

Aus Sicht der Diagnostik

Gastkommentar

Nach dem erfolgreichen Abschluss des 1000-Genome-Projektes werden Erbgutsequenzen von über 1.000 Personen der Forschung frei zur Verfügung stehen. Dass bei diesen gewaltigen Datenmengen Fehler auftreten, lässt sich kaum vermeiden, aber das ist im Bereich der Grundlagenforschung auch zu akzeptieren.

Ganz anders sieht es für Anwendungen in der Medizin aus. Wie Dr. Bernd Timmermann vom MPI für molekulare Genetik in Berlin in seinem Artikel treffend darlegt, können nur fehlerfreie Sequenzierdaten die Voraussetzung für diagnostische Entscheidungen liefern.

Das enorme Potenzial der Hochleistungssequenzierungen kann deshalb nur ausgeschöpft werden, wenn sich die Medizin auf klare Rahmenbedingungen verlassen kann, die sowohl den Ansprüchen der Industrie genügen als auch qualitativ hochwertige und vergleichbare medizinische Ergebnisse beim Anwender sicherstellen. Mit diesem Anspruch ist BIO Deutschland bereits 2006 durch Gründung der Arbeitsgruppe „Wettbewerb und Ordnungspolitik“ angetreten.

Die größte Herausforderung für den Einsatz der neuen Technologien in der täglichen Routine sehen wir darin: Genomanalysen müssen trotz aller Fortschritte noch erheblich schneller und preisgünstiger werden und gleichzeitig noch höhere Qualitätsstandards erfüllen. Die Lösung liegt in zertifizierten Sequenziersystemen, die maximale Sicherheit bei vorhersagbarem Entwicklungsaufwand bieten.



Peter Pohl

Vorstandsmitglied BIO Deutschland und
CEO der GATC Biotech AG, Konstanz



Auf dem Weg zur Next Generation Diagnostik

Technisch machbar, aber...

Die Hochdurchsatzsequenzierung ist mittlerweile fester Bestandteil der biomedizinischen Forschung. Nun beginnt der lange Weg in die Routinediagnostik. Voraussetzung sind fehlerfreie Daten und klinische Relevanz der Befunde.

Obwohl marktreife Lösungen erst seit wenigen Jahren verfügbar sind, zählt das *Next Generation Sequencing* für Molekularbiologen bereits zu den etablierten Verfahren. Die neue Gerätegeneration revolutionierte die DNA-Sequenzierung. Nämlich die erste Entzifferung des humanen Genoms mit der klassischen Sanger-Technik noch mehr als 13 Jahre in Anspruch, werden derzeit z. B. im 1000-Genome-Projekt mehr als 1.200 humane Genome in einem Zeitraum von 24 Monaten resequenziert (<http://www.1000genomes.org>).

Unter der Leitung von Hans Lehrach gehört auch das Max-Planck-Institut für molekulare Genetik in Berlin zu den insgesamt neun internationalen Zentren, die an diesem Vorhaben beteiligt sind. Das Projekt soll erstmals einen umfassenden Einblick in die natürliche genetische Variabilität des Menschen ermöglichen und alle erfassten Varianten in einer öffentlich zugänglichen Datenbank bereitstellen.

Nicht nur in der Molekularbiologie, sondern auch in der medizinischen Diagnostik findet derzeit ein technologischer Generationenwechsel bei der DNA-Sequenzierung statt. War dieser Markt bisher von klassischen Kapillarsequenzierern dominiert, finden nun zunehmend hoch parallelisierte

und miniaturisierte Sequenzierungstechnologien Anwendung. Es hat sich jedoch schnell herausgestellt, dass sich die Anforderungen in einem diagnostischen Umfeld stark von den Ansprüchen der akademischen Genomcenter unterscheiden. Zwar ist für beide Bereiche eine hohe Genauigkeit entscheidend, bei einer populationsbasierten Betrachtung wie im 1000-Genome-Projekt kann jedoch eine gewisse Fehlerrate toleriert werden. Für die Medizin ist dies dagegen nicht akzeptabel, da von den Ergebnissen die gesamte Lebensplanung eines Menschen abhängen kann. Darüber hinaus stellt die Routinediagnostik höhere Ansprüche an die Geschwindigkeit und Bezahlbarkeit.

Entscheidend ist, dass alle möglicherweise auftretenden Variationen wie SNVs (*Single Nucleotide Variations*, Einzelnucleotidaustausche), Insertionen, Deletionen und strukturelle Variationen in einem Arbeitsschritt detektiert werden und der Aufwand der Auswertung mit Verfahren der Bioinformatik in einem überschaubaren Rahmen bleibt. Die hohen Anforderungen haben dazu geführt, dass nur Geräte mit einer weiten Leselänge und extrem geringer Fehlerrate für die praktische Medizin in Betracht kommen.

Als erstes und bislang einziges kommerziell verfügbares Gerät, das all diese Anforderungen erfüllt, hält gegenwärtig der Genome Sequencer FLX von Roche/454 Einzug in diagnostische Laboratorien. Die verwendete Technologie basiert auf einem Pyrosequencing-Verfahren, bei dem pro Lauf bis zu 1,5 Millionen Sequenzen mit einer mittleren Leselänge von 400 bis 500 Basen produziert werden können. Bei einer Gerätauflaufzeit von nur zehn Stunden beträgt die verwendbare Sequenzausbeute über 600 Millionen Basen – rund ein Fünftel des gesamten menschlichen Genoms.



Picotiterplatte (PTP) des GS FLX: Die an Beads gebundenen DNA-Fragmente werden monoklonal amplifiziert und auf eine Reaktionsplatte (PTP) geladen. In diesem Beispiel wurden vier getrennte Kompartimente verwendet. Pro Sektor konnten ca. 130 MB Sequenzdaten generiert werden (Bild: MPI, Berlin).

Zu den ersten diagnostischen Applikationen, die derzeit auf diesen *Next Generation Sequencing* Geräten etabliert werden, gehört die Amplicon Sequenzierung. Bei diesem Verfahren werden die diagnostisch relevanten Zielregionen im Erbgut des Patienten zunächst mittels Polymerase-Kettenreaktion (PCR) amplifiziert. Über den Amplifikationsschritt, also während der Vervielfältigung der Erbgutsequenzen, werden zusätzlich universelle Adaptoren

eingebaut. Diese Bausteine dienen einerseits als Startsignal für den nachfolgenden Sequenzierungsschritt, andererseits sind sie für die Immobilisierung der DNA-Fragmente von Bedeutung. Um einzelne Individuen zu unterscheiden, können bei dem Verfahren so genannte *Multiple Identifier* (MIDs) als zusätzliche Elemente eingeführt werden. Hierbei handelt es sich um spezifische Sequenzabschnitte mit einer Länge von zehn Basen, die es erlauben, eine große Anzahl von Individuen zu sammeln und simultan zu sequenzieren.

Für die Analyse größerer genomischer Regionen oder umfangreicherer Gen-Gruppen empfiehlt es sich, die relevanten DNA-Sequenzen zunächst anzureichern (*Sequence Capture Approach*). Hierbei werden die Zielregionen zuerst an spezifische Sonden gebunden, anschließend in einem Elutionschritt von den Sonden wieder gelöst und in einem letzten Schritt mittels PCR amplifiziert. Das angereicherte Material kann in der Sequenzierung eingesetzt werden.

Eine Alternative zu den PCR-basierten Anreicherungsverfahren liefern spezielle High-density Oligonucleotide-Microarrays von Unternehmen wie Roche NimbleGen.

Der 384 k Array beispielsweise umfasst 385.000 Sonden und erlaubt die selektive Anreicherung von genomischen Regionen bis ca. fünf Megabasen. Der 2,1 Millionen Array erlaubt es sogar, nahezu das gesamte menschliche Exom (*Whole Exome Array*) anzureichern. So werden mehr als 180.000 humane Exome für die Analyse zur Verfügung gestellt.

Dank der neuen Technologien rückt auch die diagnostische Sequenzierung eines

kompletten Patientengenoms in greifbare Nähe. Aber schon bei einem *Whole Exome Project* lassen sich in der Regel mehr als 14.000 genetische Varianten allein in den kodierenden (exonischen) Regionen detektieren. Betrachtet man zusätzlich die angrenzenden Bereiche, die keine Proteine kodieren, sondern eher der Genregulation dienen, werden mehr als 50.000 Varianten pro Individuum nachgewiesen. Von diesen sind im Normalfall bereits mehr als 80 Prozent in den entsprechenden Datenbanken (dbSNP130) verzeichnet. Um für eine bestimmte Erkrankung wie Dickdarm- oder Brustkrebs die entsprechenden funktionellen Varianten identifizieren zu können, ist jedoch ein hohes Maß an Bioinformatik-Expertise sowie humangenetisches Fachwissen erforderlich.

Ob sich solche umfassenden Analysen als diagnostische Anwendungen in der personalisierten Medizin durchsetzen werden, wird die Zukunft zeigen. Technisch sind individuelle Genomsequenzierungen bereits machbar. Die Frage nach der physiologischen und medizinischen Interpretation von ca. 3 Millionen Varianten, in denen sich ein Individuum von der Referenzsequenz unterscheidet, ist jedoch noch völlig offen. Auch ethische Aspekte, in wie weit individuelle Daten generiert werden sollten, die für den Patienten erst in Zukunft eine medizinische Relevanz bekommen könnten, sind bisher ungelöst. 🌸



Dr. Bernd Timmermann

Next Generation Sequencing Core Facility
Max-Planck-Institut für molekulare Genetik
14195 Berlin

timmermann@molgen.mpg.de
www.molgen.mpg.de