

Big data in der Diagnostik genetischer Schwerhörigkeit

A Big Data Perspective on the Genomics of Hearing Loss



Autoren

Barbara Vona, Marcus Müller, Saskia Dofek, Martin Holderried, Hubert Löwenheim, Anke Tropitzsch

Institut

Klinik für Hals-, Nasen- und Ohrenheilkunde, Eberhard Karls Universität, Universitätsklinik Tübingen

Schlüsselwörter

Big data, Genetik, Genomik, *GJB2*, Hochdurchsatz-Sequenzierung, Genetik, Schwerhörigkeitsdiagnostik, Varianteninterpretation

Key words

Big data, genetics, genomics, *GJB2*, hearing loss diagnostics, high-throughput sequencing, variant interpretation

Bibliografie

DOI <https://doi.org/10.1055/a-0803-6149>

Laryngo-Rhino-Otol 2019; 98: S32–S57

© Georg Thieme Verlag KG Stuttgart · New York

ISSN 0935-8943

Korrespondenzadresse

Prof. Dr. med. Hubert Löwenheim

Univ. HNO-Klinik

Elfriede-Aulhorn-Straße 5

D-72076 Tübingen

hubert.loewenheim@med.uni-tuebingen.de

ZUSAMMENFASSUNG

Die vollständige Sequenzierung des menschlichen Genoms demonstriert als ein grundlegendes Beispiel eindrucksvoll die Entstehung einer großen Datenmenge (engl.: big data) in Wissenschaft und Medizin. Die Entschlüsselung des menschlichen Genoms stellt das bemerkenswerte Ergebnis multidisziplinärer Zusammenarbeit dar und gilt als eines der größten und erfolgreichsten Vorhaben der Menschheitsgeschichte. Die Bedeutung dieser Entdeckung lag nicht nur darin, die Sequenz von 3,2 Milliarden Nukleotiden des humanen Genoms zu identifizieren, sondern in Zukunft auch krankheitsassoziierte Variationen zu verstehen und dieses Wissen auf individualisierte Behandlungsansätze der personalisierten Medizin anzuwenden. Die Genomik hat sich seitdem mit bemerkenswerter Geschwindigkeit weiterentwickelt. Hierzu haben im Wesentlichen digitale, technologische Fortschritte in der Sequenzierung, Computer- und Bioinformatik wesentlich beigetragen. Die dadurch entstandenen großen genomischen Datenmengen haben den Begriff „big data“ hervorgebracht. Die

heutige bioinformatisch geleitete Einzelfallanalyse genetischer Befunde im Krankheitskontext erfordert in der Regel die Verwendung mehrerer großer Datenmengen. Diese Datenmengen liegen in Form von strukturierten genetischen Datenbanken vor und werden bspw. im Rahmen von in silico Analyseprogramme und Allel-Häufigkeitsanalysen verwendet. Die aktuellen Technologien der Hochdurchsatzsequenzierung sind in der Lage kostengünstige und qualitativ hochwertige Daten zu erzeugen. Dies reicht von der Analyse mit gezielten krankheitsassoziierten Gen-Panels, über die Exom Analyse, bis hin zur Entschlüsselung des gesamten Genoms. Diese neuen Möglichkeiten haben die Diagnostik von Erbkrankheiten revolutioniert und wirken sich auf die Diagnostik der genetischen Schwerhörigkeit aus.

Die Analyse der genetischen Grundlagen der vererbaren Form des Hörverlusts ist aufgrund großer genetischer Heterogenität und klinischer Variabilität in 2-facher Hinsicht eine besondere Herausforderung. Es sind bereits über 150 Gene bekannt, die an nicht-syndromalen und syndromalen Formen des Hörverlusts beteiligt sind. Das Mutationsspektrum eines einzelnen Hörverlust-assoziierten Gens kann mehrere zehn bis hunderte von pathogenen Varianten aufweisen. Darüber hinaus kann die Interpretation neuer Varianten eine Herausforderung darstellen, insbesondere, wenn widersprüchliche Informationen in Datenbanken hinterlegt wurden. Detaillierte und strukturierte phänotypische Informationen haben sich in der Diagnostik einiger Formen des Hörverlusts als äußerst vielversprechend erwiesen, sind aber bisher nicht für alle genetischen Formen von Schwerhörigkeit nutzbar. Während mit enormer Geschwindigkeit ständig neues Wissen sowohl im diagnostischen als auch im wissenschaftlichen Kontext entsteht, stellt diese überwältigende Menge an Informationen eine zunehmende Herausforderung für Fachärzte dar. Die fachärztliche Versorgung übernimmt hier neue Aufgaben und fungiert als Schnittstelle zwischen dem humangenetisch-diagnostischen Labor und dem Patienten. Zu diesen Aufgaben gehört die fachbezogene genetische Beratung und die klinische Einordnung von genetischen Befunden.

Diese Übersicht soll als Referenz für HNO-Ärzte dienen, die einen Einstieg in die Molekulargenetik der Schwerhörigkeit erhalten möchten. Es erfolgt die Darstellung von Schlüsselkonzepten der molekulargenetischen Diagnostik. Gerade die komplexen Prozesse, die der Identifizierung und Interpretation von genetischen Varianten zugrunde liegen, wären ohne die die enormen zur Verfügung stehenden Datenmengen nicht denkbar. Insofern sind „big data“ unabdingbare Voraussetzung, um genetische

Daten im konkreten Einzelfall zu filtern und gerade für den klinisch tätigen Arzt im Kontakt mit dem Patienten überschaubar und nutzbar zu machen.

ABSTRACT

The completion of the human genome, the most fundamental example of big data in science and medicine, is the remarkable product of multidisciplinary collaboration and is regarded as one of the largest and most successful undertakings in human history. Unravelling the human genome means not only identifying the sequence of its more than 3.2 billion nucleotide bases, but also understanding disease-associated variations and applying this knowledge to patient-tailored precision medicine approaches. Genomics has moved at a remarkable pace, with much of the propelling forces behind this credited to technological developments in sequencing, computing, and bioinformatics, that have given rise to the term "big genomics data." The analysis of genetics data in a disease context involves the use of several big data resources that take the form of clinical genetics data repositories, in silico prediction tools, and allele frequency databases. These exceptional developments have cultivated high-throughput sequencing technologies that are capable of producing affordable high-quality data ranging from targeted gene panels to exomes and genomes. These new advancements have revolutionized the diagnostic paradigm of hereditary diseases including genetic hearing loss. Dissecting hereditary hearing loss is exceptionally challenging

due to extensive genetic and clinical heterogeneity. There are presently over 150 genes involved in non-syndromic and common syndromic forms of hearing loss. The mutational spectrum of a single hearing loss associated-gene can have several tens to hundreds of pathogenic variants. Moreover, variant interpretation of novel variants can pose a challenge when conflicting information is deposited in valuable databases. Harnessing the power that comes from detailed and structured phenotypic information has proven promising for some forms of hearing loss, but may not be possible for all genetic forms due to highly variable clinical presentations. New knowledge in both diagnostic and scientific realms continues to rapidly accumulate. This overwhelming amount of information represents an increasing challenge for medical specialists. As a result, specialist medical care may evolve to take on new tasks and facilitate the interface between the human genetic diagnostic laboratory and the patient. These tasks include genetic counselling and the inclusion of genetics results in patient care.

This overview is intended to serve as a reference to otolaryngologists who wish to gain an introduction to the molecular genetics of hearing loss. Key concepts of molecular genetic diagnostics will be presented. The complex processes underlying the identification and interpretation of genetic variants in particular would be inconceivable without the enormous amount of data available. In this respect, "big data" is an indispensable prerequisite for filtering genetic data in specific individual cases and making it clear and useful, especially for clinicians in contact with patients.

Inhaltsverzeichnis

	Zusammenfassung	32
	Abstract	33
1.	Glossar	34
2.	Big data im Zeitalter der Genomik	34
2.1	Genetische Variation – gutartig oder pathogen?	35
3.	Wegbereiter für die Genomik-Revolution	35
4.	Entwicklung von Hochdurchsatz-Sequenziertechnologien	36
5.	Die Genetik des Hörverlustes	39
6.	Veränderung der Diagnostik für Hörverlust	41
6.1	Gen-Panel-Diagnostik bei Hörverlust	42
6.2	Exom-Diagnostik bei Hörverlust	42
6.3	Vor- und Nachteile von Gen-Panels und exombasierten Diagnostikansätzen	42
6.4	Diagnostische Aufklärungsraten	43
7.	Computer Ressourcen	44
8.	Hochdurchsatz-Sequenzierungsanalyse	48
9.	Beispiel einer Variantenanalyse aus <i>GJB2</i>	50
10.	Vom Genom zum Phänom	53
11.	Die Perspektiven der Hochdurchsatz-Sequenzierung	54
	Literatur	54

ABKÜRZUNGSVERZEICHNIS

A	Adenin
C	Cytosin
CADD	engl. Combined Annotation Dependent Depletion
CCIC	engl. Comprehensive Cochlear Implant Center
CNV	engl. copy number variation
<i>COL11A2</i>	engl. collagen type XI, alpha-2
ddNTP	Didesoxynukleotidtriphosphat
DFNA2A	engl. deafness, autosomal dominant 2 A
DFNA3A	engl. deafness, autosomal dominant 3 A
DFNA6/14/38	engl. deafness, autosomal dominant 6/14/38
DFNA13	engl. deafness, autosomal dominant 13
DFNB1A	engl. deafness, autosomal recessive 1 A
DFNB16	engl. deafness, autosomal recessive 16
dNTP	Desoxynukleotid
DVD	engl. Deafness Variation Database
E	Embryonales Entwicklungsstadium
EVS	engl. Exome Variant Server
ExAC	engl. Exome Aggregation Consortium Browser
G	Guanin
Gb	Gigabase
<i>GJB2</i>	engl. gap junction protein beta 2
<i>GJB6</i>	engl. gap junction protein beta 6
GME	engl. Greater Middle Eastern Variome

gnomAD	engl. genome aggregation database
HGMDB	engl. Human Gene Mutation Database
HGP	engl. Human Genome Project
HPO	engl. Human Phenotype Ontology
KCNQ4	engl. Potassium voltage-gated channel subfamily KQT member 4
LOVD	engl. Leiden Open Variation Database
MAF	engl. minor allele frequency
NIH	engl. National Institutes of Health
mRNA	engl. messenger ribonucleic acid
MYO1A	engl. myosin IA
P	Postnatales Entwicklungsstadium
PCR	Polymerase-Kettenreaktion
SHIELD	engl. Shared Harvard Inner-Ear Laboratory Database
SIFT	engl. Sorting Intolerant from Tolerant
STRC	engl. stereocilin
T	Thymin
WFS1	engl. wolframin ER transmembrane glycoprotein

1. Glossar

Abdeckung Die Sammlung von ausgerichteten Sequenzierungslesungen (engl. read) über einem Nukleotid oder einem Zielbereich.

Ablesungen Ein kurzes Fragment der Sequenz (engl. Reads).

Autosom Chromosom, das nicht zu den Gonosomen (Geschlechtschromosomen) gehören.

Baits „Köder“, Capture-Sonden, die aus Oligonukleotiden bestehen, die zu einer Zielregion für die Sequenzierung komplementär sind.

CFTR Ein Gen, das für einen Chloridkanal kodiert und an der autosomal rezessiven Mukoviszidose beteiligt ist.

Desoxynukleotid Modifiziertes Desoxynukleotid, dem eine 3' Hydroxylgruppe fehlt, um die DNA-Kettenverlängerung bei der Sanger-Sequenzierung zu hemmen.

DNA-Bibliothek Eine Sammlung von amplifizierten DNA-Fragmenten für die Hochdurchsatz-Sequenzierung.

Exom Der Teil des Genoms, der potentiell für Proteine kodiert.

Exomsequenzierung Sequenzierung aller Exons in kodierenden Genen.

Exon Eine Region eines Gens, das für ein Protein kodiert.

Genom Der komplette DNA-Satz in einem Organismus.

Gigabase 10⁹ Nukleotide.

GJB2 Ein Gen, das für das gap-junction-beta-2 (Connexin-26) kodiert, das einen autosomal-rezessiven (DFNB1A) und autosomal-dominanten (DFNB3A) Hörverlust verursacht.

Hochdurchsatz-Sequenzierung Ein skalierbares und relativ kostengünstiges Sequenzierungsverfahren, das von Gen-Panels bis zu kompletten Genomsequenzierung reichen kann.

Indel Ein Begriff für die Insertion oder Deletion einer oder mehrerer Basen in einem Genom.

In silico Paneldiagnostik Ein Berechnungsfilter, der auf Exom- oder Genomsequenzierungsdaten angewendet wird, der die Varianten für die Analyse in einer ausgewählten Teilmenge von Genen einschränkt.

In silico Vorhersageprogramme Computerwerkzeuge, die die Pathogenität von Varianten vorhersagen.

Intron Eine nicht kodierende Region eines Gens zwischen zwei kodierenden Exons.

Kilobase 1,000 Nukleotide.

Kopienzahl-Variation Deletionen oder Duplikationen von chromosomalen Regionen, die die Anzahl der Genkopien beeinflussen.

Lücken Eine Region, die während der Sequenzierung meist aus technischen Gründen schlecht abgedeckt oder übersehen wird.

Megabase 1,000,000 Nukleotide.

missense-Veränderung Eine Nukleotidsubstitution, die eine Aminosäure verändert.

Mooresches Gesetz Eine Beobachtung, dass sich die Anzahl der Transistoren auf einer dichten integrierten Schaltung alle 2 Jahre verdoppelt und damit die Kosten für Transistoren halbiert werden.

nicht-synonyme Veränderung Eine Nukleotidsubstitution, die die Aminosäuresequenz verändert.

Niedrige Allelfrequenz Die Häufigkeit des weniger verbreiteten Allels.

nonsense-Veränderung Nukleotidsubstitution, die zu einem vorzeitigen Stoppcodon während der Transkription führt.

Panel-Diagnostik Sequenzierung ausgewählter Gene, die für eine bestimmte Krankheit relevant sind.

Phänome Die umfassende Beschreibung des Phänotyps und des Krankheitsverlaufes in einem Individuum.

Sanger-Sequenzierung Eine Art der Sequenzierung, die ein Kettenverlängerungsabbruch mit chemisch modifizierten Dideoxynukleotiden verwendet. Mit dem Verfahren wird eine DNA-Sequenzierung ermöglicht.

Spleiß-Veränderung Eine Variante, die das normale Gen-Spleißen während der Übersetzung beeinflusst.

Start gain variant Eine Variante, die eine neue Übersetzungsinitiationsstelle verursacht.

Start loss variant Eine Variante, die die normale Übersetzungsinitiationsstelle stört.

Stop gain variant Eine Variante, die zu einem vorzeitigen Stoppcodon während der Transkription führt.

Stop loss variant Eine Variante, die das Terminator-Codon entfernt und zu einem verlängerten Transkript führt.

STRC Ein Gen, das für Stereocilin kodiert, ein Strukturprotein in der Stereozilien der äußeren Haarzellen des Innenohrs, das einen autosomal-rezessiven Hörverlust verursacht (DFNB16).

synonyme Veränderung Nukleotidsubstitution, die die Aminosäuresequenz nicht verändert.

Terabase 10¹² Nukleotide.

Variante Eine Abweichung von der Referenzsequenz.

Zusatzbefund Ein genetisches Testergebnis, das nichts mit der primären Krankheitsindikation zu tun hat.

2. Big data im Zeitalter der Genomik

„Big data“ ist mittlerweile ein allgegenwärtiger Begriff im Gesundheitswesen. Für die Bereiche Genetik und Genomik ist diese „große Datenrevolution“ geradezu tragend. Es erscheint mittlerweile unmöglich, aussagekräftige Ergebnisse zu gewinnen und diese zu interpretieren, ohne auf die großen Datenmengen – „big data“ – der

in zahlreichen Datenbanken gespeicherten genomischen Informationen zuzugreifen. Das grundlegendste Beispiel für „big data“ in diesem Bereich ist die menschliche Genomsequenz, die im einfachsten Sinne als Blaupause für die Aufklärung der DNA-Sequenz der mehr als 20 000 Gene im menschlichen Genom dient. Im Jahr 2003 wurde die vollständige Sequenzierung des menschlichen Genoms im Rahmen des Humangenomprojekt (engl.: Human Genome Project; HGP) offiziell erreicht. Die Vollendung des HGP stellt einen bemerkenswerten Meilenstein dar, der das Verständnis der genetischen Variation im menschlichen Genom überhaupt erst umfassend ermöglicht und weiter beschleunigt hat. Diese enorme internationale Forschungsanstrengung hatte neben den beiden im Vordergrund stehenden Zielen alle Gene des Menschen zu identifizieren und die gesamte Sequenz der 3,2 Milliarden Basenpaare der DNA des Menschen zu analysieren auch technologische Aufgaben – nämlich relevante Technologien für die Sequenzierung und Datenanalyse zu finden. Dies führte mit einer gewissen zeitlichen Verzögerung etwa ab dem Jahr 2006 zu einer sich eher im Hintergrund vollziehenden, technologischen Revolution insbesondere im Bereich der Sequenzier-Technologien. In Folge konnte von der generischen Analyse des humanen Genoms auf die Analyse einer hohen Anzahl individueller Einzelgenome und funktioneller Elemente übergegangen werden. Aus dem HGP entwickelten sich daher in Folge weitere Projekte. Seit 2003 verfolgt das Projekt ENCODE (**ENC**yclopedia **Of DNA Elements**) des National Institutes of Health (NIH) das Ziel, die funktionalen Elemente des menschlichen Genoms zu identifizieren und deren Rolle für die Genregulation zu beschreiben. Das im Jahr 2006 initiierte Projekt PGP (**P**ersonal **G**enome **P**roject) möchte allen Menschen den Zugang zu ihrem Genom verschaffen, um damit eine individualisierte medizinische Behandlung zu ermöglichen. Zunächst sind mindestens 100 000 Teilnehmer vorgesehen. Im Rahmen des Projekts 1000-Genomes konnte bereits im Jahr 2012 ein Katalog genetischer Varianten von mehr als 1000 Personen aus weltweit 14 verschiedenen Populationen veröffentlicht werden. Im Jahr 2015 erfolgte eine Erweiterung auf 2504 Personen. Im dem seit 2008 aktiven, spezifisch auf Krebserkrankungen ausgerichteten Internationalen Krebsgenom-Konsortium (International **C**ancer **G**enome **C**onsortium; ICGC, <http://icgc.org>) sollen 50 Tumorarten von 25 000 Patienten umfassend molekular analysiert werden. Alle diese Datenbanken tragen ständig neue „big data“ Bausteine bei und verbessern fortlaufend unser Verständnis der normalen und der krankheitsassoziierten genetischen Vielfalt.

Die Variationen im menschlichen Genom tragen nicht nur zu unseren einzigartigen Eigenschaften bei, sondern bestimmen häufig auch die Chancen für eine gezielte Behandlung im Krankheitsfall. Das menschliche Genom ist das Ergebnis von Generationen von Migration, Selektion und Anpassung. Natürlich auftretende Fehler während der Meiose können sowohl kleine als auch große Veränderungen, die als genetische Variation bezeichnet werden, in unser Genom einbringen. Hiervon können viele als gutartig oder polymorph angesehen werden, während andere Veränderungen auslösend für Erkrankungen und damit pathogen sein können. Diese Veränderungen können einzelne Nukleotide, auch als Basen bezeichnet (Adenin (A), Thymin (T), Guanin (G) und Cytosin (C)), oder mehrere Millionen Nukleotide im Genom betreffen. Veränderungen können aber auch ganze Chromosomen (z. B. Monosomie, Trisomie) einbeziehen oder den Austausch von genetischem Material innerhalb verschiedener

Teile eines einzelnen Chromosoms (intrachromosomale Rekombination) oder zwischen verschiedenen Chromosomen (interchromosomale Rekombination) beschreiben.

2.1 Genetische Variation – gutartig oder pathogen?

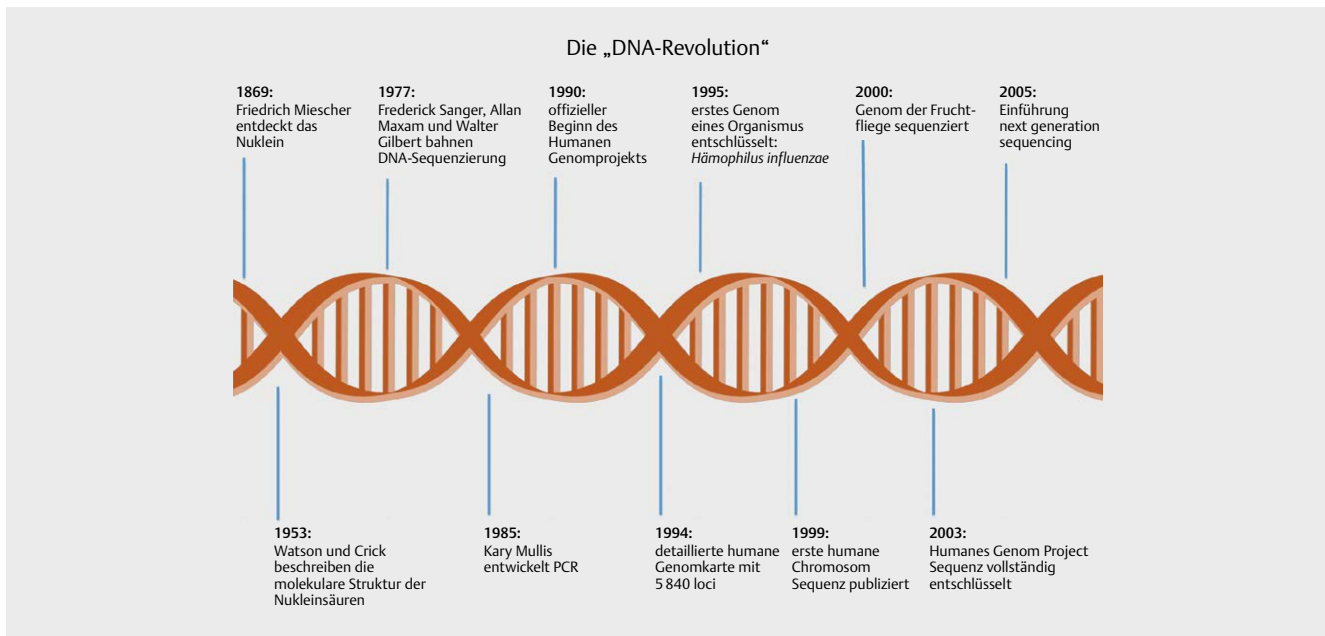
Das Exome Aggregation Consortium (ExAC) hat die Exomsequenzen (Protein-kodierende Sequenz) von 60 706 Individuen zusammengetragen und stellt diese Daten über eine Internet-basierte Plattform insbesondere auch im Hinblick auf die Frequenz und funktionelle Bedeutung von genetischen Varianten zur Verfügung. Aus dieser bisher größten Exom-basierten Sequenzierstudie wird abgeschätzt, dass beim Menschen in der Protein-kodierenden Sequenz das Äquivalent von einer Variante pro acht Nukleotidpositionen vorliegt [1]. Die Entschlüsselung von Variations-Mustern ist schwierig, da die Variation einer ungleichmäßigen Verteilung folgt und die Dichte der Variation durch Mutationseigenschaften und selektive Auswahl beeinflusst wird.

Eine weitere wichtige Form der genetischen Variation ist die Kopienzahl-Variation (engl.: Copy Number Variation; CNV). Die CNVs sind definiert als Duplikationen und Deletionen eines DNA-Abschnitts mit 50 bis zu 3 000 000 Basenpaaren. Auf Basis dieser Art der genetischen Variation können bis zu 9,5% des Genoms bei gesunden Menschen variieren und an „Gewinnen“ (Genduplikationen) oder „Verlusten“ (Gendeletionen) beteiligt sein [2]. Folglich können die 3,2 Milliarden Nukleotide, die sich normalerweise in unserem Genom befinden, um $\pm 9,5\%$ variieren. Dies zeigt eindrucksvoll die Widerstandsfähigkeit des menschlichen Genoms auch gegenüber z. T. großen Veränderungen.

Eine der grundlegenden Aufgaben der klinischen Interpretation genomischer Daten ist die Differenzierung zwischen normaler und pathogener Variation [3–9]. In den letzten 15 Jahren seit der Entschlüsselung der menschlichen Genomsequenz hat die Genomforschung damit begonnen, Toleranz und Intoleranz für Variation näher zu charakterisieren. Hierfür werden bei gesunden und erkrankten Individuen die Veränderungen der Gene verglichen, die bei Erkrankungen wiederholt betroffen sind. Da sich das Wissen auf Basis der immer schnelleren und präziseren Sequenzier-Technologien in hoher Geschwindigkeit weiterentwickelt besteht in diesem Gebiet eine erhebliche Entwicklungsdynamik.

3. Wegbereiter für die Genomik-Revolution

Der Weg zum menschlichen Genom (► **Abb. 1**) hat seine Wurzeln auch an Universität Tübingen. Der aus einer bekannten Medizinerfamilie stammende Arzt Friedrich Miescher entdeckte und isolierte hier im Jahr 1869 das Nuklein (► **Abb. 2**) aus den Kernen von weißen Blutkörperchen [10]. Miescher hatte nach seinem Medizinstudium in Basel zunächst vor eine klinische Ausbildung zu durchlaufen, dann aber aufgrund der seit seiner Kindheit bestehenden Schwerhörigkeit bewusst auf eine klinische Tätigkeit verzichtet und sich der Forschung in Tübingen zugewandt [10]. Obwohl Miescher die Bedeutung seiner Entdeckung nicht vollständig erkannte, vermutete er dennoch, dass die Substanz, die er isoliert hatte, die Moleküle der Vererbung darstellen. Dies wurde 75 Jahre später, im Jahr 1944, durch die klassischen Experimente von Avery, MacLeod und McCarty bestätigt [11, 12]. Im Jahr 1953 wurde die Struktur der DNA durch Watson und Crick aufgeklärt, die auf Daten von Rosalind Franklin und Maurice



► **Abb. 1** Die DNA-Revolution. Zeitliste ausgewählter Meilensteine aus denen die moderne Molekulargenetik hervorgegangen ist. Der Beginn datiert das Jahr 1869 mit der Entdeckung des Nukleins durch den schwerhörigen Mediziner Friedrich Miescher in Tübingen.

Wilkins aufbauten. Hierdurch entstand das Gebiet der Molekularbiologie [13, 14]. Mehr als zwei Jahrzehnte nach der Entdeckung der Struktur der DNA entstanden die ersten beiden „zuverlässigen“ Sequenzierungsmethoden. Die Maxam-Gilbert-Sequenzierung [15] basiert auf chemischer Spaltung. Diese Technik verwendet eine radioaktive Markierung von DNA-Fragmenten, die an jedem der Nucleotide (Adenin, Thymin, Guanin, Cytosin) chemisch gespalten werden, um die Reihenfolge zu bestimmen [16]. Eine alternative Form der Sequenzierung wurde nach einem der Entwickler, Frederick Sanger, benannt (Sanger-Sequenzierung) und basiert auf veränderten Ribosezuckern (► **Abb. 3a**) [17]. Diese Methode wird auch als „chain-termination“ oder „Di-deoxy-Technik“ bezeichnet, weil sie Dideoxynucleotide (ddNTPs) verwendet, denen eine 3'-Hydroxylgruppe fehlt, wodurch die Verlängerung einer wachsenden Nucleotidkette gestoppt wird. Unter Verwendung von 4 verschiedenen dNTP/ddNTP-Mischungen, von denen jede einem Nucleotid entspricht, integriert die DNA sowohl normale dNTPs, die eine Verlängerung des DNA-Strangs ermöglichen, als auch ddNTPs, die zu einem zufälligen Abbruch des DNA-Strangs führen (► **Abb. 3b**). Die Sequenzfragmente werden auf ein Gel aufgetragen und die Nucleotidabfolge kann bestimmt werden. Die Maxam-Gilbert-Sequenzierungsmethode war jahrzehntelang weit verbreitet, weil sie direkt DNA-Fragmente analysierte, während die frühen Methoden der Sanger-Sequenzierung eine klonale Amplifikation eines DNA-Fragments erforderten. Nach weiteren Entwicklungen übertraf jedoch die Popularität der Sanger-Sequenzierung die der Maxam-Gilbert-Sequenzierung, sodass sie über ein Vierteljahrhundert die am häufigsten genutzte Sequenzierungsmethode darstellte und auch heute noch wegen ihrer Zuverlässigkeit in der Validierung weit verbreitet ist.

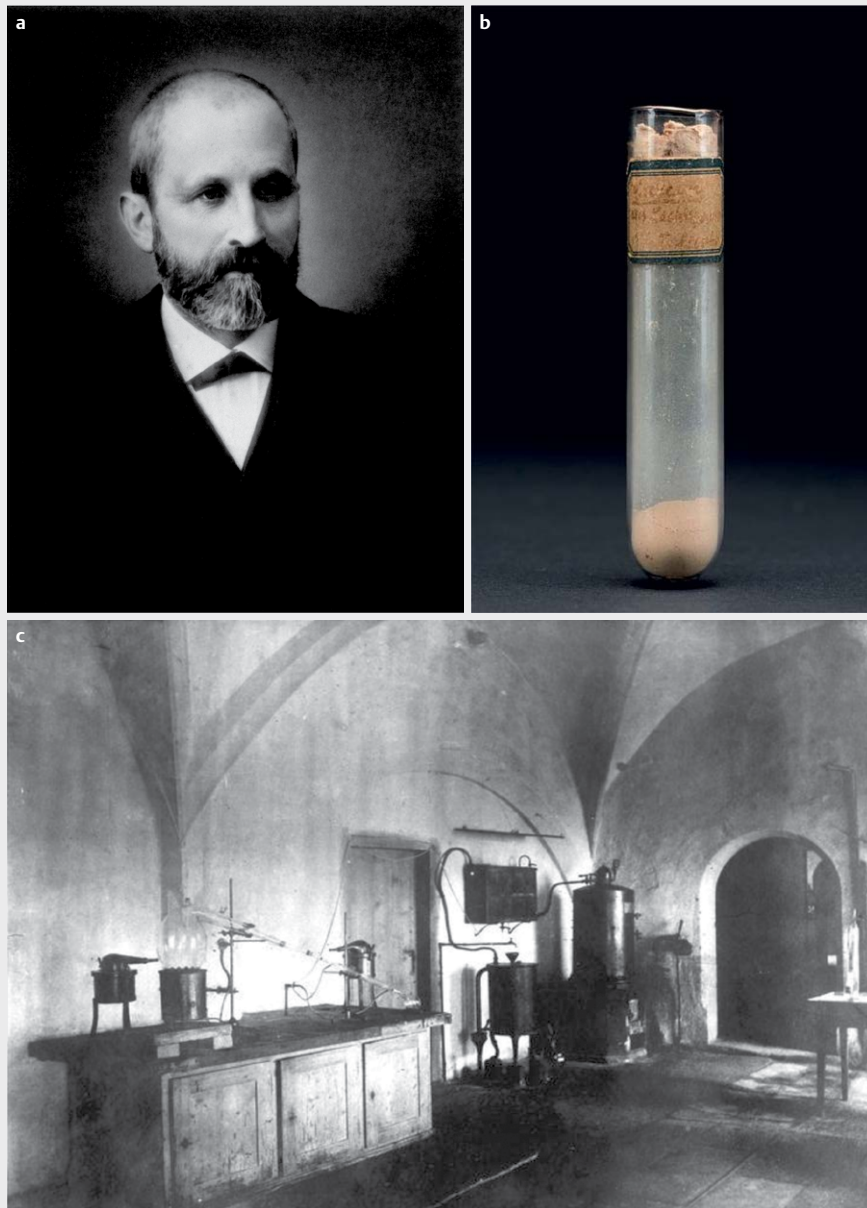
Im Jahr 1985 wurde die Polymerase-Kettenreaktion (engl: polymerase chain reaction; PCR) entwickelt. Hierbei wurden zwei Primer-Paare verwendet, die einen Abschnitt der zu kopierenden DNA flankieren. Diese Idee spielte eine Vorreiterrolle in der modernen Mole-

kularbiologie [18]. Diese Methode war auch grundlegend für das Humangenomprojekt, das Ende der 80er Jahre geplant und 1990 offiziell gestartet wurde [19]. In der damaligen westdeutschen Regierung gab es bezüglich der ethischen Fragen starken Widerstand gegen dieses Projekt [20]. Dennoch war Deutschland eines von nur 6 Ländern, die gemeinsam fast die gesamte Sequenzierung im Humangenomprojekt [21] durchführten.

Im Jahr 1994 wurde die erste Humangenomkarte mit 5 840 Loci veröffentlicht. Dies stellte einen großen Fortschritt in der genetischen Kartierung dar und unterstützte die Bemühungen zur Identifizierung von Genen erheblich [22]. Im folgenden Jahr wurde der erste Organismus, *Haemophilus influenzae*, [23] sequenziert, vier Jahre später folgte das menschliche Chromosom 22, das zweitkleinste der Autosomen [24]. Das Genom von *Drosophila melanogaster* wurde im Jahr 2000 sequenziert [25] und ebnete den Weg für die Erforschung konservierter Gene, die für Erbkrankheiten beim Menschen verantwortlich sind [26]. Die Komplettierung der menschlichen Genomsequenz im Jahr 2003 eröffnete nicht nur eine neue Ära in der Medizin, sondern beförderte auch bedeutende Entwicklungen in der DNA-Sequenzierung und den damit verbundenen Computertechnologien. Weniger als 2 Jahre später, im Jahr 2005, entstand die erste Hochdurchsatz-Sequenzierungsmethode in der Gruppe von George Church [27], die einen neuartigen zyklischen Array- und Multiplex-Sequenzierungsansatz verwendete, der die Kosten für die Sequenzierung auf etwa ein Neuntel der Kosten der Sanger-Sequenzierung senkte. Diese transformative Methode wurde 2007 zur „Methode des Jahres“ gewählt [28].

4. Entwicklung von Hochdurchsatz-Sequenzierungstechnologien

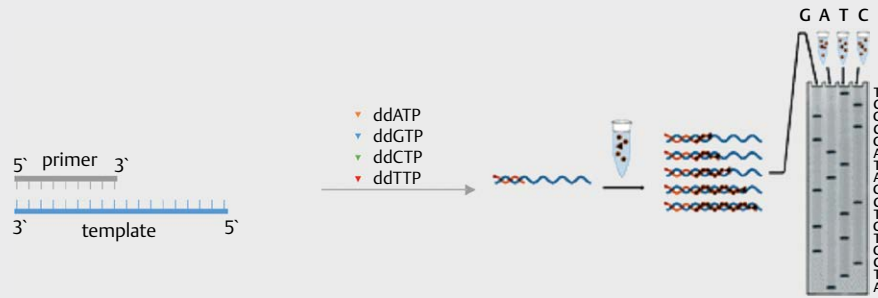
Die Sequenzierung eines menschlichen Genoms dauerte im Humangenomprojekt fast 15 Jahre und kostete \$3 Milliarden USD. Heute



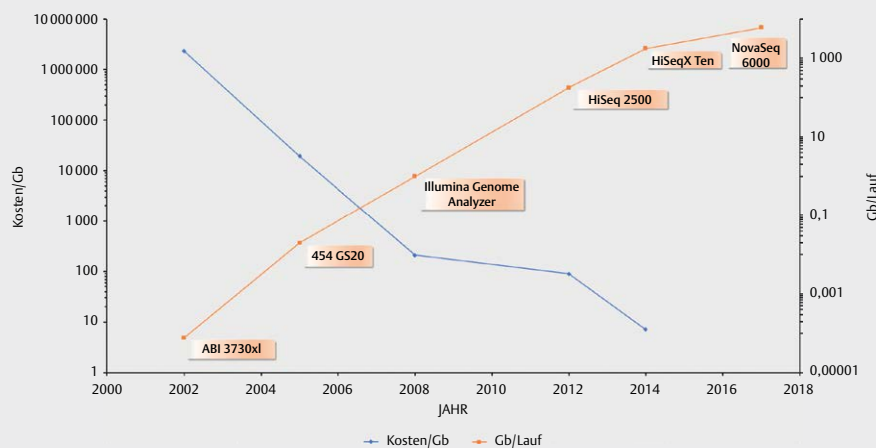
► **Abb. 2** Friedrich Miescher und die Entdeckung der Nukleinsäuren. **a** Friedrich Miescher (geboren am 13. August 1844, gestorben am 26. August 1895) war ein Schweizer Mediziner. Nach seinem Medizinstudium suchte Miescher wegen seiner Schwerhörigkeit nach einem Fach ohne Patienten-Interaktion. Er beschloss daher, seine Karriere der medizinischen Forschung zu widmen und ging aus diesem Grund nach Tübingen zu Felix Hoppe-Seyler an die „Wiege der Biochemie“. Dort entdeckte er 1869 die Nukleinsäure, den Grundstoff der Erbsubstanzen DNA und RNA. **b** Reagenzglas mit Nuklein aus Lachssperma, von Friedrich Miescher beschriftet und mit seinem Namen versehen (um 1871). **c** Schlosslabor Tübingen „Wiege der Biochemie, in dem Felix Hoppe-Seyler das Hämoglobin, und Friedrich Miescher die Nukleinsäure entdeckte. (Mit freundlicher Genehmigung des Museums der Universität Tübingen; MUT)

kann diese Aufgabe in nur 19,5 Stunden für 1000 US \$ [29, 30] ausgeführt werden. Die DNA-Sequenzieretechnologie existiert seit den 1970er Jahren und hat sich zu einer unverzichtbaren Technologie in der molekulargenetischen Diagnostik entwickelt. Die Sanger-Sequenzierung, die zur ersten Entschlüsselung des menschlichen Genoms verwendet wurde, gilt aufgrund ihrer Zuverlässigkeit und Genauigkeit von bis zu 99,999 % [31] immer noch als „Goldstandard“ der Sequenzierung. Eine moderne Version dieser Sequenzierungs-

methode findet heute noch breite Anwendung [17, 32]. Die Jahre unmittelbar nach der Entschlüsselung des menschlichen Genoms waren geprägt von der Entwicklung kommerzieller Hochdurchsatz-Sequenziergeräte (Anmerkung: Die Begriffe Hochdurchsatz-Sequenzierung, Next Generation Sequenzierung und massive parallele Sequenzierung werden synonym verwendet) (► **Abb. 4**). Diese Sequenzieretechnologien haben die generierte Datenmenge um mehrere Größenordnungen gesteigert. Diese Steigerung ermöglichte eine massive Kosten-



► **Abb. 3** Sanger-Sequenzierung. Eine Darstellung der modifizierten modernen Sanger-Methode. Ein Primer bindet an ein amplifiziertes Template und eine Erweiterung um ein einzelnes Nucleotid erfolgt. Die Erweiterung mit Standard-Deoxynucleotiden (dATP, dGTP, dCTP, dTTP, nicht dargestellt) erfolgt bis zur Integration eines fluoreszierend markierten Dideoxynucleotids (ddATP (orange), ddGTP (blau), ddCTP (grün), ddTTP (rot)), sodass die wachsende DNA-Kette abbricht. Nach mehreren Zyklen werden die DNA-Fragmente mit einem Gel nach ihrer Länge getrennt und die Reihenfolge der Nucleotide entsprechend der Sortierung der fluoreszenz-markierten Fragmente bestimmt.



► **Abb. 4** Entwicklung der Sequenzierkosten über die vergangenen Jahrzehnte. Vergleich der Kosten (in US\$) und der generierten Datenmenge pro Jahr während der Weiterentwicklung der Sequenziergeräte. Da die Gerätekapazität in Gigabasen (Gb) (Gb/Lauf, rot, rechte y-Achse) zunahm, entsprach dies einem starken Rückgang der Sequenzierkosten über die Zeit (Kosten/Gb, blau, linke y-Achse). Die Kostenschätzung ist nur bis 2014 verfügbar.

reduktion über einen relativ kurzen Zeitraum (► **Abb. 5**). Seit etwa 2007 hat die Reduktion der Sequenzierkosten sogar das Mooresche Gesetz für die Reduktion von Rechenkosten deutlich übertroffen. Das Mooresche Gesetz besagt, dass sich die Rechenleistung alle 2 Jahre verdoppelt und damit die Kosten halbiert werden. Im Jahr 1998 erzeugte der Sequenzer ABI3730xl (Thermo Fisher Scientific) 84 Kilobasen Daten pro Lauf [21]. Mit dem 2005 vorgestellten Genome Analyzer (Illumina) System konnten 1,3 menschliche Genome pro Jahr (Illumina) sequenziert werden, das waren 1 Gigabasen pro Lauf [33]. Dieser Technologiesprung wurde weiterentwickelt und brachte innerhalb von 10 Jahren um das Jahr 2008 eine Verbesserung der Sequenzleistung von 10^2 Kilobasen pro Tag auf 10^{12} Kilobasen pro Tag [34]. Ein weiterer bemerkenswerter Fortschritt war im Jahr 2014 das HiSeqX Ten System (Illumina), das 1,8 Gigabasen pro Sequenzierreaktion erzeugte und damit erstmals die 1000 US-\$-Grenze für ein menschliches Genom unterschritt. Das aktuelle NovaSeq 6000 System (Illumina) kann bis zu 6 Terabasen Sequenzdaten in weniger

als zwei Tagen erzeugen. Mit Blick auf die Zukunft scheint es sehr wahrscheinlich, dass sich diese Entwicklungen für Sequenziergeschwindigkeit und Kostenreduktion fortsetzen werden. Das Sequenzierunternehmen Illumina will das 100-Dollar-Genom in den nächsten 10 Jahren erreichen [35].

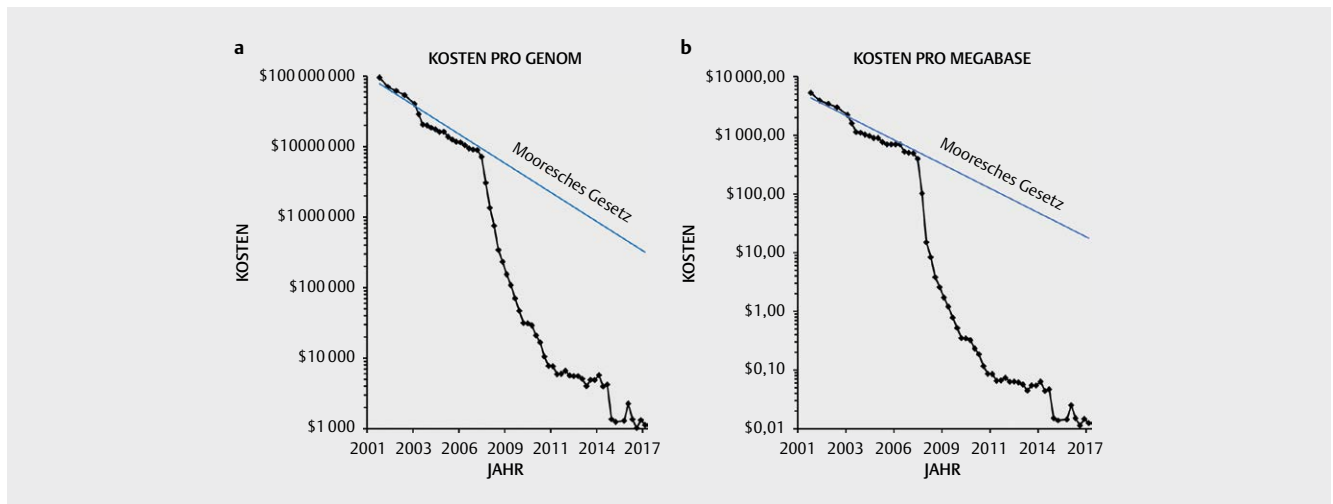
Diese Sequenzierungstechnologien gingen mit anderen Entwicklungen in der Digitalisierung wie der Datenspeicherung, dem parallelen Rechnen, der Weiterentwicklung der CPU-Architektur und der Erfindung des World Wide Web einher. Dies hat ebenfalls zur Kostensenkung beigetragen. Der sprunghafte Anstieg der Datenmenge und der starke Rückgang der Kosten machen diese Methoden allgemein für den einzelnen Patienten zugänglich und ermöglichen es der Forschung und klinischen Labors, große Datensätze von Hunderttausenden von Personen zu generieren. Diese Datensätze sind entscheidend, um neue Zusammenhänge bei Erkrankungen aufzudecken und die Annotation des gesamten Katalogs pathogener Varianten des Menschen zu unterstützen. Die Umstellung auf „big data“

in der Erforschung des Genoms hat enorme Auswirkungen auf die Diagnostik und auch die Behandlung von Patienten in allen Krankheitsbereichen. Die Schwerhörigkeit ist aufgrund der zugrundeliegenden genetischen Komplexität ein besonders interessantes und herausforderndes Beispiel.

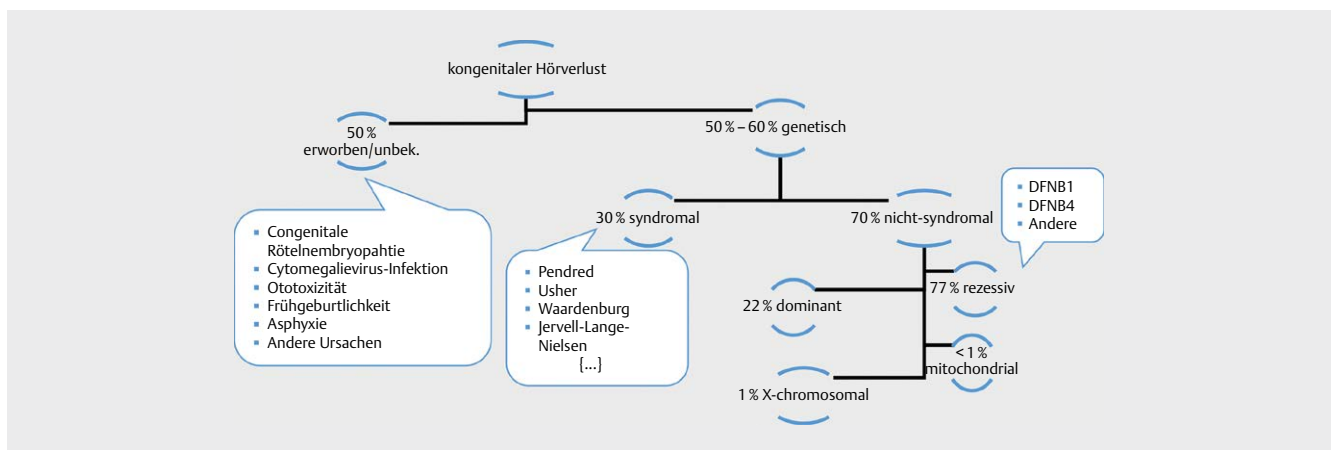
5. Die Genetik des Hörverlustes

Nach Angaben der Weltgesundheitsorganisation treten genetische Störungen mit einer Prävalenz von 10 pro 1000 Geburten auf [36] und betreffen lebenslang bis zu 1 von 17 Personen [37]. Schwerhörigkeit ist die häufigste kongenitale sensorische Störung, die ein bis 2 von 1000 Neugeborenen betrifft [38]. Mehr als die Hälfte der sensorischen Hörstörungen sind genetisch bedingt (► **Abb. 6**). Ein Hörverlust tritt überwiegend nicht-syndromal (70%) auf, kann aber auch

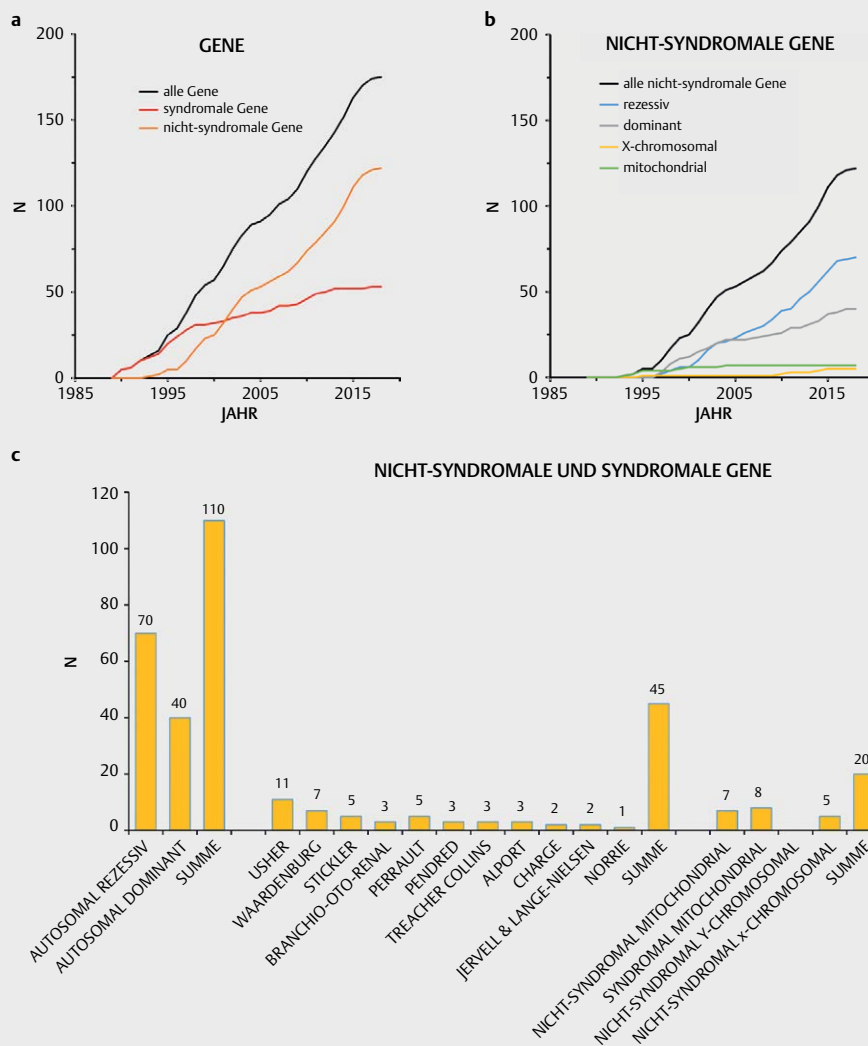
Teil eines klinischen Syndroms (30%) sein [39]. Hörverlust wird klassischerweise nach dem Vererbungsmuster der Mendelschen Regeln, autosomal-rezessiv (77%), autosomal-dominant (22%), X-chromosomal (1%) und mitochondrial (< 1%) vererbt [39]. Im Gegensatz zu manchen autosomal-rezessiven genetischen Störungen wie z. B. Mukoviszidose, die durch pathogene Varianten in nur einem einzigen Gen (*CFTR*) diagnostiziert werden können, ist der erbliche Hörverlust eine genetisch stark heterogene Erkrankung, an der mehrere hundert Gene beteiligt sind. Im Fall der Mukoviszidose sind bis zu 70% der bei Patienten mit europäischer Ethnizität entdeckten Varianten mit einer Deletion von 3 Basenpaaren verbunden, die für einen Verlust der Aminosäure Phenylalanin (p.Phe508del) kodieren [40]. Im Gegensatz hierzu haben die Bemühungen, die molekulare Genetik des Hörverlusts zu entschlüsseln, bereits zur Beschreibung von tausenden Varianten der derzeit bekannten Gene für nicht-syndromale



► **Abb. 5** Darstellung der DNA-Sequenzierungskosten und dem Mooreschen Gesetz. Das Mooresche Gesetz (engl.: „Moore's law“) besagt, dass eine Kostenreduktion alle 2 Jahre durch Verdoppelung der Rechenleistung integrierter Schaltkreise in Rechenmaschinen erfolgt. **a** Kosten pro Genom und **b** Kosten pro Megabase, von 2001 bis 2018 folgen einem nahezu identischen Profil. In beiden Analysen liegt der Kostenrückgang seit 2007 deutlich über dem Mooreschen Gesetz.



► **Abb. 6** Anteil von Umweltfaktoren und genetischen Faktoren bei angeborenem Hörverlust. Etwa bei der Hälfte der Patienten ist der Hörverlust auf eine genetische Ursache zurückzuführen.



► **Abb. 7** Anzahl der nicht-syndromalen und syndromalen Gene, die einen Hörverlust verursachen (Hereditary Hearing Loss Homepage [41]). **a** Identifizierung der 175 Gene, die im Laufe der Zeit entdeckt wurden (schwarz), in nicht-syndromale (orange) und syndromale (rot) Kategorien eingeteilt. **b** Betrachtung der Anzahl der Gene pro Jahr: nicht-syndromale Gene, die einen Hörverlust verursachen (schwarz), aufgeteilt nach Vererbungsmuster, rezessiv (blau), dominant (grau) und X-chromosomal (gelb), sowie mitochondriale Vererbung (grün). **c** Eine Übersicht über nicht-syndromale und syndromale Gene, die auf der „Hereditary Hearing Loss Homepage“ dargestellt sind. Gene, die sowohl an autosomal-dominanten und rezessiven als auch an nicht-syndromalen und syndromalen Hörverlust beteiligt sind, werden jeweils einzeln in der jeweiligen Kategorie dargestellt. N ist die Anzahl der identifizierten Gene.

[41] und syndromale [42] Formen des Hörverlusts geführt (► **Abb. 7a, b**). Die „Deafness Variation Database“ (DVD) hat mit einem Gen-Set von 152 Genen bereits über 8 000 pathogene Varianten erfasst [42]. Zum Vergleich: Die Human Gene Mutation Database (HGMD) ist eine umfassende Sammlung aller bekannter Keimbahnvarianten, die mit menschlichen Erkrankungen assoziiert sind. Diese Datenbank (HGMD Professional 2018.2) enthält derzeit rund 225 000 Varianten, von denen die überwiegende Mehrheit als pathogen klassifiziert ist [43]. Der relativ hohe Anteil an pathogenen Varianten allein für Hörverlust unterstreicht die für diese Sinneserkrankung vorliegende genetische Komplexität.

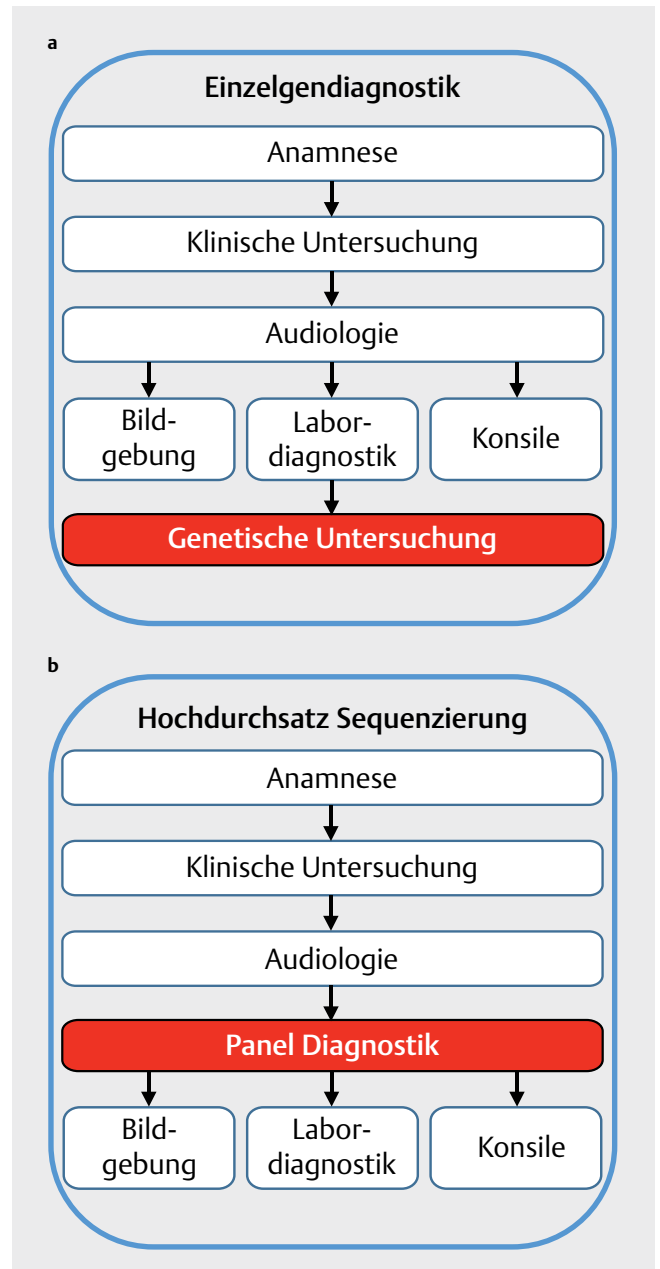
Die „Hereditary Hearing Loss Homepage“ beschreibt derzeit 175 Gene für Hörverlust (► **Abb. 7a**). Dazu gehören 161 nicht-syndromale Hörverlust-Loci, von denen 122 Gene identifiziert wurden.

Diese nicht-syndromalen Hörverlust-Gene wurden als ca. 70 autosomal-rezessive, 40 autosomal-dominante und 5 X-chromosomale Hörverlust-assoziierte Gene sowie 7 mitochondriale Hörverlust-assoziierte Varianten klassifiziert. Weiterhin sind in dieser Datenbank derzeit 53 Gene für syndromalen Hörverlust dokumentiert (► **Abb. 7c**). Die genetische Heterogenität des Hörverlusts erschwert die Interpretation der Befunde. Sowohl die hohe Anzahl der beteiligten Gene, als auch die hohe Anzahl pathogener Varianten für jedes Gen tragen zu dieser Komplexität bei. Trotz dieser Komplexität wurden erhebliche Fortschritte bei der Aufklärung genetischer Schwerhörigkeit erzielt. Dies ist der Entwicklung großer Datenbanken und Bioinformatik-Programme zu verdanken, einschließlich jener, die speziell für die Genetik von Hörverlust entwickelt wurden.

6. Veränderung der Diagnostik für Hörverlust

Ohne die in der heutigen Form verfügbaren Möglichkeiten der Hochdurchsatz-Sequenzierung beruhte die konventionelle klinische Evaluationen auf einer Reihe von klinischen Untersuchungen, um ein möglichst detailliertes phänotypisches Bild zu erhalten, das eine gezielte molekulargenetische Analyse leitete (► **Abb. 8**) [44, 45]. Im Ausschlussverfahren folgte dann häufig die molekulargenetische Einzeluntersuchung mit der Sequenzierung des Connexin-26-Gens *GJB2*. Mutationen im Gen *GJB2* sind in erster Linie mit einer autosomal-rezessiven Form des nicht-syndromalen Hörverlusts assoziiert. In Deutschland ist dieses Gen für die Diagnose von etwa jedem fünften Patienten mit genetischem Hörverlust verantwortlich [46]. In einer eigenen Studie haben wir bei ca. 17% der Cochlea-Implantat-Kandidaten, die sich einer molekulargenetischen Untersuchung auf Hörverlust unterzogen, pathogene Varianten von *GJB2* [47] identifiziert. In den vergangenen Jahrzehnten wurde der Erfolg dieses Screening-Verfahrens durch die geringe Länge des *GJB2* Gens, die eine einfache Sequenzierung ermöglicht, und die relativ hohe Diagnoserate unterstützt [48]. Wenn die klinische Evaluation auf eine bestimmte Form eines syndromalen Hörverlust hinwies, wurde versucht auf der Basis dieser klinischen Verdachtsdiagnose eine gezielte Sangersequenzierung von Kandidatengen durchzuführen. Die Leitung der genetischen Analyse durch phänotypische Daten konnte bei einer genetisch hochheterogenen und phänotypisch sehr variablen Erkrankung wie der Schwerhörigkeit jedoch nur begrenzten Erfolg bieten und beschränkte sich daher auf wenige Gene mit einer klaren Genotyp-Phänotyp-Korrelation. Die molekulargenetischen Untersuchung von Einzelgenen war langsam, arbeitsintensiv, teuer und führte oft zu wenig informativen Ergebnissen bzw. ließen die genetische Ursache ungeklärt [49]. Darüber hinaus sind die meisten Formen des Hörverlusts nicht-syndromal, das heißt klinisch konnten keine spezifischen phänotypischen Merkmale identifiziert werden. Damit war die Diagnostik nach Ausschluss einer *GJB2*-Mutation hinaus auf Basis von Einzelgenuntersuchungen eine Herausforderung und konnte nicht erschöpfend durchgeführt werden. Die phänotypische Variabilität von Schwerhörigkeit macht es nahezu unmöglich, eine prädiagnostische Hypothese durch klinische Untersuchung und audiologische Befunde aufzustellen. Das reine Screening anderer Kandidaten Gene war zu teuer, zu langwierig und oft auch methodisch auf wenige Gene beschränkt.

In zurückliegenden Jahrzehnt hat sich ein richtungsweisender Übergang von der Einzelgen-Sequenzierung zur Hochdurchsatz-Sequenzierung für die genetische Diagnostik vieler Erkrankungen, insbesondere auch des Hörverlusts, vollzogen [50]. Mithilfe dieser Technologie können bis dahin unvorstellbar großen Datenmengen generiert und ausgewertet werden. Trotz der für die Schwerhörigkeit bestehenden genetischen Heterogenität können in kurzer Zeit und zu geringen Kosten verwertbare Ergebnisse erzielt werden. Aktuelle diagnostische Ansätze nutzen entweder erkrankungsspezifische Gen-Panels oder die Exom-Sequenzierung. Das Exom umfasst die potenziell für Proteine kodierenden Regionen der DNA, die etwa 1 bis 2% des gesamten Genoms ausmachen. Der strategische Übergang zu den Methoden der Hochdurchsatz-Sequenzierung verändert das diagnostische Vorgehen in der Patientenversorgung. Die deutlich verbesserte Zugänglichkeit der genetischen Diagnostik rückt damit in im Verhältnis zu breit angelegten klinischen Untersuchungen in den Vordergrund.



► **Abb. 8 a** Klassische Abfolge der Hördiagnostik. Die Reihenfolge beginnt mit Anamnese, klinische Untersuchung, audiologische Prüfung, Bildgebung, Zusatzuntersuchung (z. B. Labordiagnostik, Schilddrüsenfunktionsprüfung, EKG), Konsiluntersuchungen (z. B. Ophthalmologie, Neurologie, Kardiologie) und endet schließlich mit der humangenetischen Untersuchung auf Basis der Einzelgendiagnostik. Genetische Ursachen können nach diesem Vorgehen per Ausschlussdiagnostik in der Regel nur angenommen werden, können aber meist nicht direkt nachgewiesen werden. **b** Zukünftige Abfolge der Hördiagnostik unter Einbeziehung einer umfassenden genetischen Diagnostik. Der direkte Nachweis einer genetischen Ursache wird bereits nach der Anamnese, klinischen Untersuchung und audiologischen Prüfung angestrebt. Der klinische Phänotyp kann bei der genetischen Befundbeurteilung wertvolle Hinweise liefern. Im Falle des Nachweises einer syndromalen Form der Schwerhörigkeit können dann gezielt weiter Zusatzuntersuchung und Konsiluntersuchungen veranlasst werden. Verändert nach Löwenheim, 2014 [45].

Die Hochdurchsatz-Sequenzierung bietet mehrere entscheidende Vorteile gegenüber Einzelgen-Ansätzen, z. B. werden zeitgleich viele Patienten in einem standardisierten Verfahren im Labor untersucht. Insbesondere können alle bekannten Gene im Zusammenhang mit Hörverlust in einer einzigen Reaktion sequenziert und parallel analysiert werden. Dies ermöglicht einen hypothesenfreien Zugang zur genetischen Diagnostik. Da Syndrome wie Usher- oder Pendred-Syndrome erst Jahre nach Beginn eines Hörverlusts klinisch als syndromale Erkrankung auffällig werden, war es bisher kaum möglich, in diesen Fällen die genetisch bedingte Hörstörung vor Auftreten der übrigen Symptome trotz eingehender klinischer Untersuchung genau zu diagnostizieren. Bei syndromaler Schwerhörigkeit kann eine frühe Diagnose vor Eintreten von Begleitsymptomen klinische Untersuchungen ersparen, ermöglicht eine prognostische Einschätzung und eine Beratung über weitere Behandlungsmöglichkeiten [50]. Bei nicht-syndromaler Schwerhörigkeit können betroffene Patienten mit gut charakterisierten pathogenen Varianten ebenfalls eine prognostische Bewertung und eine personalisierte Auswahl von Therapie und Hörrehabilitation nutzen.

6.1 Gen-Panel-Diagnostik bei Hörverlust

Gen-Panels stellen einen selektiven und spezifischen Ansatz für die molekulargenetische Diagnostik dar, da sie auf krankheitsassoziierte Gene in einem bestimmten Krankheitsgebiet fokussieren. Das Panel-Design beinhaltet die Auswahl von Genen auf der Grundlage des aktuellen Wissens für ein individuelles „Köder“-Design (engl.: bait design). Diese Köder werden aus Oligonukleotiden hergestellt, die komplementär zu denen für eine bestimmte Erkrankung in Frage kommenden Zielregionen/Exons ausgewählt werden. Anreicherung (engl.: „Targeted Genomic Enrichment“) oder „Sequence Capture“ sind Begriffe, die die Auswahl der gewünschten DNA-Regionen zur Amplifikation und Anreicherung bei der Erstellung einer sogenannten Sequenzierungsbibliothek beschreiben. Eine Bibliothek enthält den kompletten Satz von ausgewählten und angereicherten Fragmenten, die für das Krankheitsbild von Interesse sind. Bevor Gen-Panels in einem diagnostischen Bereich verwendet werden, durchlaufen sie Gen-Auswahl- und Designschritte. Darüber hinaus werden die Sequenzierungsdaten zunächst einer Validierung und Optimierung auf Qualität und Homogenität unterzogen. Dies hat zur Folge, dass die Sequenzierungsabdeckung (engl.: Coverage) über den Gen-Satz hinweg eine größere Homogenität mit weniger „Lücken“ (oder Basen mit schlechter oder keiner Abdeckung) aufweist. Der Einsatz methodisch gut abgedeckter Gene ist wichtig für eine umfassende Sequenzanalyse der Varianten, die sich in diesen Regionen befinden können. Gen-Panels besitzen also den Vorteil einer spezifisch auf eine bestehende klinische Fragestellung wie der Schwerhörigkeit gerichtete Diagnostik bei gleichzeitig hoher qualitativer Aussagekraft.

6.2 Exom-Diagnostik bei Hörverlust

Die Exom-Sequenzierung reichert alle derzeit bekannten Gene und Gen-Isoformen an und beschränkt sich nicht auf die für ein bestimmtes Krankheitsgebiet bekannten Gene. Es gibt viele kommerziell erhältlich Kits zur Bibliothekserstellung für die Exom-Sequenzierung, die ständig verbessert werden. Viele Anbieter erlauben es den Anwendern auch, den Gehalt an spezifischen „Ködern“ zu erhöhen, um die gewünschte Sequenzierung von Zielregionen zu verbessern. Auch können bekannte pathogene Varianten, die sich nicht in exomischen Re-

gionen befinden und ansonsten völlig übersehen werden würden, mit einbezogen werden. Da Exome einen wesentlich größeren Anteil der genomischen Sequenz im Vergleich zu Gen-Panels anreichern, ist die Sequenzierqualität von Exom-Datensätzen im Vergleich zu Gen-Panels tendenziell geringer. Entgegen diesem Ruf einer geringeren diagnostischen Qualität ist die exombasierten Diagnostik jedoch nach deutlicher Optimierung inzwischen erfolgreich in das klinische Umfeld integriert [51–54]. Die Sanger-Sequenzierung von eingeschränkt abgedeckten Regionen kann diese Methode gut ergänzen.

Bei der Exom-Sequenzierung wird die Analyse für die Diagnostik am effizientesten durch ein so genanntes *in silico* Gen-Panel gesteuert. Ähnlich den primären Gen-Panels wird hier die Analyse auf klinisch relevante Gene beschränkt, um Zeit zu sparen und eine Diagnose zügig stellen zu können. Dieser Genauswahlprozess ermöglicht eine Analyse, die alle Gene umfasst, die für den spezifischen Phänotyp des Patienten von klinischer Bedeutung sind und die einen bestimmten Abdeckungsgrenzwert erfüllen. Dies eröffnet verbesserte Möglichkeit zur Variation dieser Geninhalte als dies bei klassischen primären Gen-Panels der Fall ist, die auf einen festen, immer gleichen Gen-Satz beschränkt sind.

6.3 Vor- und Nachteile von Gen-Panels und exombasierten Diagnostikansätzen

Es gibt mehrere Vorteile bei der Auswahl von Gen-Panels für Hörverlust gegenüber der Exom-Sequenzierung. Ein Argument für die Auswahl von Gen-Panels gegenüber der Exom-Sequenzierung ist, dass die produzierten Daten spezifisch für die in Frage stehende Primärerkrankung sind. Dies bedeutet, dass Gene, die mit anderen Erkrankungen in Zusammenhang stehen, nicht sequenziert werden und die Analyse und die genetischen Ergebnisse auf die primäre Indikation beschränkt bleiben. Mit anderen Worten: Laboratorien, die die Gen-Panel-Diagnostik nutzen, müssen nicht das Potenzial von Sekundärbefunden berücksichtigen, die für den Patienten von klinischer Bedeutung sein können, aber nicht in Zusammenhang mit der primären Indikation stehen. Expertengruppen des American College of Medical Genetics haben Richtlinien für die Meldung von Sekundärbefunden in mindestens 59 medizinisch verwertbaren Genen bei der klinischen Genomsequenzierung empfohlen [55]. Die überwiegende Zahl dieser Gene betrifft autosomal-dominante Erkrankungen, die typischerweise erst im Erwachsenenalter symptomatisch werden. Zufällige Varianten werden unabhängig vom Alter des Patienten gemeldet, jedoch sind Erkrankungen, die zum Neugeborenen-Screening gehören, ausgeschlossen. Im Jahr 2013 veröffentlichte die Deutsche Gesellschaft für Humangenetik Richtlinien für die Rückmeldung von Zusatzbefunden, die das Zustimmungsverfahren und das Recht des Patienten, diese Ergebnisse nicht zu kennen oder nicht zu erhalten, beinhalten [56]. Diese Richtlinien spezifizieren keine sekundären Befunde, die in einem bestimmten Satz von Genen gefunden wurden, vielmehr werden vier Kategorien definiert, in die eine Variante fallen kann. Es wird v. a. ermutigt, Zusatzbefunde zu melden, für die Behandlungsmethoden existieren und über die Betroffenen informiert werden wollen. Damit geht die Exom-Sequenzierung in ihrer diagnostischen Abdeckung aber weit über die ursprüngliche Fragestellung bspw. einer bestehenden Schwerhörigkeit hinaus. Für den eine genetische Untersuchung veranlassenden Facharzt gehen die möglichen Befunde dann auch potenziell weit über das eigene Fachgebiet hinaus.

Da bei der Sequenzierung von vordefinierten Gen-Panels im Vergleich zur Exom-Sequenzierung eine kleinere und spezifisch auf eine

Erkrankung ausgerichtete Gruppe von Genen angereichert und analysiert wird, ist die Abdeckung der Gene in der Regel viel höher, und das spezifische „bait design“ kann sehr gut auf schwer zu sequenzierende Regionen abzielen (d. h. GC-reiche Regionen, repetitive DNA-Sequenzen, die als Tandem-Wiederholungen bezeichnet werden, und ungleichmäßig fragmentierte DNA-Bereiche). Dies bedeutet, dass die Sensitivität (Falsch-Negativ-Rate) und Spezifität (Falsch-Positiv-Rate) von Detektionsvarianten deutlich verbessert werden kann. Weiterhin können falsch-positive Ergebnisse aus Hochdurchsatz-Sequenzierungsdaten zuverlässig mittels Sanger-Sequenzierung entdeckt und validiert werden. Die einheitliche Abdeckung unterstützt auch eine konsistentere Detektion von Kopienzahl-Abweichungen. Die Daten, die bei der Gen-Panel-Sequenzierung generiert werden, sind deutlich kleiner als bei der Exom-Sequenzierung. Dies vereinfacht auch die Datenspeicherung wesentlich.

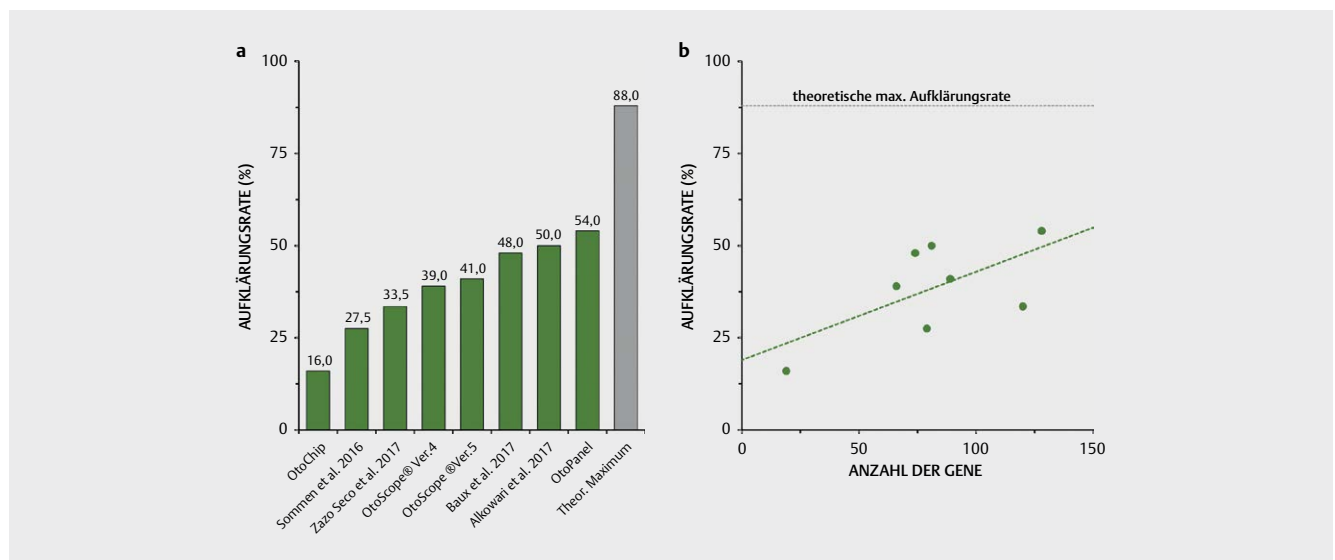
Die Exom-Sequenzierung wiederum bietet einige Vorteile gegenüber den primären Gen-Panels. So besteht die Möglichkeit die erhobenen Daten zu einem späteren Zeitpunkt erneut zu analysieren und neu-ursächliche Gene einzuschließen. Dies kann für Patienten, bei denen die Analyse der bekannten Gene ergebnislos bleibt zur Diagnosesicherung beitragen. Die Geschwindigkeit der Entwicklungen in der Genom-Diagnostik führt zur Identifikation immer neuer Gene. Gen-Panels erfordern eine regelmäßige Aktualisierung der Geninhalte und werden anschließend einer Validierung unterzogen. Erst nach diesen Schritten kann die DNA des Patienten erneut überprüft werden. Dieses Verfahren ist aufwendiger und teurer als die Generierung eines Exom-Datensatzes. Bei der Exom-Sequenzierung können molekulargenetische Untersuchungen von Patienten mit einer Vielzahl unterschiedlicher Erkrankungen parallel geprüft werden, und dadurch die Durchlaufzeiten für Laboruntersuchungen deutlich reduziert werden. Bei primären Gen-Panels sind je nach Labor und Anzahl der Testanforderungen müssen die Labors oft mehrere Wochen oder sogar Monate Wartezeit erforderlich, bis genügend DNA

von Patienten mit der Fragestellung Hörverlust eingetroffen ist, um ein Gen-Panel zu erstellen und zu sequenzieren.

6.4 Diagnostische Aufklärungsraten

Obwohl die bereits umfangreiche Liste der am Hörverlust beteiligten Gene derzeit noch unvollständig ist [57], haben sich die diagnostischen Aufklärungsraten bei genetisch bedingten Hörverlusten in Deutschland [47, 58, 59] und weltweit [60] seit der Implementierung von Hochdurchsatz-Sequenzierungsansätzen in die Routinediagnostik deutlich verbessert. Um dies zu veranschaulichen, wird eine Reihe von Studien dargestellt, die über diagnostische Aufklärungsraten einschließlich *GJB2* auf der Basis von Hochdurchsatzsequenzierung berichten (► **Abb. 9**). In diesen Studien variierte die Methodik von Gen-Panels [58, 61–63] bis hin zur Exom-Sequenzierung [54], die Anzahl der analysierten Gene war nicht einheitlich, und auch die Analyse der Kopienzahl-Variation wurde nicht in allen Studien konsequent durchgeführt. Die diagnostischen Aufklärungsraten reichten in Abhängigkeit von der Anzahl der eingeschlossenen Gene von 16% in einer Studie, die 19 Gene auf einem Oligo-Hybridisierungs-Array [64] umfasste, bis zu 54% [47] einer eigenen Untersuchung, in der ein Gen-Panel mit bis zu 128 Genen für Hörverlust verwendet wurde. Betrachtet man die in den verschiedenen Studien am häufigsten von Mutationen betroffenen Gene (► **Tab. 1**), so zeigt sich, dass sich die fünf am häufigsten betroffenen Gene z. T. unterscheiden. Die Ausnahme bleibt *GJB2*, das stets als das am häufigsten betroffene Gen identifiziert wird. Die unterschiedlichen Ergebnisse können mit dem unterschiedlichen ethnischen Hintergrund der Patienten und der variablen Anzahl der untersuchten Gene und Patienten erklärt werden.

Eine umfassende molekulargenetische Diagnostik sollte die Analyse der Kopienzahl-Variation [65] beinhalten. Eine der bisher größten Studien zu dieser Thematik ergab, dass etwa 15% der von Hörverlust betroffenen Patienten mindestens eine Kopienzahl-Variation in einem mit Hörverlust assoziierten Gen trugen. Bei Patienten mit



► **Abb. 9** Diagnostische Aufklärungsraten in ausgewählten Hochdurchsatz-Sequenzierungsstudien. **a** Übersicht über die diagnostische Aufklärungsrate bei Patienten mit Hörverlust, die sich einem Gentest unterziehen. **b** Es besteht eine positive Korrelation zwischen der diagnostischen Aufklärungsrate und der Anzahl der untersuchten Gene.

► **Tab. 1** Die am häufigsten betroffenen Gene aus ausgewählten Studien.

Autor	Hernandez et al., 2010	Sloan-Hegen et al., 2016	Zazo Seco et al., 2016	Baux et al., 2017	Alkowari et al., 2017	Sommen et al., 2016	Tropitzsch et al., 2013
Land	USA	USA	Niederlande	Frankreich	Katar	Belgien	Deutschland
Patienten	44	1119	200	207	81	160 families	154
<i>GJB2</i>		1.	1.	1.		1.	1.
<i>MYO15A</i>		5.	3.	4.		2.	3.
<i>SLC26A4</i>		3.					
<i>MYO7A</i>	1.		4.	3.		2.	2.
<i>CDH23</i>	2.		4.	5.	1.		
<i>OTOF</i>			5.		3.		
<i>USH2A</i>	3.		2.	4.			
<i>TMC1</i>					2.	2.	
<i>MYO6</i>					3.		5.
<i>TECTA</i>		4.		4.			4.
<i>STRC</i>		2.	3.	2.			
<i>LOXHD1</i>			4.				
<i>TRIOBP</i>			5.				
<i>OTOA</i>				5.	2.		
<i>GJB6</i>					4.		

gesicherter erblicher Schwerhörigkeit hatten 18,7 % eine Variation der Kopienzahl, die zu der genetischen Diagnose beitrug. Das Gen *STRC* ist das häufigste von Kopienzahl-Variationen betroffene Gen. *STRC* kodiert für das Protein Stereocilin, das in den Stereozilien der äußeren Haarzellen lokalisiert ist. *STRC* weist in der europäischen Bevölkerung eine hohe Deletions-Trägerrate von ca. 1,6 % auf, die fast so hoch ist wie die bekannte c.35delG-Trägerrate (1,89 %) für *GJB2* in der gleichen Population [48]. Folglich sollte insbesondere für das Gen *STRC* (DFNB16) eine Analyse sowohl auf Deletionen als auch auf pathogene Varianten [65–70] erfolgen. *STRC* ist nur ein Beispiel für die Notwendigkeit zum Einschluss der Kopienzahl-Variation.

Die zunehmende Verbesserung der Aufklärungsraten mit bis zu 50 % in der molekulargenetischen Diagnostik belegt, dass bei allen Patienten mit Hörverlust eine Hochdurchsatz-Sequenzierung zu empfehlen ist [49]. Patienten mit negativen genetischen Befunden und fortbestehendem klinischem Verdacht auf eine hereditäre Schwerhörigkeit sollten in den folgenden Jahren einen erneuten Test in Erwägung ziehen, da sich das Wissen über Gene und Variantentypen immer weiterentwickelt. Unter Berücksichtigung von nicht-synonymen Mutationen, Spleißstellen-Mutationen (engl.: „splice-site mutations“), Indels (Insertionen und Deletionen) und Kopienzahl-Variationen in den kodierenden Regionen eines Gen-Panels könnte in Zukunft eine diagnostische Aufklärungsrate von 88,3 % erreicht werden [49] (► **Abb. 8b**). Die aktuellen Fortschritte in der Genomik weisen darauf hin, dass auch nicht-kodierende Regionen des Genoms an einer Erbkrankheit beteiligt sind [71]. Das Ausmaß der Beteiligung von Veränderungen in den nicht-kodierenden Regionen des Genoms für Hörverlust muss noch aufgeklärt werden. Allerdings sind krankheitsassoziierte intronische Varianten mit mehr als 20 Nukleotiden aus kodierenden Exons und so genannten „deep intronic“-Varianten,

die nicht in Standard-Gen-Panels und Exomen erfasst würden, bereits in vielen Beispielen der Genetik von Schwerhörigkeit beteiligt [72, 73]. Es ist naheliegend anzunehmen, dass in Zukunft ein erheblicher Anteil der schwerhörigen Patienten mithilfe von Varianten nicht-kodierender Regionen diagnostiziert werden kann, die Auswirkungen auf die Genexpression und das normale Spleißen von Genen haben.

Die molekulargenetische Diagnostik ist auf eine qualitativ hochwertige Sequenzierung sowie effektive Analysestrategien der Bioinformatik angewiesen, die den Ausschluss nicht relevanter Varianten unterstützen und gleichzeitig relevante Varianten für die Expertenanalyse enthalten. Die Priorisierung der Varianten erfolgt typischerweise durch den Einsatz verschiedener Tools und Datenbanken, von denen viele im nächsten Abschnitt kurz beschrieben werden. Da sich die Hochdurchsatz-Sequenzierung immer weiterverbreitet, steigt auch die Anzahl der Tools und Größe der Datenbanken exponentiell an.

7. Computer Ressourcen

Die Bioinformatik setzt rechenintensive Methoden zur Aufbereitung und Analyse von Daten ein, um biologisch und medizinisch relevante Ergebnisse zu erhalten. Vorgelagerte bioinformatische Prozesse, die an Hochdurchsatz-Sequenzierungsdaten beteiligt sind, können im Rahmen dieser Übersicht nur kurz angesprochen werden. Sie beinhalten Datenvorverarbeitungsschritte, die die Ausrichtung von Lese- oder Sequenzierungsdaten auf die Referenzsequenz des menschlichen Genoms beinhalten, sowie Nachverarbeitungsschritte, die die Entfernung von Duplikaten und die Neukalibrierung der Basisqualität beinhalten. Auch diese Verfahren wurden kontinuierlich verbessert, um die Genauigkeit der Identifizierung von Varianten zu erhöhen. Das durchschnittliche Exom enthält über 20 000 Varianten, von denen 500 als

selten oder (noch) nicht in Varianten-Frequenzdatenbanken beschrieben sind [74]. Auch die Varianten, die in Gen-Panels nachgewiesen werden, können extrem selten sein oder in der Interpretation unklar bleiben. Es gibt eine Vielzahl von Programmen (engl.: „tools“), die zur Unterstützung der Analyse eingesetzt werden können, die in ► **Tab. 2** zusammengefasst sind. Der folgende Abschnitt beschreibt, wie diese Datenbanken und Programme auf Hochdurchsatz-Sequenzierungsdatensätze angewendet werden.

Mehrere Datenbanken enthalten Informationen über menschliche Gene, wie GeneCards [75] und die „Online Mendelian Inheritance in Man“ (OMIM) [76] Webseite. Diese Ressourcen enthalten Zusammenfassungen über klinische und funktionelle Informationen zu den derzeit charakterisierten Genen. In Bezug auf Hörverlust listet die „Hereditary Hearing Loss Homepage“ [41] die Orte und Gene auf, die an nicht-syndromalem Hörverlust beteiligt sind. Auch die häufigsten Syndrome mit Hörverlust sind verzeichnet. Viele Laboratorien wählen die Genlisten für ihre Gen-Panel-Inhalte für die Panel-Sequenzierung mithilfe dieser Datenbanken aus.

„Variant Frequency Database“ Repositorien wurden von großen Netzwerken internationaler Kollaborationen entwickelt, um Informationen über Häufigkeit von Varianten im Exom oder Genom darzustellen. Die Kenntnis der Häufigkeit einer Variante kann die Interpretation enorm erleichtern. Wenn z. B. ein Patient einen autosomal-dominanten Hörverlust hat und eine seltene Variante nicht nur in einem heterozygoten Zustand vorliegt, sondern bei anderen vermutlich normal hörenden Patienten homozygot auftritt, dann spricht dies gegen die Pathogenität im Rahmen eines autosomal-dominanten Erbgangs. Bei Schlussfolgerungen ist somit Vorsicht geboten, wie am weiter unten beschriebenen Beispiel von *GJB2* c.35delG gezeigt wird. Dennoch sind diese Datenbanken nützliche Werkzeuge, um die Häufigkeit einer Variante zu verstehen und damit Belege für oder gegen Pathogenität zu liefern. Zu beachten ist jedoch, dass nur, weil eine Variante häufig ist, das per se nicht unbedingt bedeutet, dass sie gutartig ist oder umgekehrt, weil eine Variante selten oder neu ist, sie nicht notwendigerweise pathogen ist.

Eine der ersten Datenbanken zur Dokumentation von genetischen Variationen war die „Database of Short Genetic Variants“, später abgekürzt dbSNP [77], die alle identifizierten genetischen Variationen wie einzelne Nukleotidpolymorphismen und Indels in den Genomen des Menschen und vieler anderer Arten dokumentieren soll. Weitere unabhängige Datenbanken haben sich im Laufe der Jahre entwickelt, wie der „exome variant server“ (EVS), der die Exomdaten von 6 500 europäischen und afroamerikanischen Individuen enthält [78]. Noch größere Datenbanken, wie der „Exome Aggregation Consortium Browser“ (ExAC), der Variantenfrequenzen aus den Exomdaten von 60 706 Individuen zeigt. Die ExAC wurde später auf 123 136 Exome und 15 496 Genome in einer erweiterten Datenbank namens „Genome Aggregation Database“ (gnomAD) [1] erweitert. In dieser Datenbank wurden die Variantenfrequenzen in vielen Teilpopulationen wie Latino, Nicht-Finnisch-Europäisch und Finnisch-Europäisch, Afrikanisch, Ashkenazi-Jüdisch, Ostasien, Südasien und „andere“ Individuen, die diesen Populationen nicht zugeordnet werden können, untersucht. Während der Entwicklung dieser Datenbanken wurde deutlich, dass es viele unterrepräsentierte Subpopulationen gab, die die Entwicklung einer Reihe anderer Datenbanken auslösten, nämlich des „Greater Middle Eastern Variome“ (GME), das die Exomdaten von 2 498 Individuen aus verschiedenen Ländern des

Nahen Ostens [79] und des „Iranome“, das die Exomdaten von 800 Individuen aus 8 verschiedenen ethnischen Gruppen im Iran enthält [80]. Weitere Anstrengungen zur Erfassung der genomischen Variation in seltenen und isolierten menschlichen Populationen werden notwendig sein, um die einzigartigen Varianten zu verstehen, die nur in diesen Populationen existieren. Von diesen seltenen Populationen kann man viel über das menschliche Genom und den Zusammenhang zwischen Varianten und Krankheiten lernen.

Eine der aussagekräftigsten Strategien zur Bestimmung der Pathogenität einer Variante ist die funktionelle Validierung und experimentelle Prüfung. Dies ist jedoch in klinischen Labors nicht möglich, die genetische Ergebnisse innerhalb eines begrenzten Zeitrahmens an ihre Auftraggeber melden müssen. Klinische Genetiker sind daher für die Analyse der Pathogenität von Missense-Varianten auf in silico-Pathogenitätsvorhersagen ausgewichen. Diese Programme verwenden Algorithmen, die für Varianten Pathogenitätswerte zuweisen, die Informationen über die evolutionäre Konservierung und den Einfluss der Aminosäuresubstitution auf die Proteinstruktur berücksichtigen [81]. Da die klinische Validierung in diesen Programmen nicht durchgeführt wird, werden in der Regel mehrere Programme parallel genutzt, um die möglichen Ergebnisse zu verstehen. Einige dieser Programme sind in ► **Tab. 2** dargestellt. Das Programm „MutationTaster“ [82] und „PolyPhen-2“ [83] untersuchen die Wirkung von Aminosäuresubstitutionen auf die Proteinstruktur, während „SIFT“ [84, 85] zusätzlich die Wirkung von Indel-Varianten auf die Struktur vorhersagt.

Messenger RNA (mRNA) Spleißen ist der Prozess des Entfernens der intronischen Sequenz, die keine Aminosäuren kodiert, und des Spleißens der kodierenden exonischen Sequenz zu einem einzigen Transkript. Genetische Varianten, die den normalen Spleißprozess stören, können die Genexpression und die resultierenden Proteinprodukte wesentlich beeinflussen und zur Entstehung genetischer Erkrankungen beitragen [86, 87]. Varianten, die das korrekte Spleißen von Genen beeinflussen, können weit entfernt von den normalen Intron-Exon-Sequenzgrenzen liegen. Es ist wichtig, die möglichen Auswirkungen dieser Varianten zu verstehen. Daher wurde eine Reihe von Programmen entwickelt, wie z. B. „Human Splicing Finder“ [88], „GeneSplicer“ [89], „MaxEntScan“ [90] und „NNSPLICE“ [91]. Diese Programme vergleichen die normale und die alterierte Sequenz auf eine Veränderung in konservierten Sequenzen, die für den Ablauf normaler Spleißmechanismen verwendet werden.

Die Interpretation von Varianten und deren Übersetzung in verwertbare klinische Ergebnisse stellt eine große Herausforderung dar. Gerade vor dem Hintergrund der Flut von genomischen Daten, die derzeit relativ kostengünstig und einfach zu generieren sind, haben klinische Genetiker die Aufgabe, viele seltene Varianten zu verstehen, die ein Mosaik normaler Variationen und potenziell krankheitsrelevanter Veränderungen darstellen. Eine Reihe von Datenbanken wie „ClinVar“ [92], „HGMD“ [43] und die „Leiden Open Variation Database“ (LOVD) [93] dokumentieren Interpretationen von Varianten im klinischen Kontext. Wenn Varianten in diesen Datenbanken enthalten sind, bieten sie in der Regel eine Interpretation und einen Link zu Publikationen, die diese Interpretation und klinische Informationen beschreiben. Viele dieser Datenbanken sind auf Experten angewiesen, die diese Informationen weitergeben, oder sie haben einen Stab von Varianten-Kuratoren, die dies für die Datenpflege tun. Es besteht die Gefahr, dass die Pathogenität der veröffentlichten Varianten nicht genau verstan-

► **Tab. 2** Computergestützte Datenbanken und Analysewerkzeuge, die häufig bei der Interpretation von genetischen Varianten verwendet werden.

Ressourcen für Gene und Phänotypen					
Datenbank	Hintergrund	Diagnosetool	Recherchetool	Referenz/URL	
GeneCards: The Human Gene Database	Integrative Datenbank, die Informationen über menschliche Gene sowie klinische und funktionelle Informationen enthält.	✓	✓	[75] https://www.genecards.org/	
Hereditary Hearing Loss Homepage	Online-Datenbank für Gene, die an erblichem Hörverlust beteiligt sind.	✓	✓	[41] http://hereditaryhearingloss.org	
Online Mendelian Inheritance in Man (OMIM)	Online Mendelschen Vererbung im Menschen Online-Ressource für Informationen über menschliche Gene und genetische Phänotypen	✓	✓	[76] https://www.omim.org/	
Allelfrequenz-Datenbanken: nützlich zum Verständnis der Häufigkeit einer Variante über verschiedene Ethnien hinweg					
Database	Hintergrund	Diagnosetool	Recherchetool	Referenz/URL	
Database of Short Genetic Variations (dbSNP)	Datenbank mit genetischen Varianten	✓	✓	[77] https://www.ncbi.nlm.nih.gov/projects/SNP/	
Greater Middle East Variome Project (GME)	Großes Variome-Projekt im Nahen Osten Allelfrequenz-Referenzsatz aus der Exomesequenz von 2.497 Individuen aus dem Nahen Osten.	✓	✓	[79] http://igm.ucsd.edu/gme/index.php	
Exome Aggregation Consortium (ExAC)	Exom Aggregations-Konsortium Allelfrequenz-Referenz-Set aus der Exomesequenz von 60.706 Individuen	✓	✓	[1] http://exac.broadinstitute.org/	
Exome Variant Server (EVS)	Exome Variantenserver Allelfrequenz-Referenz-Set aus der Exomesequenz von 6.503 Individuen	✓	✓	[78] http://evs.gs.washington.edu/EVS/	
Genome Aggregation Database (gnomAD)	Genom-Aggregationsdatenbank Allelfrequenz-Referenz-Set von Individuen, der 123.136 Exome und 15.496 Genome umfasst.	✓	✓	[1] http://gnomad.broadinstitute.org/	
Iranome	Allelfrequenz-Referenz-Set aus 800 Exomen, die verschiedene ethnische Gruppen im Iran repräsentieren.	✓	✓	[80] http://www.iranome.com/	
In silico Pathogenitätsvorhersagewerkzeuge für die Variantenanalyse					
Database	Hintergrund	Diagnosetool	Recherchetool	Referenz/URL	
Combined Annotation-Dependent Depletion (CADD)	Kombinierte Annotation Integration vieler Pathogenitätsannotationen in einen einzigen Pathogenitätswert in Form von C-Scores zur Priorisierung von Funktionsvarianten.	✓	✓	[109] https://cadd.gs.washington.edu/	
MutationTaster	Programm zur Vorhersage der Pathogenität zur Bestimmung der Auswirkungen von Varianten auf die DNA-Ebene	✓	✓	[82] http://www.mutationtaster.org/	
PolyPhen-2	Programm zur Bestimmung des Einflusses von Aminosäuresubstitutionen auf die Funktion eines Proteins	✓	✓	[83] http://genetics.bwh.harvard.edu/pph2/index.shtml	
Sorting Intolerant from Tolerant (SIFT)	Programm zur Bestimmung des Einflusses einer Aminosäuresubstitution aus Missense- und Indelvarianten auf die biologische Funktion eines Proteins	✓	✓	[84, 85] http://sift.jcvi.org http://sift-dna.org/sift4g	
Splicevarianten Vorhersage Tool					
Database	Hintergrund	Diagnosetool	Recherchetool	Referenz/URL	
Human Splicing Finder	Programm zur Vorhersage der Auswirkungen von Varianten auf die Spleißergebisse.	✓	✓	[88] http://www.umd.be/HSF3/	

► **Tab. 2** Fortsetzung.

Ressourcen für Gene und Phänotypen					
GeneSplicer	Programm zur Vorhersage der Auswirkungen von Varianten auf die Spleißer- gebnisse.	✓	✓	[89] http://www.cbcb.umd.edu/software/ GeneSplicer/gene_spl.shtml	
MaxEntScan	Programm zur Vorhersage der Auswirkungen von Varianten auf die Spleißer- gebnisse.	✓	✓	[90]	
NINSPLICE	Programm zur Vorhersage der Auswirkungen von Varianten auf die Spleißer- gebnisse.	✓	✓	[91]	
Klinisch orientierte Datenbanken zur Varianteninterpretation					
Database	Hintergrund	Diagnosetool	Recherchetool	Referenz/URL	
ClinVar	Datenbank, die genetische Varianten mit Phänotypassoziationen und unterstützenden Beweisen berichtet.	✓	✓	[92] https://www.ncbi.nlm.nih.gov/clinvar/	
The Connexin-deafness Homepage	Variantendatenbank für die Gene <i>GJB1</i> , <i>GJB2</i> , <i>GJB3</i> , <i>GJB6</i>	✓	✓	http://davinci.crg.es/deafness/index.php	
Deafness Variation Database (DVD)	Expert-kurierter Katalog der genetischen Variation in taubheitsassoziierten Genen	✓	✓	[42] http://deafnessvariationdatabase.org/	
Human Gene Mutation Database (HGMD)	Eine Datenbank, die alle bekannten Varianten annotiert, die für die menschliche Erkrankung verantwortlich sind.	✓	✓	[43] http://www.hgmd.cf.ac.uk/	
Leiden Open Variation Database 3.0 (LOVD v.3.0)	Datenbank mit einer Gen orientierten Sammlung von DNA-Varianten	✓	✓	[93] https://www.lovd.nl/3.0/home	
Tools zur Analyse der evolutionären Konservierung					
Database	Hintergrund	Diagnosetool	Recherchetool	Referenz/URL	
phyloP	Datenbank zur Konservierung von Nukleotidsequenzen	✓	✓	[104]	
Grantham distance	Eine Quantifizierung des physikalisch-chemischen Abstandes zur Messung biochemischer Unterschiede zwischen nativen und substituierten Aminosäuren.	✓	✓	[105]	
Genexpressionsdatenbanken					
Database	Hintergrund	Diagnosetool	Recherchetool	Referenz/URL	
gEAR Portal	Datenbank, die zelltypspezifische Genexpression basierend auf Microarray- Genexpression und RNAseq-Daten zeigt.	x	✓	[100] https://gear.igs.umaryland.edu/	
Shared Harvard Inner-Ear Laboratory Database (SHIELD)	Maus- und Hühner-Innenohr-Expressionsdatensätze, die RNAseq-, CHIP-Seq- und GeneChip-Daten enthalten.	x	✓	[99] Shen et al., 2015 https://shield.hms.harvard.edu	
Audiologisch basiertes Genotyp-Phänotyp-Tool					
Database	Hintergrund	Diagnosetool	Recherchetool	Referenz/URL	
Audiogene	Ein Programm zur Verwendung audiometrischer Daten, um vorherzusagen, welche Gene bei Patienten mit autosomal dominantem Hörverlust betroffen sein könnten.	x	✓	[96-98] https://audiogene.eng.uiowa.edu/	

den wird oder der Grad der Unsicherheit nicht korrekt kommuniziert wird. Somit können irrtümlich aufgenommene Varianten diese Datenbanken mit Fehlinformationen „verschmutzen“. Folglich sind diese Datenbanken dafür bekannt, dass sie falsch-positive Ergebnisse enthalten auch wenn sie eine sehr hilfreiche Quelle für klinische Genetiker sein können. Fehlinformationen können sich ungünstig auf die Diagnosestellung für die Patienten auswirken, zu einer ineffizienten Nutzung der Ressourcen führen und die Entdeckung wahrer Gen- und Variantenassoziationen behindern [94]. 8,5% der Varianten, die bei HGMD als krankheitsassoziiert gemeldet wurden, waren auch in einem Pool von über 1 000 asymptomatischen Individuen vorhanden, was darauf hindeutet, dass diese Varianten möglicherweise fälschlicherweise mit einer Krankheit in Verbindung gebracht wurden oder die Penetranz niedriger als erwartet war [8].

Die Folgen einer falschen Priorisierung von Varianten betreffen auch die Identifizierung von Genen und führen zu falschen Gen-Krankheitsassoziationen. Im Jahr 2014 wurde das Gen *MYO1A* als Gen für autosomal-dominante nicht-syndromale Schwerhörigkeit disqualifiziert. Dies gelang durch die Beobachtung einer disharmonischen Segregation von einer Missense- und 2 Nonsense-Varianten in 3 verschiedenen Familien [95]. In allen 3 Familien wurde die molekulargenetische Diagnose durch andere Gene für Schwerhörigkeit gesichert, die dem Phänotyp der Patienten entsprachen. Dies unterstreicht die Bedeutung der Analyse einer umfassenden Anzahl von mit Hörverlust assoziierten Genen, um eine Diagnose stellen zu können. In diesen Familien wurden auch bei normal hörenden Personen die vermutete pathogene Variante *MYO1A* nachgewiesen, was auch klar gegen eine Pathogenität spricht. Falsch assoziierte Gene können erhebliche Auswirkungen auf die genetische Beratung, das Krankheitsmanagement und sogar auch auf die Familienplanung haben.

Darüber hinaus verwenden klinische Laboratorien, die Varianten und klinische Informationen austauschen, unterschiedliche Interpretationskriterien, weshalb diese Informationen sorgfältig geprüft werden müssten. Die „deafness variation database“ (DVD) [42] ist die einzige von Experten kuratierte Datenbank, die der Kommentierung jeder Variante in jedem Schwerhörigkeits-Gen gewidmet ist. Die strenge Analyse der in dieser Datenbank enthaltenen Varianten hat zuvor als „pathogen“ erkannte Varianten als „gutartig“ neu repriorisiert. Dies erfolgte auf der Grundlage der Häufigkeit der gemeldeten Varianten in mehreren Populationen und der Berücksichtigung der Unterschiede dieser Varianten über mehrere Populationen hinweg. Die Studie, aus der die DVD hervorging, ergab, dass 93 Varianten von Genen für Schwerhörigkeit von „pathogen“ zu „gutartig“ re-kategorisiert wurden. Dies entsprach über 4% der identifizierten Varianten. Diese Datenbank ist auch mit einem „machine-learning“ basierten audiometrischen „Profiling-Tool“ namens „AudioGene“ [96–98] verbunden, um Genotypen aus audiometrischen Daten von autosomal-dominanten Formen von Hörverlust vorherzusagen. Datenbanken wie AudioGene nutzen Unterschiede in autosomal-dominanten Audiogrammen und berücksichtigen Alter und Fortschreiten des Hörverlusts [98]. Einige Beispiele von AudioGene werden im Zusammenhang mit der Bedeutung des „Phänoms“ beschrieben (► Abb. 12).

Das Verständnis der Genexpression im Innenohr ist im klinisch diagnostischen Umfeld weniger relevant, ist aber bei der Identifikation neuer Schwerhörigkeits-Gene von wesentlicher Bedeutung. Traditionelle Expressionsdatenbanken enthalten eine Vielzahl von Geweben, aber keine Informationen über die Expression von Genen im In-

nenohr. Um dieses Defizit auszugleichen, sind eine Reihe von Datenbanken entstanden, die sich auf die Expression von Genen im Innenohr spezialisiert haben. Die „Shared Harvard Inner-Ear Laboratory Database“ (SHIELD) [99] nutzt die RNA-Sequenzierung, um einen Überblick über die Genexpression von 4 Entwicklungsstadien (E16, P0, P4 und P7) der Maus-Cochlea und Utrikel zu erhalten. Eine weitere Datenbank namens „gEAR Portal“ [100] enthält Genexpressionsinformationen aus verschiedenen Entwicklungsstadien der Maus sowie Zebrafische. Expressionsmuster für das humane Innenohr sind nur fragmentarisch vorhanden.

8. Hochdurchsatz-Sequenzierungsanalyse

Für Fachärzte der Hals-Nasen-Ohren-Heilkunde, die keine eigenen praktischen Erfahrungen in der genetischen Datenanalyse haben, ist es eine Herausforderung, die unterschiedlichen Verfahren der Hochdurchsatzsequenzierung in ihrer diagnostischen Validität einzuordnen. Die Diskussion um falsch-positive Ergebnisberichte zeigt [101], dass die Varianteninterpretation sehr komplex ist. Der nächste Abschnitt zielt darauf ab, die wichtigsten Schritte zu entmystifizieren, zu vereinfachen und darzustellen wie Hochdurchsatz-Sequenzierungsdaten aktuell verarbeitet und analysiert werden.

Gen-Panel-, Exom- und Genom-Sequenzierungsdaten bestehen aus Millionen von Ablesungen (engl.: „reads“), die in einer FASTQ-Datei enthalten sind. Jede Probe hat 2 FASTQ-Dateien (read 1 und read 2), die die bidirektionale Ausrichtung der Sequenzierung darstellen (► Abb. 10). Diese Dateien enthalten auch Basisaufruf- und Qualitätsinformationen und werden als Sequenzeingabe für die Ausrichtung oder Abbildung auf die menschliche Referenzgenomsequenz verwendet. Bei der Ausrichtung werden die Millionen von Ablesungen von kurzen DNA-Abschnitten an der richtigen Stelle des menschlichen Referenzgenoms organisiert. Die Visualisierung der Leseausrichtung kann die Tiefe oder Abdeckung pro Basis anzeigen, d. h. die Anzahl der Ablesungen an jeder Basisposition (► Abb. 10). Nach dem Ausrichten der Ablesungen werden Varianten aufgerufen, die dann einer sogenannten Variantenfilterung unterzogen werden, die benutzerdefinierte Parameter verwendet, um die für die manuelle Analyse verbleibenden Varianten zu reduzieren (► Abb. 11).

Die Variantenfilterung kann auf eine Teilmenge von Genen beschränkt werden, z. B. auf diejenigen, die an autosomal-rezessiven oder autosomal-dominanten Hörverlust-Genen beteiligt sind, wenn das familiäre Vererbungsmuster klar genug ist, um dies zu unterscheiden. Ebenfalls von großem Interesse sind Varianten in und neben der kodierenden Sequenz, sodass in der Regel ein Filter verwendet wird, um intronische Varianten zu entfernen, die möglicherweise nicht von Interesse sind. Obwohl in den Vor- und Nachbearbeitungsschritten bereits eine Reihe von Qualitätskontrollschritten durchgeführt werden, verbleiben viele Varianten niedriger Qualität in den Daten, die durch die Anwendung von Qualitätsgrenzen entfernt werden müssen.

Ein weiterer wichtiger Schritt ist die Filterung gegen niedrige Allelfrequenzen (engl.: minor allele frequencies; MAFs). MAFs werden berechnet als die relative Häufigkeit des weniger häufigen (kleinen) Allels oder der Variante in den Allelen, die in einem Pool von sequenzierten Individuen identifiziert wurden. Zum Beispiel identifiziert eine bestimmte Population mit 50 Individuen eine Person mit einer heterozygoten Variante. Fünfzig Personen haben jeweils 2 Allele, also insgesamt 100 Allele. Die MAF würde wie folgt berechnet werden [1

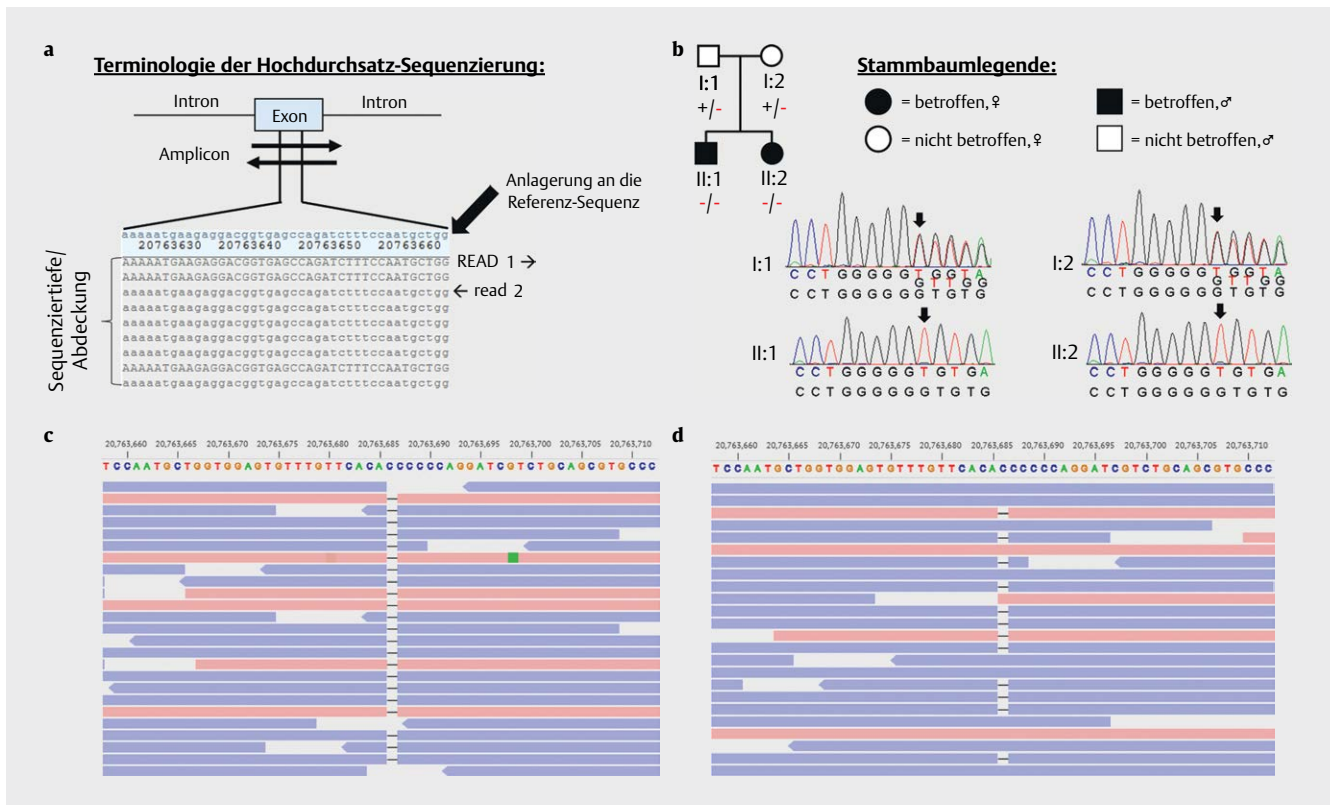


Abb. 10 Hochdurchsatz-Sequenzierung am Beispiel der *GJB2* c.35delG-Deletion. **a** Eine visuelle Darstellung einiger Merkmale der Hochdurchsatz-Sequenzierung. **b** Stammbaum einer Familie, mit normal hörenden Eltern und 2 betroffenen Kindern. Frauen werden durch Kreise dargestellt. Männer werden durch Quadrate dargestellt. Das Symbol + steht für die normale DNA-Sequenz, das Symbol – für die Deletion. +/- zeigt eine Person, die heterozygot ist und -/- zeigt eine Person, die homozygot für die c.35delG-Deletion ist. Unterhalb des Stammbaums sind repräsentative Sanger-Sequenzierungsbilder mit der heterozygoten und homozygoten Deletion dargestellt. Eine Visualisierung von c.35delG mit Sequenzierung zeigt homozygote **c** und heterozygote **d** Deletionen. Die Deletion wird durch eine Lücke in der gelesenen Farbfolge dargestellt. Diese Bilder wurden mit dem Integrative Genomics Viewer [1] von gnomAD visualisiert.

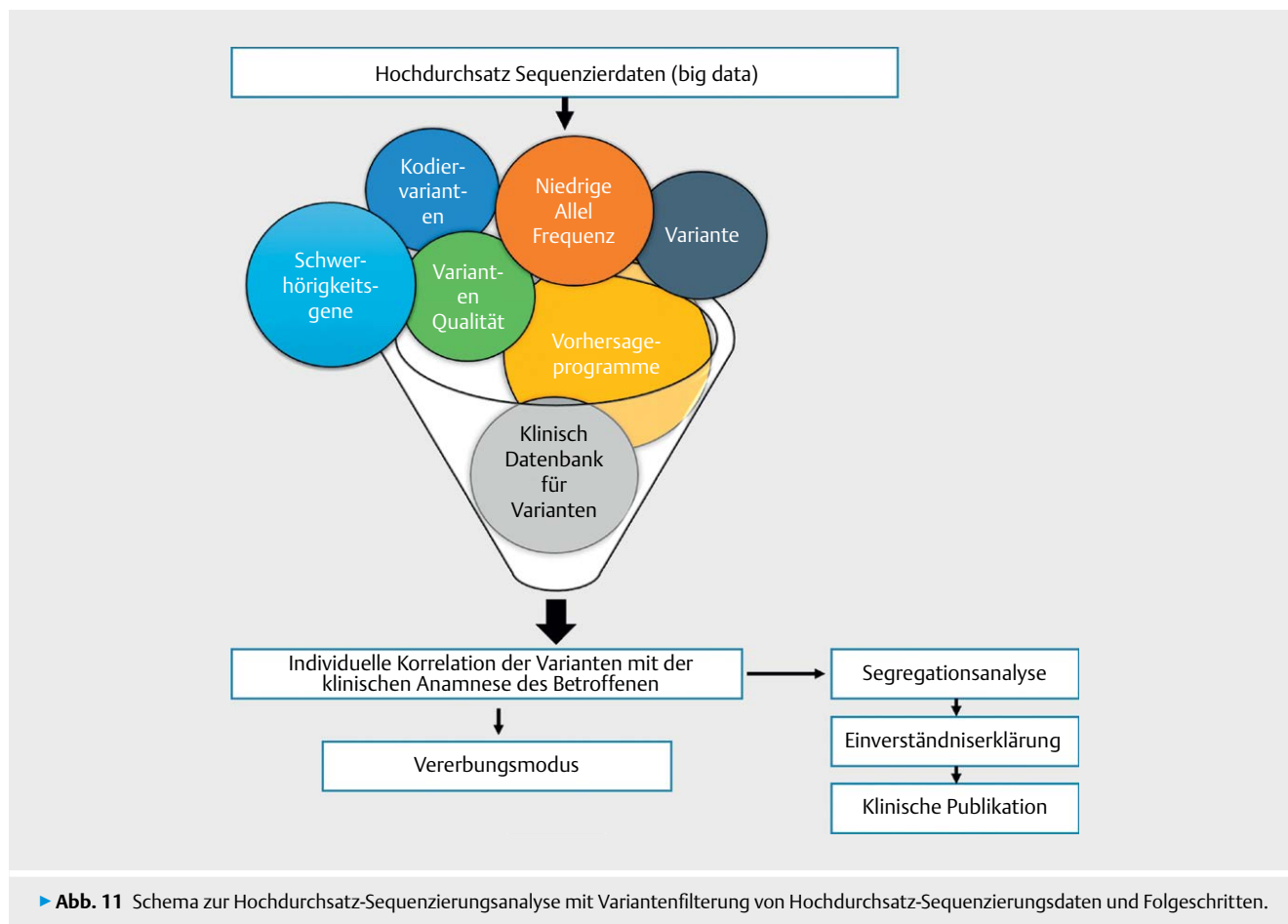
Wechselalternatives Allel]/[100 Allele gesamt] für eine Häufigkeit von 0,01 (1 %) bei den getesteten Personen. Die Einstellung optimaler MAF-Schwellenwerte ist wichtig, um häufige Varianten, die wahrscheinlich gutartig sind, deutlich zu reduzieren [42]. Optimale MAF-Schwellenwerte für Hörverlust wurden in großen Kohorten aus mehreren Labors evaluiert, die Expertenempfehlungen ermöglicht haben. MAF-Schwellenwerte werden als $\leq 0,00007$ (0,007 %) für Varianten in autosomal-rezessiven Hörverlustgenen und $\leq 0,00002$ (0,002 %) für Varianten in autosomal-dominanten Hörverlustgenen empfohlen [102].

Ein weiterer gängiger Filterparameter ist die Auswahl des Variantentyps. Durch die Auswahl von nicht-synonymen Varianten („missense, splice, indel, stop gain und stop loss, sowie start gain und start loss“) würden bspw. alle synonymen Varianten entfernt, obwohl synonyme Varianten für die Beeinflussung der „Spleißlandschaft“ von Interesse sein können, welche die Proteinfunktion stark beeinträchtigen könnte. Ein weiterer Filterschritt besteht darin, jede Variante mithilfe verschiedener Pathogenitätsvorhersage-Tools zu analysieren und zu dokumentieren, ob die Vorhersageergebnisse jede Variante als pathogen oder gutartig bewerten. Während nicht jede Variante in gut untersuchten Genen in klinisch orientierten Variantendatenbanken dokumentiert ist, werden diese Datenbanken

referenziert, um festzustellen, ob eine bestimmte Variante bereits bei einem Patienten interpretiert wurde.

Auch die Konservierung von Nukleotid- und Aminosäuresequenzen wird berücksichtigt, da Varianten, die hochkonservierte Nukleotide und Aminosäuren betreffen, a priori wahrscheinlich eher pathogen sind [103]. PhyloP-Scores messen durch den Vergleich mehrerer Arten den Grad der evolutionären Konservierung in Nukleotiden, indem sie beurteilen, ob die Substitutionsraten langsamer oder schneller als erwartet sind [104]. Die phyloP-Werte reichen von -14 (nicht konserviert) bis 6,4 (hochkonserviert). Grantham-Distanzen bewerten den evolutionären Abstand zwischen 2 Aminosäuren unter Berücksichtigung der biochemischen und physikalischen Eigenschaften von Aminosäuren und reichen von 0 bis zu einem maximalen Abstand von 215. Je weiter 2 Aminosäuren voneinander entfernt sind, desto geringer ist deren „Austauschbarkeit“ und desto höher ist die Wahrscheinlichkeit, dass ein Austausch von Aminosäuren pathogen ist [105, 106].

Schließlich kann die Festlegung von vordefinierten „Allel-Balance“-Werten (Verhältnis der Anzahl der Ablesungen mit der Variante zur Anzahl der Ablesungen mit der Referenzbasis) je nach Vererbung Varianten zeigen, die homozygot oder heterozygot erscheinen. Es wird erwartet, dass 50% der Ablesungen eine Variante zeigen würden, wenn eine heterozygote Variante vorhanden ist, obwohl



diese „Allel-Balance“ bei Ablesungen hoher Qualität hohe Abweichungen aufweisen kann. Bei einer homozygoten Variante wird erwartet, dass 100% der Ablesungen die Variante zeigen. Die Sanger-Sequenzierung wird zur Validierung von Varianten empfohlen, die „Allel-Balance“-Werte aufweisen, die von den akzeptierten Grenzen (engl: „Cutoffs“) abweichen [49].

Nachdem die Varianten im Rahmen der Anamnese des Patienten, durch Segregationstests oder durch Test von zusätzlichen betroffenen oder nicht betroffenen Familienmitglieder analysiert und erheblich reduziert worden sind, können Fehlinterpretationen von Varianten vermieden werden.

9. Beispiel einer Variantenanalyse aus *GJB2*

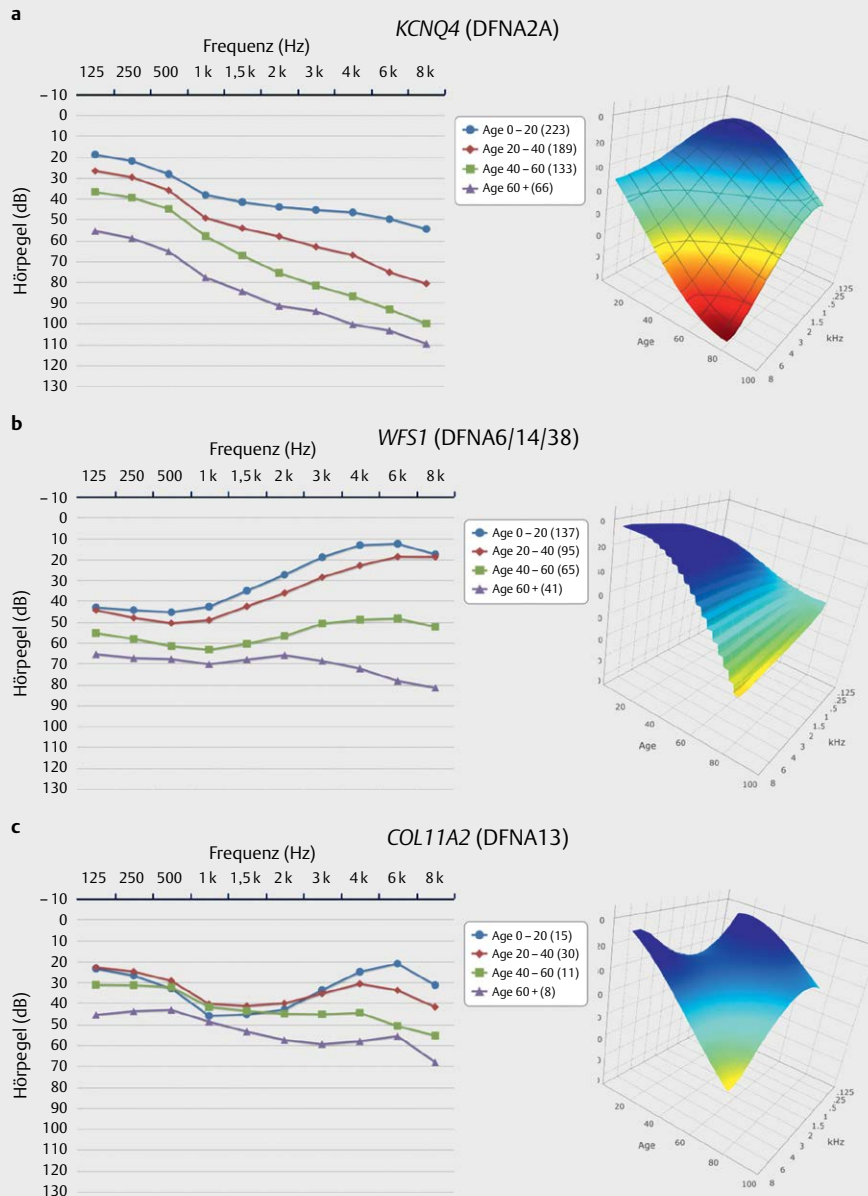
Eine europäische Familie mit nicht betroffenen Eltern fragt genetische Tests für ihre beiden Kinder an (► **Abb. 10b**), die jeweils einen angeborenen, hochgradigen Hörverlust aufweisen. Das Vererbungsmuster in dieser Familie scheint einem autosomal-rezessiven Muster zu folgen. Nach der genetischen Untersuchung stellt sich heraus, dass die Kinder im Gen *GJB2* im homozygoten Zustand eine einzige Nukleotiddeletion (c.35delG) haben, während ihre Eltern beide heterozygot sind. Diese Deletion ist die häufigste Ursache für genetisch bedingten Hörverlust in Europa. Die Hochdurchsatz-Sequenzierungsdaten zeigen, dass die betroffenen Kinder homozygot sind,

wobei 100% ihrer Ablesungen die Deletion zeigen (► **Abb. 10c**). Die nicht betroffenen Eltern sind heterozygot, wobei etwa die Hälfte ihrer Ablesungen die Deletion und die andere Hälfte die korrekte Sequenz zeigt (► **Abb. 10d**). Die *GJB2*-Genexpression ist gut untersucht, sie ist in den Stützzellen und den Haarzellen des vestibulären und cochleären Epithels während mehrerer Entwicklungsstadien der Maus vorhanden. Obwohl dieses Gen und diese Variante gut charakterisiert sind, liefert es ein gutes Beispiel dafür, wie wichtig es ist, Expertenrichtlinien für die Variantenfilterung anzuwenden, um nicht einen potenziell signifikanten Befund zu übersehen, sondern auch eine korrekte Assoziation zwischen Variante und Erkrankung zu gewährleisten.

► **Tab. 3** zeigt eine Zusammenfassung der Informationen aus den verschiedenen Ressourcen, die für die Variantenanalyse verwendet werden. *GJB2* kodiert das gap-junction-beta-2-Gen, das am besten für nicht-syndromalen Hörverlust (DFNB1A) bekannt ist, aber auch mit autosomal-dominantem nicht-syndromalem Hörverlust (DFNA3A) assoziiert ist. Darüber hinaus ist *GJB2* mit einer Reihe von autosomal-dominanten Syndromen wie dem Bart-Pumphrey-Syndrom, dem Hystrix-like-Ichthyosis-Taubheits-Syndrom, dem Keratitis-Ichthyosis-Taubheits-Syndrom, Keratoderma und Palmoplantar mit Taubheit und dem Vohwinkel-Syndrom verbunden. Es gibt derzeit über 400 bekannte Varianten für das Gen *GJB2*, die in HGMD [43] dokumentiert sind, wobei das c.35delG die häufigste Variante bei nicht-syndromalen Hörverlust ist. Diese Variante hat eine MAF von

► **Tab. 3** Ein Beispiel für eine abweichende Interpretation der homozygoten Deletion von *GJB2* c.35delG.

Datenbank	Informationen
GJB2-Informationen	
GeneCards	Gap Junction Protein Beta 2; assoziiert mit Vohwinkel-Syndrom und Keratodermie, palmoplantar mit Taubheit, sowie autosomal-dominant (DFNA3A) und autosomal-rezessiv (DFNB1A) Hörverlust.
Hereditary Hearing Loss Homepage	DFNA3A, DFNB1A
OMIM	Gap Junction Protein beta-2; beteiligt am Bart-Pumphrey-Syndrom, autosomal-dominant (DFNA3A) und autosomal-rezessiv (DFNB1A) nicht-syndromaler Hörverlust, Hystrix-like-Ichthyosis-Taubheit-Syndrom, Keratitis-ichthyosis-Taubheitssyndrom, Keratodermie und palmoplantar mit Taubheit und Vohwinkel-Syndrom.
Allelfrequenz Analyse der c.35delG-Variante	
dbSNP	MAF = 0,002; Klinische Bedeutung: pathogen
GME	Gesamtanzahl der Allele (Varianten): 5, keine homozygoten Individuen in 1.984 Allelen (992 Individuen); MAF = 0,00252
ExAC	Gesamtanzahl der Allele (Varianten): 733, einschließlich 3 homozygoter Individuen in 121.352 Allelen (60.676 Individuen); MAF = 0,00604
EVS	Gesamtanzahl der Allele (Varianten): 93, keine homozygoten Individuen in 12.425 Allelen (6.212 Individuen); MAF = 0,00748
gnomAD	Gesamtanzahl der Allele (Varianten): 1.721, darunter 10 homozygote Individuen in 275.002 Allelen (135.501 Individuen); MAF = 0,006258
Iranome	Gesamtanzahl der Allele (Varianten): 3, keine homozygoten Individuen in 1.600 Allelen; MAF = 0,00187575
Klinisch orientierte Datenbanken zur Varianteninterpretation	
ClinVar	Klinische Bedeutung: Pathogen von 26 Einsendern, keine Konflikte bei der Interpretation der unterschiedlichen Pathogenität. Bedingungen: Taubheit, autosomal-rezessiv 1A, verstümmelnde Keratodermie, Hystrix-like-Ichthyosis-Taubheit-Syndrom, autosomal-dominante Keratitis-ichthyosis-Taubheitssyndrom, Keratodermie palmoplantärer Taubheit, Knöchelpolster, Taubheit und Leukonychie-Syndrom, Taubheit, autosomal-dominant 3a, digene <i>GJB2/GJB6</i> Taubheit, nicht-syndromaler Hörverlust und Taubheit, Hörverlust, bilateraler sensorineuraler Hörverlust, bilateraler leitfähiger Hörverlust, schwerer sensorineuraler Hörverlust, nicht-syndromaler Hörverlust, rezessiver, Taubheit
The Connexin-deafness Homepage	Autosomal-rezessive nicht-syndromale Taubheit
DVD	Pathogener, autosomal-rezessiver, nicht-syndromaler Hörverlust
HGMD	Taubheit, autosomal-rezessiv 1
LOVD v. 3.0	Pathogen
In silico Pathogenitätsvorhersage Programme zur Variantenanalyse	
CADD	Score: 24.9
MutationTaster	Krankheit verursachend
PolyPhen-2	Nicht bewertet
SIFT	Nicht bewertet
Programme zur Vorhersage von Spleißvarianten	
Human Splicing Factor	Kein signifikanter Effekt vorhergesagt
GeneSplicer	Kein signifikanter Effekt vorhergesagt
MaxEntScan	Kein signifikanter Effekt vorhergesagt
NNSPLICE	Kein signifikanter Effekt vorhergesagt
Analyse Konservierung	
phyloP	Nicht bewertet
Grantham distance	Nicht bewertet
Genexpressions-Datenbanken	
gEAR Portal	Exprimiert in P0 Maushaarzellen, P1 Haarzellen, Stützzellen und nicht-sensorischen Zellen, E16.5 und P0 Maus cochleäres und vestibuläres sensorisches Epithel.
SHIELD	FACS-sortierte Haarzellen und Ganglienzellen: Expression in Utrikel und Cochlea embryonal und postnatal (E12, E13, E16, P0, P6 und P15).
Audiometrisches Profilierungswerkzeug	
AudioGene	Gen nicht enthalten
Die volle Position der <i>GJB2</i> c.35delG Löschung ist Chr13(GRCh37):g. 20763686, NM_004004.5:c.35del, p.Gly12Valfs*2	



► **Abb. 12** 2- und 3-dimensionale Darstellungen von Audiogrammen, erstellt mit AudioGene, einem Programm zur Verarbeitung von Sequenzierdaten und Audiogrammen durch maschinelles Lernen. Der Hörverlust durch die Gene **a** *KCNQ4*, **b** *WFS1* und **c** *COL11A2* manifestiert sich mit deutlich unterschiedlichen Audioprofilen. Bilder mit Genehmigung von Smith R] [96–98].

0,002 (0,2%) bis 0,007 (0,7%), je nach Datenbank. In der Literatur sind Trägerraten von bis zu 0,0189 (1,89%) in Europa beschrieben [48]. Bei Einhaltung der Expertenempfehlungen zur Filterung einer autosomal-rezessiven Erkrankung würde jede Variante mit einem $MAF \geq 0,00007$ (0,007%) aus der Analyse entfernt, wodurch auch diese wichtige Variante entfernt würde. Für die Entstehung der c.35delG-Deletion wird ein sogenannter Gründereffekt (engl.: „founder effect“) vermutet. Hier stammt die Variante von einer einzelnen Person, die sie an Nachkommen weitergegeben hat [107]. Viele Varianten von Hörverlust können als Gründermutationen klassifiziert werden, die besonders hohe MAFs aufweisen können. Diese alten Varianten, die vor vielen tausend Jahren entstanden sind, wur-

den durch Zeit und Raum getragen. Deshalb enthalten Analyseerichtlinien auch eine Liste von Genen, die MAF-Cutoff-Empfehlungen nicht einhalten. Auch *GJB2* gehört zu den Genen, für die Ausnahmen gelten. Bei genauerer Betrachtung der „gnomAD“-Varianten werden insgesamt 1 721 Varianten unter 275 002 Allelen (132 501 Individuen) mit Sequenzierungsdaten zu dieser Position gemeldet. In dieser Datenbank sind auch 10 homozygote Patienten erfasst, die keine schweren Kinderkrankheiten hatten und deren Verwandte ersten Grades gesund waren. Da der Hörverlust durch die c.35delG-Deletion sehr früh im Leben auftritt, ist zu erwarten, dass diese Personen nicht in die Datenbank aufgenommen werden. Die Entwickler der Datenbank weisen jedoch darauf hin, dass einige Personen mit schwerer

Erkrankung, die in einer geringeren Häufigkeit als in der Allgemeinbevölkerung [1] auftritt, fälschlicherweise in die Datenbank aufgenommen wurden. Dies ist eine wichtige Erkenntnis, die bei der Verwendung verschiedener Allelfrequenz-Datenbanken zu beachten ist. Sie eröffnet aber auch die Möglichkeit für mehrere Erklärungsversuche. Bei unvollständiger Penetranz kann z. B. auch bei Normalhörenden eine Deletion nachgewiesen werden. Für diese spezielle Variante ist dies unwahrscheinlich, wurde aber für 2 weitere Varianten in *GJB2* (p.Met34Thr und p.Val37Ile) [102] festgestellt. Es könnte auch bedeuten, dass ein Hörverlust bei diesen Personen nach der Rekrutierung auftrat, was nicht rückverfolgbar ist.

Die klinisch orientierten Datenbanken wie „The Connexin-deafness Homepage“ [108], die DVD [42], HGMD [43], und LOVD v.3.0 [93] stimmen darin überein, dass diese Variante pathogen ist, mit einem möglichen Vorbehalt. ClinVar zeigte 27 Einträge für diese Deletion. Elf spezifizierte autosomal-rezessive nicht-syndromale Hörverluste und 11 weitere hatten Einträge mit der Auflistung „Hörschaden“ oder „nicht vorhanden“, was bedeutet, dass die Einsender keine Bedingung oder Vererbungsart angegeben haben. Drei ClinVar-Zusender gaben an, dass diese Variante an einem autosomal-dominanten Hörverlust beteiligt sei (Einreichungen: SCV000487402.1, SCV000700274.1 und SCV000536698.1). In einem dieser Beiträge wurden mehrere autosomal-dominante syndromale Formen von Hörverlust und autosomal-rezessivem nicht-syndromalem Hörverlust erwähnt (SCV000536698.1) und ein anderer Antragsteller führte an, dass diese Variante mit autosomal-dominanten Syndromen assoziiert sei. Ein letzter Eintrag listete diese Variante als an digener Taubheit beteiligt auf (*GJB2/GJB6*) (SCV00000038810.5). Nach diesen Angaben wäre auch die Möglichkeit eines autosomal-dominanten Hörverlustes bei den Trägereltern erhöht. Für den Laien können solche ClinVar-Einträge Verwirrung in der Interpretation stiften.

Nur 2 der 4 beschriebenen Pathogenitätsvorhersage-Tools sind in der Lage, diese Deletion zu bewerten. PolyPhen-2 und SIFT liefern Vorhersagen über Substitutionen. Der CADD-Score [109] für diese Deletion beträgt 24,9, d. h. er liegt maximal bei etwa 0,5% der schädlichen Variationen im menschlichen Genom. „MutationTaster“ hat diese Deletion als krankheitsverursachend eingestuft. Es wird nicht vorhergesagt, dass die Spleißung durch diese Änderung signifikant beeinträchtigt wird. Die Analyse der Konservierung des Nukleotids (phyloP) und der Aminosäure (Grantham-Distanz) kann nur Substitutionen und nicht Deletionen beurteilen, sodass diese nicht in der Lage sind, bei der Interpretation zu helfen.

Natürlich können Kliniker anhand von Literatur und klinischen Dokumenten über den Zusammenhang der c.35delG-Deletion bei *GJB2* mit Hörverlust die Kinder mit *GJB2*-assoziiertem Hörverlust und die Eltern als Träger sicher diagnostizieren, was für Wiederholungsrechnungen bei erneutem Kinderwunsch hilfreich sein kann.

10. Vom Genom zum Phänom

Nach der außergewöhnlichen Leistung des Humangenomprojektes, die Referenzsequenz des menschlichen Genoms zu liefern, ergaben sich viele Herausforderungen in Bezug auf die effektive Anwendung dieses Wissens auf Erbkrankheiten. Die Kenntnis der „Anatomie“ des menschlichen Genoms kann keine direkte Aussage über die in den Genotypen kodierten Phänotypen machen. Da jedoch ein Großteil der Theorie und Praxis der Medizin mit einem Phänotyp beginnt, war

es sinnvoll, das Wort „Phänom“ kurz nach der Beschreibung des Genoms einzuführen [110].

Die Phänomik erfasst den natürlichen Verlauf einer Krankheit und beschreibt das genaue Spektrum der Krankheitssubtypen und -komplikationen sowie weitere phänotypische Informationen [111]. Analog dazu zielt die Phänomik darauf ab, die gleichen zentralisierten, gut etablierten, vernetzten und konsolidierten Strategien zur Beschreibung des natürlichen Verlaufs aller Phänotypen, die die Genomik bereits für Annotationen, Methoden und Standards für die genaue Beschreibung jedes genomischen Elements besitzt, zur Verfügung zu stellen [112]. Die effektive Umsetzung von phänombasierten Forschungsansätzen erfordert neue Informatik- und Datenanalysestrategien [113]. Die Entwicklung der „Human Phenotype Ontology“ (HPO) Datenbank bietet eine standardisierte Terminologie phänotypischer Anomalien zur Rationalisierung der „phänotypischen“ Differenzialdiagnostik [114]. Die HPO hat derzeit über 13 000 Begriffe und über 156 000 Einträge zu Erbkrankheiten und hat sich als leistungsfähiges Werkzeug zur Verbesserung der Exom- und Genomanalyse erwiesen. Durch die Integration von HPO-Terminologie, die eine rationalisierte „tiefe Phänotypisierung“ (engl.: deep phenotyping) von Patienten ermöglichte, konnten im NIH „Undiagnosed Disease Program“ und im „Undiagnosed Diseases Network“ die molekulare Diagnose verbessert werden. Die erneute Analyse von Exom-Sequenzierungsdaten von bisher „nicht diagnostizierbaren“ Patienten hatte die diagnostische Lösung von zusätzlichen 10–20% der Patienten zur Folge [115, 116]. Die HPO-Datenbank enthält derzeit über 1 600 Ergebnisse mit dem Wort „Hörverlust“. Eine spezifische Terminologie könnte die Liste der genetischen Erkrankungen mit Hörverlust schnell von 1 600 auf mehrere Dutzend eingrenzen. Die Kenntnis des gesamten Phänotyps versehen mit einer optimierten Terminologie kann die Genomanalyse erheblich erleichtern. HPO-Begriffe werden derzeit in Bioinformatik-Pipelines der Hochdurchsatz-Sequenzierung integriert, um die Analysegeschwindigkeit bei Patienten mit pathogenen Varianten in bereits identifizierten und charakterisierten Genen deutlich zu erhöhen.

Hörverlust hat die besondere Herausforderung einer ausgeprägten klinischen Heterogenität des Phänotyps. Auch bei Personen in der gleichen Familie, die die gleiche Variante segregieren, kann ein heterogener Phänotyp die genaue Charakterisierung des Hörverlusts erschweren. Durch die Untersuchung einer großen Anzahl von Patienten mit Hörverlust mit pathogenen Varianten im gleichen Gen konnten mehrere Gene identifiziert werden, die robuste Assoziationen aufweisen. Dies wurde insbesondere bei autosomal-dominantem Hörverlust untersucht und hat zur Entwicklung eines Tools „AudioGene“ geführt [96, 98]. Dieses auf „machine learning“ basierende Programm analysiert audiologische Profile von Patienten mittels eines computergestützten Clustering-Algorithmus und priorisiert die wahrscheinlichsten Gene für autosomal-dominanten Hörverlust für das Mutationsscreening. Beim Vergleich der prädiktiven Leistung von „AudioGene“ mit einer Expertengruppe, die die wahrscheinlichsten autosomal-dominanten Gene auflistet, die an Patienten mit verfügbaren Audiogrammdaten beteiligt sein könnten, übertraf „AudioGene“ die Expertengenvorhersage um 33% [96]. Die Weiterentwicklung von „AudioGene“ fügt dem Audioprofil nun eine dritte Dimension hinzu: das Alter [98]. Dieses zusätzliche Merkmal ist von klinischer Bedeutung für den autosomal-dominanten Hörverlust, da die meisten For-

men des dominanten Hörverlustes progressiv sind. Das Alter wird leicht durch Farbe auf einer 3-dimensionalen Oberfläche visualisiert.

Die drei in ► **Abb. 12** gezeigten Beispiele zeigen die zwei- und dreidimensionale Darstellung der Gene *KCNQ4* (DFNA2A), *WFS1* (DFNA6/14/38) und *COL11A2* (DFNA13). Die AudioGene Profile in 2- und 3-dimensionalen Formen des Gens *KCNQ4* zeigen einen charakteristisch progressiven, hochfrequenten Hörverlust (► **Abb. 12a**). Vergleicht man dies mit den Audioprofilen für die Gene *WFS1* mit progressivem, niederfrequentem Hörverlust (► **Abb. 12b**) und *COL11A2* mit relativ stabilem Mittelfrequenz- und progressivem, hochfrequentem Hörverlust (► **Abb. 12c**), kann man sich vorstellen, wie stark dies für die Vorhersage der zugrundeliegenden genetischen Faktoren sein kann, die bei einer genetisch heterogenen Erkrankung wie Hörverlust hilfreich sind. Dieses Tool bietet eine diagnostische Strategie zur Unterstützung genauer genetischer Tests und ist ein Beispiel für die Verknüpfung von audiologischen „big data“ mit der Genetik.

11. Die Perspektiven der Hochdurchsatz-Sequenzierung

„Big data“ hat das Gebiet der Diagnostik genetischer Schwerhörigkeit revolutioniert. Moderne Sequenzierverfahren haben das Potenzial, wichtige medizinische Diagnosen zu entschlüsseln, und so die Patientenversorgung erheblich zu beeinflussen und die personalisierte Medizin zu unterstützen. Das Feld zeichnet sich aktuell durch ständige große Fortschritte aus und nicht jede Variante in unserem Genom ist derzeit bekannt und wird richtig verstanden. Die Aussage von Cynthia C. Morton's American Society of Human Genetics Presidential Address im Jahr 2014 betonte, dass wir uns im gegenwärtigen Zustand der Genetik gerade an einem besonderen Punkt befinden: „we find ourselves building the plane as we are flying it“ [117]. Es ist einfach die Unsicherheiten des Gebiets von „big data“ als Zeichen zu interpretieren, dass das Gebiet wenig Nutzbares zu bieten hat. Aber die Wahrheit ist, dass die Genomik wahrscheinlich eine immer größere Rolle in der Patientenversorgung spielen wird. Eines Tages werden wir hoffentlich in der Lage sein, fast alle Patienten auf molekulargenetischer Grundlage zu diagnostizieren, sogar solche mit höchst-seltenen genetischen Störungen.

Es ist schwer vorauszusehen, ob in der Zukunft jedes Neugeborene bei der Geburt einem genetischen Screening unterzogen wird, um Betroffene zu identifizieren, bevor die Symptome auftreten. Das Screening kann von gezielten Gen-Panels bis hin zur Genomsequenzierung reichen, um die falsch-positive-Rate zu reduzieren, die z. B. beim Stoffwechsel- und Hörscreening von Neugeborenen auftreten kann. Die Befürworter der Genomik sehen in dieser Technologie ein großes Potenzial, das die Voraussetzungen für eine lebenslange personalisierte medizinische Versorgung schafft. Dies könnte zusätzliche Informationen bei gefährdeten Personen unter bestimmten Bedingungen bieten. Eine rigorose Erforschung der medizinischen und ethischen Implikationen wird hoffentlich die vorteilhaftesten Wege aufzeigen und gleichzeitig die Wünsche und Rechte der Patienten respektieren.

Die genetische Landschaft des Hörverlustes ist noch lange nicht vollständig charakterisiert und für jede „Bekannt“ scheint es eine lange Liste von „Unbekannten“ zu geben. Patienten, profitieren von der molekulargenetischen Diagnose, gewonnen aus der massenhaften Zusammenführung multidisziplinärer großer Datenmengen. Pa-

tienten ohne genetische Diagnose trotz genetischer Diagnostik können in der Zukunft die genetische Untersuchung wiederholen. Mit der Weiterentwicklung der Genomdaten und „big data“ Methoden können zunehmend Gene identifiziert und charakterisiert werden. Eines Tages kann es dadurch möglich sein, dass jeder Betroffene routinemäßig seine genetische Diagnose erhält.

FAZIT FÜR DIE PRAXIS

Eine genetische Untersuchung in Form einer Gen-Diagnostik sollte nach Anamnese, körperlicher Untersuchung und audiologischer Untersuchung für die Diagnose einer Schwerhörigkeit erfolgen. Eine Gen-Diagnostik kann nachfolgende diagnostische Verfahren, die invasiv sein könnten, vermeiden. Die Gen-Diagnostik erlaubt es, den Patienten und seine Familie bezüglich Therapieoptionen und Familienplanung zu beraten. Diese bilden die Grundlage für die Entwicklung einer personalisierten Medizin und in Zukunft einer möglicherweise maßgeschneiderten Pharmakotherapie oder einer individuellen molekularen Therapie.

Danksagung

Die Autoren danken Nora Knoblich für die Unterstützung bei den Abbildungen und Tabellen. Wir danken Prof. R.J.H. Smith (University of Iowa, USA) und Prof. E. Seidl (University Tübingen, Germany) für die Bereitstellung der Daten für Abbildung 2 und 12.

Interessenkonflikt

Die Autoren geben an, dass kein Interessenkonflikt besteht.

Literatur

- [1] Lek M, Karczewski KJ, Minikel EV et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; 536: 285–291
- [2] Zarrei M, MacDonald JR, Merico D et al. A copy number variation map of the human genome. *Nat Rev Genet* 2015; 16: 172–183
- [3] Xue Y, Chen Y, Ayub Q et al. Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. *Am J Hum Genet* 2012; 91: 1022–1032
- [4] Sudmant PH, Rausch T, Gardner EJ et al. An integrated map of structural variation in 2,504 human genomes. *Nature* 2015; 526: 75–81
- [5] Narasimhan VM, Hunt KA, Mason D et al. Health and population effects of rare gene knockouts in adult humans with related parents. *Science* 2016; 352: 474–477
- [6] Narasimhan VM, Xue Y, Tyler-Smith C. Human Knockout Carriers: Dead, Diseased, Healthy, or Improved? *Trends Mol Med* 2016; 22: 341–351
- [7] Consortium UK, Walter K, Min JL et al. The UK10K project identifies rare variants in health and disease. *Nature* 2015; 526: 82–90
- [8] Cassa CA, Tong MY, Jordan DM. Large Numbers of Genetic Variants Considered to be Pathogenic are Common in Asymptomatic Individuals. *Hum Mutat* 2013; 34: 1216–1220

- [9] Sulem P, Helgason H, Oddson A et al. Identification of a large set of rare complete human knockouts. *Nat Genet* 2015; 47: 448–452
- [10] Dahm R. Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum Genet* 2008; 122: 565–581
- [11] Maderspacher F. Rags before the riches: Friedrich Miescher and the discovery of DNA. *Current Biology* 2004; 14: R608–R608
- [12] Avery OT, MacLeod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* Type III. *J Exp Med* 1944; 79: 137–158
- [13] Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 1953; 171: 737–738
- [14] Zallen DT. Despite Franklin's work, Wilkins earned his Nobel. *Nature* 2003; 425: 15
- [15] Efstratiadis A, Kafatos FC, Maniatis T. The primary structure of rabbit beta-globin mRNA as determined from cloned DNA. *Cell* 1977; 10: 571–585
- [16] Shendure J, Balasubramanian S, Church GM et al. DNA sequencing at 40: past, present and future. *Nature* 2017; 550: 345–353
- [17] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977; 74: 5463–5467
- [18] Saiki RK, Scharf S, Faloona F et al. Enzymatic Amplification of Beta-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle-Cell Anemia. *Science* 1985; 230: 1350–1354
- [19] Rose EA. Applications of the polymerase chain reaction to genome analysis. *FASEB J* 1991; 5: 46–54
- [20] Dickman S. West-Germany Voices Objections to European Genome Project. *Nature* 1988; 336: 416–416
- [21] Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. *Science* 2003; 300: 286–290
- [22] Murray JC, Buetow KH, Weber JL et al. A comprehensive human linkage map with centimorgan density. Cooperative Human Linkage Center (CHLC). *Science* 1994; 265: 2049–2054
- [23] Fleischmann RD, Adams MD, White O et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995; 269: 496–512
- [24] Dunham I, Shimizu N, Roe BA et al. The DNA sequence of human chromosome 22. *Nature* 1999; 402: 489–495
- [25] Adams MD, Celniker SE, Holt RA et al. The genome sequence of *Drosophila melanogaster*. *Science* 2000; 287: 2185–2195
- [26] Bier E. *Drosophila*, the golden bug, emerges as a tool for human genetics. *Nat Rev Genet* 2005; 6: 9–23
- [27] Shendure J, Porreca GJ, Reppas NB et al. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 2005; 309: 1728–1732
- [28] [Anonymous]. Method of the year. *Nat Methods* 2008; 5: 1
- [29] Sevilla G. New GUINNESS WORLD RECORDS™ Title Set for Fastest Genetic Diagnosis. In: Rady Children's Hospital San Diego; 2018
- [30] Vence T. \$1,000 Genome at Last? In, *The Scientist* 2014
- [31] Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol* 2008; 26: 1135–1145
- [32] Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci, U S A* 1977; 74: 560–564
- [33] Illumina. An Introduction to Next-Generation Sequencing Technology. In: Illumina, Inc. ; 2017
- [34] Mardis ER. A decade's perspective on DNA sequencing technology. *Nature* 2011; 470: 198–203
- [35] [Anonymous]. Illumina Introduces the NovaSeq Series—a New Architecture Designed to Usher in the \$100 Genome. In: *Business Wire* 2017
- [36] [Anonymous]. Genes and Human Disease. In: World Health Organization. 2018;
- [37] Davies SC. Annual Report of the Chief Medical Officer 2016. In, *Generation Genome* 2017
- [38] Morton CC, Nance WE. Newborn hearing screening – a silent revolution. *N Engl J Med* 2006; 354: 2151–2164
- [39] Morton NE, Shields DC, Collins A. Genetic epidemiology of complex phenotypes. *Ann Hum Genet* 1991; 55: 301–314
- [40] Sosnay PR, Siklosi KR, Van Goor F et al. Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. *Nat Genet* 2013; 45: 1160–1167
- [41] Van Camp G, Smith RJH. Hereditary Hearing Loss Homepage. In: Shearer AE, Sommen M eds 2018;
- [42] Shearer AE, Eppsteiner RW, Booth KT et al. Utilizing ethnic-specific differences in minor allele frequency to recategorize reported pathogenic deafness variants. *Am J Hum Genet* 2014; 95: 445–453
- [43] Stenson PD, Mort M, Ball EV et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet* 2017; 136: 665–677
- [44] Bitner-Glindzic M. Hereditary deafness and phenotyping in humans. *Br Med Bull* 2002; 63: 73–94
- [45] Löwenheim H. [Zukunft der Hördiagnostik] [Article in German]. *Zeitschrift für Audiologie Audiological Acoustics* 2014; 20: 62–65
- [46] Bartsch O, Vatter A, Zechner U et al. GJB2 mutations and genotype-phenotype correlation in 335 patients from Germany with nonsyndromic sensorineural hearing loss: evidence for additional recessive mutations not detected by current methods. *Audiol Neurootol* 2010; 15: 375–382
- [47] Tropitzsch A, Friese N, Michels L et al. Next-generation Sequencing in der Diagnostik der genetischen Schwerhörigkeit. 30 Wissenschaftliche Jahrestagung der Deutschen Gesellschaft für Phoniatrie und Pädaudiometrie. 2013Bochum, Germany: DocV43
- [48] Mahdiah N, Rabbani B. Statistical study of 35delG mutation of GJB2 gene: a meta-analysis of carrier frequency. *Int J Audiol* 2009; 48: 363–370
- [49] Shearer AE, Black-Ziegelbein EA, Hildebrand MS et al. Advancing genetic testing for deafness with genomic technology. *J Med Genet* 2013; 50: 627–634
- [50] Shearer AE, Smith RJ. Massively Parallel Sequencing for Genetic Diagnosis of Hearing Loss: The New Standard of Care. *Otolaryngol Head Neck Surg* 2015; 153: 175–182
- [51] Sie AS, Prins JB, van Zelst-Stams WA et al. Patient experiences with gene panels based on exome sequencing in clinical diagnostics: High acceptance and low distress. *Clin Genet* 2015; 87: 319–326
- [52] Sheppard S, Biswas S, Li MH et al. Utility and limitations of exome sequencing as a genetic diagnostic tool for children with hearing loss. *Genet Med* 2018 [Epub ahead of print]
- [53] Guan Q, Balciuniene J, Cao K et al. AUDIOME: a tiered exome sequencing-based comprehensive gene panel for the diagnosis of heterogeneous nonsyndromic sensorineural hearing loss. *Genet Med* 2018 [Epub ahead of print]
- [54] Zazo Seco C, Wesdorp M, Feenstra I et al. The diagnostic yield of whole-exome sequencing targeting a gene panel for hearing impairment in The Netherlands. *Eur J Hum Genet* 2017; 25: 308–314
- [55] Kalia SS, Adelman K, Bale SJ et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med* 2017; 19: 249–255
- [56] [Anonymous]. [Opinion of the German Society for Human Genetics on additional genetic findings in diagnostics and Research] [Article in German]. In, *Guidelines and Statements of the German Society for Human Genetics* 2013;

- [57] Bowl MR, Simon MM, Ingham NJ et al. A large scale hearing loss screen reveals an extensive unexplored genetic landscape for auditory dysfunction. *Nat Commun* 2017; 8: 886
- [58] Sommen M, Schrauwen I, Vandeweyer G et al. DNA Diagnostics of Hereditary Hearing Loss: A Targeted Resequencing Approach Combined with a Mutation Classification System. *Hum Mutat* 2016; 37: 812–819
- [59] Vona B, Muller T, Nanda I et al. Targeted next-generation sequencing of deafness genes in hearing-impaired individuals uncovers informative mutations. *Genet Med* 2014; 16: 945–953
- [60] Yan D, Tekin D, Bademci G et al. Spectrum of DNA variants for non-syndromic deafness in a large cohort from multiple continents. *Hum Genet* 2016; 135: 953–961
- [61] Baux D, Vache C, Blanchet C et al. Combined genetic approaches yield a 48% diagnostic rate in a large cohort of French hearing-impaired patients. *Sci Rep* 2017; 7: 16783
- [62] Sloan-Heggen CM, Bierer AO, Shearer AE et al. Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. *Hum Genet* 2016; 135: 441–450
- [63] Alkowari MK, Vozzi D, Bhagat S et al. Targeted sequencing identifies novel variants involved in autosomal recessive hereditary hearing loss in Qatari families. *Mutat Res* 2017; 800-802: 29–36
- [64] Hernandez AL, Cox S, Kothiyal P et al. The Otochip sequencing array for hearing loss and Usher syndrome. *International Symposium on Usher Syndrome and Related Diseases*; 2010; Valencia, Spain
- [65] Shearer AE, Kolbe DL, Azaiez H et al. Copy number variants are a common cause of non-syndromic hearing loss. *Genome Med* 2014; 6: 37
- [66] Moteki H, Azaiez H, Sloan-Heggen CM et al. Detection and Confirmation of Deafness-Causing Copy Number Variations in the STRC Gene by Massively Parallel Sequencing and Comparative Genomic Hybridization. *Ann Otol Rhinol Laryngol* 2016; 125: 918–923
- [67] Vona B, Hofrichter MAH, Neuner C et al. DFNB16 is a frequent cause of congenital hearing impairment: implementation of STRC mutation analysis in routine diagnostics. *Clinical Genetics* 2015; 87: 49–55
- [68] Plevova P, Paprskarova M, Tvrda P et al. STRC Deletion is a Frequent Cause of Slight to Moderate Congenital Hearing Impairment in the Czech Republic. *Otology & Neurotology* 2017; 38: E393–E400
- [69] Francey LJ, Conlin LK, Kadesch HE et al. Genome-wide SNP genotyping identifies the Stereocilin (STRC) gene as a major contributor to pediatric bilateral sensorineural hearing impairment. *American Journal of Medical Genetics Part A* 2012; 158a: 298–308
- [70] Amr SS, Murphy E, Duffy E et al. Allele-Specific Droplet Digital PCR Combined with a Next-Generation Sequencing-Based Algorithm for Diagnostic Copy Number Analysis in Genes with High Homology: Proof of Concept Using Stereocilin. *Clinical Chemistry* 2018; 64: 705–714
- [71] Ren C, Liu F, Ouyang ZY et al. Functional annotation of structural ncRNAs within enhancer RNAs in the human genome: Implications for human disease. *Scientific Reports* 2017; 7: 15518
- [72] Nakano Y, Kelly MC, Rehman AU et al. Defects in the Alternative Splicing-Dependent Regulation of REST Cause Deafness. *Cell* 2018; 174: 536–548 e521
- [73] Khan AO, Becirovic E, Betz C et al. A deep intronic CLRN1 (USH3A) founder mutation generates an aberrant exon and underlies severe Usher syndrome on the Arabian Peninsula. *Sci Rep* 2017; 7: 1411
- [74] Frebourg T. The challenge for the next generation of medical geneticists. *Hum Mutat* 2014; 35: 909–911
- [75] Stelzer G, Rosen N, Plaschkes I et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics* 2016; 54: 1.30.31–31.30.33
- [76] [Anonymous]. Online Mendelian Inheritance in Man, OMIM®. In: Baltimore, MD, USA: McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University; 2018
- [77] Sherry ST, Ward MH, Kholodov M et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001; 29: 308–311
- [78] [Anonymous]. NHLBI Exome Sequencing Project (ESP) Exome Variant Server. In; 2018
- [79] Scott EM, Halees A, Itan Y et al. Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat Genet* 2016; 48: 1071–1076
- [80] Akbari MR, Fattahi Z, Beheshtian M et al. Iranome: A human genome variation database of eight major ethnic groups that live in Iran and neighboring countries in the Middle East. 67th Annual Meeting of The American Society of Human Genetics, 2017; 2017; Orlando, FL, USA
- [81] Walters-Sen LC, Hashimoto S, Thrush DL et al. Variability in pathogenicity prediction programs: impact on clinical diagnostics. *Mol Genet Genomic Med* 2015; 3: 99–110
- [82] Schwarz JM, Cooper DN, Schuelke M et al. MutationTaster2: Mutation prediction for the deep-sequencing age. *Nat Methods* 2014; 11: 361–362
- [83] Adzhubei IA, Schmidt S, Peshkin L et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010; 7: 248–249
- [84] Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* 2009; 4: 1073–1082
- [85] Vaser R, Adusumalli S, Leng SN et al. SIFT missense predictions for genomes. *Nat Protoc* 2016; 11: 1–9
- [86] Xiong HY, Alipanahi B, Lee LJ et al. RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science* 2015; 9: 1254806
- [87] Majoros WH, Holt C, Campbell MS et al. Predicting Gene Structure Changes Resulting from Genetic Variants via Exon Definition Features. LID - 10.1093/bioinformatics/bty324 [doi]. *Bioinformatics* 2018 [Epub ahead of print]
- [88] Desmet FO, Hamroun D, Lalonde M et al. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009; 37: e67
- [89] Perteu M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* 2001; 29: 1185–1190
- [90] Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004; 11: 377–394
- [91] Reese MG, Eeckman FH, Kulp D et al. Improved splice site detection in Genie. *J Comput Biol* 1997; 4: 311–323
- [92] Landrum MJ, Lee JM, Benson M et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* 2016; 44: D862–D868
- [93] Fokkema IF, Taschner PE, Schaafsma GC et al. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat* 2011; 32: 557–563
- [94] Akle S, Chun S, Jordan DM et al. Mitigating false-positive associations in rare disease gene discovery. *Hum Mutat* 2015; 36: 998–1003
- [95] Eisenberger T, Di Donato N, Baig SM et al. Targeted and Genomewide NGS Data Disqualify Mutations in MYO1A, the “DFNA48 Gene”, as a Cause of Deafness. *Hum Mutat* 2014; 35: 565–570
- [96] Hildebrand MS, DeLuca AP, Taylor KR et al. A contemporary review of AudioGene audioprofiling: a machine-based candidate gene prediction tool for autosomal dominant nonsyndromic hearing loss. *Laryngoscope* 2009; 119: 2211–2215
- [97] Taylor KR, Booth KT, Azaiez H et al. Audioprofile Surfaces: The 21st Century Audiogram. *Ann Otol Rhinol Laryngol* 2016; 125: 361–368
- [98] Taylor KR, DeLuca AP, Shearer AE et al. AudioGene: predicting hearing loss genotypes from phenotypes to guide genetic screening. *Hum Mutat* 2013; 34: 539–545

- [99] Shen J, Scheffer DI, Kwan KY et al. SHIELD: an integrative gene expression database for inner ear research. *Database (Oxford)* 2015; 2015: bav071
- [100] Hertzano R, Orvis J. gEAR Portal. In; 2018
- [101] Tandy-Connor S, Gultinan J, Krempely K et al. False-positive results released by direct-to-consumer genetic tests highlight the importance of clinical confirmation testing for appropriate patient care. LID – 10.1038/gim.2018.38 [doi]. *Genet Med* 2018 [Epub ahead of print]
- [102] Oza A, DiStefano M, Hemphill S et al. Expert Specification of the ACMG/AMP Variant Interpretation Guidelines for Genetic Hearing Loss. *bioRxiv* 2018. DOI: 10.1101/313734
- [103] Hu H, Huff CD, Moore B et al. VAAST 2.0: Improved Variant Classification and Disease-Gene Identification Using a Conservation-Controlled Amino Acid Substitution Matrix. *Genet Epidemiol* 2013; 37: 622–634
- [104] Pollard KS, Hubisz MJ, Rosenbloom KR et al. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 2010; 20: 110–121
- [105] Grantham R. Amino acid difference formula to help explain protein evolution. *Science* 1974; 185: 862–864
- [106] Tang H, Wyckoff GJ, Lu J et al. A universal evolutionary index for amino acid changes. *Mol Biol Evol* 2004; 21: 1548–1556
- [107] Van Laer L, Coucke P, Mueller RF et al. A common founder for the 35delG GJB2 gene mutation in connexin 26 hearing impairment. *J Med Genet* 2001; 38: 515–518
- [108] Ballana E, Ventayol M, Rabionet R et al. The Connexin-deafness Homepage. In. September 8, 2018 ed; 2018
- [109] Kircher M, Witten DM, Jain P et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014; 46: 310–315
- [110] Scriver CR. After the genome – the phenome? *J Inher Metab Dis* 2004; 27: 305–317
- [111] Oetting WS, Robinson PN, Greenblatt MS, Cotton RG et al. Getting ready for the Human Phenome Project: the 2012 forum of the Human Variome Project. DOI: 10.1002/humu.22293
- [112] Deans AR, Lewis SE, Huala E et al. Finding Our Way through Phenotypes. *Plos Biol* 2015; 13: e1002033
- [113] Poldrack RA, Congdon E, Triplett W et al. A phenome-wide examination of neural and cognitive function. *Sci Data* 2016; 3: 160110
- [114] Köhler S, Vasilevsky NA-O, Engelstad M et al. The Human Phenotype Ontology in 2017. DOI: 10.1038/sdata.2016.110
- [115] Vasilevsky NA, Foster ED, Engelstad ME et al. Plain-language medical vocabulary for precision diagnosis. *Nat Genet* 2018; 50: 474–476
- [116] Gall T, Valkanas E, Bello C et al. Defining Disease, Diagnosis, and Translational Medicine within a Homeostatic Perturbation Paradigm: The National Institutes of Health Undiagnosed Diseases Program Experience. *Front Med* 2017; 4: 62
- [117] Morton CC. 2014 Presidential Address: The Time of Our Lives. *Am J Hum Genet* 2015; 96: 347–351