of genome data for cancers, see Schlomm/Biernath, Chapter 2, and for the importance of rare diseases, see RieB/Schulze-Hentrich, Chapter 1.

19 For this, see Jungkunz et al., Chapter 4.

bined with the associated clinical data. The data must be archived in such a way that the protection of sensitive data is guaranteed and at the same time pseudonymized, dataset-relevant clinical parameters of individual samples can be used for the research.

3.3 Challenges of centralized data use: data protection and informed consent

The uniform storage and availability of omics data require extensive legal, technical, and organizational measures. Data have to be archived in a way that is legally secure and protected from misuse, and at the same time, they must be efficiently searchable and usable on a lasting basis. To this end, genome data must be linked with additional (meta)data of the donor(s) (patients or research participants), such as age, biological gender, and disease information. Due to the sensitivity of these personal data – health data and genome data are among the “special categories” of personal data in accordance with Art. 9(1) of the General Data Protection Regulation (GDPR) – these data are particularly in need of protection. Genome datasets collected for research purposes are usually pseudonymized²⁰, i.e., they cannot be easily identified with a particular person. Individual genome segments may nevertheless lead to re-identification under certain circumstances, particularly by linking them with other data sources such as publicly accessible genealogy databases (Gymrek et al., 2013). For that reason, pseudonymized data are only made available, even to researchers, with strict access restrictions and security precautions. Human genome research data must be archived in protected form, e.g., in the European Genome-phenome Archive, EGA²¹ (see Section 3.4.3). The human sequence data archived there are not publicly accessible and are only ever stored and transmitted in encrypted form. Only authorized researchers with legitimate research projects are given controlled access to such data (Shabani et al., 2016). For this purpose, an access application must be submitted to the person or institute responsible for the data. A data access committee or similar body checks the application and decides whether access is to be granted. Genome data and metadata can only be viewed and used through contractually regulated, controlled processes in order to meet all the requirements mandated by the GDPR.

The preceding informed consent of the data donors includes clear, transparent instruction on the goals and methods, but also the risks of a study as well as data approval. This consent generally forms the legal basis for accessing the data. It also ensures that patients are informed of the use of their genetic information, give their voluntary consent, and can revoke their consent at any time. The extensive aspects of legal and ethical issues in connection with genome data will be discussed in the coming chapters.

²⁰ With pseudonymization, direct, personal information such as names or dates of birth are replaced by (usually randomly generated) sequences of numbers and letters. This prevents any direct identification of the data with individual persons. Information on traceability in order to inform data subjects of the findings of the genome analysis is kept separate and is not shared with others.

²¹ See: <https://ega-archive.org/> [23.10.2023].

3.4 Structures for archiving and controlled access

3.4.1 Challenges and solutions in establishing a secure genome data infrastructure

As already mentioned above, the secure archiving of genome data gathered for diagnostic purposes with controlled access requires a very well-planned data infrastructure. The first task of a genome data infrastructure is to be able to receive data from various sources, institutions, and studies (data submission), obtain them in an ordered manner (with metadata), archive and securely collate (link) them in order to allow them to be searched and used on a platform with controlled access. With a network of comprehensive IT security measures (cyber security), steps must be taken in the process to ensure that the data are protected from unauthorized access and data loss. At the same time, authorized clinicians and researchers must be given controlled access to the data released to them (upon application) within a GDPR-compliant legal framework. The data infrastructure must operate on a lasting basis and must be scalable and adaptable to changed requirements (e.g., growing data volumes, supplementary (meta)data, alternative data formats, and analytical methods). Private cloud infrastructures, in particular, based on scalable object stores (e.g., operated by public institutions) provide this flexibility on a technical level and can be easily combined with powerful computing infrastructures (CPUs, GPUs)²². While the data are physically distributed in this infrastructure, it is important to pool the metadata and responsibility for operations on a national level centrally. Any such data infrastructure would have to meet the following requirements:

- Ensure compliance with strict guidelines and procedures for user identification and authorization
- Employ standardized processes for administering access inquiries, ensuring secure access via encrypted downloads and GDPR-compliant data processing
- Establish a central access point that can quickly assess data usage inquiries
- Support users in taking the steps required to submit data, search for datasets, apply for access, and obtain access to datasets as simply as possible
- Implement the FAIR principles²³ (Wilkinson et al., 2016): easy findability, accessibility, interoperability, and reusability for better secondary use of data

These requirements are met by the German Human Genome-Phenome Archive (GHGA)²⁴ (the GHGA Consortium, 2023) (see Section 3.4.3) in Germany. The GHGA also takes on the role of a national data hub as part of the Federated European Genome-phenome Archive (fEGA) (see Section 3.4.4). Further structures are also planned in the genome sequencing pilot project under the aegis of the BfArM²⁵ and RKI²⁶. Solutions for an integrated and FAIR

²² CPU stands for “central processing unit” and designates the central computing and control unit of a computer. GPU stands for “graphics processing unit” and designates the graphics processor of a device.

²³ The FAIR principles (Findable, Accessible, Interoperable, Reusable) represent a guide to improving the secondary use of data. Findable: Data should be easy to find with the aid of detailed descriptions and unambiguous designations. Accessible: Data should be available via clear, controlled access routes. Interoperable: Data should be structured in such a way that they can be combined and integrated with other data sources without problem. Reusable: Data should be optimally prepared for secondary use and reuse through detailed metadata.

²⁴ See: <https://www.ghga.de/> [29.04.2024].

²⁵ Federal Institute for Drugs and Medical Devices.

²⁶ Robert Koch Institute.

use of genome and omics data from research and patient care are currently being discussed and implemented. The national genome initiative genomDE has presented a series of practical suggestions for this purpose which define the framework for the infrastructures required, and among other things set out what a logistical and legally secure link between sequencing centers and (one) genome data center might look like.

3.4.2 FAIR data use and standardized workflows

With the establishment of a genome data infrastructure, (meta)data models and data management processes which make the data findable in the infrastructures and facilitate a standardized analysis of different datasets, must be implemented. With regard to FAIR data use, the link with ontologically clearly defined technical and individual metadata is crucial in making omics data not only findable but also interpretable and integrable. Metadata deliver much more than mere information on the characteristics of the datasets. Ideally, they also describe the experimental protocols and instruments used, the software used for the processing as well as the origin of the samples (cell or tissue type), the status of the disease, and deidentified details of the data donors such as their age or biological gender. Harmonization of the data description is of great importance and should be achieved by using established and widely distributed ontologies and vocabulary. The implementation of clearly defined, uniform metadata standards facilitates the shared use of data, enabling the interoperable exchange of data across national archives. This makes it easier for clinical users and researchers to fulfill their primary role of contextualizing not only their own data, but also to better understand and utilise third party data to gain significant insights. Ideally, standardized analysis workflows should be used for processing the genome data and for their subsequent interpretation. Such workflows should also be based as far as possible on optimized analytical methods and global standards – including with regard to comparability – and should be adaptable on a modular basis in order to facilitate reproducible and efficient research and diagnostics. It is especially necessary to process the datasets uniformly for cross-study comparisons and the joint analysis of several cohorts.

3.4.3 Data archives: current status and prospects

Examples of how a central, national data archive could be efficiently organized can be derived from the European research data infrastructure. Efforts have already been made to make omics data openly and FAIRly available – but in compliance with the necessary technical and organizational security precautions – including across national borders. In this context, mention should be made of the European Genome-phenome Archive (EGA) which is jointly operated by the European Bioinformatics Institute (EBI; Hinxton, England) and the Centre for Genomic Regulation (CRG; Barcelona, Spain). This European human genome archive stores personal genetic and phenotype data from research projects in Europe, and acts as an intermediary for controlled access. In order to reflect the different interpretations and implementations of the GDPR in EU member states and to take account of national data protection provisions, it was decided to supplement the EGA infrastructure in 2022 through an internationally federated infrastructure²⁷ – the Federated European

²⁷ See: <https://blog.ega-archive.org/the-federated-ega-network> [24.04.2024].

Genome-Phenome Archive (fEGA). National hubs within the fEGA archive and administer data in compliance with the relevant national data protection provisions, while at the same time, they facilitate the controlled exchange of data within the federated structure. In the process, only metadata on studies and datasets are transmitted to EGA's central metadata archive; the sequencing data and personal metadata remain in the relevant national hubs. This enables data to be found across countries and all hubs.

In Germany, the German Human Genome-Phenome Archive (GHGA)²⁸ is the national data hub in the fEGA. The GHGA is funded as part of the NFDI initiative (National Research Data Infrastructure)²⁹ of the German Research Foundation (DFG). Work on a genome data infrastructure is already far advanced. In the future, human genome data from Germany generated in research and published in accordance with the FAIR principles will/can be archived and administered – in compliance with national data protection provisions (the GHGA Consortium, 2023). In the parts in which consent to use for research purposes is available, the genome data from the genome sequencing pilot project can also be administered in the GHGA. In later phases, the GHGA is planning to offer additional services such as the GHGA Cloud (for analyses) and the GHGA Atlas (for a more in-depth interpretation of genome data) in addition to archiving.

As a hub within the fEGA, the GHGA also offers links and opportunities for exchange with other European initiatives in which Germany is involved such as the 1+ Million Genomes Initiative (1+MG)³⁰ and the Genomic Data Infrastructure (GDI)³¹ project. In addition, collaboration with other networks for personalized or genome medicine, e.g. via data analysis networks such as ELIXIR³², would be possible such as the NHS Genomic Strategy³³ (England), the France Genomic Medicine 2025 Initiative³⁴ (France), Genomic Medicine Sweden³⁵ (Sweden), the Danish National Genome Center³⁶ (Denmark) and the Precision Medicine Infrastructure associated with Science and Technology (Spain).

3.5 Conclusion

The comprehensive sequencing of human genomes will change healthcare in medicine and medical research on a lasting basis. The genome sequencing pilot project under § 64e SGB V reflects this expected development and will strengthen healthcare, as well as Germany's scientific role in the medically important field of human genome research. The establishment of a sustainable, widely usable, secure genome data infrastructure, driven by high

28 German Human Genome-Phenome Archive; see: <https://www.ghga.de/> [29.04.2024].

29 See: <https://www.nfdi.de/?lang=en> [29.04.2024].

30 See: <https://digital-strategy.ec.europa.eu/en/policies/1-million-genomes> [29.04.2024].

31 See: <https://gdi.onemilliongenomes.eu/> [24.10.2023].

32 See: <https://elixir-europe.org/> [16.11.2023].

33 See: <https://www.england.nhs.uk/genomics/genomics-strategy/> [16.11.2023].

34 See: <https://pfgm2025.aviesan.fr/> [24.04.2024].

35 See: <https://genomicmedicine.se/en/> [16.11.2023].

36 See: <https://eng.ngc.dk/> [16.11.2023].

standards, is essential for this assignment. The essential building blocks of this data infrastructure are already available and great energy should be devoted to its rapid expansion, combination, and consolidation. In this process, it is important not to lose sight of the fact that the establishment of high data standards and the central archiving of genome data are indispensable both for the optimum use and interpretation of the data for basic medical care and for further-reaching research. Genomes are not self-explanatory and their use and interpretation require a high degree of professional expertise. With its partners in health centers, the genomDE initiative forms an excellent framework for generating this connection for enhanced healthcare and further development through research. Operationally clear structures are important prerequisites – in order to guarantee the connection with international genome data infrastructures and make a visible contribution to programs such as 1+MG and GDI.

3.6 References

Fehse, B. et al. (Ed.) (2022): Im Fokus: RNA. Eine aktuelle Bestandsaufnahme der Arbeitsgruppe Gentechnologiebericht. Berlin Institute of Health at Charité, Berlin. DOI: 10.17169/refubium-36831.

Gymrek, M. et al. (2013): Identifying personal genomes by surname inference. In: *Science* 339(6117): 321–324. DOI: 10.1126/science.1229566.

International Human Genome Sequencing Consortium (2001): Initial sequencing and analysis of the human genome. In: *Nature* 409(6822): 860–921. DOI: 10.1038/35057062.

Liao, W.-W. et al. (2023): A draft human pangenome reference. In: *Nature* 617(7960): 312–324. DOI: 10.1038/s41586-023-05896-x.

Shabani, M. et al. (2016): Controlled access under review: Improving the governance of genomic data access. In: *PLoS Biology* 13(12): e1002339. DOI: 10.1371/journal.pbio.1002339.

Smetana, J./Brož, P. (2022): National genome initiatives in Europe and the United Kingdom in the era of whole-genome sequencing: a comprehensive review. In: *Genes* 13(3): 556. DOI: 10.3390/genes13030556.

The GHGA Consortium (2023): The German Human Genome-Phenome Archive – GHGA Brochure. Zenodo. DOI: 10.5281/zenodo.8359328.

Venter, J. C. et al. (2001): The sequence of the human genome. In: *Science* 291(5507): 1304–1351. DOI: 10.1126/science.1058040.

Walter, J./Gasparoni, N. (2021): Themenbereich Epigenetik: von Zellidentitäten bis hin zu Krankheiten und Therapien. In: Fehse, B. et al. (Ed.): Fünfter Gentechnologiebericht. Sachstand und Perspektiven für Forschung und Anwendung. Nomos, Berlin: 93–113. Unter: <https://www.nomos-elibrary.de/10.5771/9783748927242-93/3-themenbereich-epigenetik-von-zellidentitaeten-bis-hin-zu-krankheiten-und-therapien?page=1> [19.02.2024].

Walter, J./Schickl, H. (Ed.) (2019): Einzelzellanalyse in Forschung und Medizin – Eine Stellungnahme der interdisziplinären Arbeitsgruppe Gentechnologiebericht. Berlin Brandenburgische Akademie der Wissenschaften, Berlin. Unter: https://edoc.bbaw.de/files/3278/BBAW_Einzelzellanalyse_Walter_Schickl.pdf [19.02.2024].

Wilkinson, M. D. et al. (2016): The FAIR guiding principles for scientific data management and stewardship. In: *Scientific data* 3(1): 1–9. DOI: 10.1038/sdata.2016.18.

4. The ethical framework for the use of genomic data in research

Martin Jungkunz, Christoph Schickhardt and Eva C. Winkler

4.1 Introduction

The significance of modern genome sequencing procedures has risen sharply in the last few decades.¹ Biomedical research is inconceivable today without the deployment of genomic analytical procedures – in both basic research and clinical research. Genome analyses are also already being used in clinical care, e.g., in human genetic diagnostics in children with developmental delays and cancer medicine² in order to identify the genetic causes of a disease and inform treatment strategies. Due to the fact that the cost of genome sequencing and analysis has fallen considerably and that it is already clear today that genomic research and diagnostics are delivering visible benefits, it can be expected that genomic analytical procedures will become even more significant. Projects such as the genomDE³ model project demonstrate that the expansion of genomic analytical procedures will be incorporated into standard care in the foreseeable future. This model project envisages deploying genome analyses under defined standard conditions to diagnose rare and oncologic diseases and identify therapies.

The data created by genome analyses, genomic data, are extremely valuable for biomedical research – usually in combination with other clinical data. As large datasets of genomic data are required to discover rare correlations, there are numerous initiatives aimed at sharing genomic datasets and establishing genomics networks for research purposes.⁴ Some of these initiatives are aimed at establishing and providing an infrastructure in which data are archived and/or linked, and which are intended to provide researchers with access to large volumes of data. As of now, the data generally remain within the sphere of influence of the data generators (researchers, hospitals) who also have a say in deciding third parties' access to the data. Besides linking data with the aid of large infrastructure projects, data are also shared in and between research consortia. In both cases, genomic data, together with clinical data, which originate from healthcare or are directly collected in research projects, are usually shared.

¹ For further information on modern genome sequencing technologies, see Rieß/Schulze-Hentrich, Chapter 1.

² For genomic data in cancer diagnostics, see Schlomm/Biernath, Chapter 2.

³ See: <https://www.genom.de/> [15.12.2023].

⁴ See, for example, the "German Human Genome Phenome Archive" (GHGA), but also the European "1+ Million Genomes" initiative, and the "Global Alliance for Genomics and Health" (GA4GH). For ethical and legal framework conditions for infrastructure projects, see Schickhardt et al., 2023.

As substantial as the benefits are, it is equally important that research involving genomic data is conducted within a framework that takes account of ethical aspects and guarantees trustworthy governance.⁵ Starting from the ethical challenges posed by the use of genomic data for research, this chapter aims to outline such a framework.

4.2 Ethical framework conditions for genomic data used in research

The ethical challenges posed by the use of genomic data for research are closely linked to certain basic, specific properties of this kind of data. Here, “genomic data” refers to extensive data that originate from whole-genome or whole-exome sequencing, as well as data collected as part of very comprehensive gene panels.

On the one hand, genomic data are characterized by the fact that they are specific to each individual (with the exception of identical twins who share the same genetic material). In principle, therefore, it is possible to re-identify a person on the basis of their unique genomic data (Gymrek et al., 2013), although certain preconditions have to be met. A person can only be identified on the basis of their genomic dataset if third parties also have access to other personal data such as their real name or address together with the genetic data of this person. It is not possible to anonymize genomic data. Reference to a person is, at least potentially, inherent to the data. However, it is possible, and standard in research, to decouple the data from directly identifying information (e.g., by removing the name, date of birth, or address or replacing them with codes).

A further specific property of genomic data is their predictive power. This means that genomic data not only contain information on characteristics of the data subject that have already manifested (e.g., hair or eye color) or a disease that has already been diagnosed. Genomic data can also permit probability statements to be made about the potential emergence of diseases in the future.

A third specific property of genomic data is that they also permit statements on relatives of the data subjects, both in terms of manifest characteristics and with regard to potential predisposition to diseases. Genomic data can also serve to identify relatives.

4.2.1 Best usability and protection of patients through trustworthy governance

If we look at the three properties of genomic data mentioned (reidentification, predictive power, relevance to relatives), it becomes clear that these data are to be classified as highly sensitive. Consequently, their protection is of great importance. As with all procedures in which data are processed, data subjects’ right to informational self-determination, in our case patients or test subjects from whom the data originate, must be respected.⁶ This refers to the individual’s “authority to, in principle, decide themselves on the disclosure and use of their personal data” (BVerfGE 65, 1, 43). With regard to research with genomic data (or personal data in general), this right means firstly that the data can in principle only be used if the

⁵ GA4GH, for example, aims to establish such a governance framework (Rehm et al., 2021).

⁶ For the legal framework, see Müller-Terpitz/Straub, Chapter 5.

data subject has been informed of the particular use and has given their consent. Secondly, the right to informational self-determination includes further (micro)rights that remain and can be exercised after the beginning of the data use, e.g., the right to receive a copy of the data, to revoke their consent, or the right to have the data deleted.

4.2.1.1 Protection of the right to informational self-determination

The gold standard for participating in research is the consent of the subjects after they have been specifically informed about the research purposes and data processing by the researchers. When it comes to genomic research and other forms of data-driven medical research, such a procedure would be extremely complicated if data were to be used not only for a single project but also for further research projects, some of which are not even known at the time when the data are collected. One method of giving data subjects the chance to agree to any use of the data for every single research project even with data-intensive research – which to date has been the subject of predominantly theoretical discussion –, is the so-called “dynamic consent” model. In this model, prior to each use of their data by a specific project, data subjects receive information about that project electronically via an app and they can give or refuse their consent to such use. However, dynamic consent is very time-consuming to implement, makes it harder to use the data and entails data protection risks.⁷

In order to take account of data subjects’ wishes and thereby respect their right to informational self-determination without excessively obstructing research, so-called broad consent is therefore sometimes applied.⁸ Here, the data subjects are informed in advance of the genomic analysis that their data are to be used for research purposes and they can consent to such use (cf. e.g., EURAT, 2015). There is no specific information provided on individual research projects.

The use of broad consent is considered research-friendly, but it also means that data subjects are not able to exercise their informational self-determination in a fully informed state. In order to compensate for this restriction, it is ethically warranted to establish a supporting governance framework that creates the organizational, technical, legal, ethical, and procedural framework for data access and supervisory regulation in such a way that research with genomic data enjoys justified trust. In the process of informing data subjects and obtaining their consent, the risks and implications of using genetic data must be appropriately addressed and explained in addition to the potential benefits of the research. Sample texts for informing data subjects of the use of genomic data for secondary research can be found, for example, in the information material of the German Genome Archive GHGA (Bruns et al., 2022). This can be supplemented by an additional range of (online) information as developed by the authors for the Medical Informatics Initiative (MII).⁹

⁷ For different consent models for secondary use of data, see Wiertz/Boldt, 2022 as well as for their evaluation Jungkunz et al., 2022.

⁸ In a representative survey of cancer patients on the use of their clinical data in Baden-Württemberg, broad consent received the largest level of approval among the various consent models (Königeter et al., 2022).

⁹ See <https://www.medizininformatik-initiative.de/de/genetische-daten> [15.12.2023].

Consent is only part of the right to informational self-determination. In order to guarantee the aforementioned micro rights or their exercise (right of objection, data erasure, receipt of a copy of the data, etc.), additional technical, and organizational measures have to be put in place. This relates to specific, technical structures and the definition of responsibilities that facilitate the realization of the rights specified on a technical and organizational level. For example, there must be a defined procedure for deleting or releasing data at the request of the data subjects.¹⁰

4.2.1.2 Research-friendly data access with concurrent privacy protection

On the one hand, the governance framework for the use of genomic data for research purposes should, as mentioned, promote justified trust in research, but on the other, maintain the research benefit of the data at as high a level as possible. To achieve this, it is important to make the data accessible to as many researchers as possible with simple rules, but at the same time to guarantee the highest possible protection of the data. The FAIR principles act as a guideline for guaranteeing the highest possible research benefit of data: data are to be “findable,” “accessible”, “interoperable” and “reusable”. One initial step for making data findable is to provide access to the data via a central body to which researchers can turn if they wish to conduct research with genomic data. The second step is to provide high-quality metadata. This refers to a description of the data themselves, i.e., information on where they originate, by whom they were obtained, and for what use they have been approved. Good, i.e., extensive metadata are essential to enable researchers to identify data for their research projects and can be made available through an online portal, for example.

After researchers have identified the data with which they wish to do research, they must apply for access to the actual data. With respect to the authorization for data access, clear criteria must be communicated in advance which are verified when an application for data access and use is processed. Examples of such criteria include the approval of a research ethics committee and a legal basis for the use of the data. It must also be plausible that the requested data are required in accordance with the planned design of the study. Various arrangements are conceivable for the procedure for verifying and approving an application. For example, applications for data access are processed by the actual bodies releasing the data themselves in GHGA. However, in principle, it is also conceivable for applications to be verified by a central body. With both arrangements, it is conceivable that a data access committee will be used to process applications and indeed this is often standard practice as envisaged by the Medical Informatics Initiative (MII)¹¹ – an infrastructure for the secondary use of data from clinics. The data access committee should be comprised of representatives of different relevant scientific disciplines (e.g., genomics, biometrics, ethics, legal), as well as of data protection officers and representatives of the institution to which the committee is answerable (e.g., the relevant hospital).

¹⁰ For the release of raw genomic data, cf. EURAT, 2019.

¹¹ See: <https://www.medizininformatik-initiative.de/de/start> [15.12.2023].

4.2.2 Handling additional genetic findings

A further important aspect regarding the use of genomic data for research purposes is the possible discovery of so-called additional findings. These are findings with potential significance for health and/or reproduction that are discovered in the course of genetic or genomic analyses, but which are not connected to the intended question of the genome analysis and which were not actively sought (EURAT, 2023). According to the definition of the term, in the case of genomic data obtained in a purely research context, findings are deemed to be additional findings if they are not related to the original research question. With respect to the secondary use of genomic data from human genetics for research purposes, additional findings can occur if the existing data are analyzed for research purposes with a question or methodology other than those applied when they were collected in the context of the original patient care. Additional findings are ethically challenging because, in advance of any genome analysis, only very general information can be provided on different categories of additional findings, as in principle a large number of possible disease-relevant genetic variants might be discovered about which it would be simply too complicated to provide information in advance. This generic information conflicts with the human genetic principle of only reporting findings about whose characteristics and consequences the concerned person has been informed in advance, in order to enable him or her to decide either for or against reporting.

There has been a long-standing discussion in bioethics as to whether additional findings which arise in a research context, are to be reported or not and if so, which ones and how exactly. There is a large measure of agreement that prior to the genomic testing, subjects must be clearly informed whether additional findings will be reported and if so, what categories of additional findings. However, recommendations on how to deal with additional findings sometimes differ greatly, especially in an international comparison (Miller et al., 2022; GfH, 2023).

Actual additional findings should be reported through the attending physicians in each case as they are already in contact with the patients and know whether the findings in question may already be known. They also have to ensure that the finding is clinically validated and they can offer a genetic counseling which is urgently required and prescribed by law if genetic findings are reported. Depending on the infrastructure in which the genome research occurs, the reporting of additional findings in practice may turn out to differ widely. For example, if the research is carried out close to patient care, i.e., if there is a reasonable opportunity for researchers to send an additional finding to the attending physician, it is advisable to offer to report additional findings that reveal a predisposition to a treatable disease. In the case of large international research consortia, there may be no such opportunity to contact the attending physician, making reporting very difficult and thus unreasonable for researchers.

However, if it is possible to report additional findings within the particular research context, the reporting practice should be based on the patient's wishes (Winkler/Schickhardt, 2014; EURAT, 2015) or in the case of underage patients on the child's best interest (EURAT, 2023). The reporting decision to be made during the process of patient information and obtaining consent should be the responsibility of the patients (for underage children usu-

ally of the parents, taking the child's best interest into account). No additional findings should be "forced upon them" or withheld from them against their will. On the one hand, this focus on the wishes of the patient is warranted as this is the only way in which they can exercise their right to know or not know. On the other, any answer to the question of whether an additional finding should be reported or not always depends on an assessment of the benefits of reporting weighed against any possible harm caused. Even if empirical data strongly suggest that the psychological harm caused by genetic knowledge in general (Crozier et al., 2015) or additional findings in particular (Nambot et al., 2021) is usually less than may intuitively appear to be the case, nevertheless it can be assumed that any such potential harm may vary greatly from person to person and (adult) patients can best weigh up the potential harm against the benefit themselves.

4.2.3 Responsibility of researchers in genome research

The question must be posed as to what responsibilities researchers have when handling genomic data. Genome researchers take on an important role insofar as by analyzing genomic data, they have access to knowledge that may be of relevance to the health of the patient concerned. In some respects, therefore, researchers obtain medical knowledge without necessarily being bound by any professional medical ethics. By contrast, research ethics are based on different principles from medical ethics. Physicians have an obligation towards the individual patient and must treat them in accordance with good clinical practice. Researchers, on the other hand, have an obligation towards good scientific practice which puts the focus on the search for knowledge, and not on the responsibility for individual patients. This raises the question, to what extent researchers working with genomic data and thus potentially in possession of knowledge about a test subject's health, must meet certain aspects of medical ethics besides their professional scientific ethics. It makes sense to define the responsibilities of genome researchers in a code of conduct. This serves the purpose not only of making the rules of conduct available to everyone in written form. Depending on how such a code of conduct is implemented in an institution, it may gain normative, legal weight, e.g., if a research institute implements the code as a general instruction that researchers must adhere to. The EURAT Group from Heidelberg has drawn up a proposal for such a code (EURAT, 2015). This proposal defines the most important ethical principles on which researchers' actions in relation to patients (data subjects) should be based. These include:

- respect for autonomy
- beneficence and non-maleficence
- diligence
- non-discrimination and justice
- protection of privacy
- no rise to financial gains
- good scientific practice
- pursuit of public benefit
- protection of future generations

Besides these general principles, researchers should commit to certain important procedures. These include:

- risk assessment and communication with data subjects,
- obtaining informed consent and approval of ethics committee
- pseudonymization of data and samples
- establishment of access restrictions for the data and
- obligation to request that any project partners follow ethical guidelines.

4.3 Outlook

Against the background of the most important ethical framework conditions specified for the trustworthy use of genomic data for research purposes, it is particularly interesting to take a look at the current draft legislation. The Health Data Utilization Act (GDNG) and the Digital Act (DigiG) show the relevance of the subject of the secondary use of patient data. The aim of the GDNG is to collate different patient data, including genomic data, and to make them available for research. To do so, the GDNG does not provide for a consent model but instead an opt-out solution. Consequently, a large part of the responsibility shifts to a trustworthy governance framework, making it all the more relevant.

4.4 Conflict of interest

Both authors, C. Schickhardt and E. C. Winkler, have been active in the Medical Informatics Initiative (MII) and the German Human Genome and Phenome Archive (GHGA) for many years. M. Jungkunz states that there is no conflict of interest.

4.5 References

- Bruns, A. et al. (2022):** Consent modules for data sharing via the German Human Genome-Phenome Archive (GHGA). See: <https://www.ghga.de/resources/consent-modules> [15.12.2023].
- Crozier, S. et al. (2015):** The psychological impact of predictive genetic testing for Huntington's disease: a systematic review of the literature. In: *J Genet Couns* 24(1): 29-39. DOI: 10.1007/s10897-014-9755-y.
- EURAT (2015):** Cornerstones for an ethically and legally informed practice of whole genome sequencing, 2nd, updated Edition. Forum Marsilius Kolleg, Heidelberg.
- EURAT (2019):** On the release of raw genomic data to patients and study participants. Forum Marsilius Kolleg, Heidelberg.
- EURAT (2023):** Position paper on the return of additional genetic findings in minors. Forum Marsilius Kolleg, Heidelberg.
- GfH (2023):** Stellungnahme der Deutschen Gesellschaft für Humangenetik zu Zusatz- und Zufallsbefunden in der genetischen Diagnostik. In: *Medizinische Genetik* 35(4): 313-321. DOI: 10.1515/medgen-2023-2060.
- Gymrek, M. et al. (2013):** Identifying personal genomes by surname inference. In: *Science* 339(6117): 321-324. DOI: 10.1126/science.1229566.
- Jungkunz, M. et al. (2022):** Stellungnahme zur Etablierung der sekundären Forschungsnutzung von Behandlungsdaten in Deutschland. Ergebnisse des Verbundprojekts LinCDat: "Learning from clinical data. Ethical, social and legal aspects". Forum Marsilius Kolleg, Heidelberg.
- Köngeter, A. et al. (2022):** Patients' willingness to provide their clinical data for research purposes and acceptance of different consent models: Findings from a representative survey of cancer patients in Germany. In: *J Med Internet Res* 24(8): e37665. DOI: 10.2196/37665.

- Miller, D. T. et al. (2022):** ACMG SF v3.1 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). In: *Genet Med* 24(7): 1407-1414. DOI: 10.1016/j.gim.2022.04.006.
- Nambot, S. et al. (2021):** Incidental findings in a series of 2500 gene panel tests for a genetic predisposition to cancer: Results and impact on patients. In: *Eur J Med Genet* 64(5): 104196. DOI: 10.1016/j.ejmg.2021.104196.
- Rehm, H. L. et al. (2021):** GA4GH: International policies and standards for data sharing across genomic research and healthcare. In: *Cell Genom* 1(2). DOI: 10.1016/j.xgen.2021.100029.
- Schickhardt, C. et al. (2023):** Dateninfrastrukturen für die Gesundheitsforschung. Ethische Rahmenbedingungen und rechtliche Umsetzung. In: *Bundesgesundheitsbl* 66(2): 160-167. DOI: 10.1007/s00103-022-03648-2.
- Schoot, V. V. et al. (2021):** The impact of unsolicited findings in clinical exome sequencing, a qualitative interview study. In: *Eur J Hum Genet* 29(6): 930-939. DOI: 10.1038/s41431-021-00834-9.
- Wiertz, S./Boldt, J. (2022):** Evaluating models of consent in changing health research environments. In: *Med Health Care Philos* 25(2): 269-280. DOI: 10.1007/s11019-022-10074-3.
- Winkler, E. C./Schickhardt, C. (2014):** Ethische Herausforderungen der Genomsequenzierung in der translationalen Forschung und Antworten aus dem EURAT-Projekt. In: *LaboratoriumsMedizin* 38(4): 211-220. DOI: 10.1515/labmed-2014-0027.

5. The legal framework for the use of patient details and genome data – protection and benefits from a legal medical perspective

Ralf Müller-Terpitz and Hannah Lilly Straub

5.1 Protection of patients and their data by the GenDG (German Genetic Diagnostics Act)

In July 2021, the Fifth Volume of the Social Code Book (SGB V) was supplemented by the legal basis for a pilot project for comprehensive diagnostics and identification of the correct therapy by means of genome sequencing in the case of rare oncologic diseases.¹ The new § 64e (2) Sentence 2, (15) SGB V now stipulates that in the case of genome sequencing, the provisions of the Genetic Diagnostics Act (GenDG)² must be observed. The GenDG serves in particular to meet the state's duty to protect human dignity (Art. 1(1) GG) and the right to informational self-determination (Art. 2 (1) in conjunction with Art. 1 (1) GG).³ To do so, it attaches strict conditions to the conduct of genetic tests.

5.1.1 Informed consent through the provision of information

In order to protect patients' autonomy and their informational right of self-determination, the conduct of genetic tests and communication of the results obtained are subject to the consent (§ 8 (1) Sentence 1 and 2 GenDG) of the persons concerned. The condition for issuing valid consent is that patients must have been informed beforehand in accordance with Section 9 GenDG.⁴ Such information must put patients in a position to understand the test and potential consequences so as to enable them to weigh up the risks and benefits for themselves.⁵ § 9 (1 and 2) No. 1 GenDG therefore makes it a condition that the information provided must include the essence, significance and implications of the test as well as, in particular, its purpose, nature, scope, informative value and possible results. Potential findings extending beyond the purpose of the test (incidental findings) are also to be included.⁶ Informing the subject in advance of genome sequencing therefore represents a challenge. Genome sequencing reveals such a wealth of information that it is difficult to provide comprehensive information on all conceivable implications or even every single

¹ For genome sequencing in the context of cancers, see Schlomm/Biernath, Chapter 2.

² Genetic Diagnostics Act dated 2009/07/31 BGBl. I 2529 (Federal Law Gazette), last amended by Art. 15(4) of the Act dated 2021/05/04, BGBl. I 882.

³ For the right to informational self-determination from an ethical perspective, see Jungkunz et al., Chapter 4.

⁴ Stockter, in: Prütting, MedR, § 8 GenDG Note 12; Kern, in: Kern, GenDG, § 9 Note 1.

⁵ Guideline from GEKO (Genetic Diagnostics Committee) for the demands made on the substance of the information provided, 963(964 f.).

⁶ Guideline from GEKO for the demands made on the substance of the information provided, 963(965 f.).

possible result of the test (Molnár-Gábor, 2017: 168 ff.; EURAT, 2015: 70). Possible incidental findings are incalculable, ranging from carrying a tendency or predisposition to certain diseases to precise, predictive statements on the outbreak and course of later diseases. The challenge, therefore, is to find a practicable solution that in spite of the wealth of genetic information, guarantees that patients receive an idea of the test and its consequences and are able to base their informed consent upon it. The information must naturally cover more aspects than would be the case for a targeted genetic test. For example, attention should be drawn to the fact that the amount of information obtained is many multiples of what is required for diagnostic purposes. However, it should suffice to give concrete shape to the wealth of this information by way of examples, by categorizing possible results (e.g., treatable/non-treatable diseases) (Molnár-Gábor et al., 2014: 81/92; Fleischer, 2018: 86 f.; Duttge et al., 2019: 129/139). In this way, the patient's consent can also include whether they would like to be informed of incidental findings, and if so, what type of findings. However, it should be emphasized that it is neither practically possible nor legally necessary to specifically inform the patient of all possible results of genome sequencing (Rüping et al., 2022: 663/665 ff.; GEKO, 2013: 17). According to the case law established by the Federal Court of Justice (BGH), the only requirement is that patients "largely" understand what they are consenting to.⁷ As a secondary option, the purpose of the test can be extended, by way of the patient's consent, to finding genetic characteristics other than those of relevance to the original diagnosis. In this case, some additional gene sequences would be analyzed that otherwise would have been ignored due to their lack of relevance to the original diagnosis. However, that would then represent a separate genetic test for which all the requirements in terms of information, consent and counseling would have to be met. If this was a predictive, genetic test, the offer of genetic counseling (§ 10 GenDG) would also be mandatory (Rüping et al., 2022: 663/665).

5.1.2 Test results

Genome sequencing aims to produce findings for the diagnostic purpose targeted. As the original aim and reason for the test, they form part of the process of informing the patient and obtaining their consent. If effective consent has been obtained and has not been revoked, the results can be communicated to the patient in accordance with § 11 (1) and (4) in conjunction with § 8 (1 and 2) GenDG.

As already explained (5.1.1), however, numerous other diagnostic results can also be deduced from genome sequencing which might be of interest to the patient. According to the provisions of the GenDG quoted, such results may only be communicated if such communication is covered by valid consent. As the test is restricted by the consent, all other surplus information must therefore be hidden (Rüping et al., 2022: 663/666). To protect the so-called right of ignorance that follows from the right of informational self-determination, no genetic information may be forced upon patients. Any disclosure of incidental findings in spite of consent to their communication being originally refused may only be considered – if at all – in exceptional cases of serious but treatable diseases or diseases that put the

⁷ BGH, NJW 1977, 337; 1971, 1887; 1984, 1397 (1398); 2000, 1784 (1786).

general public in danger.⁸ If the original information provided has excluded the subject of incidental findings in contravention of obligations, hypothetical consent to communication of the incidental finding can possibly be assumed on these tightly defined conditions (serious disease that is hard to treat or that puts the general public in danger).⁹

5.1.3 Retention of results of genetic tests

Pursuant to § 12 (1) GenDG, the results of genetic tests must be retained for ten years. There is legitimate criticism that this retention period is very short compared with the informative value of genetic data (Vossenkuhl, 2013: 140). The results can be of benefit to the patients themselves or their dependents for diagnostic or therapeutic purposes even many years later. They are also quickly available and can help to prevent incurring sequencing costs.

Test results can nevertheless be retained for a longer period pursuant to § 12 (1) Sentence 3 GenDG if otherwise patients' sensitive interests would be compromised or patients so request. If there is hope that patients might be able to benefit from future research results, the simplest course of action against this background would be to obtain their consent for longer retention of the results. Otherwise, the responsible medical practitioner must come to the assessment that the deletion of test results will compromise patients' sensitive interests and therefore such results should not be destroyed.

5.2 Protection of genome data through legal data protection regulations

According to § 64e (2) Sentence 2 SGB V, data protection regulations are to be respected in addition to the regulations of the Genetic Diagnostics Act (GenDG). These regulations are relevant first and foremost for the processing of genome data for research purposes as processing for medical purposes is comprehensively covered by the GenDG. The data from genome sequencing constitute genetic data as defined by Art. 4 No. 13 of the General Data Protection Regulation (GDPR)¹⁰ and they therefore fall under the special category of personal data as defined by Art. 9 (1) GDPR.

5.2.1 Processing for research purposes

The processing of genomic data for research purposes can be legitimized on the one hand by explicit consent (Art. 9 (2)(a) GDPR) from the data subject; on the other, there are standards which permit processing for research purposes even without the consent of the data subject (so-called statutory processing facts – Art. 9 (2)(j) GDPR).

In the case of research-related data processing on the basis of consent, the question arises as to whether such consent can refer to several research projects not yet specified in

⁸ This is the – admittedly not uncontroversial – view of Duttge et al., 2019: 129/139; See also recommendations of the BMBF project group "Right of ignorance" (Duttge/Lenk, 2016: 402).

⁹ EURAT, 2023: 45 in the event of danger to life and limb if the parents of underage children have not been informed.

¹⁰ The (EU) 2016/679 Regulation of the European Parliament and the Council dated April 27, 2016 on the protection of natural persons with regard to the processing of personal data, the free movement of such data and repealing Directive 95/46/EC, ABI.EU No L 119 dated 04.05.2016.

detail (broad consent). Because under Art. 6 (1) (a) GDPR, consent must refer to “one or more specified purposes” (purpose limitation principle); and Art. 9 (2) (a) GDPR also states that the data subject must have consented to the processing of their genomic data “for one or more specified purposes”. As it is often not possible to fully identify the purpose of the research at the time when the data are collected, GDPR Recital No. 33 which can be used to interpret the relevant standards provides for a loosening of this purpose limitation principle to the extent that consent to “certain areas of scientific research” is to be permitted. This opportunity for “broad consent” is criticized, however, in sections of the legal literature as according to such critics, it does not permit the right of informational self-determination to be effectively exercised due to a lack of sufficient information. This procedure, they argue, must therefore remain restricted to tightly defined exceptions (Fröhlich/Spieker, 2022: 346/352 f.; Bronner/Wiedemann, 2023: 77/81).¹¹ As a general rule, however, consent to the processing of personal data for more closely defined areas of research should rightly suffice (Schreiber, 2019: 111 f. with further references).¹² The right of disposal over one’s own data, protected under the constitution, must also include the right to approve them for more broadly defined purposes provided this is done in the knowledge of gaps in information and details (Fleischer, 2018: 287 with further references) and consent can be revoked at any time.

On the basis of Art. 9 (2) (j) GDPR, national legislators also created legal grounds for processing. Freedom of research enjoys protection under the constitution both on a national and EU level (Art. 5 (3) GG [Constitution], Art. 13 CFR [Charter of Fundamental Rights of the European Union]). Against this, we have data subjects’ right to informational self-determination also protected in member state and EU law (Art. 2 (1) in conjunction with Art 1 (1) GG, Art. 8 CFR). If researchers have recourse to a statutory research clause, the processing of personal data for scientific purposes can be legitimized by weighing up the balance of interests which under certain conditions may turn out in favor of the research interest. In accordance with Art. 89 (1) GDPR, however, “suitable guarantees for the rights and freedoms of data subjects” must always be ensured in this case, for example by the pseudonymization of the data used.¹³

5.2.2 “Patchwork” of data protection research clauses

This data protection status quo is rightly criticized as lacking transparency due to the applicability of different regulatory systems. The GDPR offers a framework under European law which is given concrete shape in many areas by national provisions. In Germany, the question of which national regulations are applicable depends on who is processing the data and where this work will be performed. Depending on the answer to this question, either the provisions of the Federal Data Protection Act (BDSG) or those of each of the 16 state data protection laws must be observed. The aforementioned research clauses are not uniformly defined in these legal frameworks but differ in their preconditions. While § 27 (1) Sentence 1

¹¹ Also critical of the possibility of broad consent, Jungkunz et al., 2022: 92.

¹² The Medical Informatics Initiative sponsored by the BMBF drew up a consent template in 2020 (see also: <https://www.medizininformatik-initiative.de/de/muster-text-zur-patienteneinwilligung> [28.11.2023]). This template was also accepted by the Conference of Independent National and State Data Protection Officers (see DSK [Data Protection Conference] press release dated 24.04.2020).

¹³ Similar to the formulation in Art. 9 (2) (j) GDPR.

BDSG stipulates that the research interests must “significantly” outweigh the interests of the data subject, § 13 (1) LDSG BW¹⁴ dispenses with this requirement for significance. § 17 (1) DSG NRW¹⁵, on the other hand, states that sensitive issues pertaining to the data subject may not outweigh research interests and § 13 NSDG¹⁶ requires that sensitive interests may not obstruct or outweigh the public interest in research. This legal fragmentation and the fact that the preconditions for balancing research and the protection of personal data were not further outlined by legislators, lead to legal uncertainty when it comes to applying the standard (Weichert, 2020: 18 ff.; Specht-Riemenschneider/Radbruch, 2021: 1358/1359; Buchner, 2022: 555/559 ff.; Jungkunz et al., 2022: 95 ff.). The consequence is that researchers tend to bypass the system (Graf von Kielmansegg, 2022: 341/342; AKEK, 2023: 1 f.), although this reluctance seems to be unfounded most of the time.

Against this background, the EU is working on improving the legal situation. For example, the EU Commission published a draft for a European Health Data Space – EHDS in May 2022. The aim of this initiative, among other things, is to facilitate and simplify the primary and secondary use of health data for research purposes.¹⁷ Health system protagonists are to commit to making certain categories of electronic health data available, including genetic and genomic data.¹⁸ Approval to use the data is then to be obtained from an access body within a short period of time.¹⁹ Scientific research in the health or care sector is also specified as a potential processing purpose which applicants can pursue.²⁰ The planned regulation is not intended to cancel requirements of the GDPR, but rather to create a legal basis for the data processing as defined by Art. 9 (2) (j) GDPR. All other preconditions which the GDPR sets for the processing of personal data (including genomic data) are to remain unchanged.²¹ The draft legislation is still in its infancy, which is why it remains to be seen when and how it will take shape. Although the draft opens up the potential for a simpler and more uniform use of genome data for research purposes aside from declarations of consent (Tschammler/Uecker, 2022: 01265; AKEK, 2023: 2), nevertheless uncertainties might remain, e.g., in terms of its status vis-a-vis the GDPR (Raji, 2023: 3/6 f.). Following on from the draft EHDS, an initial draft from the Federal Ministry of Health (BMG) for a Health Data Utilization Act (GDNG) has been in place at a national level since March 2023.

5.3 Conclusion: the potential of genome data within the framework of legal requirements

Genomic tests can be used with the consent of patients for diagnostic and therapeutic purposes. To do so, the strict requirements of the Genetic Diagnostics Act (GenDG) must

¹⁴ State Data Protection Act of Baden-Württemberg.

¹⁵ Data Protection Act of North Rhine Westphalia.

¹⁶ Data Protection Act of Lower Saxony.

¹⁷ cf. Recital No. 1 EHDS-E.

¹⁸ Art. 33(1)(e) EHDS-E.

¹⁹ Art. 46(1 and 3) EHDS-E.

²⁰ Art. 34(1)(e) EHDS-E.

²¹ Recital No. 37 EHDS-E.

be observed, and present a particular challenge in terms of providing information the proper manner. Patients may only be informed of secondary findings that reveal themselves in addition to the originally intended results if they have at least previously consented to being informed of the relevant category of findings. There is also an option to extend the purpose of the test from the outset to further characteristics beyond the original purpose of the diagnosis.

As a general rule, the retention period for test results is ten years. A longer period of retention is also possible with the consent of the patients or if there is a prospect of the patients benefiting from new research insights.

Research with genome data can be conducted on the basis of patient consent, or based on statutory research clauses. The concept of “broad consent” can be relied upon for the consent, and for that reason the consent does not have to relate to specific research projects, but the data can be used more broadly for defined research areas. If the processing of genomic data is based on statutory research clauses, there are legal uncertainties caused among other things by the legal fragmentation of data protection regulations.

The possibility of simplifying the use of health data as part of medical research is currently on the political agenda. It remains to be seen whether the draft legislation on a European and national level will facilitate research in this field.

5.4 References

- AKEK = Arbeitskreis Medizinischer Ethikkommissionen (2023):** Stellungnahme zur Sekundärnutzung von Patientinnen- und Patienten-Daten zu Forschungszwecken in Deutschland, 09.08.2023. Unter: https://www.akek.de/wp-content/uploads/Stellungnahme-EHDS_V7.1.pdf [28.11.2023].
- Bronner, P./Wiedemann, F. (2023):** Rechtmäßigkeit der Datenverarbeitung bei wissenschaftlicher Forschung an staatlichen Hochschulen. In: ZD 2: 77–85.
- Buchner, B. (2022):** Forschungsdaten effektiver nutzen. In: DuD 9: 555–560. DOI: 10.1007/s11623-022-1658-8.
- DSK = Datenschutzkonferenz (2020):** Pressemitteilung der Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder vom 27.04.2020. Unter: https://www.datenschutzkonferenz-online.de/media/dskb/20200427_Beschluss_MII.pdf [14.12.2023].
- Duttge, G. et al. (2019):** Die genomische Hochdurchsatzsequenzierung. In: Duttge, G. et al. (Ed.): Next Generation Medicine. Mohr Siebeck, Tübingen: 129–146.
- Duttge, G./Lenk, C. (2019):** Empfehlungen zum anwendungspraktischen Umgang mit dem “Recht auf Nichtwissen”. Ergebnisse einer rechtsethischen Grundlagenanalyse auf erfahrungswissenschaftlicher Basis in den Anwendungsfeldern von Humangenetik und Psychiatrie. In: Duttge, G./Lenk, C. (Ed.): Das sogenannte Recht auf Nichtwissen. Brill, Leiden: 253–269. DOI: 10.30965/9783957437792_018.
- EURAT = Projektgruppe EURAT (2015):** “Ethische und Rechtliche Aspekte der Ganzgenomsequenzierung des menschlichen Genoms”: Stellungnahme Eckpunkte für eine Heidelberger Praxis der Ganzgenomsequenzierung. 2. Auflage, Heidelberg. Unter: https://www.uni-heidelberg.de/md/totalsequenzierung/stellungnahme_2.pdf [14.12.2023].
- EURAT = Projektgruppe EURAT (2023):** “Ethische und rechtliche Aspekte der Translationalen Medizin”: Stellungnahme zur Rückmeldung genetischer Zusatzbefunde Minderjähriger. Heidelberg. Unter: <https://journals.ub.uni-heidelberg.de/index.php/forum-mk/issue/view/5962> [14.12.2023].
- Fleischer, H. (2018):** Rechtliche Aspekte der Systemmedizin. LIT Verlag, Berlin.
- Fröhlich, W./Spiecker, I. (2022):** Die breite Einwilligung (Broad Consent) in die Datenverarbeitung zu medizinischen Forschungszwecken – der aktuelle Irrweg der MII. In: GesR 6: 346–353. DOI: 10.9785/gesr-2022-210605.

- GEKO = Gendiagnostik-Kommission (2013):** Erster Tätigkeitsbericht der Gendiagnostik-Kommission. Unter: https://www.rki.de/DE/Content/Kommissionen/GendiagnostikKommission/Taetigkeitsbericht/Taetigkeitsbericht_01.pdf?__blob=publicationFile [14.12.2023].
- Graf von Kielmansegg, S. (2022):** § 27 BDSG und die medizinische Forschung. In: GesR 6: 341-346. DOI: 10.9785/gesr-2022-210604.
- Jungkunz, M. et al. (2022):** Stellungnahme der Etablierung der sekundären Forschungsnutzung von Behandlungsdaten in Deutschland. Ergebnisse des Verbundprojekts LinCDat, Heidelberg. Unter: <https://journals.ub.uni-heidelberg.de/index.php/forum-mk/article/view/91697> [26.10.2023].
- Kern, B.-R. (2021):** Gendiagnostikgesetz Kommentar. C.H. Beck, München.
- Medizin-Informatik-Initiative (2020):** Mustereinwilligung (16.04.2020). Unter: <https://www.medizininformatik-initiative.de/de/mustertext-zur-patienteneinwilligung> [28.11.2023].
- Molnár-Gábor, F. et al. (2014):** Das Ende der Zufallsbefunde in der genetischen Diagnostik: Die Ganzgenomsequenzierung und die Erwartbarkeit zusätzlicher Befunde. In: Zeitschrift für Medizin-Ethik-Recht 2: 81-105.
- Molnár-Gábor, F. (2017):** Die internationale Steuerung der Biotechnologie am Beispiel des Umgangs mit neuen genetischen Analysen. Duncker & Humblot, Berlin.
- Prütting, D. (2022):** Medizinrecht Kommentar. 6. Auflage. Luchterhand, Köln.
- Raji, B. (2023):** Datenräume in der Europäischen Datenstrategie am Beispiel des European Health Data Space. In: ZD 1: 3-8.
- Rüping, U. et al. (2022):** Ärztliche Aufklärungspflichten über die Möglichkeit von Nebenbefunden in der genomischen Medizin. In: MedR 40: 663-667. DOI: 10.1007/s00350-022-6273-z.
- Schreiber, M. (2019):** Die medizinische Forschung mit abgetrennten Körpersubstanzen Minderjähriger. LIT Verlag, Münster.
- Specht-Riemenschneider, L./Radbruch, A. (2021):** Forschung braucht Daten. In: Deutsches Ärzteblatt 27-28: A 1358-1361.
- Tschammler, D./Uecker, P. (2022):** Verordnungsentwurf zur Schaffung eines Europäischen Gesundheitsdatenraums – erster Überblick mit Fokus auf die Nutzung von Gesundheitsdaten für Forschungs- und Entwicklungsvorhaben. In: ZD-Aktuell: 01265.
- Vossenkuhl, C. (2013):** Der Schutz genetischer Daten. Springer, Heidelberg.
- Weichert, T. (2020):** Die Forschungsprivilegierung in der DS-GVO. In: ZD 1: 18-24.

Outlook: Charting the course to forward-looking human genome diagnostics and genome research

Jörn Walter

The 2024 genome sequencing pilot project marks a turning point in human genome research in Germany.¹ For the first time, there is a major opportunity for Germany to take on a pioneering role in genome research by, on the one hand, establishing a direct link between genome sequencing and healthcare and, on the other, putting these comprehensive genome data (consent permitting) to use for researching and investigating complex diseases, as well as for the development of new drugs and diagnostic procedures. This will result in synergies and far-reaching opportunities for modern, forward-looking genome diagnostics and functional medical genome research.

This development comes at an opportune moment. Thanks to the deployment of the latest technologies (primarily 3rd generation sequencing equipment), genomes can be analyzed more precisely and more comprehensively than previously. Moreover, these new technologies allow reading and mapping genetic and epigenetic changes at the same time. This extends the opportunities for diagnostics and research by essential parameters. Besides these technical advancements, new network data infrastructure will be generated facilitating coordinated, secure archiving and data use across Germany. This coordinated data structure will enhance the opportunities for genome research in Germany.²

The coordinated archiving of genomes enables comprehensive, population-related comparisons to be made for the first time, thereby promising a marked improvement in the identification of genetic variants that cause or boost diseases. The planned and urgently needed networking of a German archive with other national archives on a European level will significantly expand these opportunities.³

In the future, a qualified evaluation of genetic alterations/mutations will require the inclusion and use of functional genome data, such as single-cell, epigenomic, proteomic, and transcriptomic data. The generation of a national human genome atlas, i.e. a compendium

1 See: <https://www.bundesgesundheitsministerium.de/en/en/international/european-health-policy/genomde-en> [18.03.2024]; <https://www.aerzteblatt.de/nachrichten/144447/Modellvorhaben-zur-Genomsequenzierung-soll-2024-starten> [18.03.2024]; <https://www.genom.de> [18.03.2024].

2 See: <https://www.bundesgesundheitsministerium.de/en/en/international/european-health-policy/genomde-en> [18.03.2024]; <https://www.ghga.de/about-us/mission> [18.03.2024].

3 See: <https://www.youtube.com/watch?v=wYDJXEkzjw> [18.03.2024].

of population-related genome data that supports complex interpretations would be an ideal place for such additional data. In addition, clinical “phenome” data relevant for research will be added to this atlas (with consent and for controlled use). In this way, the atlas will open up a new level of in-depth analytical opportunities for the clinical diagnosis of individual patients. The comprehensive atlas data should be accessible through a data-secure space, e.g. a secure cloud, allowing to perform complex AI-assisted operations to decode the effect of mutations and hence to support personal diagnosis and treatment.

The basis for all these options is that personal genome data are obtained voluntarily and on the foundation of informed consent and securely stored in standardized form in controlled, non-commercial data archives. Such protected networks must enable data to be stored in standardized formats over the long term. Security can be guaranteed by various levels of encryption and by a surveilled network architecture. At the same time, the usability of the data should not be restricted in order to harness their value added for biomedical research. Clear rules are required for this such as controlled mechanisms and conditions for access and usage but also rules and controls for handling data. This requires a continuous supporting development of a legal and ethical framework.

New specialist personnel trained in interdisciplinary skills will be needed for the production, processing, and evaluation of complex genetic data. The training of corresponding specialist genomic expertise, particularly in the field of specialist medical training (i.a. human genetics, oncology) and medical bioinformatics (database management, AI-assisted analyses), must be quickly and decisively expanded.

It is now time to communicate and discuss the need for modern, responsible, trailblazing genome diagnostics and research more broadly than has previously been the case. The findings made by genome research over the last 20 years have fundamentally extended the overall view of the genome and its significance for complex physiological, biochemical, and medical processes. This knowledge should be used and deployed for the diagnosis and treatment of diseases of genetic origin, but also for general healthcare. It is of great importance here to promote an interdisciplinary dialog between patients and patients’ representatives, clinicians, healthcare providers, scientists, and official regulators on different levels in order to thereby discuss substantive, legal, and ethical aspects to support the process and if necessary, to make adjustments.

In charting the course to modern genome diagnostics, there is also a need for a continuous, supporting dialog with the general public and society as a whole. The Working Group *Gene Technology Report* would like to contribute toward such a dialog with this brochure. The successful implementation and acceptance of the pilot project will depend on transparent educational public relations work geared to increasing social acceptance of genome diagnostics and genome research and allaying any (remaining) fear of the dangers of a “gläserner Mensch” (“transparent/unprotected human”). One central condition for securing the trust of patients and the public is the long-term guarantee of data security and data protection in a technical, legal, and ethical context.

Diese Bestandsaufnahme ist online abrufbar unter:
<https://www.gentechnologiebericht.de/publikationen>
This stocktaking is available online at:
<https://www.gentechnologiebericht.de/en/publications>

ISBN: 978-3-00-078676-1