

medgen 2014 · 26:21–27
DOI 10.1007/s11825-014-0433-0
Online publiziert: 21. Februar 2014
© Springer-Verlag Berlin Heidelberg 2014

T. Bettecken¹ · A. Pfeufer^{2, 3, 4} · R. Sudbrak^{5, 6} · R. Siddiqui⁷ · A. Franke⁸ ·
T.F. Wienker⁹ · M. Krawczak^{7, 10}

¹ Max-Planck-Institut für Psychiatrie, München

² Institut für Bioinformatik und Systembiologie, Helmholtz-Zentrum, München

³ Praxis für Humangenetik, Isar Medizin Zentrum, München

⁴ Myriad GmbH, Martinsried

⁵ Abteilung Analyse des Vertebratengenoms, Max-Planck-Institut für Molekulare Genetik, Berlin

⁶ Alacris Theranostics GmbH, Berlin

⁷ TMF - Technologie- und Methodenplattform für die vernetzte medizinische Forschung e. V., Berlin

⁸ Institut für Klinische Molekularbiologie, Christian-Albrechts-Universität Kiel

⁹ Abteilung Molekulare Humangenetik, Max-Planck-Institut für Molekulare Genetik, Berlin

¹⁰ Institut für Medizinische Informatik und Statistik, Christian-Albrechts-Universität Kiel

Next Generation Sequencing in der diagnostischen Praxis

Von der Variante zum Befund

Hintergrund

Unter dem Begriff Next Generation Sequencing (NGS) werden all jene neueren Verfahren der DNA-Sequenzierung zusammengefasst, die auf der massiven Parallelisierung aller hierfür notwendigen enzymatischen Reaktionen in sehr kleinen Volumina beruhen. Die Entwicklung der NGS-Technologie hat weite Teile der Medizin nicht nur wissenschaftsmethodisch verändert. So lässt sich dank NGS ein menschliches Genom in einer der Sanger-Sequenzierung nahe kommenden Qualität derzeit schon für etwa € 10.000 (Summe aus Biochemie und Bioinformatik) sequenzieren. Über Jahre hinweg haben sich die relativen Kosten der Erzeugung von DNA-Sequenzdaten in immer kürzeren Zeiträumen halbiert, und dieser (relative) Preisverfall hat längst den Abwärtstrend bei elektronischen Speichermedien überholt (■ **Abb. 1**). Mit dem dadurch in greifbare Nähe gerückten „1000-Dollar-Genom“ schickt sich das NGS nun auch an, die (genetische) Diagnostik zu durchdringen. Nach Abklingen manch

anfänglicher Euphorie hat sich jedoch gezeigt, dass ein Einsatz des NGS in der Patientenversorgung auf lange Sicht nur für hinreichend penetrante genetische Varianten sinnvoll ist. So steht das NGS derzeit v. a. in 2 Anwendungsbereichen vor dem Durchbruch in die klinische Routine:

- in der Onkologie zur molekularpathologischen Subtypisierung klinisch nicht unterscheidbarer Tumortypen und zur Therapiesteuerung im Rahmen einer sog. personalisierten Medizin sowie
- in der Humangenetik zur Diagnose monogener Erkrankungen, die für sich betrachtet zwar selten sind, in ihrer Gesamtheit aber eine Vielzahl von Patienten und ihre Familien schwer und oftmals über Jahre hinweg belasten.

Diagnostische Anwendungen außerhalb dieser beiden Paradigmen sind derzeit noch nicht umfassend etabliert. In der vorliegenden Arbeit soll dementsprechend nur der zweite Anwendungsbereich, d. h. die molekularpathologische

Bewertung von Keimbahnvarianten, in seiner aktuellen Entwicklung näher beleuchtet werden.

Die NGS-basierte Diagnostik monogener Erkrankungen zielt entweder auf die Gesamtheit des Genoms ($3 \cdot 10^9$ Basenpaare; „whole genome sequencing“, WGS), dessen proteinkodierende Bereiche (laut ENSEMBL/GENCODE Version 19 etwa 21.000 Gene, entsprechend etwa $5 \cdot 10^7$ Basenpaaren; „whole exome sequencing“, WES) oder eine von der klinischen Fragestellung abhängige Auswahl von Genen („gene panels“). Beim WES und der Sequenzierung von „gene panels“ werden die Zielbereiche zunächst in einem vorgelagerten Schritt angereichert. Diese Anreicherung ist jedoch mit zusätzlichen Kosten verbunden und birgt die Gefahr zusätzlicher Fehler, sodass es auch bei einer gezielten Sequenzierung sinnvoll sein kann, zunächst das Gesamtgenom zu sequenzieren und die anschließende klinische Interpretation der Sequenzdaten auf das Exom oder das „gene panel“ zu beschränken. Nach wie vor werden kleinere „gene panels“ auch

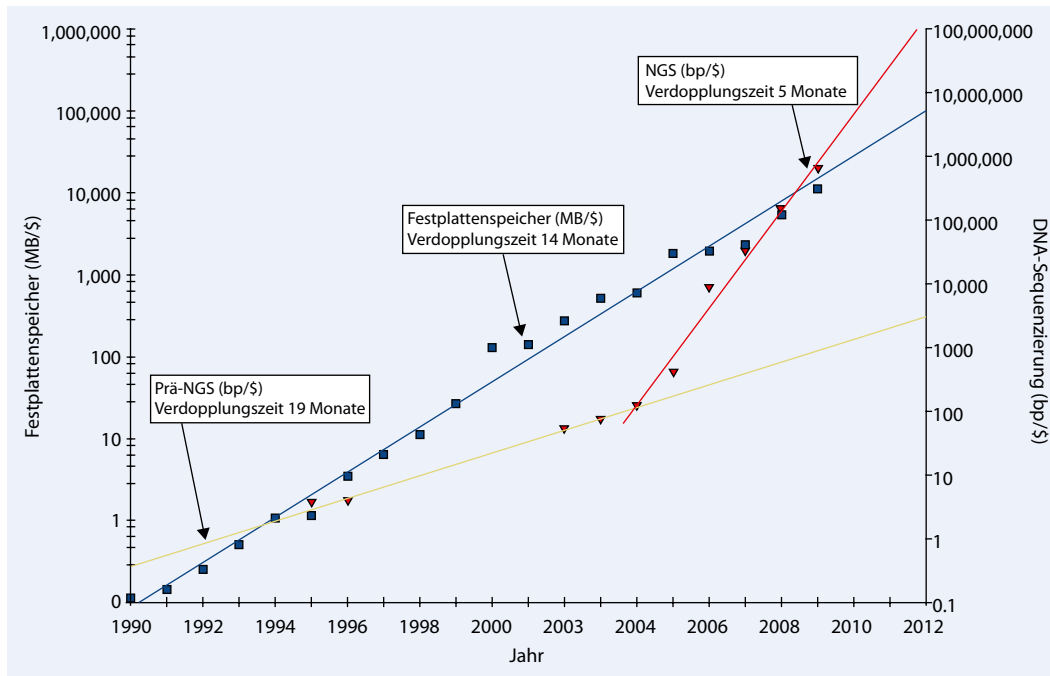


Abb. 1 Entwicklung der relativen Kosten für DNA-Sequenzierung (Sanger-Sequenzierung, Next Generation Sequencing) und Festplattenspeicher. (Adaptiert nach [11])

mit der herkömmlichen Methodik nach Sanger sequenziert.

Problem: Befunderstellung für neue genomische Varianten

Das menschliche Genom ist höchst variabel und die allermeisten darin befindlichen Varianten sind ohne pathologische Relevanz. Dieses Faktum macht die Ergebnisbewertung einer genomweiten Mutationssuche – von wenigen Ausnahmen abgesehen – äußerst kompliziert. Selbst viele exonische Varianten sind nach heutigem Wissensstand schwer oder gar nicht klinisch interpretierbar. Bei der diagnostischen Anwendung des NGS in der Humangenetik ist also die Ergebnisbewertung und nicht die Ergebnisgewinnung das wesentliche Problem. Viele Einrichtungen, die dennoch bereits heute NGS in der Diagnostik einsetzen, haben an dieser Stelle einen Kompromiss in Form abgestufter Befunde entwickelt. Der „herkömmliche“ Befund enthält nur Varianten, die in bekannten (d. h. im Onlinekatalog Mendelian Inheritance in Man, OMIM, aufgeführten) Krankheitsgenen lokalisiert und einer anerkannten funktionalen Bewertung zugänglich sind. Dabei handelt es sich um Genvarianten, die auch durch die Sanger-Sequenzierung von Kandida-

tengen entdeckt worden wären. Der „NGS-Befund“ enthält zusätzlich grundlagenwissenschaftliche Daten (z. B. aus In-vitro-Experimenten oder Tiermodellen), die sich mit den verbleibenden Mutationen durch eine teils spekulative molekularpathologische Interpretation in Verbindung bringen lassen.

In der humangenetischen diagnostischen Praxis wird die klinische Interpretation von NGS-Daten wohl noch einige Zeit hinter den technischen Möglichkeiten zurückbleiben, maßgeblich weil das Wissen um die Bedeutung der meisten Varianten unzureichend ist oder schlichtweg fehlt. Daher wird im Folgenden das Konzept einer Datenbank vorgestellt, die reichhaltige Informationen zu Genotyp und Phänotyp einer Vielzahl von Patienten mit monogenen Erkrankungen systematisch, zentral und unter Berücksichtigung aller ethischen und rechtlichen Rahmenbedingungen zusammenführen soll. Durch diese Datenbank ließe sich im Lauf von etwa 2–3 Jahren die Evidenzbasis schaffen, ohne die eine sachgerechte Befunderstellung auf der Grundlage von NGS-Daten nach Ansicht der Autoren nicht möglich ist.

DNA-Sequenzierung in der molekulargenetischen Diagnostik

Nach ersten Anfängen in der Frühzeit der Molekularbiologie nahm die molekulargenetische Diagnostik mit der Entwicklung der Polymerasekettenreaktion (PCR, ab dem Jahr 1985) und der nichtradioaktiven DNA-Sequenzierung nach Sanger (4-Farben-Kapillarsequenzierung, ab dem Jahr 1994) jeweils einen großen Aufschwung.

Sanger-Sequenzierung versus Next Generation Sequencing

Herkömmlicherweise geht die molekulargenetische Diagnostik von einer Auswahl möglicher Krankheitsgene aus, die sich aus der jeweiligen klinischen Verdachtsdiagnose ergibt. Diese Liste lässt sich gemäß Erfolgsaussichten der Diagnostik grob priorisieren, wobei neben Kausalitätserwägungen auch die Größe der einzelnen Gene eine Rolle spielt. Dieser Reihung folgend werden die Gene dann Exon für Exon mittels Sanger-Sequenzierung nach auffälligen Mutationen durchsucht. Die (ätiologische) Bewertung der dabei aufgefundenen Varianten erfolgt durch einen Abgleich mit Datenbanken bekannter Mutationen oder unter Verwendung bioinformatischer Verfahren, wie

z. B. der In-silico-Vorhersage des Einflusses der Mutation auf das entsprechende Genprodukt. Darauf aufbauend wird ein abschließender Befund erstellt.

Durch die NGS-Technologie wird dieses Vorgehen quasi auf den Kopf gestellt. Die Parallelisierung des Sequenzierungsprozesses und der daraus resultierende hohe Durchsatz erlauben es, die Diagnostik statt mit einer sorgfältig priorisierten Liste nun mit einer geeigneten Vorauswahl von Genen („gene panels“) zu beginnen oder sich gleich für die Sequenzierung des gesamten Exoms oder Genoms des Patienten zu entscheiden. Die Probleme der NGS-basierten Diagnostik stehen also nicht am Anfang sondern am Ende der labortechnischen Untersuchung – und zwar in Form einer Vielzahl potenziell krankheitsverursachender Varianten, die es zu interpretieren und in einen abschließenden medizinischen Befund bezüglich des klinischen Phänotyps zu überführen gilt.

Molekulargenetische Diagnostik mit NGS-Technologien

Ein Blick auf die technischen Abläufe der NGS zeigt, dass es sich hierbei um einen weitestgehend hypothesenfreien Ansatz handelt, dessen Überlegenheit sich erst beim parallelen, ungezielten Durchsuchen möglichst vieler genomischer Bereiche zeigt. Durch den sinnvollen Einsatz der NGS werden daher nicht unbedingt die Kosten der einzelnen Diagnostik gesenkt, sondern es wird eher die Untersuchung einer ungleich größeren Zahl von Genen als bisher zu vergleichbaren Kosten ermöglicht.

Sequenzierung des kompletten Genoms („whole genome sequencing“, WGS)

Die Sequenzierung eines kompletten menschlichen Genoms ist nicht nur theoretisch machbar, sie wurde auch bereits in größerem Umfang, z. B. im Rahmen des 1000-Genomes-Projekts [1], praktisch durchgeführt. Beim WGS wird die genomische DNA eines Individuums in Fragmente von 100–500 bp zerlegt (die optimale Fragmentgröße ist plattformabhängig), anschließend amplifiziert, in Fragmentbibliotheken überführt und so in

den Sequenzierungsprozess eingebracht. Dabei fallen je nach Technologie für ein Genom Hunderte von Millionen „reads“ an (d. h. Sequenzfragmente von 50–500 Basenpaaren), von denen zwar die DNA-Sequenz, nicht aber die Lage im Genom bekannt ist. Diese muss vielmehr bioinformatisch durch Abgleich mit einer Referenzsequenz ermittelt werden. Idealerweise sollte am Ende das gesamte Genom gleichmäßig mit „reads“ abgedeckt sein. In der Realität variiert diese Abdeckung („coverage“) jedoch erheblich.

Sequenzierung des kompletten Exoms („whole exome sequencing“, WES)

Wegen des hohen Durchsatzes bei NGS ist eine diagnostische Beschränkung auf wenige Gene allein zur Steigerung der Kosteneffizienz nicht sinnvoll. Trotzdem können die Arbeitsabläufe auch bei NGS so an die Möglichkeiten der Technologien angepasst werden, dass die Datenerzeugung aus ein und derselben Probe fallabhängig optimiert wird. So lassen sich z. B. durch eine Einschränkung der Sequenzierung auf das Exom die verfahrensabhängigen Kosten pro Individuum erheblich senken und dennoch die klinisch relevanten Sequenzinformationen größtenteils gewinnen. Eine solche Einschränkung wird in der Praxis durch die gezielte Anreicherung der Exone in den Fragmentbibliotheken ermöglicht. In einem 2. Schritt könnte dieselbe Fragmentbibliothek des Patienten aber auch für eine WGS verwendet werden.

Sequenzierung krankheitsbezogener „gene panels“

Selbst beim Einsatz von NGS kann es gute Gründe geben, sich im konkreten Fall auf eine überschaubare Liste von Kandidatengen zu beschränken. Sie ist z. B. sinnvoll, wenn die klinische Symptomatik genügend Hinweise auf spezifische genetische Ursachen liefert. Für eine Reihe von Erkrankungen wurden daher in Abhängigkeit von der jeweiligen NGS-Plattform krankheitsbezogene „gene panels“ zusammengestellt. Diese enthalten jeweils etwa 10–50 Gene. Nach gezielter Anreicherung der Exone, Spleißstellen und Kontrollregionen werden diese gemeinsam in einem einzigen Sequenzierungsvorgang unter-

Hier steht eine Anzeige.

 Springer

sucht. Eine Priorisierung ist hierbei nicht erforderlich, da alle Panelgene im selben Arbeitsgang sequenziert werden.

Interpretation und Befunderstellung

Die über viele Jahre in der molekularen Diagnostik monogener Krankheiten gesammelten Erfahrungen haben das langvermutete Phänomen der genetischen Heterogenität klinisch identischer Krankheiten in unerwartetem Maße bestätigt. Sowohl die Zahl als auch die Variabilität der für eine phänotypisch mehr oder wenig klar definierte Krankheit verantwortlichen Gene ist i. d. R. erheblich grösser als bis dato angenommen. Dies wirft für die molekulargenetische Diagnostik viele neue Probleme auf. So wird nur in Einzelfällen eine in einem bestimmten Fallkontext identifizierte neue Sequenzvariante allein dadurch, dass sie in einer einschlägigen Datenbank aufgeführt ist, als krankheitsverursachend zu identifizieren sein. Stattdessen ist eher damit zu rechnen, dass sich die meisten Listen bekannter Krankheitsgene als unvollständig erweisen. Dies gilt insbesondere für solche Krankheiten, die eine variable Symptomatik und unterschiedliche klinische Verläufe aufweisen, wie z. B. mentale Retardierung, generalisierte Entwicklungsverzögerung, gestörte Sprachentwicklung und Autismus. Viele dieser Krankheiten, die in der humangenetischen Praxis eine erhebliche Bedeutung haben, werden überhaupt erst durch eine umfassende Nutzung von NGS molekulardiagnostisch fassbar.

Zweifellos werden die allermeisten krankheitsspezifischen Kataloge ursächlicher genetischer Varianten schon allein wegen der zeitlichen Progredienz der Einführung des NGS zunächst unvollständig bleiben. Zumindest bei den monogenen Erkrankungen kann man allerdings davon ausgehen, dass eine einmal als pathogen erkannte Variante für jeden weiteren Fall gleicher Symptomatik ebenfalls ätiologisch relevant sein dürfte. Zudem wurde erst kürzlich nachgewiesen, dass Neumutationen ursächlich für einen erheblichen Teil aller schwerwiegenden kindlichen Entwicklungsstörungen sind [2]. Angesichts der stark eingeschränkten Reproduktion der davon Betroffenen ist dieser

medgen 2014 · 26:21–27 DOI 10.1007/s11825-014-0433-0
© Springer-Verlag Berlin Heidelberg 2014

T. Bettecken · A. Pfeufer · R. Sudbrak · R. Siddiqui · A. Franke · T.F. Wienker · M. Krawczak
Next Generation Sequencing in der diagnostischen Praxis. Von der Variante zum Befund

Zusammenfassung

Next Generation Sequencing (NGS) erlaubt die Untersuchung des kompletten Exoms oder Genoms eines Patienten mit vertretbarem zeitlichem und finanziellem Aufwand. Dieser diagnostische Quantensprung hat neben rechtlichen, ethischen und ökonomischen Aspekten auch vielfältige Auswirkungen auf die Patientenversorgung. Die weitgehende Einführung des NGS in die Routinediagnostik steht momentan jedoch noch vor vielen Hürden. Insbesondere ist zu erwarten, dass mittels NGS eine Vielzahl seltener Varianten bei einem Patienten entdeckt wird, die nach derzeitigem Wissensstand klinisch nicht hinreichend klar interpretierbar sind. Als einen ersten Schritt zur Lösung die-

ses Problems wird das Konzept einer Datenbank vorgestellt, die systematisch genotypische und phänotypische Informationen aus dem Versorgungskontext in Deutschland integrieren soll. Die so entstehende Ressource wäre nicht nur von großem wissenschaftlichem Interesse. Sie böte vordringlich den klinisch tätigen Humangenetikern die notwendige Evidenzbasis für eine zuverlässige Bewertung ihrer patientenbezogenen Sequenzierungsdaten.

Schlüsselwörter

Datenbank · Diagnose · DNA · Mutation · Genetische Polymorphismen

Next generation sequencing in diagnostic practice. From variant to diagnosis

Abstract

Next generation sequencing (NGS) allows whole exome or whole genome sequencing for a given patient to be performed timely and at reasonable cost. This diagnostic quantum leap not only has various legal, ethical and economical aspects but will naturally also impact upon patient care. Currently, however, the wide-spread introduction of NGS into routine diagnostics is facing many obstacles. In particular, it is to be expected that NGS will identify a large number of rare variants in a given patient that are of (yet) unknown clinical significance. As a first step to-

wards solving this problem, we introduce the concept of a database that will systematically integrate genotypic and phenotypic information from the German health care context. Not only will this resource be of great scientific value, but the database shall also provide human geneticists with the evidence base necessary for the reliable evaluation of their patient-related sequencing data.

Keywords

Database · Diagnosis · DNA · Mutation · Polymorphism, genetic

Zusammenhang nicht weiter verwunderlich. Dank NGS sind die benannten Fälle jedoch endlich auch diagnostisch zugänglich geworden, und die Beurteilung der pathogenetischen Relevanz von Neumutationen dürfte ebenfalls erheblich von einer verlässlichen, frei verfügbaren Datenbasis genetischer Varianten profitieren.

Eine Datenbank für genomische Varianten und klinische Phänotypen

Der Zugang zu einer stets aktuellen und verlässlich kuratierten Datenbank pathogenetisch relevanter Sequenzvarianten ist für die Sicherung einer hinrei-

chenden Qualität und Effizienz der molekulargenetischen Diagnostik unabdingbar. Kommerzielle Anbieter entwickeln gegenwärtig Geschäftsmodelle zum NGS, die auf proprietärem Wissen über die entsprechenden Genotyp-Phänotyp-Beziehungen beruhen. Es steht jedoch zu befürchten, dass die daraus resultierende Konkurrenzsituation und Abgrenzung zwischen verschiedenen (öffentlichen und privaten) Datenressourcen dem Gesamtziel einer verbesserten Diagnostik eher abträglich sein werden. Vordringliches Ziel einer von der Technologie- und Methodenplattform für die vernetzte medizinische Forschung (TMF) e. V. koordinierten Initiative ist es deshalb, eine

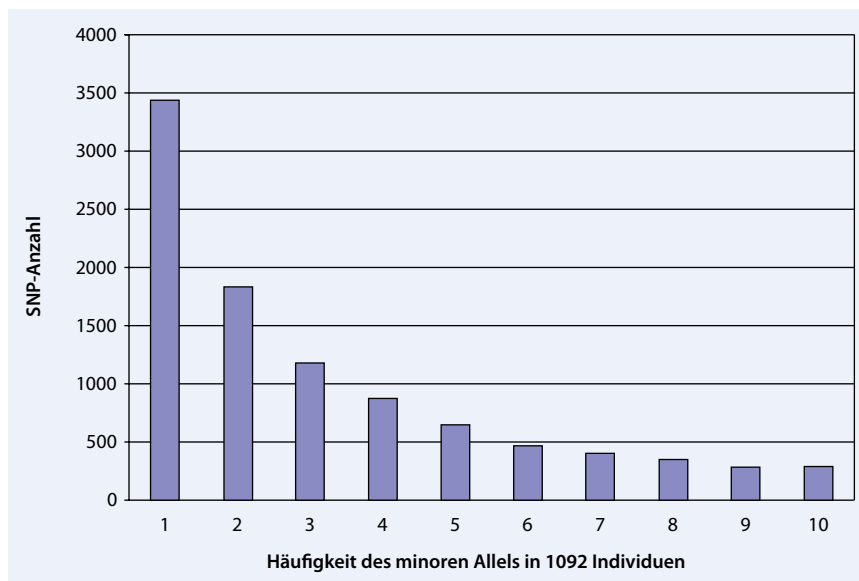


Abb. 2 ▲ Minore Allelhäufigkeit (MAF) seltener SNP in einer repräsentativen 1-Mb-Region auf Chromosom 16p. Etwas mehr als die Hälfte (9976/17580, entspricht 56,7%) aller SNP zeigte in den 1092 Individuen des 1000-Genomes-Projekts eine MAF unter 1%; mehr als 45% (7988) wiesen eine MAF <0,5% auf. Mehr als ein Drittel dieser seltenen SNP (3449/9976, entspricht 34,5%) kam sogar nur auf einem einzigen von 2184 Chromosomen vor. Diese Zahl schließt allerdings die meisten Sequenzierungsfehler ein

Datenbank pathogenetisch relevanter Sequenzvarianten zu etablieren, die frei von finanziellen, wissenschaftlichen oder politischen Partikularinteressen möglichst alle Interessenvertreter aus der NGS-basierten Diagnostik einbindet. Die Lücke zwischen bekannten und aktuell nicht (oder nur schwer) interpretierbaren Varianten wird sich nachhaltig nur schließen lassen, wenn alle einschlägigen NGS-Befunde zeitnah und vollständig einem qualifizierten Interessentenkreis verfügbar gemacht werden. Da dies selbstverständlich unter Wahrung aller Belange der Datensicherheit und des Datenschutzes zu geschehen hat, müssen Art und Umfang der Datenerfassung und der Datenkommunikation auf das zur Erfüllung des diagnostischen Zwecks notwendige Minimum beschränkt bleiben.

Kurzer historischer Exkurs

Seit der Entwicklung einer reproduzierbaren Darstellung von Chromosomen in der Metaphase [3] sind Varianten des menschlichen Genoms molekular charakterisiert und dokumentiert worden – sehr früh z. B. die Trisomie 21 [4]. In der Folge wurden zwar immer mehr und immer neue numerische und strukturelle

Chromosomenaberrationen entdeckt, aber erst Jahrzehnte später gelang es, sie systematisch und mit klinischen Phänotypinformationen versehen zu katalogisieren [5]. Mit dieser damals noch gedruckten Datenbank erhielten Zytogenetiker und genetische Berater erstmals die Möglichkeit, wiederholt auftretende chromosomale Aberrationen zuverlässig diagnostisch zu bewerten. Vor etwa 20 Jahren wurde dann die Human Gene Mutation Database (HGMD) [6] mit dem Ziel etabliert, alle in der wissenschaftlichen Literatur publizierten Mutationen zusammenzutragen, die sachlich begründet als ursächlich für eine monogene Erkrankung angesehen werden konnten. Eine Recherche in HGMD erlaubt somit die diagnostische Einordnung von Genveränderungen mit hohem positiv prädiktivem, aber nur geringem negativ prädiktivem Wert. Der mit „allelic variants“ bezeichnete Eintrag in der OMIM-Datenbank (<http://www.omim.org>, Zugriffen: 24. Januar 2014) verfolgt letztendlich das gleiche Ziel wie HGMD. Eine ähnliche, vermutlich aber deutlich umfangreichere Aufgabe wird auch die hier vorgeschlagene Datenbank genomischer Varianten aus NGS erfüllen, mit dem Unterschied, dass dieser Katalog schon von Anbeginn nicht

mehr in gedruckter Form wird publiziert werden können.

Inhalt und Umfang der NGS-Datenbank

In seiner aktuellen Fassung („build 137“) enthält die Single-nucleotid-polymorphism(SNP)-Datenbank (dbSNP) des National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/SNP>, Zugriffen: 24. Januar 2014) mehr als 53 Mio. Varianten des menschlichen Genoms, und jede Sequenzierung wird zum Nachweis eines nicht unerheblichen Teils dieser Varianten bei einem einzelnen Individuum führen. Allerdings kommen für eine autosomal-rezessive Erkrankung vermutlich nur Varianten mit einer Allelfrequenz unter 5% als kausal in Betracht; bei autosomal-dominanten Erkrankungen wird die Grenze sogar unter 0,5% liegen. Dieses Filterkriterium dürfte aber die Anzahl der potenziellen Datenbankeinträge nicht wesentlich verringern. Ein Blick in eine repräsentative 1-Mb-Region auf Chromosom 16p (■ **Abb. 2**) zeigt, dass mehr als ein Drittel aller SNP eine minore Allelhäufigkeit von unter 1% aufweist. Von diesen wiederum trat mehr als die Hälfte nur auf einem einzigen Chromosom der 1092 Probanden des 1000-Genomes-Projekts [1] auf. Daher sollen in die hier vorgestellte NGS-Datenbank nur solche Varianten aufgenommen werden, deren Kausalität nach einhelliger Einschätzung der Mitteleiler und der Kuratoren hinreichend belegt ist. Dies schließt durchaus und ausdrücklich auch Mutationen ein, die erstmalig bei einem Patienten gefunden wurden. Schließlich besteht für solche Einträge im Bedarfsfall, d. h. bei erneutem Nachweis der gleichen oder einer überlappenden Variante in einem anderen Patienten, die Möglichkeit des gezielten Kontakts mit dem Mitteleiler im Sinne einer ärztlichen Konsultation.

Neben Varianten, die nur eine oder wenige Basenpaare betreffen, enthält das menschliche Genom auch größere Deletionen, Insertionen und Duplikationen, die z. T. auch als Copy-number-Varianten (CNV) bezeichnet werden, sowie umfangreiche Inversionen. Trotz der geringen Wahrscheinlichkeit, mit der eine mehrfache Identifikation solcher genetischen

scher Veränderungen als kausal für unabhängige Erkrankungsfälle ohne verwandtschaftlichen Zusammenhang zu erwarten ist, sollten auch sie in die Datenbank aufgenommen werden. Selbst wenn eine 100%ige Übereinstimmung der Sequenzveränderung bei rekurrenten Neumutationen nicht zu erwarten ist, so enthalten Position und Umfang größerer submikroskopischer Veränderungen u. U. wertvolle Information für die Interpretation benachbarter oder überlappender Mutationen.

Rechtliche und ethische Aspekte

Die Durchführung medizinisch indizierter gendiagnostischer Maßnahmen ist in Deutschland seit 2012 durch das Gendiagnostikgesetz (GenDG) geregelt. Danach ist eine genetische Untersuchung zu Zwecken der Krankenversorgung allein den dafür qualifizierten Personen und Laboren vorbehalten (§5, §7) und sie ist grundsätzlich an die Aufklärung (§9) und Einwilligung (§8) der Betroffenen geknüpft. Die Ergebnisse der Untersuchung sind normalerweise 10 Jahre aufzubewahren und nach Ablauf dieser Frist zu löschen, solange die Betroffenen nicht schriftlich etwas anderes verfügen (§12). Dies bedeutet, dass die Übertragung eines genetischen Befunds in eine Datenbank wie die hier vorgeschlagene ebenfalls der aufgetragenen Einwilligung bedarf. Bei den dafür in Frage kommenden Patienten wird es sich in vielen Fällen um nichteinwilligungsfähige Minderjährige handeln. Die Einwilligung in die Speicherung von deren NGS-Daten muss also zunächst durch die Erziehungsberechtigten erfolgen und gegebenenfalls bei Erreichen der Einwilligungsfähigkeit durch die Patienten erneuert werden.

Die in der NGS-Datenbank zusammengeführten Informationen sollen primär die Befunderstellung aus NGS-Daten sukzessive auf eine immer breitere und zuverlässigere Basis stellen. Daneben werden sich die Daten jedoch auch als wertvoll für die Beantwortung wissenschaftlicher Fragen erweisen. Zweifellos dient die Datenbank an dieser Stelle ausschließlich Dritinteressen. Da die Daten aber im Versorgungskontext erhoben wurden und dort zum Zeitpunkt der Übernahme bereits

zweckdienlich waren, und da wegen der vorgesehenen Anonymisierung die Risiken für die Betroffenen minimal sind, stellt die NGS-Datenbank nach Ansicht der Autoren kein Instrument der fremdnützigen Forschung im Sinne der Deklaration von Helsinki dar. Vielmehr ist in ihrer Verwendung für die Diagnosestellung im Falle ebenfalls Erkrankter und in der Erforschung der Genotyp-Phänotyp-Beziehungen ein nicht anders zu erzielender Gruppennutzen zu sehen, der diese Verwendung ethisch rechtfertigt (s. z. B. [12]).

Eine öffentlich zugängliche Sammlung von NGS-Varianten, wie die hier vorgeschlagene, muss der besonderen Schutzwürdigkeit genetischer Daten in jedem nur denkbaren Umfang Rechnung tragen. Die verfügbar gemachten genetischen und klinischen Informationen werden zwar vollständig anonymisiert. Allerdings lässt sich nicht ausschließen, dass allein durch die Detailliertheit der phänotypischen Beschreibung, die für eine sinnvolle klinische Nutzung der Datenbank erforderlich sein wird, für einzelne Patienten ein Risiko der Re-Identifizierung besteht. Aus diesem Grund bedarf es für den Betrieb der vorgeschlagenen Datenbank eines besonderen Datenschutzkonzepts, das dieser Gefahr wirksam vorbeugt. Denkbar sind Erweiterungen generischer Datenschutzkonzepte, wie z. B. desjenigen der TMF e. V., das bereits an mehreren Stellen erfolgreich an den Einsatz im molekulargenetischen Forschungskontext angepasst wurde [7].

Internationale Aktivitäten

Angesichts des weltweit zunehmenden Einsatzes von NGS für diagnostische Zwecke verwundert es nicht, dass sich auch in anderen Ländern Initiativen gebildet haben, um die daraus hervorgehenden Informationen zu katalogisieren und für Dritte nutzbar zu machen. Beispielhaft sei hier zunächst das sog. ClinVar-Projekt genannt, das sich zum Ziel gesetzt hat, ein frei zugängliches öffentliches Archiv für Varianten des menschlichen Genoms mit Bezug zum klinischen Phänotyp zu etablieren (<http://www.ncbi.nlm.nih.gov/clinvar/>). Zugriffen: 24. Januar 2014, [8]). Aktuell gibt es Bestrebungen der National

Institutes of Health (NIH), ClinVar mit einer Förderung von 25 Mio. US-Dollar zu einer „clinical genome resource“ (ClinGen) zu erweitern. Diese soll dann (ähnlich wie im hier vorgestellten Konzept einer deutschen NGS-Datenbank) US-amerikanischen Diagnostiklaboren als zentrale Referenz für medizinisch relevante genetische Varianten dienen [13]. Das wie ClinVar ebenfalls in den USA angesiedelte Projekt eines Exom-variant-Servers (EVS) ist aus dem Exome-sequencing-Projekt (ESP) hervorgegangen, einer vom National Heart, Lung and Blood Institute (NHLBI) geförderten Forschungsinitiative (<http://evs.gs.washington.edu/EVS/>). Zugriffen: 24. Januar 2014, [9]). Der EVS enthält derzeit exomweite Sequenzdaten von mehreren tausend Patienten. Das in England angesiedelte und von der Europäischen Union geförderte Projekt „Gen2Phen“ verfolgt neben der Datensammlung und der Etablierung der notwendigen Informationstechnologieinfrastruktur auch das Ziel, die für die Interpretation humaner Genomvarianten relevanten Informationen aus Modellorganismen in die Datenbank mit einzubeziehen [10].

Zweifellos wird es in Zukunft sinnvoll sein, all diese nationalen Aktivitäten international übergreifend zusammenzuführen. Auch die hier skizzierte deutsche Initiative versteht ihre Planungen in diesem Sinne. Allerdings sollte die Einbringung sensibler und schutzbedürftiger genetischer Daten aus verschiedenen nationalen Quellen nach Ansicht der Autoren auf nationaler Ebene koordiniert werden. Nur so lässt sich die Einhaltung der einschlägigen rechtlichen und ethischen Vorgaben garantieren. Außerdem wird eine gut etablierte nationale Datenbank einen deutlich signifikanteren Beitrag zu umfangreicheren internationalen Initiativen leisten können, als viele kleine Insellösungen. Und nicht zuletzt ist anzumerken, dass alle genannten Datenbankinitiativen entweder noch im Aufbau begriffen sind, wie etwa ClinGen, und ihre Nutzbarkeit und Nützlichkeit somit noch nicht absehbar ist, oder dass sie maßgebliche Defizite aufweisen. Ein Abgleich mit HGMD hat beispielsweise einen sehr geringen negativ prädiktiven Wert, da der Datenbank „nur“ eine Literaturrecherche

zugrunde liegt und viele krankheitsverursachende Mutationen nicht mehr in prominente wissenschaftliche Publikationen eingehen. Zudem enthält HGMD nur sehr basale Phänotypinformationen und wäre daher für eine klinische Anwendung nicht spezifisch genug. Das ESP-Projekt bezieht sich mit dem EVS wiederum nur auf Herz-, Lungen und Bluterkrankungen und enthält zudem relativ kryptische Phänotypinformationen.

Fazit für die Praxis

- Zweifellos steht NGS erst am Anfang eines qualifizierten Einsatzes in der molekulargenetischen Routinediagnostik. Für die klinische Interpretation einer Vielzahl der dabei aufgedeckten Varianten fehlt (noch) die notwendige Evidenzbasis.
- Um diesen Mangel schnell, sachgerecht und nachhaltig zu beheben, sind richtungsweisende Entscheidungen hinsichtlich der Schaffung einer deutschen NGS-Datenbank und ihrer internationalen Vernetzung erforderlich. Eine solche Datenbank würde nicht nur die diagnostische Sicherheit erhöhen, sie würde auch dazu beitragen, den Einsatz der NGS in der Krankenversorgung möglichst kosteneffizient für die Gemeinschaft der Krankenversicherten zu gestalten. Und nicht zuletzt wäre sie ein herausragendes Translationsprojekt zwischen Wissenschaft und Krankenversorgung.
- Die rechtlichen Rahmenbedingungen einer klinischen NGS-Datenbank sind in Deutschland weitestgehend durch das Gendiagnostikgesetz und die Datenschutzgesetzgebung abgesteckt. Daher wird es in den kommenden Jahren v. a. humangenetischer, epidemiologischer und bioinformatischer Anstrengungen bedürfen, um durch die vorgeschlagene Datenbank die Anwendung von NGS in der humangenetischen Diagnostik und der personalisierten Medizin in Deutschland auf eine nachhaltige inhaltliche Grundlage zu stellen.

Korrespondenzadresse

Dr. T. Bettecken

Max-Planck-Institut für Psychiatrie
Kraepelinstr. 2–10, 80804 München
bettecken@mpipsykl.mpg.de

Ergänzender Ausblick. Zur Thematik des Artikels werden die Autoren gemeinsam mit der TMF e. V. einen Workshop als Satellitenveranstaltung der Jahrestagung der Deutschen Gesellschaft für Humangenetik in Essen organisieren. Termin: Mittwoch 19. März 2014, 9–12 Uhr.

Einhaltung ethischer Richtlinien

Interessenkonflikt. A. Pfeufer ist Angestellter der Myriad GmbH, Martinsried. T. Bettecken, R. Sudbrak, R. Siddiqui, A. Franke, T. F. Wienker und M. Krawczak geben an, dass kein Interessenkonflikt besteht.

Dieser Beitrag beinhaltet keine Studien an Menschen oder Tieren.

Literatur

1. The 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65
2. De Ligt J, Veltman JA, Vissers LE (2013) Point mutations as a source of de novo genetic disease. *Curr Opin Genet Dev* 23:257–263
3. Hsu TC (1952) Mammalian chromosomes in vitro – the karyotype of Man. *J Heredity* 43:167–172
4. Lejeune J, Gautier M, Turpin R (1959) Etude des chromosomes somatiques de neuf enfants mongoliens. *C R Hebd Seances Acad Sci* 248:1721–1722
5. Schinzel A (1984) Catalogue of unbalanced chromosome aberrations. Walter de Gruyter & Co, Berlin
6. Krawczak M, Cooper DN (1997) The human gene mutation database. *Trends Genet* 13:121–122
7. Mathieu N, Löhnhardt B, Grütz R et al (2013) Ethische und rechtliche Implikationen der Speicherung humaner Genomdaten. *Medizin Genet* 25:278–283
8. Baker M (2012) One-stop shop for disease genes. *Nature* 491:171
9. Tennessen JA, Bigham AW, O'Connor TD et al (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337:64–69
10. Webb AJ, Thorisson GA, Rookes AJ, GEN2PHEN Consortium (2011) An informatics project and online „Knowledge Centre“ supporting modern genotype-to-phenotype research. *Hum Mutat* 32:543–550
11. Stein LD (2010) The case for cloud computing in genome informatics. *Genome Biol* 11: 207
12. Stellungnahme der Zentralen Ethikkommission bei der Bundesärztekammer zum Schutz nichteinwilligungsfähiger Personen in der medizinischen Forschung (1997). *Dt. Ärzteblatt* 94: A1011–1012 (http://www.aerzteblatt.de/pdf/94/15/a1011_5.pdf). Zugriffen: 24. Januar 2014
13. NIH seeks better database for genetic Diagnosis (2013). *Science* 342: 27 (<http://www.sciencemag.org/content/342/6154/27.full>). Zugriffen: 24. Januar 2014