

Next Generation Sequencing (NGS) von DNA

Mehr Informationen durch mehr Daten?

T. Eggermann, L. Soellner, I. Kurth, K. Eggermann und M. Begemann

Institut für Humangenetik, Universitätsklinikum der RWTH Aachen

Next Generation Sequencing (NGS) ermöglicht als Hochdurchsatzverfahren die umfassende Analyse des menschlichen Genoms. Entwicklung und diagnostischer Einsatz des NGS sind aber mit bioinformatischen und ethischen Herausforderungen verbunden. Der Beitrag beschreibt die Methodik, erläutert Chancen und Herausforderungen und zeigt anhand der nicht-invasiven pränatalen Testung (NIPT) und eines klinischen Fallbeispiels aktuelle spezifische Anwendungen und Probleme des NGS auf.

Stichwörter: Humangenetik, *Next Generation Sequencing* (NGS), Exom, Big Data, nicht-invasive pränatale Testung (NIPT), seltene Erkrankungen, personalisierte Medizin, humangenetische Beratung

1 Einleitung

Die Verwendung der in den letzten Jahren entwickelten *Next Generation Sequencing* (NGS)-Strategien in der Grundlagenforschung hat zu einem enormen Wissenszuwachs beigetragen, auf dem Gebiet der Humangenetik insbesondere bei der Entdeckung neuer krankheitsrelevanter Gene und krankheitsverursachender genetischer Veränderungen (Mutationen). Diese neuen Erkenntnisse erlauben Einsichten in das Zustandekommen von angeborenen Erkrankungen, sie helfen aber auch, grundlegende biologische Prozesse zu verstehen: Mutationen zeigen nämlich, wie Abläufe richtig funktionieren sollten.

Ebenso revolutioniert der Einsatz des NGS die genetische Diagnostik: Während bis vor einigen Jahren durch Verwendung der klassischen Verfahren (z.B. Chromosomendiagnostik, Sanger-Sequenzierung) nur große oder ausgewählte Veränderungen erfasst werden konnten, ermöglicht das NGS nun die Erfassung aller Veränderungen im Erbgut. Dieser Schritt ist eine wesentliche Grundlage für die sogenannte personalisierte bzw. Präzisions-Medizin.

1.1 Aufbau des menschlichen Genoms

Das menschliche Kerngenom ist mit Ausnahme weniger Zellen diploid und besteht aus etwas mehr als sechs Milliarden Basenpaaren (bp), die in den Körperzellen auf 46 Chromosomen (22

Autosomenpaare und die beiden Geschlechtschromosomen X und Y) verteilt sind. Keimzellen sind dagegen haploid. Insgesamt ist die Erbinformation des Menschen auf $3,27 \times 10^9$ bp = 3,27 Gigabasenpaaren (Gbp) = 3.270 Megabasenpaaren (Mbp) = 3.270.000 Kilobasenpaaren (kbp) gespeichert. Allerdings sind von diesem Genom nur ca. 2,5% relevant für ca. 19.000 proteinkodierende Gene (Abb. 1).

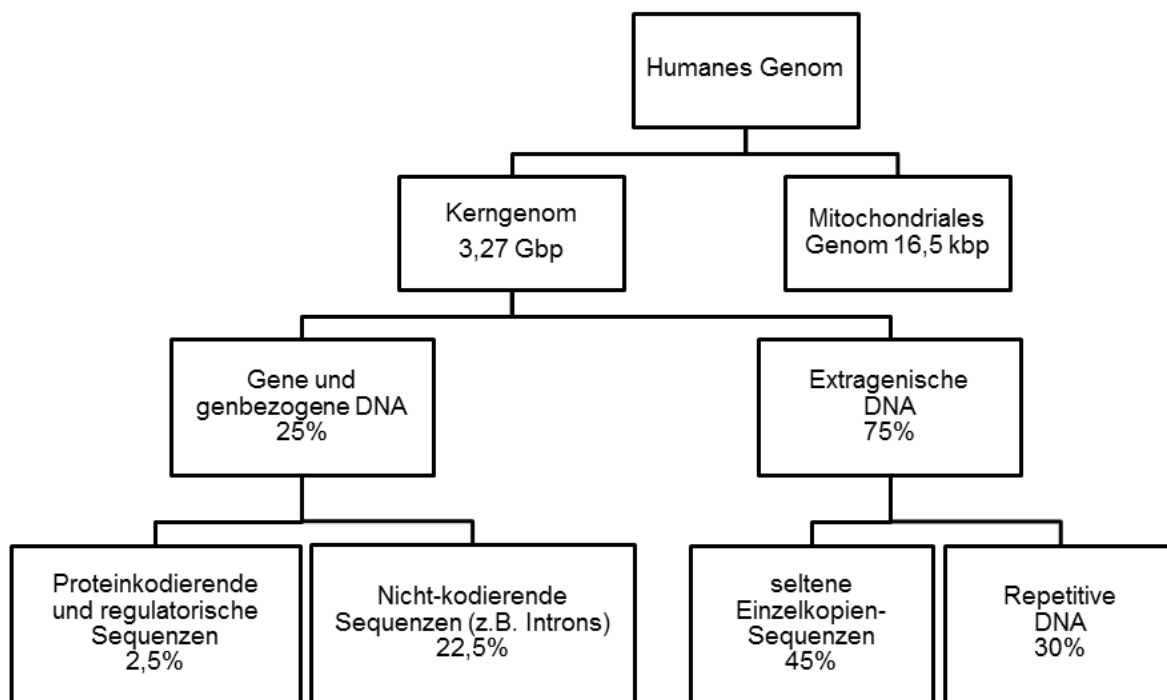


Abb. 1: Zusammensetzung des menschlichen Genoms. Der Anteil der proteinkodierenden und regulatorischen Sequenzen beträgt nur 2,5% (angelehnt an: http://geneticssuite.net/files/Two%20human%20genomes_0.jpg).

Krankheitsverursachende (pathogene) Veränderungen sind derzeit in ca. 4900 Genen berichtet. Es wird geschätzt, dass sie zu ca. 85% in den Exons lokalisiert sind [1], so verändern sie die Proteinstruktur und -funktion. Ungefähr 15% der Mutationen betreffen regulatorische Abschnitte in nicht-kodierenden Genabschnitten, insbesondere die für das Spleißen notwendigen DNA-Erkennungsstellen an den Intron-Exon-Übergängen sowie Promotoren.

Die humangenetische Diagnostik ist derzeit auf Exons und zugehörige regulatorische Abschnitte, also das Exom, ausgerichtet, da dort die Mehrzahl der bekannten pathogenen Punktmutationen lokalisiert und die Interpretation von Veränderungen in diesen Bereichen einfacher ist.

1.2 Bisherige Methoden der humangenetischen Diagnostik

Während die Chromosomendiagnostik und die DNA-Chip-Analyse den Zugewinn oder Verlust nur größerer DNA-Abschnitte erfassen, werden in der molekularen humangenetischen Diagnostik

Veränderungen von Einzelbasenpaaren erfasst. Die bisher eingesetzten Methoden basieren meist auf einer Polymerasekettenreaktion (PCR) und umfassen entweder Tests, die auf eine spezifische Mutation ausgelegt sind (s. [2]), oder solche, die die Analyse der Einzelbasenpaarabfolge in einem ausgewählten Abschnitt erlauben.

In der molekularen Routinediagnostik hat sich zunächst die DNA-Sequenzierung nach Sanger durchgesetzt: Diese auch als Kettenabbruch- oder Didesoxymethode bezeichnete Technik stellt den Goldstandard zur Aufklärung einer DNA-Sequenz und zum Nachweis von Punktmutationen dar (s. [2; 3]). Die zu analysierende Sequenz wird i.d.R. auf der Basis eines spezifischen PCR-Produkts erzeugt. Je nach Sequenzierautomat können Sequenzen von 300 bis 1000 bp Länge analysiert werden. Selbst Geräte, die 96 Proben in einer Stunde parallel sequenzieren können, würden für ein gesamtes menschliches Exom, d.h. für alle 19.000 Gene, etwa 13 Tage benötigen. Jedes der Exons müsste mit mindestens einem spezifischen Primerpaar in Einzelansätzen mittels PCR amplifiziert und dann sequenziert werden, die Kosten würden sich auf mehrere Millionen Euro belaufen. Entsprechend aufwändiger ist die vollständige Sequenzierung des menschlichen Genoms: Tatsächlich kostete die 2001 abgeschlossene Sequenzierung der durch James Watson und Craig Venter mittels Sanger-Methode charakterisierten Genome mehrere Milliarden Dollar und dauerte mit einer Vielzahl von Sequenzierautomaten mehrere Jahre.

2 *Next Generation Sequencing* (NGS): Ein Hochdurchsatzverfahren

Mit der Entwicklung verschiedener Verfahren, die unter dem Begriff NGS zusammengefasst werden (vgl. Tab. 1), kann die bisherige Begrenzung der Sequenzierkapazität umgangen werden. Mittlerweile steht eine Vielzahl von NGS-Plattformen und -Verfahren zur Verfügung, denen allen die schnelle und kostengünstige Analyse einer großen Zahl von DNA-Fragmenten gemeinsam ist. Damit können ganze Genome innerhalb kurzer Zeit mit vertretbaren Kosten sequenziert werden. Dieser Quantensprung wird besonders deutlich an den rapide sinkenden Preisen für Sequenzierleistungen: Kosteten diese mittels Sanger-Sequenzierung im Jahre 2000 noch 3,7 Milliarden US-\$ pro Genom, so fiel der Preis mit den ersten NGS-Geräten im Jahre 2006 auf zehn Millionen US-\$, und aktuell betragen die Sequenzierkosten für ein Genom weniger als 1000 US-\$.

NGS wird mittlerweile für viele Fragestellungen in der Forschung eingesetzt, dabei wird nicht nur DNA analysiert, sondern über RNA-Analytik und weitere sogenannte „Omic“-Technologien können z.B. auch (indirekt) Genprodukte und Gen-Protein-Interaktionen untersucht werden (zur Übersicht: [4]). In der humangenetischen Diagnostik steht derzeit der Nachweis krankheitsverursachender Mutationen im Vordergrund (s. Fallbeispiel in Kasten 2), eine weitere zunehmend relevante Anwendung ist die nicht-invasive vorgeburtliche Diagnostik an zellfreier fetaler DNA aus mütterlichem Blut (s. Abschnitt 3).

	Ziel-Sequenz	Vorteile	Nachteile
<i>Whole Genome Sequencing</i>	Komplettes humanes Genom: 3,27 x 10 ⁹ bp	<ul style="list-style-type: none"> - das gesamte Genom wird erfasst, also auch die nicht-kodierenden Abschnitte - neue Gene und krankheitsassoziierte Sequenzen können identifiziert werden 	<ul style="list-style-type: none"> - Wahrscheinlichkeit für Zusatzbefunde ist sehr hoch - Wahrscheinlichkeit für Nachweis unklarer Ergebnisse ist sehr hoch - riesige Datenmengen müssen ausgewertet, interpretiert und gespeichert werden
<i>Whole Exome Sequencing</i>	Komplettes humanes Exom: ca. 19.000 Gene, 180.000 Exons, 30 x 10 ⁶ bp (1% des Genoms)	<ul style="list-style-type: none"> - alle proteincodierenden Bereiche des Genoms werden erfasst - auch neue Gene können identifiziert werden - besonders für Krankheiten mit unspezifischen klinischen Merkmalen geeignet 	<ul style="list-style-type: none"> - Wahrscheinlichkeit für Zusatzbefunde ist sehr hoch - Wahrscheinlichkeit für Nachweis unklarer Ergebnisse ist erhöht - nicht-kodierende Abschnitte werden i.d.R. nicht analysiert - große Datenmengen müssen ausgewertet und interpretiert werden
<i>Clinical Exome Sequencing</i>	Klinisches Exom: 4813 Gene, 62.000 Exons	<ul style="list-style-type: none"> - gezielte Charakterisierung von Genen, in denen krankheitsverursachende Mutationen bekannt sind - nicht nur krankheitsspezifische Gene, sondern auch solche von Krankheiten mit ähnlichen Merkmalen werden untersucht (Differentialdiagnose) - besonders für Krankheiten mit unspezifischen klinischen Merkmalen geeignet 	<ul style="list-style-type: none"> - Wahrscheinlichkeit für Zusatzbefunde ist erhöht - Wahrscheinlichkeit für Nachweis unklarer Ergebnisse ist erhöht - neue Gene/Mutationen werden nicht erfasst - nicht-kodierende Abschnitte werden i.d.R. nicht analysiert - große Datenmengen müssen ausgewertet und interpretiert werden
<i>Multigen-Panel</i>	Kodierende Exons ausgewählter Gene, die krankheits-spezifisch sind	<ul style="list-style-type: none"> - gezielte Charakterisierung von Genen, in denen Mutationen bei der spezifischen Erkrankung bekannt sind - geringe Wahrscheinlichkeit für Zusatzbefunde 	<ul style="list-style-type: none"> - kaum flexibel: sollen neue Gene eingeschlossen werden, muss das Panel angepasst werden - Mutationen in Genen, die zu ähnlichen klinischen Merkmalen führen (Differentialdiagnosen), werden nicht erfasst - nicht-kodierende Abschnitte werden i.d.R. nicht analysiert

Tab. 1: Übersicht über verschiedene NGS-Formate, die im Rahmen der humangenetischen Forschung und Diagnostik eingesetzt werden oder verwendet werden könnten (wichtige Begriffe siehe auch Glossar in Kasten 2).

2.1 Methodisches Vorgehen beim NGS

Beim NGS sind die Laborabläufe (2.1.1) und die Datenanalyse (2.1.2) zu unterscheiden, beide Phasen bestehen aus mehreren Schritten und nehmen jeweils mehrere Tage in Anspruch (Abb. 2).

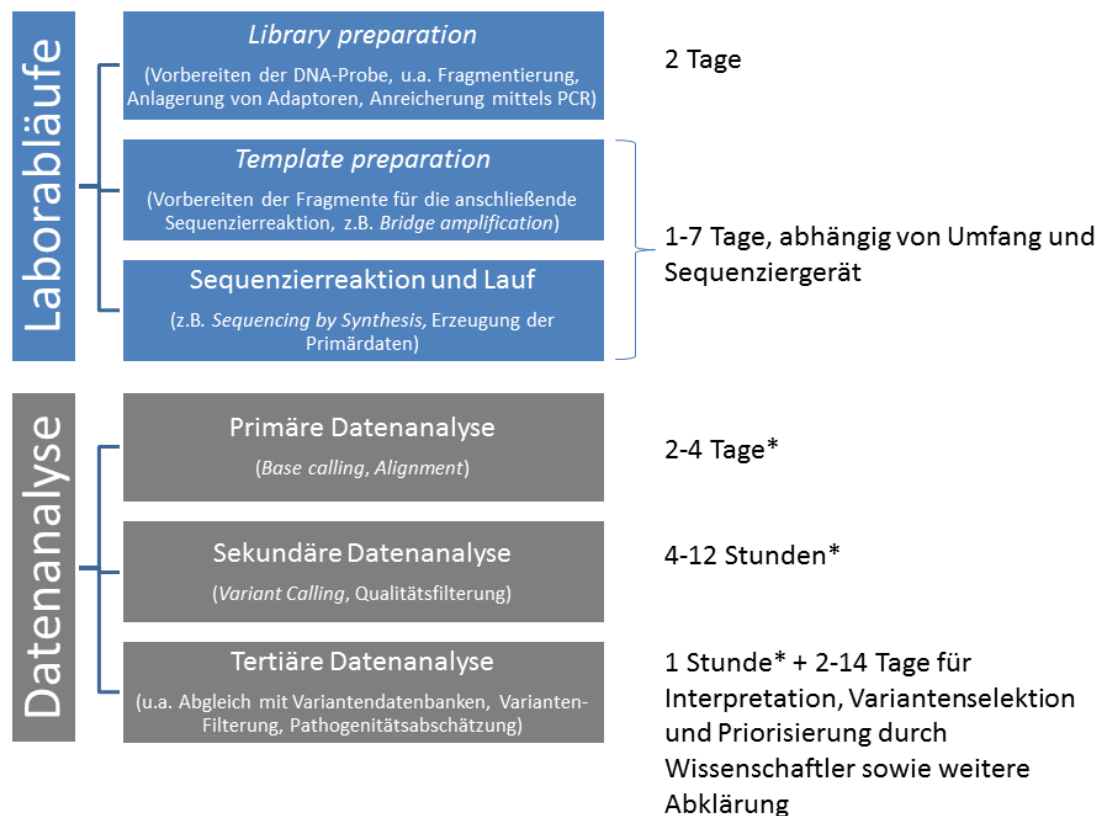


Abb. 2: Gesamtübersicht über die Labor- und Auswerteschritte beim *Next Generation Sequencing* am Beispiel eines Exoms. Die Darstellung ist exemplarisch: Im Labor können auch andere Anreicherungs- und Sequenzierstrategien verwendet werden; die Dauer der einzelnen Schritte bei der Datenanalyse ist jeweils abhängig von der Rechenkraft und von der Größe des Datensatzes (siehe *). Erläuterungen im Text (Abschnitt 2.1), wichtige Begriffe im Glossar (Kasten 1).

2.1.1 Laborabläufe

Die verfügbaren NGS-Strategien unterscheiden sich zwar in den verwendeten Reagenzien und Laborprotokollen (zur Übersicht: [5, 6]), umfassen aber in der Regel die gleichen wesentlichen drei Schritte, die hier beispielhaft skizziert werden (Abb. 2 und 3).

a) Vorbereitung der DNA-Proben (*Library Preparation*)

Die zu untersuchende DNA-Probe wird zunächst enzymatisch oder mechanisch in kleine Fragmente geschnitten (fragmentiert; Abb. 3a), wobei die Fragmentlänge von der entsprechenden Anwendung abhängt. Anschließend werden künstliche kurze DNA-Stücke an die DNA-Fragmente angehängt, die als Adaptoren bezeichnet werden. Sie setzen sich u.a. aus

Primersequenzen zusammen, die für das NGS als Startpunkte dienen. Weiterhin können sie individuelle Sequenzen, sogenannte Barcodes oder Indices (vgl. Kasten 2) beinhalten, die die gleichzeitige Analyse von DNAs verschiedener Personen erlauben (Abb. 3a).

Zur Anreicherung der DNA können u.a. PCR-basierte Verfahren eingesetzt werden, z.B. wenn ausgewählte Gene analysiert werden sollen (Multigen-Panel, vgl. Tab. 1). Dabei werden Tausende PCR-Reaktionen in einem Ansatz parallel durchgeführt. Je nach Zahl der Zielfragmente bietet es sich aber auch an, diese durch spezifische Sonden, bei denen es sich um Oligonukleotide („Oligos“) handelt, zu binden, um sie gezielt aus dem ursprünglichen Fragmentierungsgemisch anzureichern.

Liegt die DNA fragmentiert, mit Adaptoren versehen und spezifisch angereichert vor, spricht man von einer DNA-Bibliothek (*Library*).

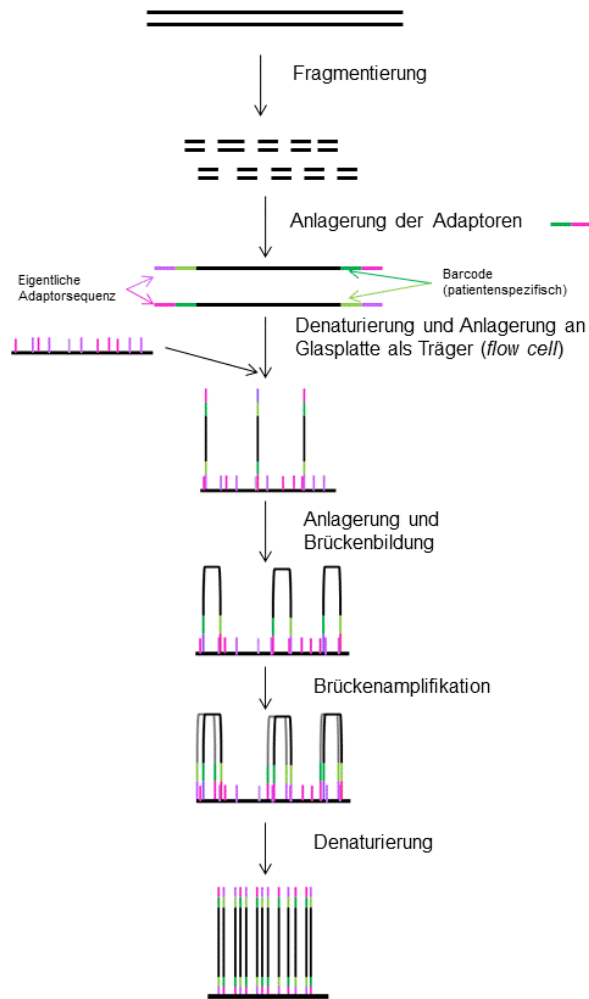
b) Vorbereitung der Sequenzierproben (*Template preparation*)

Zur Vorbereitung der erzeugten *DNA Library* für die anschließende Sequenzierreaktion werden einzelsträngige Vorlagen in großer Zahl generiert. Von den derzeit verfügbaren Methoden wird hier exemplarisch die sogenannte Brückenamplifikation (*Bridge Amplifikation*, Abb. 3a) beschrieben. Die geschnittenen DNA-Fragmente werden denaturiert und an eine Glasplatte, eine sogenannte Durchflusszelle (*Flow cell* oder *Flow chip*, vgl. Kasten 2) angelagert, indem einer der beiden an die DNA-Fragmente angebrachten Adaptoren an komplementäre kurze DNA-Sonden auf der Glasplatte bindet (Abb. 3a). Dieses Oligonukleotid dient dann als Primer für die anschließende Amplifikation, es entsteht ein neuer Gegenstrang. Der Ursprungsstrang wird durch Denaturierung entfernt. Der Neustrang stellt nun die Verlängerung des an die *Flow cell* gebundenen Oligos dar. Sein freies Ende kann nun ebenfalls wieder an ein komplementäres Oligonukleotid auf der *Flow cell* binden. Wie zuvor dient dieses als Primer und wird verlängert, es kommt zur Brückenbildung (Abb. 3a). Das entstandene doppelsträngige PCR-Produkt wird denaturiert, und diese Schritte werden zyklisch wiederholt. Durch die fortlaufende Brückenamplifikation der zuvor gebundenen Einzelstränge entstehen nach und nach sogenannte Cluster aus DNA-Fragmenten gleicher Sequenz, die schließlich als Vorlage für die eigentliche NGS-Reaktion dienen. Der beschriebene Vorgang wird als Clustergenerierung (*Cluster generation*) bezeichnet.

c) Sequenzierreaktion und Lauf

Die Sequenzierreaktion kann mittels verschiedener Methoden erfolgen, im vorgestellten Laborablauf ist dies das sogenannte *Sequencing by Synthesis* (Abb. 3b; Kasten 2): Der gegen einen Teil der Adaptorsequenz gerichtete Primer lagert sich an das an die *Flow cell* gebundene einzelsträngige DNA-Fragment an, anschließend erfolgt wie bei einer PCR der Einbau des komplementären Nukleotids, wobei jedes der vier Nukleotide mit einem individuellen Fluoreszenzfarbstoff markiert ist (Abb. 3b).

a) Library & Template preparation



b) Sequenzierreaktion: Sequencing by Synthesis

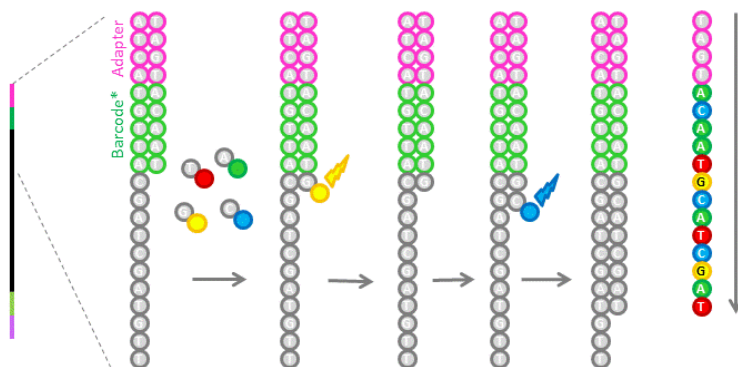


Abb. 3: Laborablauf (vgl. Abb. 2) beim *Next Generation Sequencing* am Beispiel der Brückenamplifikation (siehe a) und des *Sequencing by Synthesis* (siehe b). Es sei darauf verwiesen, dass auch andere Strategien zur Verfügung stehen. (* Der Barcode wird vom Primer/Adapter ausgehend mit sequenziert.)

Je nach eingebautem Nukleotid wird der jeweilige Farbstoff zur Fluoreszenz angeregt und das Lichtsignal vom Gerät detektiert. Danach wird das Fluoreszenzsignal abgespalten, und das nächste Nukleotid kann eingebaut werden. Diese Vorgänge werden über viele Zyklen wiederholt.

Beim beschriebenen Prinzip spricht man von sogenannten „reversiblen Terminatoren“. Sie erlauben einen Einbau von aufeinanderfolgenden fluoreszenzmarkierten Nukleotiden in den DNA-Strang. Hier besteht der grundsätzliche Unterschied zur klassischen Sanger-Sequenzierung, bei der ein Einbau eines fluoreszenzmarkierten Nukleotids zum Kettenabbruch führt.

Die Abfolge der Lichtsignale wird digital gespeichert und dient als Grundlage für die weitere (bio)informatische Analyse.

2.1.2 Datenanalyse

Die von der NGS-Plattform erzeugten digitalen Daten stellen die Primärdaten dar (Abb. 2), die in eine Nukleotidabfolge übersetzt werden, indem eine Software die auf der *Flow cell* pro Zyklus detektierten Lichtsignale ermittelt und einem Nukleotid zuordnet. Man bezeichnet dies als *Base calling* (Kasten 2). Da die DNA-Cluster und damit die Fluoreszenzsignale teilweise sehr eng bei einander liegen, kann es zu Überlagerungen von Lichtsignalen und damit zu einer schlechteren Auswertbarkeit kommen. Dies ist ein Qualitätsparameter beim NGS (vgl. Abschnitt 2.2.1).

Ergebnis des *Base calling* ist eine Abfolge von Nukleotiden, wobei für jedes einzelne Nukleotid die Angabe erfolgt, mit welcher Wahrscheinlichkeit dieses tatsächlich vorliegt. Die gesamte Nukleotidabfolge eines Fragments heißt *Read* (Kasten 2).

Nach digitaler Erzeugung der *Reads* ist zunächst unklar, welchen Teilen des Genoms sie jeweils entsprechen. Im Rahmen der **primären Datenanalyse** (Abb. 2) werden die einzelnen *Reads* mit Hilfe von Computerprogrammen ihrer Stelle im humanen Referenzgenom, welches online unter <https://www.genome.gov/> öffentlich zugänglich ist, zugeordnet. Man nennt dies *Alignment*.

Da der gleiche DNA-Abschnitt aus hunderten oder tausenden Zellen vorliegt, ist das gleiche Nukleotid in hunderten bis tausenden von *Reads* vorhanden. Man spricht hier von Sequenziertiefe oder *Coverage* (Kasten 2). Dieses Potential des NGS wird auch als massives paralleles Sequenzieren oder Ultratief-Sequenzieren (*Ultradeep sequencing*, auch *Deep sequencing*, Kasten 2) bezeichnet und ist insbesondere für den Nachweis von geringen Anteilen veränderter DNA in einer Probe wichtig, z.B. bei der nicht-invasiven pränatalen Testung (NIPT, s. Abschnitt 3).

Im Rahmen der **sekundären Datenanalyse** (Abb. 2) wird ermittelt, an welchen Positionen sich die untersuchte DNA vom Referenzgenom unterscheidet. Bei diesem *Variant calling* müssen Sequenzierartefakte herausgefiltert werden, die z.B. anhand der geringen Qualität eines Nukleotidsignals identifiziert werden können. Verschiedene Programme und Algorithmen stehen auch hier zur Verfügung. Nach dieser Qualitätsfilterung bleiben bei Analyse eines Exoms bis zu

40.000 Varianten übrig, die dann im Hinblick auf ihre Pathogenität, d.h. auf einen Zusammenhang mit den klinischen Merkmalen des Patienten, untersucht werden müssen. Bei dieser **tertiären Datenanalyse** werden die Varianten eines Patienten mit Variantendatenbanken - *Single nucleotide polymorphism* (SNP)-Datenbanken - verglichen. Auf diese Weise werden häufige, pathogenetisch nicht relevante Varianten herausgefiltert (Abb. 2 und 4; vgl. Abschnitt 2.2.1). Nach dieser Filterung bleiben nur noch ca. 2% der Varianten übrig. Die weitere Filterung richtet sich nach der Plausibilität, also nach der Frage, ob die Mutation ein klinisches Bild erklärt (vgl. Abb. 3 sowie Abschnitt 2.2).

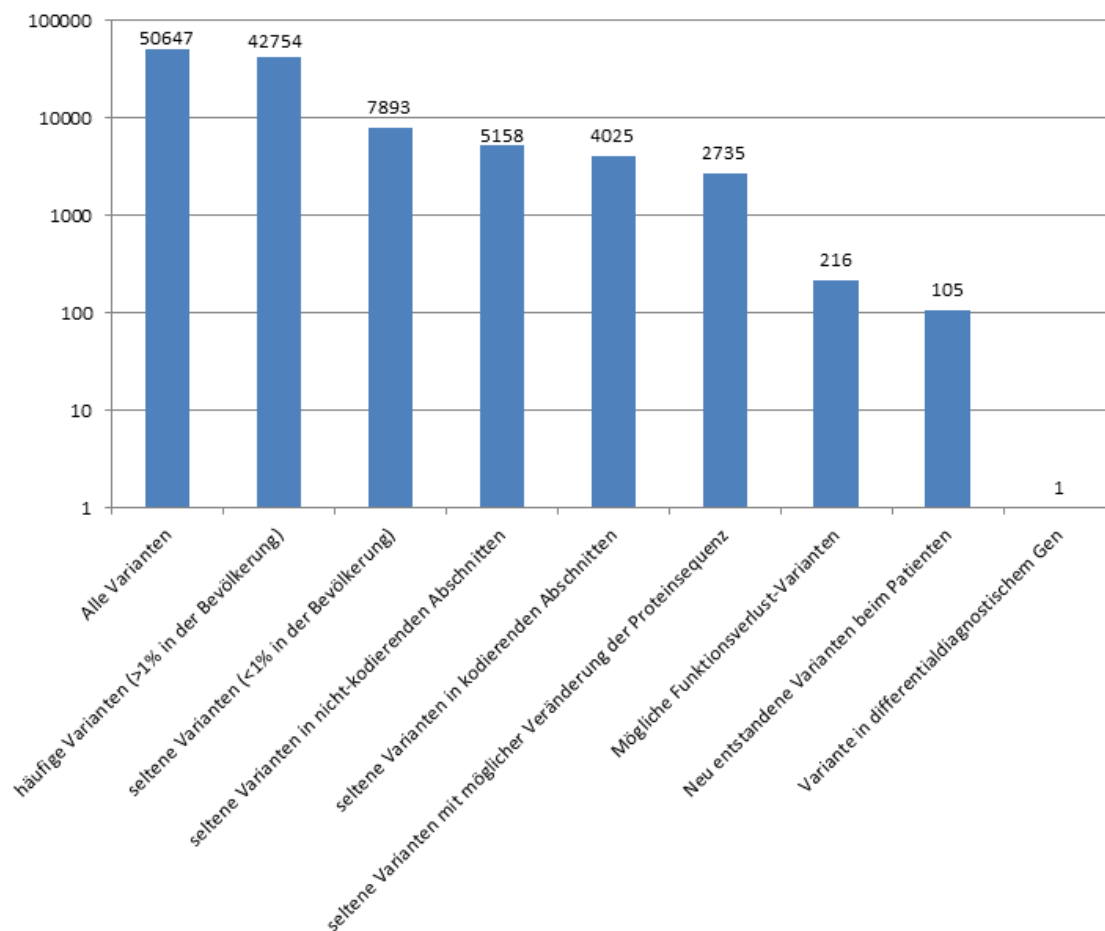


Abb. 4: Beispielhafte Darstellung für Variantenzahlen nach einer *Whole Exome Sequencing*-Analyse (vgl. Tab. 1) und verschiedenen bioinformatischen Filterungen.

Kriterien bei dieser Pathogenitätsabschätzung sind unter anderem:

- der funktionelle Schweregrad der Variante: Funktionsverlust-Mutationen (*Loss-of-function*-Mutationen, z.B. Leseraster- oder Stopmutationen) werden i.d.R. von vorneherein als schwerwiegender eingestuft als Varianten, die funktionell gleichartige Aminosäuren austauschen (neutrale Mutationen) oder die Proteinsequenz gar nicht verändern (stille Mutationen). Von den durchschnittlich 250 bis 300 Funktionsverlust-Mutationen, die jedes

Individuum trägt, finden sich 50 bis 100 Varianten in Genen, für die ein Zusammenhang mit einem Krankheitsbild bekannt ist. Für die funktionelle Charakterisierung von Varianten stehen zahlreiche digitale Werkzeuge zur Verfügung [7].

- die Rolle des betroffenen Gens im Stoffwechsel: Es ist zu prüfen, ob das Gen zu einem bestimmten Merkmal beiträgt (z.B. Wachstumsgen: Mutationen führen möglicherweise zum Kleinwuchs) und/oder ob sogar Mutationen in diesem Gen als Ursache der Erkrankung bekannt sind.
- der zugrundeliegende Erbgang: Sind mehrere Geschwister betroffen, aber beide Eltern gesund, spricht dies für einen autosomal-rezessiven Erbgang, elterliche Blutsverwandtschaft ist hierfür ein weiterer Hinweis. Beim autosomal-dominanten Erbgang sollte ein Elternteil betroffen sein, oder man geht von einer Neumutation (*de novo*) aus. Häufig kann hier die Analyse nicht nur des Patienten, sondern auch seiner Eltern und Geschwister hilfreich sein. Bei jedem Menschen finden sich 30 bis 40 neu entstandene Varianten (*de novo variants*) [8], die nicht bei den Eltern nachweisbar sind.

<i>Barcoding</i>	Individuelle Markierung von vorbereiteten Proben verschiedener Patienten für den NGS-Lauf, so dass diese zwar gemeinsam auf dem Gerät analysiert, aber dann getrennt ausgewertet werden können (anderes Wort: <i>Indexing</i>).
<i>Base calling</i>	Computerbasierter Prozess, bei dem die im NGS-Lauf erzeugten Primärdaten in Form digitaler Bilder in eine Nukleotidfolge umgewandelt werden.
<i>Coverage</i>	Abdeckung eines Fragments bei einer gewählten Sequenziertiefe (in %).
<i>Deep sequencing</i>	NGS-Anwendung, bei der die Zahl der <i>Reads</i> wesentlich höher ist als die Länge der zu untersuchenden Sequenz.
<i>Flow cell / Flow chip</i>	Glasträger, auf dem die zu sequenzierenden DNA-Fragmente durch Bindung an komplementäre Adaptoren immobilisiert werden.
<i>Indexing</i>	Anderes Wort für <i>Barcoding</i> .
<i>Read</i>	Nukleotidsequenz, die beim NGS erzeugt wird.
<i>Sequencing by Synthesis</i>	Methode des NGS, bei dem die Sequenzinformation während der Neusynthese eines komplementären Stranges ausgelesen wird.

Kasten 1 : Glossar von Begriffen, die im Zusammenhang mit *Next Generation Sequencing* häufig verwendet werden. Weitere Begriffe, die sich auf verschiedene NGS-Formate beziehen, sind in Tab. 1 aufgeführt.

2.2 Herausforderungen der Methode

2.2.1 Humangenetische Forschung

Während die Laborabläufe (*Wet Lab*) mittlerweile schon gut etabliert sind, ist die Erzeugung, Speicherung, Auswertung und Interpretation der NGS-Daten (derzeit noch) mit großen Herausforderungen verbunden, welche insbesondere bioinformatische und auch finanzielle Aspekte betreffen.

Die Anschaffungskosten für eine NGS-Plattform liegen derzeit zwischen ca. 100.000 € und 500.000 €. Welches Gerät gewählt wird, hängt v.a. vom Probenaufkommen und von der Fragestellung ab: Ein kleineres Gerät ist für die Untersuchung eines klinischen Exoms (vgl. Tab. 1) ausreichend, dagegen wäre ein vollständiges humanes Genom damit nur sehr teuer und zeitaufwändig zu analysieren. Bei Großgeräten ist der Preis pro untersuchter Base wesentlich günstiger, lohnt sich aber erst bei einem hohen Probenaufkommen.

Die Masse der Primärdaten, der Abgleich mit Referenzsequenzen sowie die Speicherung der digitalen Daten (*Big Data*) erfordert immense IT-Kapazitäten [9]. Mit der Menge der erzeugten Daten steigen die Kosten für deren Berechnung und Archivierung, dies kann den Anschaffungspreis eines NGS-Gerätes um ein Vielfaches übersteigen.

Das Nadelöhr bei der Durchführung von NGS-Experimenten ist derzeit noch die Interpretation und Auswertung. Hierbei muss man sich vor Augen führen, dass die Menge der auswertbaren Daten exponentiell zunimmt: Wertete man früher pro Patient nur einzelne Gene aus, so können es mittels NGS Hunderte bis Tausende Gene werden. Diesem Quantensprung trägt die im Vergleich zur Sanger-Sequenzierung veränderte Auswertung Rechnung, bei der nicht mehr Base für Base im Elektropherogramm betrachtet wird, sondern eine Liste von Varianten unter anderem über Parameter wie *Coverage* und Anzahl der *Reads* (vgl. Kasten 1) beurteilt wird.

Von besonderer Bedeutung ist die Qualitätskontrolle der Primärdaten (vgl. Abschnitt 2.1.2): Sie ist einerseits Voraussetzung für die Sicherheit der Ergebnisse, aber auch Grundlage für die Vergleichbarkeit von Ergebnissen über einen längeren Zeitraum und zwischen Laboratorien.

Für die klinisch-genetische Interpretation ist die hohe genetische Variabilität des menschlichen Genoms eine erhebliche Herausforderung (vgl. Abschnitt 2.1.2): Hier muss der Humangenetiker entscheiden, ob und welche genetische Variante tatsächlich ursächlich für eine Erkrankung oder ohne klinische Bedeutung für den Träger ist [10]. Um das Ausmaß der natürlichen Variabilität des Menschen zu ermitteln, wurden in den letzten Jahren eine Vielzahl von Studien durchgeführt und die Daten in öffentliche Datenbanken eingespeist. Eines dieser Projekte ist das 1000-Genome-Projekt [11], ein internationales Projekt, das im Oktober 2012 die Erbgutdaten von mehr als 1.000 Menschen veröffentlicht hat. Dabei wurden u.a. etwa 15 Millionen SNPs berichtet. Solche Einzelnukleotid-Varianten, aber auch Veränderungen ganzer DNA-Blöcke sind jedoch nicht

unbedingt krankheitsverursachend; es gibt viele Beispiele für nicht-pathogene Veränderungen in proteinkodierenden Exons, die sogar die Aminosäureabfolge eines Proteins verändern, dessen Funktion aber nicht beeinflussen (vgl. Abschnitt 2.1.4).

2.2.2 Humangenetische Diagnostik und Beratung

Es ist zu erwarten, dass mit der rasanten Zunahme an genetischen Daten und Wissen viele der beschriebenen technischen und bioinformatischen Probleme gelöst werden; Erfahrungen mit früheren Screeningtests (z.B. DNA-Chips [3]) legen dies nahe. Immer mehr an Bedeutung gewinnt nun die Interpretation und Übermittlung der Befunde durch den Humangenetiker.

Zunächst muss klar sein, dass bei weitem nicht für alle genetisch bedingten Erkrankungen die Ursache molekular geklärt werden kann: Mit den derzeitigen bioinformatischen Werkzeugen sind nicht alle Arten von Veränderungen nachweisbar, z.B. werden Deletionen oder Duplikationen je nach Größe noch nicht verlässlich erfasst.

Außerdem beschränkt sich die Diagnostik i.d.R. auf Bereiche des Genoms mit bekannter Funktion, also auf das klinische Exom, das mittels *Clinical exome sequencing* untersucht wird (vgl. Tab. 1), oder auf bestimmte Gene, in denen Mutationen als krankheitsverursachend bekannt sind. Die hier eingesetzten Multigen-Panels (vgl. Tab. 1) sind für bestimmte Erkrankungen oder Erkrankungsgruppen spezifisch (z.B. Epilepsie-Panel).

Hierin unterscheidet sich die diagnostische Anwendung des NGS von der in der Forschung, wo es gerade darum geht, durch die Sequenzierung ganzer Exome oder sogar Genome neue krankheitsrelevante Gene und Mutationen zu identifizieren. So könnten in der Diagnostik z.B. Patienten mit einer bestimmten Kleinwuchsform auf Mutationen im *IGF2*-Gen untersucht werden, da für dieses Gen Mutationen bekannt sind. Dagegen sind Untersuchungen des Gens für den zellulären Rezeptor von IGF2 (*IGF2R*) derzeit nicht sinnvoll, da Mutationen in *IGF2R* zwar funktionell zu Kleinwuchs führen könnten, aber noch nicht beschrieben wurden.

Zur wesentlichen Aufgabe eines Humangenetikers gehört es, die Pathogenität einer Variante, insbesondere von *Single nucleotide polymorphisms* (SNPs, auch *Single nucleotide variants* bzw. SNVs genannt), abzuschätzen, also zu ermitteln, ob eine identifizierte Variante tatsächlich die Erkrankung eines Patienten erklärt oder nicht (Plausibilitätsprüfung oder Pathogenitätsabschätzung, vgl. Abschnitt 2.1.2). Zwar besteht die Möglichkeit, diese Veränderungen nach möglichem Schweregrad einzuteilen [12], dies hilft aber nicht immer weiter, und die Diagnose bleibt dann unklar.

Einen Hinweis kann die Häufigkeit einer Variante geben: Wird sie in mehr als 1% der Normalbevölkerung beobachtet, handelt es sich bei einer seltenen Erkrankung des Patienten meist um eine apathogene Veränderung. Allerdings gibt es einzelne krankheitsverursachende Varianten, die häufiger sind als 1%, umgekehrt sind auch die meisten seltenen SNPs (und CNVs,

Copy number variations) apathogen. Die Häufigkeit einer Variante ist auch abhängig von der ethnischen Herkunft einer Person: Während z.B. die Mutation p.Phe508del („ $\Delta F508$ “) im CFTR-Gen (Ursache der autosomal-rezessiv erblichen Mukoviszidose) mit einem Anteil von 4-5% in der deutschen Bevölkerung relativ häufig ist, kommt sie in Finnland viel seltener vor [2].

Zur Pathogenitätsbewertung einer Variante wird neben ihrer Häufigkeit auch ihre evolutionäre Konservierung berücksichtigt, d.h., ob im gleichen Gen verschiedener Taxa (z.B. einzelner Tierarten oder ganzer Tierstämme) an einer bestimmten Stelle immer die gleiche Base auftritt, oder ob es Variationen gibt. Bioinformatische Programme, die die genannten Parameter berücksichtigen, helfen bei der Plausibilitätsabschätzung.

Ein wichtiger Aspekt in der humangenetischen Diagnostik und Beratung sind Zusatzbefunde: Da man in vielen Genen nach Veränderungen schaut, erhöht sich nicht nur die Wahrscheinlichkeit, die ursächliche Mutation für eine diagnostizierte Erkrankung zu finden, sondern auch die Möglichkeit, Ergebnisse zu erhalten, die nicht mit der ursprünglichen Fragestellung in Verbindung stehen, aber dennoch Bedeutung für die untersuchte Person und/oder ihre Verwandten haben können. Daher ist eine umfassende Aufklärung über „Wesen, Bedeutung und Tragweite“ von Zusatzbefunden im Vorfeld von humangenetischen Untersuchungen vorgesehen [13], aufgrund des Umfangs der Testung aber oft kaum umsetzbar. Es wäre insbesondere bei Exom- und Genom-Untersuchungen nahezu unmöglich, die Betroffenen über alle denkbaren Zusatzergebnisse im Vorfeld einer Untersuchung aufzuklären. Dieses Problem und mögliche Lösungen werden in Fachkreisen derzeit intensiv diskutiert [14]. Das stets geltende Recht auf Nicht-Wissen muss auch und insbesondere im Zusammenhang mit der möglichen Erhebung von Zusatzbefunden gewahrt werden.

Mit der Entwicklung und dem Einsatz von NGS-basierten Tests ist man aber auch dem Ziel einer personalisierten Medizin erheblich näher gekommen. In der personalisierten Medizin soll jeder Patient die Therapie erhalten, die für ihn spezifisch angepasst ist. Das betrifft z.B. die individuelle Reaktion auf Arzneimittel: So können genetische Veranlagungen bedingen, dass ein Narkosemittel vertragen wird oder Nebenwirkungen hervorruft, oder ob ein Medikament wirkt oder nicht, neben Brustkrebs oder Leukämien z.B. auch bei Mukoviszidose (Cystische Fibrose, CF): Hier wird derzeit untersucht, ob die Substanz Ataluren bei Patienten, die die Mutation p.G542X im *CFTR*-Gen tragen, zur Behandlung eingesetzt werden kann (zur Übersicht: [15]). In Zellkulturrexperimenten wurde nachgewiesen, dass Ataluren diese Stopmutation unterdrücken kann, so dass wieder ein funktionelles Protein entsteht.

Kritisch zu betrachten ist die Kommerzialisierung von Gentests [16]: Immer mehr Firmen bieten international frei verkäufliche und profitorientierte Gentests direkt an (*direct-to-consumer genetic testing*, DCT), ohne dass möglicherweise eine klinische Notwendigkeit besteht und ohne dass eine kompetente Aufklärung und Ergebniserläuterung erfolgen [17]. In diesem Zusammenhang sei auch auf die Berichterstattung in den Medien verwiesen, die einerseits auf den ersten Blick

aufklärerisch und um Differenzierung bemüht erscheint, andererseits aber auch unkritisch oder oberflächlich sein kann.

Wissenschaftlich fundierte Informationen lassen sich im Internet ebenso leicht finden wie populärwissenschaftliche, journalistische, von Laien verfasste oder kommerzielle Quellen. Titel wie „An agenda for personalized medicine“ (<https://www.nature.com/articles/461724a>) und „Growth of genome screening needs debate“ (<https://www.nature.com/articles/476027a>) zeigen die Beschäftigung der internationalen Wissenschaftlergemeinschaft mit den Herausforderungen von und den Anforderungen an das NGS. Auf der Suche nach seriösen und aktuellen Informationen zum Stand der Forschung und zur Gesetzeslage auch in Deutschland und insbesondere nach Beiträgen von Fachleuten zur gesellschaftlichen Diskussion, auch in Form von Positionspapieren, sei außerdem die Homepage der Deutschen Gesellschaft für Humangenetik unbedingt empfohlen (<https://www.gfhev.de>).

3 Spezifische Anwendungen des NGS

3.1 Nicht-invasive Pränatale Testung (NIPT)

Da mit zunehmendem mütterlichen Alter die Wahrscheinlichkeit für ein Kind mit Trisomie 21, aber auch mit anderen numerischen Chromosomenstörungen (Trisomien der Chromosomen 13 und 18) steigt, wird in der Bundesrepublik Deutschland insbesondere Schwangeren ab 35 Jahre eine vorgeburtlich (pränatale) Testung angeboten. Diese besteht derzeit i.d.R. in der mikroskopischen Chromosomendarstellung aus Fruchtwasser (Amniozentese). Nachteile dieser invasiven Methode sind vor allem das Risiko, dass es durch den Eingriff selbst zu einer Fehlgeburt kommen kann, und der späte Zeitpunkt der Diagnose, falls Konsequenzen wie ein Schwangerschaftsabbruch gezogen werden sollten.

Aus diesem Grunde wurde in der Vergangenheit das nicht-invasive Ersttrimester-Screening entwickelt. Dieses leitet aus Ultraschallbefunden (z.B. Nackenfaltentransparenz) und Hormonwerten im Blut der Schwangeren eine Wahrscheinlichkeit für das Vorliegen z.B. einer Trisomie beim Fet ab. Eine sichere Aussage ist aber nur über weitere Methoden (Amniozentese, weiterer Ultraschall) möglich. Der Anteil falsch-positiver Testergebnisse liegt bei 3-5%.

In den letzten Jahren findet nun auch zunehmend das NGS Anwendung in der pränatalen Diagnostik (vgl. Tab. 2). Seine hohe Sensitivität ermöglicht es, zellfreie fetale DNA, die aus abgestorbenen Zellen des Feten bzw. der Plazenta stammt und aus dem venösen Blut der Schwangeren gewonnen wird, molekular zu charakterisieren. Derzeit können Trisomien, zukünftig auch Einzelgen-Mutationen bestimmt werden. Insbesondere das massive Parallelsequenzieren durch NGS erlaubt es, einzelne fetale Moleküle vor dem Hintergrund mütterlicher DNA zu identifizieren. Verschiedene Strategien sind hierfür entwickelt worden (zur

Übersicht: [16]). So kann mit Hilfe von SNPs zwischen mütterlichem und fetalem Genom unterschieden und auch die relative Menge fetaler DNA bestimmt werden.

Jahr	Entwicklung
1997	Nachweis zellfreier DNA eines Feten im Blut einer Schwangeren
1999	Nachweis, dass sich diese zellfreie fetale DNA nach der Geburt abbaut
2000	Nachweis zellfreier fetaler RNA im mütterlichen Blut
2004	Nachweis, dass zellfreie DNA von Mutter und Fet unterschiedliche Länge haben
2007	Entwicklung eines PCR-Ansatzes zum Trisomie 21-Nachweis an mütterlichem Blut
2008	Entwicklung von NGS-Tests zum allg. Trisomie-Nachweis
2010	Sequenzierung eines gesamten fetalen Genoms aus zellfreier DNA
2011	Einführung des NGS-basierten NIPT in die Routineanwendung

Tab. 2: Entdeckungen und Entwicklungen, die zur nicht-invasiven pränatalen Testung (NIPT) und deren Anwendung beigetragen haben.

Die Vorteile des NIPT im Vergleich zu invasiven Verfahren der Pränataldiagnostik wie einer Fruchtwasseranalyse sind der frühe Zeitpunkt der Analyse (eine NIPT kann bereits vor der 10. Schwangerschaftswoche erfolgen) und das Fehlen des Risikos, durch den Eingriff selbst eine Fehlgeburt auszulösen. Die werdende Mutter hat damit die Möglichkeit, zu einem frühen Zeitpunkt der Schwangerschaft Informationen über eine mögliche angeborene Behinderung ihres Kindes zu erhalten und selbstbestimmt eine Entscheidung zu treffen.

Demgegenüber stehen insbesondere die ethischen Aspekte des NIPT, da das Verfahren zu einem sehr frühen Zeitpunkt der Schwangerschaft prinzipiell eine sehr umfassende Aussage zu den genetischen Merkmalen des ungeborenen Kindes erlaubt: Für welche angeborenen oder erblichen Erkrankungen soll eine NIPT angeboten werden? Wie erfolgt die Aufklärung der Ratsuchenden? Wie ist der Umgang mit Zusatzbefunden (vgl. Abschnitt 2.2.2)? Wie wird mit der Anlageträgerschaft für autosomal-rezessive Erkrankungen oder autosomal-dominante Erkrankungen umgegangen, die erst später im Leben auftreten (z.B. Brustkrebs)? Wie kann ein Missbrauch verhindert werden (z.B. frühzeitige Geschlechtsbestimmung)?

Eine Befürchtung bei Einführung der NIPT besteht weiterhin in der möglichen gesellschaftlichen Neubewertung von Behinderungen und gesundheitlichen Störungen sowie in der Möglichkeit, dass Druck auf werdende Mütter ausgeübt wird, eine NIPT durchführen zu lassen. („Das kann man doch vorher testen.“, „Das muss doch heute nicht mehr sein.“)

Vor diesem Hintergrund sollte man sich die Häufigkeit angeborener Behinderungen und den Anteil genetischer Ursachen bewusst machen: 3-5% aller Neugeborenen weisen eine Behinderung auf, aber in nur 0,8-1% der Fälle ist eine Chromosomen- oder Genveränderung die Ursache. Damit hat der größte Teil der Behinderungen des Menschen nicht genetische Ursachen und entzieht sich der Erfassung durch die NIPT.

3.2 Klinisches Fallbeispiel einer seltenen Erkrankung

Der zehnjährige Nils (Name geändert) wird zur Abklärung der möglicherweise genetischen Ursache seiner klinischen Merkmale in der genetischen Beratung vorgestellt. Er ist derzeit 127 cm groß und wiegt 30 kg, damit ist er wesentlich kleiner als gleichaltrige Jungen. Bereits seine Geburtsmaße lagen weit unter dem Durchschnitt. Daneben zeigt er eine große Stirn, ein leicht dreieckiges Gesicht. Seine Schneidezähne sind auffällig groß, einige andere Zähne sind dafür gar nicht ausgebildet. Er sieht und hört schlecht und hat einen Herzfehler. Seine geistige Entwicklung ist leicht verzögert. Er wurde schon in mehreren Krankenhäusern und Instituten vorgestellt. Aufgrund der klinischen Merkmale wurden verschiedene Syndrome vorgeschlagen (z.B. Noonan-Syndrom, Silver-Russell-Syndrom).

Diese Odyssee hat Nils und seine Familie immer wieder fragen lassen, was er hat, wie die Prognose ist, ob es nicht andere Kinder mit ähnlichen Symptomen und auch gleichen Erfahrungen gibt. Bei dem aktuellen Besuch in der humangenetischen Beratung wird die neue Möglichkeit angesprochen, mittels NGS an einer Blutprobe von Nils alle Exons im Hinblick auf eine ursächliche Mutation hin zu untersuchen. Die Familie erklärt sich hiermit einverstanden. Das NGS des klinischen Exoms ergab, dass Nils heterozygoter Träger einer zwei Basenpaare-Deletion in Exon 10 des *ANKRD11*-Gens ist. Mutationen in diesem Gen werden autosomal-dominant vererbt, sie führen zum sog. KBG-Syndrom. Dabei handelt es sich um ein seltenes Kleinwuchssyndrom, für das bisher nur ca. 20 Patienten in der Literatur beschrieben wurden. Damit gehört es zur Gruppe der seltenen Erkrankungen.

Diese Fallbeschreibung zeigt die Probleme für Familien, die von einer solchen seltenen Erkrankung betroffen sind: Da diese Erkrankungen selten sind, sind sie meist unbekannt und werden nicht oder erst sehr spät diagnostiziert. Auch gerichtete Behandlungen sind nur für einzelne der Erkrankungen bekannt, da u.a. die Erfahrung hierzu fehlt. Die Familie von Nils wurde in einer abschließenden genetischen Beratung über das Ergebnis des NGS informiert, dabei wurde auch die Bedeutung für die potentielle Familienplanung von Nils und seinen Verwandten erklärt. Da das KBG-Syndrom autosomal-dominant vererbt wird, hätten Kinder von Nils eine Wahrscheinlichkeit von 50%, ebenfalls an diesem Syndrom zu erkranken. Für die Eltern und Geschwister von Nils besteht dagegen keine erhöhte Wahrscheinlichkeit, da die Veränderung bei Nils neu entstanden war.

Anschrift der Verfasser

Prof. Dr. rer. nat. Thomas Eggermann, Dr. rer. nat. Lukas Söllner, Prof. Dr. med. Ingo Kurth, Dr. rer. nat. Katja Eggermann, Dr. rer. nat. Matthias Begemann; Institut für Humangenetik, Pauwelsstr. 30, D-52074 Aachen, Tel.:0241 8080178, Fax: 0241 8082394, Email (für die Verfasser): teggermann@ukaachen.de

Literatur

- [1] Majewski, J., Schwartzentruber, J., Lalonde, E., Montpetit, A. and Jabado, N. (2011): What can exom sequencing do your your? Journal of Medical Genetics: 48: S. 580-589; online: <https://www.ncbi.nlm.nih.gov/pubmed/21730106> (13.12.2018)
- [2] Eggermann, K., Deerberg, J., Mehta, S. und Eggermann, T. (2015): PCR und Sequenzierung in der humangenetischen Diagnostik. PdN-BioS 2/64: S. 33-38.
- [3] Eggermann, T. und Heil, I. (2013): Hochdurchsatztechnologien in der Genomanalyse. PdN-BioS 6/62: S. 16-23.
- [4] Buermans, H.P., den Dunnen, J.T. (2014): Next generation sequencing technology: Advances and applications. Biochimica et Biophysica Acta 1842: S. 1932-1941; online: <https://www.ncbi.nlm.nih.gov/pubmed/24995601> (13.12.2018)
- [5] Neveling, K. und Hoischen, A. (2014): Einführung in die Grundlagen der Hochdurchsatzsequenzierung. Medizinische Genetik 26: S. 231-238.
- [6] Mardis, E.R. (2008): Next-generation DNA sequencing methods. Annual Review Genomics and Human Genetics 9: S. 387-402; online: <https://www.ncbi.nlm.nih.gov/pubmed/18576944> (13.12.2018)
- [7] Weißmann, R. und Gilissen, G. (2014): NGS-Datenanalyse und Qualitätskontrolle. Medizinische Genetik 26: S. 239-245.
- [8] Francioli, L.C., Polak, P.P., Koren, A., Menelaon, A., Chun, S., Renkens, I., Genome of the Netherlands Consortium, van Duijn, C.M., Swertz, M., Wijmenga, C., van Ommen, G., Slagboom, P.E., Boomsma, D.I., Ye, K., Guryev, V., Arndt, P.F., Kloosterman, W.P., de Bakker, P.I. and Sunyaev, S.R. (2015): Genome-wide patterns and properties of de novo mutations in humans. Nature Genetics 47: S. 822-826; online: <https://www.ncbi.nlm.nih.gov/pubmed/25985141> (13.12.2018)
- [9] Mardis, E. R. (2016): The challenges of big data. Disease Models and Mechanisms 9: S. 483-5; online: <https://www.ncbi.nlm.nih.gov/pubmed/27147249> (13.12.2018)
- [10] Stankiewicz, P. and Lupski, J.R. (2010): Structural Variation in the Human Genome and its Role in Disease. Annual Review Medicine 61: S. 437-55; online: <https://www.ncbi.nlm.nih.gov/pubmed/20059347> (13.12.2018)

- [11] 1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E. and McVean, G.A. (2010): A map of human genome variation from population-scale sequencing. *Nature* 467: S. 1061-73; online: <https://www.ncbi.nlm.nih.gov/pubmed/20981092> (13.12.2018)
- [12] Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K. and Rehm, H.L. (2015): ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine* 17: S. 405-424; online: <https://www.ncbi.nlm.nih.gov/pubmed/25741868> (13.12.2018)
- [13] https://www.gfhev.de/de/leitlinien/LL_und_Stellungnahmen/2013_05_28_Stellungnahme_zu_genetischen_Zufallsbefunden.pdf (13.12.2018)
- [14] Henn, W. (2014): Multiparameter-Genanalyse und informationelle Selbstbestimmung. *Medizinische Genetik* 26: S. 273-277
- [15] Quon, B.S. and Rowe, S.M. (2016): New and emerging targeted therapies for cystic fibrosis. *British Medical Journal* 30: S. 352:i859. <https://www.ncbi.nlm.nih.gov/pubmed/27030675> (13.12.2018)
- [16] http://www.gfhev.de/de/leitlinien/LL_und_Stellungnahmen/2011_12_02_GfH-Stellungnahme_DTC-Genests.pdf (13.12.2018)
- [17] Heil, I. und Sowa, E. (2013): Gentests aus dem Internet? Eine Aufgabe zum selbstständigen Recherchieren und Urteilen. *PdN-BioS* 6/62: 11-13
- [18] Wong, A.I. and Lo, Y.M. (2015): Noninvasive fetal genomic, methylomic, and transcriptomic analyses using maternal plasma and clinical implications. *Trends in Molecular Medicine* 21: S. 98-108. <https://www.ncbi.nlm.nih.gov/pubmed/25618775> (13.12.2018)