

4. Themenbereich genetische Diagnostik: Das humane Genom in der medizinischen Diagnostik

4.1 Das Humangenomprojekt – ein grundlegender Durchbruch in der Biologie des Menschen

Vor 20 Jahren wurden die ersten Referenzsequenzen¹ des humanen Genoms öffentlich präsentiert. Damit war der Anspruch verbunden, dass nun das menschliche Genom entschlüsselt sei. Es dauerte weitere Monate bis die internationalen Arbeitsgruppen ihre Ergebnisse in den angesehenen Fachzeitschriften *Nature* (International Human Genome Sequencing Consortium, 2001) und *Science* (Venter et al., 2001) veröffentlichten.² Die Kosten des Humangenomprojekts waren enorm. Sie wurden insgesamt mit ca. einer Milliarde US\$ veranschlagt. Das Projekt war mit großem internationalen Aufwand durchgeführt worden und involvierte multiple Konsortien weltweit. Das Humangenomprojekt wurde aber auch auf politischer Ebene als *der* große Durchbruch der Biologie gefeiert. So präsentierte Präsident Bill Clinton den ersten Arbeitsentwurf der Sequenz des humanen Genoms in einer Pressekonferenz zusammen mit namhaften amerikanischen Wissenschaftlern, darunter auch Francis Collins, Leiter des Humangenomprojekts, und Craig Venter, Präsident von Celera, mit den folgenden Worten: „[das Humangenom ist] die schönste Karte der Geschichte des Menschen, [...] die Sprache, in

1 Als Sequenzen bezeichnet man die Abfolge der ca. drei Milliarden Basenpaare auf den Chromosomen der menschlichen DNA. Die 2001 präsentierten Genomsequenzen des Menschen wurden als Referenzsequenzen bezeichnet, da diese nun als Vergleichssequenzen für die Entschlüsselung weiterer menschlicher Genome dienen konnten.

2 Es gab zwei verschiedene Konsortien: Das 1990 gegründete Internationale Humangenomprojekt, ein Zusammenschluss öffentlicher Forschungsgruppen vieler Länder, und eine privatwirtschaftliche Initiative der Firma Celera um den Forscher Craig Venter. Letzterer arbeitete mit einer anderen Methodik, aber in Kenntnis der Fortschritte des öffentlichen Humangenomprojekts. Beide präsentierten ihre Ergebnisse gemeinsam der Öffentlichkeit, publizierten dann aber separat.

der Gott das Leben erschaffen hat“ (Bill Clinton, 2000; Übersetzung des Autors).³ Allerdings entsprach diese erste Arbeitsversion des humanen Genoms mit im Schnitt einem Sequenzierfehler alle 1.000 Basen und noch rund 150.000 Lücken nur zu knapp 30 % der ersten finalen Version vom 14. April 2003.⁴ Diese Sequenz und die im Laufe der Jahre immer weiter verbesserten Versionen entsprechen dem heute verwendeten Referenzgenom. Wobei es inzwischen mehrere unterschiedliche Referenzgenome von Menschen aus verschiedenen Regionen gibt,⁵ was Rückschlüsse auf zwischenmenschliche Variationen erlaubt.

Der Anspruch und das mit diesen Daten in Verbindung gebrachte Versprechen waren groß. Die Kenntnis des menschlichen Genoms sollte zu einem Paradigmenwechsel in der Medizin führen und diese in eine nächste Phase der Personalisierung, Prävention und akkuraten Vorhersage bringen (Collins/McKusick, 2001). Für die meisten Patienten haben sich diese Versprechungen bislang nicht materialisiert. Dies trifft insbesondere auf häufige Erkrankungen wie Diabetes oder koronare Herzerkrankungen zu, bei der die genomischen Daten zwar einen Erkenntnisgewinn, nicht aber den erwünschten Durchbruch in Therapie und Prävention gebracht haben. Ganz anders sieht es bei seltenen Erkrankungen mit Mendelscher Vererbung aus – also solche, die von Mutationen in einzelnen Genen hervorgerufen werden – wie erbliche Krebsformen und genetisch bedingte Entwicklungsstörungen bei Kindern. Hier hat das Wissen über die humane Genomsequenz für diese Patienten sowohl in Hinblick auf die Diagnostik wie auch auf zukünftige Therapien zu einem wahren Durchbruch und zu einer Flut an Informationen geführt. Gleiches gilt für die Anwendung der Genomsequenzierung auf Tumoren. Hier haben die Sequenzierung und Analyse von Krebs verursachenden Muta-

3 Zitat von einer Pressekonferenz Bill Clintons im Weißen Haus am 26. Juni 2000 mit der Präsentation einer Arbeitsfassung der ersten Referenzsequenz des menschlichen Genoms: Clinton bezeichnete die Entschlüsselung des Genoms im Original als „the most important, most wondrous map ever produced by humankind [...]. Today, we are learning the language in which God created life“ und verglich sie mit der Mondlandung. Am 12. Februar 2001 wurden die beiden Referenzsequenzen dann offiziell präsentiert und kurz darauf publiziert. Siehe unter: <https://www.genome.gov/about-nhgri/Director/genomics-landscape/Feb-12-2021-twenty-year-anniversary-of-draft-human-genome-sequence-publications#twenty-year> [09.03.2021]; sowie unter: <https://www.genome.gov/human-genome-project/Timeline-of-Events> [09.03.2021].

4 Siehe unter: <https://www.genome.gov/25520492/online-education-kit-2003-human-genome-project-completed> [17.03.2021].

5 Im Februar 2021 wurden 64 verschiedene Referenzgenome von 32 Menschen aus verschiedenen Regionen publiziert. Die Zahl der Referenzgenome ist doppelt so hoch wie die der Menschen, weil jeder Mensch Träger von zwei Sätzen des Humangenoms ist, von der Mutter und vom Vater. Siehe unter: <https://science.sciencemag.org/content/early/2021/02/24/science.abf7117?versioned=true> [09.03.2021].

tionen ganz entscheidend zu einem Wechsel in der Therapie hin zu einer personalisierten Versorgung geführt. Es steht außer Frage, dass diese Entwicklung erst am Anfang steht und sich der Erkenntnisgewinn und daraus folgende direkte Anwendungen in der Medizin weiter beschleunigen werden (vgl. Shendure et al., 2019).

Die ersten Analysen des humanen Genoms brachten jedoch auch manche Überraschung mit sich. So war das menschliche Genom überraschend klein, war es doch mit ca. 3,4 Milliarden Basenpaaren nur unwesentlich größer als das Genom z. B. der Maus. Noch überraschender war allerdings, dass sich nur ca. 20.000 Gene im Genom finden ließen, definiert als solche Sequenzen, die in ein Protein übersetzt werden. Diese Zahl war nur unwesentlich größer oder sogar gleich zu der Zahl der Gene, die in vermeintlich einfachen Organismen wie der Fruchtfliege (14.000) oder dem Fadenwurm, *C. elegans*, (20.000) gefunden wurden. Zudem machte der Anteil der kodierenden Gene nur ca. 1,5 % des ganzen Genoms aus. Zuvor war man davon ausgegangen, dass die Gene alle wichtigen Informationen enthalten. Warum machten sie also nur einen so kleinen Prozentsatz des Genoms aus? Diese Frage beschäftigt Wissenschaftler bis heute, ihre Erforschung ist eine der zentralen Herausforderungen der neuen Genomik (siehe Rheinberger/Müller-Wille, Kap. 11, zum Verständnis und der Rolle von Genen in Gegenwart und Vergangenheit).

4.2 Technologischer Fortschritt in der Genomanalyse

Bevor die humane Genomsequenz bekannt wurde, mussten grundlegende Informationen über die Sequenz und die genomische Lokalisation von einem Gen bzw. einer Mutation durch aufwendige Prozesse der Klonierung gewonnen werden. Die in den 1990er Jahren entwickelte Methode des „positional mapping“⁶ machte es einfacher, Gene, die Krankheiten verursachen, zu identifizieren. Hierbei wurde die DNA von mehreren Familienmitgliedern mit derselben Krankheit verglichen, um eine ungefähre chromosomale Position der Mutation herauszufinden. Die anschließende Suche involvierte das Klonieren von kurzen chromosomalen Fragmenten mit Enzymen, die dann in Bakterien vermehrt wurden, um sie weiter analysieren zu können. Die hierfür angewandten Verfahren basierten im Wesentlichen auf der sogenannten Sanger-Sequenzierung, die weiter unten erläutert wird (siehe hierzu Abschnitt 4.2.2). Die Identifizierung von mutierten Abschnitten im Genom und deren Analyse war mit diesem Verfahren äußerst aufwändig und ineffektiv. Da die hierfür notwendigen Technologien nur von wenigen

6 „Positional Mapping“ ist eine Methode zur Bestimmung des Ortes im Genom (auf dem Chromosom und in der Chromosomenregion), an dem ein zu untersuchendes Gen sich befinden muss.

Laboren durchgeführt werden konnten, waren die Ursachen der meisten Mendelschen Erkrankungen unbekannt. Entsprechend war eine Diagnose dieser seltenen Erkrankungen in den meisten Fällen nur klinisch möglich anhand von Symptomen oder bestimmten ausgeprägten Merkmalen und setzte eine große Expertise und Erfahrung seitens der beteiligten Ärzte und Wissenschaftler voraus. Die Diagnosestellung war daher oft dem Zufall überlassen, je nachdem, ob der Patient dem richtigen Experten vorgestellt wurde oder nicht. Über die Jahre hat sich die genetische Diagnostik immer weiter verfeinert, sodass wir mittlerweile in der Lage sind, verlässliche Diagnosen für eine große Anzahl von genetisch bedingten Erkrankungen stellen zu können. Dieser Fortschritt, der in der Medizin seinesgleichen sucht, wäre ohne das Humangenomprojekt und die Entwicklung von genomischen Analyseverfahren nicht denkbar (Clausnitzer et al., 2020). Im Folgenden werden die hierfür verwendeten Technologien und Methoden näher erläutert.

4.2.1 Die Anfänge der genomischen Diagnostik – Zytogenetik

Die erste genetische Diagnostik, die im Labor gestellt werden konnte, beruhte auf der Analyse von Chromosomen, gewonnen im Allgemeinen aus weißen Blutzellen (Lymphozyten). Mit dieser Technologie konnten schon ab den 1960ern Veränderungen insbesondere der Zahl der Chromosomen, wie z. B. bei Trisomie 21 – ein Down-Syndrom – festgestellt werden. Die Zytogenetik analysiert Chromosomen in einem kondensierten, d. h. verdichteten und in sich aufgerollten, Zustand, in dem sie sich abschnittsweise in ihrer Färbung unterscheiden, wenn man sie mit dafür geeigneten Farbstoffen färbt. Durch diese besondere Färbung entsteht ein im Mikroskop sichtbares Bandenmuster (Bänderung), das für jedes Chromosom spezifisch ist und somit die Zuordnung individueller Chromosomenabschnitte ermöglicht.

Durch die Analyse der Bänderung können sogenannte strukturelle Varianten des Genoms identifiziert werden. Hierzu gehören neben den numerischen Veränderungen ganzer Chromosomen der Austausch zwischen zwei Chromosomen (Translokation) und der Verlust von genetischem Material (Deletion), sowie der Zugewinn von genetischem Material (Duplikation/Verdoppelung oder Amplifikation/Vermehrung). Die Technik benötigt lebende Zellen, aus denen die Chromosomen gewonnen werden, die dann mikroskopisch untersucht werden. Dadurch ist die Methode relativ arbeitsintensiv und kann schlecht automatisiert werden. Zudem ermöglicht sie keine Analyse auf molekularer Sequenzebene, da sie eine mikroskopiebasierte Technologie ist. Auf der anderen Seite ermöglicht die Zytogenetik eine Analyse auf zellulärer Ebene, sodass hiermit auch Veränderungen erfasst werden können, die nur in einem Teil der Zellen vorliegen

(wenn es sich bei dem Organismus um ein Mosaik aus genetisch verschiedenen Zelltypen handelt). Die Zytogenetik war bis in die 90er Jahre die einzige diagnostische Methode, die der Humangenetik zur Verfügung stand. Heute ist die Zytogenetik immer noch Teil der diagnostischen Routine, insbesondere zur Detektion von chromosomalen Veränderungen, aber ihre Bedeutung hat durch die Entwicklung neuer diagnostischer Verfahren abgenommen.

4.2.2 Die Identifikation von Sequenzvarianten mittels Sanger-Sequenzierung

Während sich mit der Zytogenetik ganze Chromosomen visuell darstellen lassen, wird die Sequenzierung genutzt, um einzelne Abschnitte des Genoms in ihrer Basenfolge zu analysieren. In der Didesoxymethode nach Sanger, auch Kettenabbruchmethode genannt, werden zunächst mit einer DNA-Polymerase⁷ komplementäre DNA-Stränge, also Kopien der Einzelstränge, erzeugt. Für diese Reaktion werden aber neben normalen Nukleotiden auch Nukleotide verwendet, die zum Abbruch der Synthese führen. Dies führt zu Strängen unterschiedlicher Länge, die mit markierten Nukleotiden versehen sind, die dann ausgelesen werden können. Mit der Sanger-Methode wurde das komplette Humangenom erstellt. Mit zunehmender Identifikation von Krankheitsgenen und ihren Mutationen konnte die Sanger-Sequenzierung auch Einzug in die Diagnostik halten. Allerdings ist eine Analyse mittels Sanger nur bei relativ kleinen DNA-Abschnitten von ca. 400–800 Basenpaaren der DNA (bp) möglich. Zudem muss der Analyse eine Vervielfältigung des entsprechenden Abschnitts aus dem Genom mittels Polymerase-Kettenreaktion (PCR) vorausgehen, da der zu untersuchende Abschnitt in ausreichender Menge und rein vorliegen muss. Mit der Kenntnis von Mutationen und Krankheitsgenen ermöglichte die Sanger-Sequenzierung eine völlig neue Qualität in der genetischen Diagnostik, indem eine klinische Verdachtsdiagnose mittels Analyse der entsprechenden Gene verifiziert werden konnte. Mithilfe dieser Technologie konnten immer mehr Erkrankungen diagnostiziert werden, da im Laufe der Jahre immer mehr Krankheitsgene identifiziert wurden. Die Technologie kam aber auch rasch an ihre Grenzen, da die Analyse großer Gene oder einer großen Anzahl von Genen nur

⁷ Eine DNA-Polymerase ist ein Enzym, welches DNA synthetisieren kann. Bei der hier beschriebenen Kettenabbruchmethode werden die beiden komplementären DNA-Stränge getrennt und für die Neusynthese des nun fehlenden zweiten Stranges werden nicht nur die normalen Bausteine der DNA (Nukleotide) zugegeben und von der DNA-Polymerase eingebaut, sondern auch besonders veränderte Nukleotide, die zum Abbruch der Synthese führen. So entstehen verschieden lange DNA-Fragmente, die man in einem Gel auftrennen kann. Durch die radioaktive Markierung der entstehenden Schnipsel entsteht auf einem Röntgenfilm ein Bandenmuster. Daraus kann auf die Sequenz zurückgeschlossen werden.

mit immensem Aufwand möglich ist. Ein Screening auf Mutationen in vielen Genen ist daher mit der Sanger-Sequenzierung nur bedingt möglich. Dies stellt ein besonderes Problem dar, wenn das klinische Bild nicht ganz eindeutig ist und hierfür eine Reihe von Genen verantwortlich sein kann.

4.2.3 Array-CGH – genomweites Screening auf Veränderungen der Kopienzahl

Die komparative (vergleichende) Genom-Hybridisierung (CGH) misst quantitative Sequenzänderungen in einem Genom, indem zwei DNA-Proben bzgl. ihrer Kopienzahl miteinander verglichen werden. Hierbei findet eine kompetitive Hybridisierung⁸ an eine Zielsequenz statt, die dann z. B. über Fluoreszenz-Farbstoffe ausgelesen werden kann. Die Array-CGH ermöglicht die genomweite Suche nach sogenannten „copy number variants“ (CNVs), das sind Strukturvarianten, die mit einem Verlust (Deletionen) oder Zugewinn (Duplikationen) von genetischem Material verbunden sind. Eine Detektion von balancierten Veränderungen, die keinen quantitativen Effekt haben, ist nicht möglich. Die Array-CGH ermöglicht je nach Auflösung auch die Detektion von relativ kleinen Varianten (> 50.000 bp), die zytogenetisch nicht detektierbar wären. Die Array-CGH ist somit ein ideales Werkzeug für das Screening von Genomen auf pathologische Veränderungen. So können z. B. mithilfe der Array-CGH in ca. 10–20 % der Fälle pathogene Veränderungen bei Patienten mit Entwicklungsstörungen gefunden werden (Visser et al., 2005). Diese Methodik hat die genomische Analyse maßgeblich verändert, weil hiermit erstmals ein Screening des ganzen Genoms auf Veränderungen in hoher Auflösung vorgenommen werden konnte anstelle der Überprüfung eines vorherigen Verdachts. Die Methode hat in weiten Teilen die Zytogenetik abgelöst, da die Auflösung besser ist, das Verfahren sich automatisieren lässt und sich die Größe und Position der Veränderung und somit auch die Bruchpunkte⁹ relativ genau bestimmen lassen. Die Array-CGH ist heute die Methode der Wahl zur Abklärung unklarer Fälle v. a. bei struktureller Veränderung der Chromosomen mit Veränderung der Kopienzahl.

8 Als Hybridisierung bezeichnet man in diesem Kontext die Anlagerung von synthetisch hergestellten DNA-Sonden (DNA-Abschnitten, die komplementär zu der zu untersuchenden DNA sind und eine detektierbare Markierung enthalten) an die zuvor durch Wärmezufuhr in ihre beiden Stränge aufgetrennte Ziel-DNA. Beim Abkühlen lagern sich die Basen der Stränge wieder aneinander an, wobei die Sonden im Wettbewerb mit dem zweiten Strang der DNA stehen, weshalb die Methode als kompetitiv bezeichnet wird. Es entstehen damit Hybride aus DNA-Sonde und Original-DNA. Die gebundenen Sonden können dann detektiert werden.

9 Bruchpunkte sind die Regionen auf den Chromosomen, an denen die detektierte Veränderung stattgefunden hat.

Die oben genannten Technologien führten nach der Veröffentlichung des Humangenoms zu einer Flut von Informationen über die Ursache von genetischen Erkrankungen. Während man zu Beginn der 90er Jahre vielleicht von einer Handvoll Erkrankungen mit genetischer Ursache wusste, führte das zunehmende Wissen über das humane Genom sowie die breite Anwendung der Sanger-Sequenzierung, Zytogenetik und Array-CGH zu der Identifikation von mehreren Tausend Krankheitsgenen. Der Einfluss dieser Entwicklung auf die Diagnose und Behandlung von Patienten mit seltenen Erkrankungen lässt sich kaum überschätzen. Erstmals war eine verlässliche Diagnosestellung im großen Maßstab möglich. Dies ermöglichte eine intensive Erforschung der Pathologie der jeweiligen Erkrankungen und die Entwicklung von Therapiestrategien. Während Letzteres ein langer Weg ist, ist die Möglichkeit einer Diagnosestellung mittels genetischer Techniken für das Patientenwohlfinden von großer Bedeutung; ist sie doch die Voraussetzung dafür, mit der Krankheit leben zu können, eine Prognose über deren Verlauf zu haben und die Beschwerden einordnen zu können.

Bei allem Fortschritt, der bis zu diesem Punkt erreicht war, war doch der eigentliche Gamechanger die Entwicklung der neuen Sequenziermethoden, des sogenannten Next-Generation-Sequencing (NGS), da sich mit dieser Methode ganze Genome und nicht nur einzelne Gene analysieren lassen (Shendure et al., 2017). Dies ermöglichte die Identifikation potenziell krankheitsverursachender Varianten im ganzen Genom mit einer bisher nicht dagewesenen Effizienz und Geschwindigkeit (Chong et al., 2015). Entsprechend hatte die Entwicklung dieser Methode auch den größten Einfluss auf die Diagnostik von genetischen Erkrankungen (Ng et al., 2010). So konnte die Zahl der Mendelschen Erkrankungen mit einem bekannten genetischen Defekt von 1.257 im Jahre 2001 auf 6.009 (OMIM Datenbank)¹⁰ zum jetzigen Zeitpunkt (Stand März 2021) erhöht werden. Die Datenbank enthält weitere ca. 3.200 Einträge von Krankheiten mit bisher unbekannter Ursache. Mithilfe der neuen Technologien des NGS können Patienten nun rasch einer Diagnose zugeführt werden und die Diagnostik erhalten, die sie dringend benötigen. Diese Form der Diagnostik ermöglicht auch ein personalisiertes Management und wird in Zukunft auf die jeweilige Mutation zugeschnittene personalisierte Therapien möglich machen. Dies ist z. B. jetzt schon bei der zystischen Fibrose möglich, bei der eine auf bestimmte Mutationen abgestimmte Therapie erfolgt.¹¹

10 Siehe unter: www.OMIM.org [09.03.2021].

11 Siehe hierzu auch das Interview mit Hans Clevers zu Organoiden im gleichnamigen Themenband der IAG (Bartfeld, 2020: 65–76).

4.3 Die neuen Sequenziermethoden – Eine Revolution in der genomischen Analyse und genetischen Diagnostik

4.3.1 Next-Generation-Sequencing

Die neuen Verfahren der Hochdurchsatzsequenzierung, auch Next-Generation-Sequencing (NGS) genannt, beruhen auf der parallelen Analyse von vielen Millionen DNA-Fragmenten zur selben Zeit (Goodwin et al., 2016). Die derzeit am häufigsten verwendete Methode der Firma Illumina nutzt klonal amplifizierte DNA-Fragmente, also abschnittsweise vervielfältigte DNA-Stücke, die auf einem Träger (sogenannten „flow cells“) fest an diesen gebunden sind (immobilisiert). Dabei baut die Methode auf die Tatsache auf, dass die sich im DNA-Doppelstrang gegenüberliegenden Basen komplementär sind und aneinander binden,¹² sodass sich aus der Sequenz eines Stranges die Basenabfolge des anderen Stranges ergibt. Bei der Methode dient einer der beiden Stränge als Vorlage für die Synthese eines neuen Doppelstranges. An den DNA-Fragmenten erfolgt eine Strangsynthese mit Fluoreszenz markierten Nukleotiden, die von einer speziellen Kamera ausgelesen werden. Über die Fluoreszenz kann detektiert werden, welche Basen eingebaut wurden. So ergibt sich Nukleotid für Nukleotid eine Sequenz an jedem einzelnen DNA-Fragment, bis schließlich eine Sequenz von 50–120 bp erreicht wird. Dadurch, dass alle immobilisierten DNA-Fragmente parallel auf diese Art und Weise sequenziert werden, kann eine Analyse der kompletten ins Gerät gegebenen DNA vorgenommen werden. Hierdurch unterscheidet sich die Technologie grundsätzlich von der Sanger-Sequenzierung, bei der immer nur ein Fragment für sich alleine sequenziert werden kann. Jedoch ergibt sich hieraus auch ein Problem der Methodik, da die einzelnen sequenzierten DNA-Fragmente erst wieder ihren genomischen Abschnitten zugeordnet werden müssen. Dies erfolgt über entsprechende Computerprogramme, bei denen jede Sequenz auf ihre Position und Einzigartigkeit im Genom hin untersucht und dann an die entsprechende Stelle ausgerichtet wird. Die Algorithmen dieser Programme ermöglichen es, auch relativ kurze DNA-Sequenzen, wie sie bei dieser Methode produziert werden (um die 100 bp), sicher im Genom zu verorten, außer es handelt sich um repetitive Sequenzen, die mehrfach vorkommen. Dadurch, dass mit dieser Technologie Sequenzabschnitte in großer Menge produziert werden, ist es auch möglich, eine sehr präzise Quantifizierung der eingesetzten DNA zu erlangen, indem die Abdeckung („coverage“) für den jeweiligen Abschnitt des Genoms quantifiziert wird. Diese Metho-

12 Die Basen bilden dabei Paare. Dabei paart sich A (Adenin) mit T (Thymin) und C (Cytosin) mit G (Guanin).

dik wird z. B. eingesetzt, um die verschiedenen RNAs¹³ in einer Zelle zu quantifizieren. Hierbei wird die gesamte RNA in DNA überschrieben und dann sequenziert. Die produzierten Sequenzen werden auf das Genom kartiert und über die Anzahl der einzelnen Reads (Ablesungen) wird eine Quantifizierung der Genexpression vorgenommen.

Ein Nachteil der Methodik ist jedoch, dass die Gesamtheit der DNA, die in das Gerät gegeben wird, parallel sequenziert wird. Wenn also nur bestimmte Teile des Genoms analysiert werden sollen, bedarf es einer weiteren Technik der Anreicherung. Dies wurde erfolgreich mit dem sogenannten Capture-Verfahren erreicht.

4.3.2 Exon- und Panelsequenzierung mittels Capture-Verfahren

Um nur bestimmte Teile des Genoms mittels NGS sequenzieren zu können, muss eine Anreicherung der gewünschten Sequenzen erfolgen. Dies kann zum einen durch die amplikonbasierte Technologie erfolgen. Hierbei werden bestimmte Zielsequenzen, z. B. die Exons, also die kodierenden Abschnitte mehrerer Gene, parallel in einem speziellen PCR-Verfahren vervielfältigt (amplifiziert). Diese PCR-Produkte werden dann im NGS-Verfahren sequenziert und anschließend wieder an die entsprechenden Abschnitte im Genom ausgerichtet. In diesem Verfahren werden also die zu untersuchenden Genabschnitte mittels PCR vervielfältigt und dann analysiert. Im Gegensatz hierzu beruht das Capture-Verfahren darauf, dass die zu untersuchenden Genomabschnitte durch komplementäre Bindung an Sonden angereichert und anschließend amplifiziert werden. Die hierfür verwendeten Sonden bestehen im Allgemeinen aus synthetisch hergestellten RNA-Molekülen, die an die Zielsequenz binden und dann über ein Magnetverfahren vom restlichen Genom getrennt werden können. Hierdurch ergibt sich eine starke Anreicherung der Zielsequenz, die dann mittels NGS sequenziert werden kann. Dadurch, dass die Sonden synthetisch hergestellt werden, lässt sich eine beliebige Zielsequenz anreichern und dann sequenzieren. Die so erreichte Flexibilität erlaubt es, die Zielsequenz an die jeweiligen diagnostischen Fragestellungen anzupassen.

So ist es möglich, eine größere Zahl an Krankheitsgenen zusammenzustellen, mit denen ganze Krankheitsgruppen in einem Test analysiert werden können. Dies ist z. B. relevant, wenn eine größere Zahl von Genen, die sich mit der Sanger-Sequenzierung nicht effektiv analysieren lassen, für eine Krankheit oder Krankheitsgruppe verantwortlich sein können. Als Beispiel zu nennen sind die ca. 60 derzeit bekannten Gene, die

13 RNAs sind wie DNA Ribonukleinsäuren und enthalten Erbinformationen. Sie entstehen beim Ablesen der DNA (Transkription). Im Gegensatz zur DNA sind sie einzelsträngig, instabiler, tragen einen anderen Zuckerbaustein und die Base Thymin (T) ist in Uracil (U) umgewandelt. Messenger RNAs (mRNA) sind Botenmoleküle, die zur Herstellung von Proteinen (Translation) genutzt werden.

eine Kardiomyopathie (Herzmuskelerkrankung) auslösen können. Entsprechend sind auch andere sogenannte Gen-Panels¹⁴ zusammengestellt worden, um Krankheitsgruppen wie Muskelerkrankungen, Lipid- und Stoffwechselstörungen, Hörstörungen etc. mit einem Test untersuchen zu können.

Alternativ lässt sich auch eine Anreicherung aller kodierenden Sequenzen, des sogenannten Exoms, mit dem Capture-Verfahren durchführen. Hierbei werden die Sonden auf alle Exons und deren flankierende Sequenzen gerichtet, sodass alle 20.000 Gene auf einmal angereichert, sequenziert und analysiert werden können. Das Verfahren der Exomsequenzierung hat sich als ein extrem effektives Werkzeug zur raschen Analyse des Genoms entwickelt. Mit diesem Verfahren wurden viele neue Gendefekte aufgedeckt, sodass eine umfassende Kartierung des Genoms in Bezug auf Krankheitsgene und Mutationen möglich war (Posey et al., 2019). Stammbaumbasierte genomische Studien und Analysen seltener Varianten in Familien mit v. a. genetischen Erkrankungen haben zur Aufklärung Hunderter neuer Krankheitsgene geführt. Eine Reihe von Mutationsmechanismen konnte in ihrer Bedeutung für die Pathogenese von seltenen Krankheiten untersucht werden. Hierzu gehören De-novo-Mutationsereignisse, somatische Variationen (Mosaik), die nicht-onkologischen Merkmalen zugrunde liegen, unvollständig penetrante Allele, Phänotypen mit hoher Locus-Heterogenität sowie oligogene Vererbung (Bamshad et al., 2011).¹⁵ Durch diese Studien konnte die Entwicklung von bioinformatischen Werkzeugen und Analysemethoden sowie eine phänotypische funktionelle Zuordnung (Annotation) aller Varianten in menschlichen Genen vorangetrieben werden. Neben dem Informationsgewinn in Bezug auf Krankheiten haben diese Untersuchungen auch zu neuen, grundlegenden Erkenntnissen der Biologie unseres Genoms beigetragen. Die Forschung an seltenen Krankheiten hat somit ganz wesentlich die funktionelle Annotation des menschlichen Genoms vorangetrieben und bildet die Grundlage für zukünftige Präzisionsmedizin.

Zusätzlich wird dieses Verfahren auch in der Diagnostik von Erkrankungen eingesetzt, bei denen keine eindeutige klinische Diagnose zu stellen ist. Abhängig von der

14 Ein Gen-Panel umfasst mehrere Sets von klinisch relevanten Genen für ein bestimmtes Krankheitsbild, die parallel analysiert werden können.

15 De-novo-Mutationsereignisse sind neu auftretende Mutationen in Eizellen, Samenzellen oder im frühen Embryo. Somatische Mutationen sind Veränderungen im Genom von Körperzellen zu einem späteren Zeitpunkt. Sie sind daher nicht in allen Zellen vorhanden. So können z. B. Mosaik entstehen. Unvollständig penetrante Mutationen sind solche, bei denen die beeinflussten Merkmale nicht in jedem Fall ausgeprägt werden, etwa weil Umwelteinflüsse eine wichtige Rolle spielen. Phänotypen mit Locus-Heterogenität sind solche, bei denen Mutationen in einer Reihe verschiedener Gene einen ähnlichen oder identischen Phänotyp erzeugen. Oligogene Vererbung liegt vor, wenn mehrere Gene einen Phänotyp beeinflussen.

Sequenziertiefe ist es möglich, mit hoher Sensitivität zuverlässig Varianten zu detektieren, je nach Methodik schon bei einer relativ geringen Mutationslast, also z. B. auch bei Mosaiken.

4.3.3 Variantenanalyse und Interpretation der Daten

Das Ergebnis jeder NGS-Sequenzierung ist eine große Menge von bis zu mehreren Millionen NGS-Sequenzabschnitten, den Reads. Für eine Exomsequenzierung („whole exome sequencing“, WES) entspricht dies ca. 40–50 Millionen Reads in einer Länge von 70–120 bp. Die erste bioinformatische Herausforderung besteht darin, diese Reads auf die Referenzsequenz abzubilden. Dabei werden die Reads mit der Referenzsequenz verglichen, um die Lage der Sequenzen zu ermitteln. Dies geschieht mit Blick auf das gesamte Genom (globales Allinieren). Da das humane Genom bis auf die Geschlechtschromosomen diploid (doppelter Chromosomensatz) ist, wird im Vergleich zum Referenzgenom entweder die gleiche Base, eine andere oder zur Hälfte eine andere Base erwartet. Dies wird dann jeweils als Homozygotie oder Heterozygotie vom Referenzgenom bezeichnet. Diese Unterschiede in der Sequenz bestimmen letztendlich die Individualität jedes Einzelnen und sind auch die relevanten Faktoren in der Suszeptibilität, d. h. der Empfindlichkeit für, bzw. der Ursache von Erkrankungen. Die Herausforderung der Analyse ist es, die krankheitsassoziierten Varianten mit einer hohen Sensitivität von Polymorphismen (nicht für die Krankheiten relevante Variationen) und von Artefakten (durch menschliche oder technische Einflüsse künstlich erzeugte falsche Ergebnisse) zu trennen.

Mit der Fertigstellung des humanen Referenzgenoms und der weiteren Sequenzierung vieler Genome stellte sich heraus, dass es weit mehr Varianten im humanen Genom gibt, als ursprünglich angenommen. Grundsätzlich lassen sich die Varianten in drei Klassen einteilen. Zum einen gibt es große Chromosomenaberrationen, die mikroskopisch und daher zytogenetisch erkannt werden können. Die nächste Kategorie sind Strukturvarianten, die Deletionen, Duplikationen, Inversionen, Translokationen und Insertionen umfassen. Strukturvarianten sind genomische Veränderungen, die größer als 50 bp sind und sowohl balanciert (kein Verlust von genetischem Material) als auch unbalanciert (Verlust von genetischem Material) sein können. Die dritte Kategorie sind die Variationen von Einzelbasen, die auch als „single nucleotide polymorphisms“ (SNPs, gesprochen „Snips“) bezeichnet werden. Für jede Kategorie sind entsprechende Methoden und Analyseverfahren notwendig.

Bei Einzelbasenvarianten kann es sich um eine Insertion, eine Deletion oder eine Substitution handeln, die zum Austausch einer einzelnen Base oder mehrerer Basen

führt. Die Beschreibung dieser Varianten erfolgt in einem standardisierten Format, einer sogenannten VCF-Datei, wodurch ein automatisierter Vergleich ermöglicht wird. Es erfolgt zunächst ein Abgleich mit Datenbanken bzgl. der Frequenz der identifizierten Varianten. Hier wird ein Frequenzfilter implementiert, mit dem häufige Varianten als Ursache für seltene Erkrankungen weitgehend ausgeschlossen werden können. Im Anschluss muss eine Bewertung der Varianten nach ihrer möglichen Pathogenität erfolgen, also nach ihrem Einfluss auf die Krankheitsentstehung. Dies ist ein komplexer Vorgang, der in einer Mischung aus bioinformatischen Algorithmen und klinischer Bewertung erfolgt.

Bei der Untersuchung des Panels oder von Exomen wird nach der Untersuchung der Frequenz durch bioinformatische Vorhersageprogramme geprüft, ob die Sequenzvariante zu einer Veränderung der Aminosäuresequenz („missense“ Mutation) oder zu einem vorzeitigen Stopp („nonsense“ Mutation) führt, oder ob Varianten zu einem veränderten Spleißen (dem Zusammenfügen von Exons und der Entfernung von Introns) führen. Weitere Vorhersageprogramme können überprüfen, ob die Veränderung der Aminosäuresequenz möglicherweise toleriert wird oder nicht, ob sie also keine Auswirkung auf die Funktion des Proteins hat oder doch. Dies geschieht z. B. über einen Vergleich der Aminosäuresequenz vieler Spezies, um zu untersuchen, ob die veränderte Aminosäure über die Evolution konserviert ist. Eine solche Konservierung deutet auf eine geringe Toleranz in der Variabilität an dieser Stelle des Proteins hin. Entsprechend ist eine Sequenzveränderung an einer solchen Stelle eher als pathogen zu bewerten. Ein weiterer Filter, der in der Vorhersage benutzt wird, ist die Frage, ob die genetische Variante neu – *de novo* – aufgetreten ist. So ist eine neu aufgetretene Sequenzvariante mit fraglicher Pathogenität, die in einem mit der jeweiligen Erkrankung assoziierten Gen gefunden wird, als wahrscheinlich pathogen zu bewerten. Wenn hingegen die Variante von einem gesunden Elternteil vererbt wurde, spricht dies eher für eine nicht krankheitsverursachende Variante. Weitere Verfahren verknüpfen die Variantenbewertung mit dem Phänotyp des Patienten. Wird dieser mit einem digitalen System wie der Human Phenotype Ontology (HPO) erfasst, so können die Varianten, das betroffene Gen und der Phänotyp auf eine pathogenetische Übereinstimmung abgeglichen werden. Dies führt zu einer weiteren Verbesserung der Vorhersagekraft (Zemajt et al., 2014). Neu entwickelte Algorithmen verbinden diese verschiedenen Methoden der Variantenbeurteilung, um so zu einer umfassenden Vorhersage zu kommen (Holtgrewe et al., 2020).

Trotz aller Fortschritte bleibt die Interpretation der Varianten eine Herausforderung. Dies liegt an der hohen genetischen Variabilität zwischen Individuen. Keine zwei Menschen sind genetisch identisch. Sogar eineiige Zwillinge zeigen genetische Unter-

schiede, hervorgerufen durch Mutationen, die während der frühen Entwicklung auftreten. Untersuchungen im *1.000 Genomes Project*¹⁶ haben gezeigt, dass ein Individuum typischerweise zwischen 4 und 5 Millionen Unterschiede in der Sequenz im Vergleich zum Referenzgenom aufweist, was ca. 0,6 % der Gesamtanzahl an Basen entspricht. Fast alle dieser Stellen (99,9 %) entsprechen kleinen Unterschieden entweder einzelner Basenaustausche oder kleiner Insertionen oder Deletionen in der genetischen Sequenz. Strukturvarianten sind seltener, aber aufgrund ihrer Größe für mehr Variationen verantwortlich als die SNPs und haben daher oft stärkere Auswirkungen. Strukturvarianten können vielfältige Auswirkungen auf die Funktionen des Genoms haben. Zum einen können sie die Gen-Dosis (die Anzahl der Genkopien) verändern und somit zu einem Zuviel oder Zuwenig an exprimierten Genen führen. Zum anderen können Strukturvarianten Gene verkürzen (trunkieren) oder es kann zu Fusionen von Genen kommen. Manche Varianten können auch zum Austausch von regulatorischen Elementen und zur Genfehlregulation führen. Die Detektion und Interpretation von Strukturvarianten stellt weiterhin eine große Herausforderung dar (Alkan et al., 2011; Spielmann et al., 2018).

Was die bisher detektierbare Zahl von Varianten pro Genom betrifft, so ist anzumerken, dass immer noch größere Teile des Genoms nicht vollständig zusammengesetzt (assembliert) sind, was insbesondere hochrepetitive (sich oft wiederholende) Bereiche des Genoms betrifft. Es ist anzunehmen, dass die Variabilität in diesen Bereichen noch wesentlich höher ist als in den bisher untersuchten. Ob diese Bereiche allerdings maßgeblich zur Krankheitsentstehung beitragen, lässt sich derzeit noch nicht sagen.

Mithilfe dieser großen Anzahl an Varianten lässt sich eine Vielzahl von Aussagen treffen. Zum einen wird diese Information intensiv genutzt, um Assoziationen mit bestimmten häufigen Erkrankungen wie Diabetes herzustellen. Diese sogenannten GWAS-Studien (Genomweite Assoziationsstudien) haben eine Vielzahl von Varianten im Genom identifiziert, die mit erhöhter oder verminderter Suszeptibilität (Empfänglichkeit) für solche Erkrankungen einhergehen. Obwohl solche Untersuchungen wichtige Informationen zur möglichen Pathogenese (Krankheitsentstehung) beisteuern, sind sie für den diagnostischen Bereich derzeit nicht verwertbar. Dies liegt an dem sehr kleinen Effekt der jeweiligen Veränderungen, die auch zusammengenommen nur eine geringe Vorhersagekraft haben.

Für seltene genetische Erkrankungen hingegen hat diese Information einen riesigen Schub an Wissen gebracht. Mit den jetzt vorhandenen Daten können Sequenzver-

16 Siehe unter: <https://www.internationalgenome.org/home> [12.03.2021]. Das Projekt sequenziert Einzelgenome und vergleicht sie miteinander.

gleiche vorgenommen und für eine effektive Filterung der Varianten genutzt werden. Hierdurch können potenziell pathogene Varianten mit hoher Effektivität identifiziert werden. Diese Datenlage ist die Grundlage für die derzeit durchgeführte molekulare Diagnostik bei seltenen Erkrankungen.

Die Sequenzierung von Panels, Exomen oder Genomen von Tumoren setzt neue Maßstäbe für den Einsatz zielgerichteter Therapeutika und befördert die technologische Entwicklung der NGS-Technologie (Khotskaya et al., 2017). Die Interpretation von NGS-Daten von Tumorgewebe ist jedoch mit enormen Herausforderungen verbunden. So müssen somatische, also im Tumor auftretende, Varianten von denen der Keimbahn unterschieden werden. Die Sequenzierung des Tumors ergibt zudem oft eine große Zahl an Varianten, befördert durch Reparaturdefekte im Tumor, von denen aber die meisten in Bezug auf die Tumorphathogenese neutral sind („Passager“-Mutationen). Im Gegensatz zur Mutationssuche bei genetischen Erkrankungen, bei denen die *eine* verursachende pathogene Variante gefunden werden muss, wird hier nach spezifischen sogenannten Treibermutationen¹⁷ gesucht, die therapeutisch angreifbar sind. NGS-Daten werden von multidisziplinären Teams interpretiert, damit der behandelnde Arzt Prioritäten setzen kann, ob und wie die Therapie auf den jeweiligen Tumor bzw. die Mutationen abgestimmt werden kann. So soll in Zukunft eine auf den jeweiligen Patienten und sein einzigartiges Genom zugeschnittene Krebstherapie angeboten werden (siehe Fehse, Kap. 6, zur Gentherapie).

4.4 Gendiagnostik in der Gesundheitsversorgung

4.4.1 Exomdiagnostik bei seltenen Erkrankungen und unklarer Diagnose

Seltene Erkrankungen treten in einer Häufigkeit von weniger als 1:2.000 auf. Aufgrund der hohen Zahl an unterschiedlichen seltenen Erkrankungen (derzeit ca. 8.000) sind sie aber insgesamt häufig. So leiden derzeit alleine in der Bundesrepublik Deutschland geschätzt mehr als vier Millionen Menschen an einer seltenen Erkrankung. Aufgrund ihrer Seltenheit sind diese Erkrankungen schwer zu diagnostizieren. Erst eine Diagnose beendet Unsicherheit und weitere unnötige Diagnostik, und ermöglicht eine Prognose und eine verlässliche Aussage bzgl. eines Wiederholungsrisikos. Oft ist nur mit einer Diagnose eine kausale Therapie möglich und ohne molekulare Diagnose kann keine Forschung zu neuen Therapien durchgeführt werden. Entscheidend für die

¹⁷ Treibermutationen sind solche, die die Erkrankung auslösen oder beschleunigen, die also etwa für das bösartige Wachstum von Tumoren verantwortlich sind.

Verbesserung der Gesundheit von Menschen mit seltenen Krankheiten ist daher eine frühzeitige Diagnose. Im Bereich von seltenen Erkrankungen wird die überwiegende Mehrheit der Krankheiten durch bisher schwer zu identifizierende, proteinkodierende Genmutationen verursacht. Die Senkung der Kosten für Gentests und die Fortschritte bei der klinischen Anwendung der Genomsequenzierung haben zur Verbesserung dieser Problematik geführt und ermöglichen inzwischen eine Diagnose bei allen, auch bei ganz seltenen Erkrankungen (Sawyer et al., 2016). Dies beruht ganz wesentlich auf der Einführung der neuen Hochdurchsatzsequenzierertechnologie. Erst mit dieser Technologie war es möglich, größere Abschnitte des Genoms oder auch das ganze Genom zu analysieren. Durch Weiterentwicklung der Sequenzierertechnologien ist es auch zu einer massiven Senkung der Sequenzierkosten gekommen, von etwa 100 Millionen US\$ im Jahre 2007 auf ca. 3.000 US\$ im Jahr 2012 und jetzt auf weniger als 1.000 € pro Genom. Hierdurch ist der Einsatz dieser Technologie in der Routinediagnostik möglich geworden. Natürlich decken die reinen Sequenzierkosten nur einen Teil des Aufwands ab. Erhebliche weitere Kosten entstehen durch die bioinformatische Analyse und Infrastruktur, das Speichern der Daten und insbesondere durch die Analyse der Varianten.

Deutschland hat sich über viele Jahre schwergetan, diese Art der Diagnostik in die Routineversorgung einzubinden. Gleichzeitig stellt die Hochdurchsatzsequenzierung das ultimative Tool in der Diagnostik von seltenen genetischen Erkrankungen dar. In einer aktuellen Studie konnten wir jetzt die Anwendung der Exom-Diagnostik bei Patienten mit unklarer Diagnose in einem bundesweiten Konsortium testen. In diesem Projekt (TranslateNAMSE)¹⁸ wurden vorwiegend pädiatrische Patienten aus verschiedenen Zentren für seltene Erkrankungen untersucht, bei denen auch unter Hinzuziehung von Experten verschiedener Fachrichtungen keine Diagnose gestellt werden konnte. In diesem Falle wurde eine Exomsequenzierung in einem von vier beteiligten Zentren durchgeführt. Die Ergebnisse zeigen, dass, unabhängig von dem durchführenden Zentrum, eine Diagnose in ca. einem Drittel der Fälle erreicht werden kann.¹⁹ Diese hohe Erfolgsquote konnte interessanterweise unabhängig von der jeweiligen Patientenkohorte vor Ort erzielt werden, was die Verlässlichkeit und Reproduzierbarkeit der Methode unterstreicht. In vielen Fällen konnten Diagnosen bei Patienten gestellt werden, die jahrelang immer wieder bei den verschiedensten Experten vorstellig waren, ohne dass ihnen geholfen werden konnte. Oft ist erst durch die richtige Diagnose eine Zugangsmöglichkeit zur Behandlung des Patienten möglich. Der Erfolg dieses Projektes

18 NAMSE steht für „Nationales Aktionsbündnis für Menschen mit Seltenen Erkrankungen“, siehe unter: <https://www.namse.de> [12.03.2021].

19 Siehe unter: <https://translate-namse.charite.de/> [12.03.2021].

hat dazu beigetragen, dass die Krankenkassen die Exomsequenzierung nach besonderer Indikationsstellung ab Januar 2021 als Kassenleistung übernehmen werden.

4.4.2 Grenzen der Exomdiagnostik

Die Einführung der Exomdiagnostik in die Routine-Patientenversorgung ist als Erfolgsgeschichte zu werten. Trotzdem kann, je nach Kohorte, bei 50 bis 70 % der Patienten ohne Diagnose auch mit dieser umfangreichen Diagnostik keine Diagnose gestellt werden. Die Gründe hierfür sind vielfältig und werden die humangenetische Forschung über die nächsten Jahre, wenn nicht Jahrzehnte, beschäftigen. Mögliche Gründe für eine fehlende Diagnose nach Exomsequenzierung können z. B. übersehene, derzeit nicht interpretierbare oder falsch interpretierte Varianten im kodierenden Genom sein. Um diese Fälle doch noch zu erfassen, muss eine regelmäßige Re-Evaluation von ungelösten Fällen und nicht eindeutig einzustufenden Varianten erfolgen. Es ist davon auszugehen, dass die Bewertung mit zunehmender Dichte der Sequenzierdaten immer besser, schneller und effizienter erfolgen kann, sodass der Prozentsatz der schlecht oder nicht interpretierbaren Varianten immer weiter sinken wird. Eine wesentliche Lücke der Exomsequenzierung liegt natürlich im nicht-kodierenden restlichen Genom, was immerhin ca. 98 % des ganzen Genoms ausmacht. Dieser Bereich wird in der derzeitigen Analyse, mit Ausnahme der Spleißvarianten, weitgehend ausgeblendet, da Varianten im nicht-kodierenden Genom derzeit noch äußerst schwierig zu beurteilen sind. Derzeitige Studien gehen davon aus, dass der Großteil des nicht-repetitiven nicht-kodierenden Genoms eine regulatorische Funktion hat und somit für die Regulation der Gene in verschiedenen Phasen der Entwicklung und Zelldifferenzierung verantwortlich ist (Consortium et al., 2020). Wie dies geschieht und wie Varianten im Genom die Genexpression beeinflussen, ist derzeit Ziel intensiver Forschung. Viele weitere Erklärungsmöglichkeiten sind denkbar. So können z. B. auch epigenetische Effekte, die nicht den Genotyp (die DNA-Sequenz) verändern, sondern eine Modifikation der DNA-Struktur vornehmen, krankheitsverursachend sein.

4.4.3 Ganzgenomsequenzierung in der Diagnostik

Es ist davon auszugehen, dass auch die Sequenzierung des ganzen Genoms („whole genome sequencing“, WGS) in naher Zukunft in der Diagnostik eingesetzt werden wird. Die Sequenzierung des ganzen Genoms hat gegenüber der Exomsequenzierung mehrere Vorteile, da sich hiermit nicht nur die kodierenden, sondern auch die nicht-kodierenden Bereiche des Genoms analysieren lassen. Zudem können Strukturvarianten zu

einem hohen Maß detektiert werden. Gleichzeitig bietet WGS eine uniforme Abdeckung des gesamten Genoms (es werden also alle Bereiche des Genoms repräsentiert), sodass die Anreicherung über Capture-Verfahren überflüssig wird (Lelieveld et al., 2015). Für die genomweite Detektion von SNPs ist diese Technologie sehr gut geeignet.

Die Detektion von Strukturvarianten mittels Illumina-Sequenzierung bleibt allerdings eine Herausforderung. Durch die Kürze der bei dieser Methode entstehenden Reads ist es oft nicht möglich, eine präzise Bestimmung von Bruchpunkten vorzunehmen. Zudem entstehen Bruchpunkte vorrangig in repetitiven Bereichen, was wiederum ein Problem darstellt, wenn relativ kurze Sequenzen von 100 Basen auf dem Genom verortet werden müssen. Eine Verbesserung der Algorithmen erlaubt hier eine zunehmend präzisere Analyse, aber die genaue vollständige Detektion bleibt weiterhin unerreicht. Alternative Technologien versprechen hier zusätzliche Informationen.

4.4.4 „Long read single molecule“-Sequenzierung

Als Third-Generation-Sequencing werden Technologien der Einzelmolekülsequenzierung bezeichnet, wie sie von den Firmen PacBio und Oxford Nanopore entwickelt wurden (Midha et al., 2019). Bei der „single molecule real-time“-Technologie (SMRT) der Firma PacBio werden einzelne DNA-Moleküle in einer sogenannten SMRT-Cell gelesen, wodurch zusammenhängende Sequenzen von mehreren 10.000 bp entstehen. Diese langen Sequenzen stellen einen potenziellen Vorteil bei der Detektion von Strukturvarianten dar. Auf der anderen Seite hat diese Technologie jedoch den Nachteil, dass sie eine geringere Lesegenauigkeit hat als die Illumina-Technologie. Dieser Nachteil wird durch ein neues Verfahren („high fidelity“) ausgeglichen, bei dem ein Molekül immer wieder sequenziert wird und so die Lesefehler korrigiert werden. Allerdings ist dieses Verfahren relativ teuer und die Leselänge wird geringer. Die SMRT-Technologie kann auch erfolgreich eingesetzt werden, um besonders GC-reiche Regionen, in denen die Basen C (Cytosin) und G (Guanin) häufig vorkommen, und Repeats (Wiederholungen von Gensequenzen) zu sequenzieren, die mit anderen Technologien nicht oder nur sehr schwer zu analysieren sind. Ob und inwieweit diese Technologie Einzug in die Routinediagnostik nehmen wird, muss sich erst noch zeigen.

Die Nanopore-Technologie ist ein weiteres vielversprechendes Sequenzierverfahren, welches allerdings auf einer völlig anderen Methode beruht (Midha et al., 2019). Hierbei wird ein DNA-Molekül durch eine sogenannte Nanopore geschickt, eine winzige (im Nanometerbereich liegende) Pore. Sie besteht aus einem Protein, das in ein Trägermaterial eingebettet ist und an das Spannung angelegt wird. Tritt die DNA hindurch, wird die Spannung verändert, was für jede Base spezifisch ist. Dies ermöglicht

eine Sequenzierung in Echtzeit, wobei die Basenfolge direkt am Computer abgelesen werden kann. Die Technologie ist mit einer relativ hohen Fehlerrate behaftet, weshalb sie derzeit zur direkten Mutationsdetektion nicht eingesetzt werden kann. Neben der enormen Schnelligkeit ist aber ein weiterer Vorteil dieser Methode, dass wie bei der PacBio-Technologie sehr lange Sequenzen (10 kB und länger) erstellt werden können. Dies ermöglicht den Einsatz der Technologie zur raschen Detektion von Strukturvarianten z. B. in Krebszellen (Mahmoud et al., 2019). Es ist davon auszugehen, dass diese Technologie aufgrund der Einfachheit der Anwendung und der extrem schnellen Analysezeit eine Anwendung in der Diagnostik finden wird.

Generell gilt, dass die Fähigkeit, Reads von über 10.000 bp Länge mit einer Genauigkeit zu generieren, die an die von Short-Read-Sequenzierungstechnologien heranreicht, die Genomik revolutionieren wird. Mit diesen Plattformen können einige der schwierigsten Regionen des menschlichen Genoms in ihrer Struktur aufgeschlüsselt werden; bisher nicht detektierbare strukturelle Varianten können erkannt werden und eine Assemblierung ganzer Chromosomen ist möglich. Long-Read-Sequenzierungstechnologien können so helfen, das gesamte Spektrum der menschlichen genetischen Variation aufzudecken, und werden damit zur Entdeckung neuer Krankheitsmechanismen beitragen (Logsdon et al., 2020).

4.5 Grenzen, Probleme und ethische Aspekte der Gendiagnostik

Wie oben ausgeführt, hat die Gendiagnostik dank neuer Technologien enorme Fortschritte gemacht, die es uns erlauben, ganze Genome in ihrer Sequenz zu erfassen und diagnostisch zu untersuchen. Trotzdem ist die Genetik in den meisten Fällen weit davon entfernt, präzise Aussagen zum Auftreten von Krankheiten zu machen. Grundsätzlich kann man sagen, dass die Aussagefähigkeit direkt mit der Effektgröße der Variante korreliert. Mutationen mit starkem Effekt, wie z. B. bei monogenen Erkrankungen, ermöglichen eine relativ präzise Aussage, ob eine Person betroffen ist, bzw. zum Krankheitsverlauf und -beginn. Aber selbst in dieser Situation gibt es eine Vielzahl an modulierenden Faktoren, die die Präzision der Prognose einschränken. So gibt es bei fast allen genetischen Erkrankungen eine hohe Variabilität der Ausprägung, sogar bei gleicher Mutation. Dies bedeutet, dass selbst innerhalb einer Familie einzelne Mutationsträger unterschiedlich betroffen sein können. Bei vielen Mutationen kommt es zu einer reduzierten Penetranz, d. h. nicht alle Mutationsträger sind überhaupt betroffen. So erhöhen z. B. Mutationen in einem der Brustkrebsgene die Wahrscheinlichkeit, an Brustkrebs zu erkranken, aber eine Aussage darüber, wann die Erkrankung auftritt

und ob überhaupt lässt sich nur statistisch treffen. Somit basieren die meisten Aussagen der genetischen Präzisionsmedizin auf Wahrscheinlichkeiten. Für viele Menschen sind Zahlen, die die Wahrscheinlichkeit bestimmen, dass z. B. eine Erkrankung eintritt, jedoch schwierig zu interpretieren und werden auch individuell nach persönlicher Einstellung und/oder Lebenslage unterschiedlich bewertet. Neben diesen Problemen produzieren genetische Tests auch sehr häufig Ergebnisse, die uneindeutig sind, weil die entsprechende Variante nach dem heutigen Wissensstand nicht eindeutig zu bewerten ist. In diesem Fall muss die Beratung entscheiden, was mitgeteilt wird und was nicht. Aus diesen Gründen ist es unerlässlich, dass genetische Diagnostik mit einer Beratung verbunden wird.

Viele Varianten im Genom sind mit Erkrankungen assoziiert, d. h. sie kommen häufiger bei Erkrankten als bei Gesunden vor. Die Effekte solcher Varianten sind jedoch im Allgemeinen klein und aus diesen Gründen diagnostisch nicht verwertbar. Trotzdem werden sie als Tests angeboten, z. B. um das Risiko für bestimmte häufige Erkrankungen zu bestimmen. Nach der derzeitigen Datenlage ist diese Indikation für eine Testung aufgrund ihrer geringen Aussagekraft medizinisch jedoch nicht indiziert. Weiterhin werden insbesondere in der Lifestyle-Genetik SNP-basierte Tests oder sogar eine Genomsequenzierung von diversen Firmen als „Direct-to-consumer“-Tests angeboten, die vorhersagen sollen, welche Lebensmittel besser vertragen bzw. gemieden werden sollten und vieles mehr. Diese Art der Testung ist wissenschaftlich nicht begründet, auch wenn dies in den Anzeigen der entsprechenden Anbieter suggeriert wird.

Genetische Tests, die mehr testen als die Gene, die direkt mit der in Frage stehenden Krankheit assoziiert sind, bergen immer die Gefahr, dass zusätzliche Ergebnisse erzeugt werden, die mit der ursprünglichen Fragestellung nichts zu tun haben. So kann eine Exomsequenzierung, durchgeführt bei einem Kind zur Abklärung einer Fehlbildung, eine Mutation in einem Elternteil aufdecken, die mit einem erhöhten Krebsrisiko verbunden ist. Diese sogenannten „Zufallsbefunde“ entstehen durch bioinformatische Suchalgorithmen und sind abhängig von der Tiefe und Art der Untersuchung. Der Umgang mit diesen Ergebnissen ist ein breit diskutiertes Problem in der Genetik (siehe Winkler/Prainsack, Kap. 17). In der derzeitigen Vorgehensweise werden Personen, bei denen ein genetischer Test durchgeführt werden soll, innerhalb eines Beratungsgesprächs gefragt, ob sie Zufallsbefunde mitgeteilt bekommen möchten oder nicht. Mitgeteilt werden nur solche Varianten, die mit einer direkten medizinischen Konsequenz verbunden sind bzw. bei denen medizinisches Eingreifen die Krankheit verhindern oder zumindest die Situation verbessern kann. Die Liste dieser Gene²⁰ beinhaltet

20 Siehe unter: <https://www.ncbi.nlm.nih.gov/clinvar/docs/acmg/> [22.03.2021].

u. a. Mutationen, die zu einem erhöhten Krebsrisiko führen und bei denen man durch entsprechende Maßnahmen das Risiko zu erkranken senken kann, oder die z. B. zu Herzrhythmusstörungen führen, eine Erkrankung, die mit einem Herzschrittmacher behandelt werden kann. Alternativ kann diese Liste von „actionable genes“ auch diagnostisch analysiert werden. In diesem Falle wären Ergebnisse in diesen Genen keine Zufallsbefunde, sondern wären systematisch nach Analyse entstanden und somit „Zusatzbefunde“.

Es ist somit klar, dass genetische Testung nicht nur Probleme lösen kann, sondern auch neue entstehen können, von denen die getestete Person vorher nichts wusste. Solche Ergebnisse können einen erheblichen Einfluss auf die weitere Lebensplanung haben und stellen somit ein potenzielles Risiko dar, das vorher bedacht werden muss. Aufgrund dieses Risikos ist bei allen umfassenden genetischen Tests eine klinisch begründete und restriktive Indikationsstellung notwendig, die unnötige Tests vermeidet. Inwieweit eine Testung von „actionable genes“ auf Wunsch sinnvoll ist, hängt von der eigenen Einstellung und einem verantwortungsvollen Umgang mit den Daten und deren Interpretation ab. Immer muss genetische Diagnostik den Nutzen der Diagnostik gegenüber dem Risiko abschätzen. Immer hat die individuelle Selbstbestimmung Vorrang.

4.6 Gesetzliche Regelungen

In Deutschland ist genetische Diagnostik durch das sogenannte Gendiagnostikgesetz (GenDG) aus dem Jahre 2009 geregelt.²¹ Das Gesetz regelt die Durchführung genetischer Untersuchungen zu medizinischen Zwecken, zur Abklärung der Abstammung, sowie im Versicherungsbereich und im Arbeitsleben, aber nicht in der Forschung. Ein wesentliches Grundprinzip des Gesetzes ist die informationelle Selbstbestimmung, woraus sich das Recht auf Kenntnis der eigenen Befunde wie auch das Recht auf Nichtwissen ergibt. Ein weiterer wichtiger Punkt ist das sogenannte Benachteiligungsverbot. Hierbei wird geregelt, dass niemand aufgrund seiner genetischen Eigenschaften bzw. aufgrund der Durchführung oder Nicht-Durchführung eines Testes oder der Ergebnisse benachteiligt werden darf. Zu den wesentlichen Punkten des Gesetzes gehört die notwendige schriftliche und rechtswirksame Einwilligung zu jeder Form von genetischer Diagnostik.

Eine genetische Beratung ist bei der Diagnostik von erkrankten Personen optional, als prädiktiver Test bei Gesunden, also zur Abklärung des Risikos einer erst zukünftig

21 Siehe unter: <http://www.gesetze-im-internet.de/genDG/index.html> [22.03.2021].

auftretenden Erkrankung, jedoch zwingend vorgeschrieben. Die Beratung kann von einem ausgebildeten Humangenetiker erfolgen oder durch einen Arzt mit der Qualifikation zur fachgebundenen genetischen Beratung. Diagnostische genetische Untersuchungen dürfen von allen Ärzten durchgeführt werden, während die sogenannten prädiktiven genetischen Untersuchungen hingegen nur von Fachärzten für Humangenetik durchgeführt werden dürfen und von solchen, die sich hierfür im Rahmen ihres Fachgebietes qualifiziert haben.

Das Ergebnis einer genetischen Untersuchung darf nur der betroffenen Person und nur durch den verantwortlichen Arzt mitgeteilt werden, es sei denn es werden zusätzliche Personen benannt bzw. andere Regelungen vereinbart. Ferner darf das Ergebnis nicht mitgeteilt werden, wenn die betroffene Person sich dagegen entschieden hat (Recht auf Nichtwissen).

Vorgeburtliche Untersuchungen dürfen nur zu medizinischen Zwecken vorgenommen werden. Ferner sind vorgeburtliche genetische Untersuchungen nicht erlaubt, die nach dem allgemeinen anerkannten Stand der medizinischen Wissenschaft erst nach Vollendung des 18. Lebensjahres ausbrechen. Diese Regelung macht eine vorgeburtliche Diagnostik auch von schweren Erkrankungen wie z. B. der Huntington-Erkrankung²² unmöglich.

Insgesamt lässt sich sagen, dass das GenDG bestimmt zur Akzeptanz genetischer Untersuchungen in der Bevölkerung beigetragen hat. Durch den seit Inkrafttreten des Gesetzes stattgefundenen technologischen Fortschritt erscheinen allerdings einige der Vorgaben als zu rigide. Die im Gesetz ursprünglich vorgesehene Untersuchung einzelner weniger Gene wird jetzt durch Exomsequenzierung oder Ganzgenomsequenzierung abgelöst. Hieraus ergeben sich neue Herausforderungen, insbesondere für die Aufklärung des Patienten und seiner Angehörigen. So wird bei der Suche nach einer krankheitsverursachenden Mutation insbesondere bei unklarer Diagnose das ganze Exom bzw. Genom untersucht. Bei diesem Vorgehen können auch Mutationen identifiziert werden, die nicht mit der zu untersuchenden Krankheit in Zusammenhang stehen. Diese Veränderungen werden als „Zufallsbefunde“ bezeichnet, die nur mitgeteilt werden, wenn dies in einer vorherigen Einwilligung ausdrücklich gewünscht wird. Die mögliche Identifikation von Mutationen, die nichts mit der eigentlichen diagnostischen Fragestellung zu tun haben, macht jede genomweite Untersuchung zu einem prädiktiven Test nach GenDG und erfordert daher eine humangenetische Beratung. Dies

²² Chorea Huntington (auch Veitstanz) ist eine unheilbare neurodegenerative monogene Erbkrankheit, bei der Betroffene zunehmend die Kontrolle über ihre Muskeln und mentale Funktionen verlieren. Sie bricht meist um das 35. bis 45. Lebensjahr aus und führt individuell unterschiedlich schnell zum Tod. Siehe unter: <https://www.dhh-ev.de/Was-ist-Huntington> [21.03.2021].

erscheint angesichts der immer weiter zunehmenden genetischen Diagnostik als nicht praktikabel. Grundsätzlich herrscht aber in der Humangenetik die Meinung vor, dass die Regelungen vom Prinzip her richtig sind, aber an die neuen diagnostischen Möglichkeiten angepasst werden müssten.

4.7 Abrechnungsmodalitäten und Zugang zu genetischer Diagnostik im deutschen Gesundheitswesen

Im deutschen Gesundheitssystem wird zwischen ambulanter und stationärer Versorgung unterschieden. Beide Bereiche unterliegen einer unterschiedlichen Abrechnungssystematik. Im stationären System werden erbrachte medizinische Leistungen über sogenannte „disease related groups“ (DRGs, krankheitsbezogene Gruppen) abgerechnet und erstattet. Zudem stehen Zusatzentgelte für bestimmte Indikationen und Patientengruppen zur Verfügung. So auch zur zytogenetischen und molekulargenetischen Untersuchung von kranken Neugeborenen. Da die DRGs Pauschalen sind, ist die zumeist teure genetische Diagnostik hierüber nur unzureichend abgedeckt. Deshalb wird fast alle genetische Diagnostik im ambulanten Bereich durchgeführt. Hier erfolgt die Erstattung durch sogenannte EBM-Ziffern (Einheitlicher Bewertungsmaßstab), die einzelne bestimmte Leistungen abbilden. Die klassischen diagnostischen Maßnahmen der Genetik wie Zytogenetik, Array-CGH, und Sanger-Sequenzierung werden mit entsprechenden Ziffern abgedeckt und können so mit den Kassen abgerechnet werden. Seit einigen Jahren ist auch die gezielte Untersuchung von Genen bis zu einer Gesamtlänge von 25 kb mittels NGS erlaubt. Diese Obergrenze wurde kürzlich aufgehoben, so dass jetzt Gen-Panels von beliebiger Größe sequenziert und analysiert werden können, sofern hiermit eine bestimmte klinische Fragestellung bzw. Indikation abgedeckt wird.

Die Anwendung der Exom- bzw. Ganzgenomsequenzierung wurde bisher von den Krankenkassen nicht erstattet und war über die Routinediagnostik nicht möglich. Allerdings ist diese Form der Testung mittlerweile in jeder größeren Region der Welt verfügbar oder wird verfügbar werden. Dazu gehören sowohl Länder mit hohem Einkommen und etabliertem Genomprogrammen wie die USA und Großbritannien als auch eine wachsende Zahl an Ländern mit mittlerem und höherem Einkommen (Phillips et al., 2021). In Deutschland ergeben sich seit diesem Jahr neue Möglichkeiten. Eine umfangreiche Diagnostik von Genen ist nun auch ohne Beschränkung möglich und wird als diagnostische Leistung von den Kassen erstattet. Die Anwendung der Exomdiagnostik ist unter selektiver Indikationsstellung über spezifische Verträge innerhalb der Zentren für seltene Erkrankungen möglich.

4.8 Ausblick: Eine Diagnose für alle seltenen Erkrankungen

Hunderte Millionen von Menschen sind von einer der ca. 8.000 genetisch bedingten seltenen Erkrankungen betroffen. Einzeln betrachtet sind die Zahlen gering, aber für alle seltenen Erkrankungen zusammen werden sie zu einem erheblichen Problem. Bei den meisten dieser Patienten dauert es lange, teilweise Jahre, bis eine korrekte Diagnose gestellt wird. Es ist das Ziel verschiedener politischer Aktivitäten, so auch des Rare Disease Research Consortiums (IDDiRC), diese Situation zu verbessern (Austin et al., 2018). Das Consortium möchte alle Menschen mit einer seltenen Erkrankung innerhalb eines Jahres nach Auftreten der Symptomatik einer Diagnose und einer entsprechenden Therapie zuführen. Da die große Mehrzahl der seltenen Erkrankungen genetisch bedingt ist, ist dies ganz wesentlich eine Frage der effizienten und vollständigen genomischen Diagnostik. Ohne Frage wird die Exomsequenzierung hierbei eine zentrale Rolle spielen. Wie schon erwähnt, bleiben aber trotz dieser sehr umfangreichen und erfolgreichen Diagnostik zwischen 50 und 70 % der Betroffenen ohne Diagnose. Einen wesentlichen Fortschritt kann hier die Verknüpfung von Phänotyp mit Genotyp z. B. mittels der Datenbank Human Phenotype Ontology (HPO) (Kohler et al., 2017) erbringen, um auch sehr seltene Varianten bzw. Erkrankungen identifizieren zu können. Sie ermöglicht durch ein standardisiertes Vokabular für die einheitliche Beschreibung verschiedener phänotypischer Merkmale genetischer Erkrankungen eine bessere Zuordnung von Phänotyp und Genotyp und erleichtert so auch die Diagnosestellung. Große Phänotyp-Genotyp-Datenbanken, die einen internationalen Austausch ermöglichen, können so helfen, sehr seltenen genetischen Veränderungen eine Krankheit zuzuordnen (Brookes/Robinson, 2015). Leider wird diese Form der Kommunikation bisher nur unzureichend genutzt. Dies liegt zum einen an Problemen der Datensicherheit und der unterschiedlichen Voraussetzungen in einzelnen Ländern, zum anderen aber auch daran, dass die allermeisten Exomdaten nicht zugänglich gemacht werden bzw. nicht mit anderen Laboren geteilt werden. Hier sind dringend Regelungen notwendig, die z. B. eine Erstattung der Kosten durch die Kassen nur dann zulassen, wenn die Daten in entsprechende Datenbanken eingepflegt werden. Eine Zersplitterung des Systems ohne zentrale Vorgaben und/oder eine rein kommerzielle Ausrichtung der Labore sind hierbei sicher nicht hilfreich.

Jenseits dieser strukturellen Herausforderungen bleibt die immer noch bestehende diagnostische Lücke ein wissenschaftliches Problem, das in den nächsten Jahren durch intensive Forschung angegangen werden muss. Hierbei ist insbesondere die Bedeutung der neuen Technologien und Methoden hervorzuheben, die einen wesentlich tieferen Einblick in die Funktion und Zusammensetzung unseres Genoms erlauben. So hat das ENCODE-Consortium sich zum Ziel gesetzt, das nicht-kodierende Genom in seiner Funk-

tion aufzuklären (Skipper et al., 2012). Eine umfassende und vollständige genomische Analyse aller Varianten im Genom sollte helfen, den Prozentsatz der diagnostizierbaren Erkrankungen zu erhöhen. Hierbei wird eine entscheidende Rolle spielen, dass es gelingt, die Funktionen des nicht-kodierenden Genoms und seiner Varianten besser zu verstehen. Es ist anzunehmen, dass andere Erklärungsmöglichkeiten jenseits der genomischen Sequenz wie epigenetische Mechanismen (also biochemische Modifikationen, welche die DNA-Struktur und ihr Ablesen beeinflussen; siehe Walter/Gasparoni, Kap. 3) und oligogene Vererbung, die Vererbung durch mehrere Gene, hier eine Rolle spielen.

4.9 Literaturverzeichnis

- Alkan, C. et al. (2011): Genome structural variation discovery and genotyping. In: *Nat. Rev. Genet.* 12: 363–376.
- Austin, C. P. et al. (2018): Future of rare diseases research 2017–2027: An IRDiRC perspective. In: *Clin Transl Sci* 11(1): 21–27.
- Bamshad, M. J. et al. (2011): Exome sequencing as a tool for Mendelian disease gene discovery. In: *Nature Reviews Genetics* 12(11): 745–755.
- Bartfeld, S. (2020): Das Potenzial von Organoiden realisieren. Ein Interview mit Hans Clevers. In: Bartfeld, S. et al. (Hrsg.): *Organoid. Ihre Bedeutung für Forschung, Medizin und Gesellschaft.* Nomos, Baden-Baden: 65–76.
- Brookes, A. J./Robinson, P. N. (2015): Human genotype-phenotype databases: aims, challenges and opportunities. In: *Nat Rev Genet* 16(12): 702–715.
- Chong, J. X. et al. (2015): The genetic basis of Mendelian phenotypes: Discoveries, challenges, and opportunities. In: *Am J Hum Genet* 97(2): 199–215.
- Clausnitzer, M. et al. (2020): A brief history of human disease genetics. In: *Nature* 577(7789): 179–189.
- Collins, F. S./McKusick, V. A. (2001): Implications of the Human Genome Project for medical science. In: *JAMA* 285(5): 540–544.
- Consortium, E. P. et al. (2020): Expanded encyclopaedias of DNA elements in the human and mouse genomes. In: *Nature* 583(7818): 699–710.
- Goodwin, S. et al. (2016): Coming of age: ten years of next-generation sequencing technologies. In: *Nat Rev Genet* 17(6): 333–351.
- Holtgrewe, M. et al. (2020): VarFish: comprehensive DNA variant analysis for diagnostics and research. In: *Nucleic Acids Res* 48(W1): W162–W169.
- International Human Genome Sequencing Consortium (2001): Initial sequencing and analysis of the human genome. In: *Nature* 409(6822): 860–921.

- Khotskaya, Y. B. et al. (2017): Next-generation sequencing and result interpretation in clinical oncology: Challenges of personalized cancer therapy. In: *Annu Rev Med* 68: 113–125.
- Kohler, S. et al. (2017): The human phenotype ontology in 2017. In: *Nucleic Acids Research* 45(D1): D865–D876.
- Lelieveld, S. H. et al. (2015): Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. In: *Hum Mutat* 36(8): 815–822.
- Logsdon, G. A. et al. (2020): Long-read human genome sequencing and its applications. In: *Nat Rev Genet* 21(10): 597–614.
- Mahmoud, M. et al. (2019): Structural variant calling: the long and the short of it. In: *Genome Biol* 20(1): 246.
- Midha, M. K. et al. (2019): Long-read sequencing in deciphering human genetics to a greater depth. In: *Hum Genet* 138(11–12): 1201–1215.
- Ng, S. B. et al. (2010): Massively parallel sequencing and rare disease. In: *Hum Mol Genet* 19(R2): R119–R124.
- Phillips, K. A. et al. (2021): Availability and funding of clinical genomic sequencing globally. In: *BMJ Glob Health* 6(2): e004415.
- Posey, J. E. et al. (2019): Insights into genetics, human biology and disease gleaned from family based genomic studies. In: *Genetics in Medicine* 21(4): 798–812.
- Sawyer, S. L. et al. (2016): Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. In: *Clinical Genetics* 89(3): 275–284.
- Shendure, J. et al. (2017): DNA sequencing at 40: past, present and future. In: *Nature* 550(7676): 345–353.
- Shendure, J. et al. (2019): Genomic medicine—progress, pitfalls, and promise. In: *Cell* 177(1): 45–57.
- Skipper, M. et al. (2012): Presenting ENCODE. In: *Nature* 489(7414): 45.
- Spielmann, M. et al. (2018): Structural variation in the 3D genome. In: *Nat Rev Genet* 19(7): 453–467.
- Venter, J. C. et al. (2001): The sequence of the human genome. In: *Science* 291(5507): 1304–1351.
- Vissers, L. E. et al. (2005): Identification of disease genes by whole genome CGH arrays. In: *Hum Mol Genet* 14(2): R215–R223.
- Zemojtel, T. et al. (2014): Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. In: *Science Translational Medicine* 6(252): 252ra123.