

Chancen und Herausforderungen von Big Data in der Onkologie

M. Schlesner

Als Big Data bezeichnet man Daten, die in dreierlei Hinsicht groß sind:

- ▶ großes Volumen
- ▶ große Geschwindigkeit, mit der die Daten generiert werden und daher auch verarbeitet werden müssen
- ▶ große Bandbreite von Datentypen

In der Onkologie werden große Datenmengen zum einen durch hochauflösende Analysen – Genomanalysen, Proteomics (das Inventar aller Proteine einer Zelle) oder Metabolomics (alle Metabolite) – sowie Hochdurchsatz-Bildgebungsverfahren für radiologische und zunehmend auch histologische Untersuchungen erzeugt. Zum anderen werden vermehrt Gesundheitsdaten gesammelt und ausgewertet – etwa in speziell angelegten Studien wie der Nationalen Kohorte, die Daten zum Gesundheitszustand von über 200 000 Menschen zwischen 10 und 69 Jahren über einen Zeitraum von bis zu 20 Jahren erfasst.

Für die Big-Data-Anwendungen in der Medizin nicht weniger bedeutsam sind aber Gesundheitsdaten, die außerhalb von Studien erhoben werden. Hierzu zählen die elektronische Patientenakte, aber

auch Daten von Fitness-Apps, von Wearables oder aus sozialen Netzwerken [1, 2]. Mit Big-Data-Technologien lassen sich diese Daten gemeinsam auswerten, um Korrelationen und Muster zu identifizieren und dadurch medizinisch oder klinisch relevante Zusammenhänge aufzudecken. Beispiele sind die individualisierte Krebsmedizin, die Bestimmung von Risikofaktoren und tumorprotektiven Faktoren durch die Verknüpfung von Lifestyle-Daten und Patientenakten sowie die Identifizierung genetischer Prädispositionen aus genomischen Daten und Patientendaten.

Genomsequenzierung in der Krebsmedizin

Die Genomsequenzierung hat in den letzten Jahren rasante Fortschritte erlebt. Während die erste vollständige Sequenzierung eines Krebsgenoms vor gerade einmal acht Jahren noch als Meilenstein gefeiert wurde, bieten manche Tumorzentren die Sequenzierung des Tumorgenoms schon in der klinischen Routine an. Durch massive Technologie-Innovationen in der DNA-Sequenzierung ist die Durchsatz-

kapazität der Sequenziergeräte stark gestiegen und der Preis pro sequenzierter Base auf einen Bruchteil gefallen. Mit dem HiSeq X10-System des Marktführers Illumina lassen sich jährlich 18 000 menschliche Genome für ca. 1 000 \$ pro Genom sequenzieren.

Während derzeit häufig noch Genpanels zur Mutationsanalyse verwendet werden, ist die Exom-Sequenzierung (also die Sequenzierung aller Protein-kodierenden Bereiche) auf dem Weg in die Routineanwendung. Und es ist bereits absehbar, dass in der Onkologie die Exomsequenzierung schon bald durch die Ganzgenomsequenzierung (whole genome sequencing, WGS) abgelöst wird. Denn nur WGS ermöglicht auch die Analyse von Mutationen im nicht-kodierenden Bereich des Genoms, deren Bedeutung zunehmend verstanden und als potenziell klinisch relevant erkannt wird [3, 4]. Die Datenmenge steigt dabei von einigen 100 MByte bei der Panelsequenzierung über 5–10 GByte bei der Exomsequenzierung auf über 200 GByte pro Patient für WGS.

Ein Beispiel für die klinische Anwendung der Genomsequenzierung in der Onkologie ist das INFORM-Register (INdividualized Therapy FORe Relapsed Malignancies in Childhood) [5]. Ein Konsortium von Ärzten und Wissenschaftlern aus kinderonkologischen Studiengruppen untersucht das Tumorgenom pädiatrischer Patienten mit rezidivierter oder progredienter Krebserkrankung, für die kein etabliertes Behandlungskonzept mehr zur Verfügung steht. Bei diesen Patienten wird das Genom auf Veränderungen untersucht, die ein Ansprechen auf bestimmte (häufig zielgerichtete) Therapien versprechen. Die gewonnenen Informationen können die behandelnden Ärzte dann in ihre Therapieentscheidung einfließen lassen. Pilotprojekte zur personalisierten Behandlung von erwachsenen Tumorkranken werden beispielsweise im Rahmen des Precision Oncology Programs des Nationalen Centrums für Tumorerkrankungen (NCT POP) und des



Big Data: das Generieren und Auswerten großer Datenmengen birgt Chancen für die Medizin – aber auch Risiken (Symbolbild).

Heidelberg Center for Personalized Oncology (DKFZ-HIPO) durchgeführt.

Chancen von Big Data



Verbesserte Diagnosestellung und Patientenstratifizierung

Das Anwendungsspektrum von Big-Data-Technologien in der Onkologie ist vielfältig. So können durch die Kombination von (histo-)pathologischen Daten mit genomweiten molekularen Daten sehr sichere und sehr differenzierte Diagnosen gestellt werden, die eine weitaus präzisere Klassifizierung von Tumoren ermöglichen. Ein Beispiel für eine molekulare Subklassifizierung mit therapeutischer Relevanz ist die Unterteilung diffus großzelliger B-Zell-Lymphome in die Subtypen „germinal-center like“ (GCB) und „activated B-cell-like (ABC) lymphoma“. Auch für eine Gruppe kindlicher Hirntumore (primitive neuroectodermal tumors of the central nervous system, CNS-PNETs) konnte durch die Kombination verschiedener genomweiter molekularer Analysen eine differenziertere Klassifizierung erreicht und 4 neue Tumorentitäten definiert werden [6].

Die präzise Diagnose ist die Grundlage für eine Stratifizierung der Patienten und damit die Auswahl der optimalen Behandlungsstrategie für jeden Patienten. Ziel ist hierbei die individualisierte Krebsmedizin, in der jeder Patient eine optimal auf ihn abgestimmte Therapie erhält. Möglichst genaue Vorhersagen des Tumoransprechens sollen ineffektive Therapien vermeiden und zusammen mit Vorhersagen von Nebenwirkungen und Therapierisiken die Basis für eine gemeinsame (partizipative) Therapieentscheidung bilden.

Personalisierte Therapie

Zielgerichtete Therapien versuchen genetische Unterschiede zwischen Tumorzellen und gesunden Zellen zu nutzen, um die Tumorzellen mit hoher Spezifität anzugreifen. Grundlage für die Wahl einer zielgerichteten Therapie sind daher genomische Analysen, die tumorspezifische Vulnerabilitäten offenbaren. Die ersten zielgerichteten Therapien waren auf aktivierte Onkogene gerichtet, die durch spezifische Inhibitoren (z.B. Tyrosinkinase-Inhibitoren) gehemmt werden.

Ein neuerer Ansatz ist synthetische Letalität, die den Verlust von Tumorsuppressorgenen ausnutzt (z.B. PARP-Inhibitoren).

Die Tumorzellen, die schon das Tumorsuppressorgen verloren haben, überleben die zusätzliche Hemmung eines weiteren, mit dem Tumorsuppressorgen in Beziehung stehenden Proteins nicht, wohingegen gesunde Zellen die Hemmung alleine überstehen. Eine Erweiterung dieser Strategie ist die derzeit diskutierte „Collateral Lethality“, die statt auf den Verlust von Tumorsuppressorgenen auf den Verlust anderer, für den Tumor direkt gar nicht relevanter Gene abzielt [7]. Dadurch ergibt sich ein wesentlich breiteres Spektrum von Vulnerabilitäten, die potenziell als Angriffspunkte für zielgerichtete Therapien ausgenutzt werden können.

Prävention

Die präventive Onkologie hat ihre Wurzeln in der Epidemiologie und daher schon seit jeher mit großen Fallzahlen gearbeitet. Die Kombination von Gesundheitsdaten und Lifestyle-Daten von Tausenden Individuen, die von Smartphone-Apps, Wearables und in sozialen Netzwerken außerhalb von Studien erfasst werden, bieten nun neue Möglichkeiten Faktoren zu identifizieren, die das Risiko für Tumorerkrankungen positiv oder negativ beeinflussen. Genauso ermöglicht die Verknüpfung von Gesundheitsdaten mit Genomdaten die Identifikation von genetischen Risikovarianten. Risikogruppen können so genauer definiert und somit einem engmaschigen Screening unterzogen werden, um Tumore oder prä-maligne Vorstufen frühzeitig zu erkennen.

Entwicklung neuer Therapien

Auch die Entwicklung neuer Medikamente kann durch Big-Data-Technologien unterstützt werden. Analysen von Genom, Proteom u.a. ermöglichen es, tumorspezifische Vulnerabilitäten und damit potenzielle neue Angriffspunkte für Therapien zu entdecken. Zusätzlich hilft die verbesserte Stratifizierung der Patienten durch molekulare Klassifikation, Wirkkeitsstudien effektiver und kostengünstiger zu machen. Durch die Auswahl entsprechender Patientenkollektive kann die Wirksamkeit leichter gezeigt oder bei Wirkstoffen, die nur in Patientensubgruppen wirksam sind, überhaupt nachgewiesen werden.

Generell erfordert die personalisierte Medizin ein Umdenken bei der Bewertung der Therapiewirksamkeit. Der bisherige Ansatz, die Wirksamkeit einer Prüfsub-

stanz über das durchschnittliche Ansprechen einer Patientenkohorte nachzuweisen, ist bei individuell auf einzelne Patienten abgestimmten Therapien naturgemäß nicht möglich. Statt in wenigen großen Studien muss die Wirksamkeit individualisierter Therapien durch eine Vielzahl sogenannter N-of-1 trials, also Studien mit nur einem Teilnehmer, nachgewiesen werden [8]. Ein Schlüssel zum Erfolg der individualisierten Medizin ist daher die Überführung der klinischen Routinebehandlung in N-of-1 trials inklusive standardisierter und umfassender Datenerhebung.

Outcomes Research

Auf Basis der so erfassten Daten können Big-Data-Technologien auch in der Versorgungsforschung neue Möglichkeiten eröffnen. Derzeit werden nur etwa 3% aller Krebspatienten in klinische Studien eingeschlossen. Das bedeutet, dass die Ergebnisse der klinischen Studien nicht notwendigerweise für die Gesamtheit der Krebspatienten unter den tatsächlichen Behandlungsbedingungen repräsentativ sind. Durch die verstärkte elektronische Verfügbarkeit von Patientendaten, u.a. aus „N-of-1 trials“, können mit Big-Data-Technologien große, heterogene Patientenkohorten retrospektiv ausgewertet werden. Dadurch lassen sich Fragen zu Behandlungseffektivität und Langzeitergebnis, aber auch zu Korrelationen mit dem Lebensstil und genetischer Prädisposition in bisher unerreichter Genauigkeit beantworten. Die kontrollierte Umgebung von klinischen Studien wird hier zwangsweise durch große heterogene Kollektive und Bedingungen ersetzt. Signifikante Aussagen werden aus diesen Kollektiven durch sehr große Fallzahlen sowie eine möglichst vollständige Erfassung aller Variablen ermöglicht.

Die American Society of Clinical Oncology (ASCO) entwickelt gemeinsam mit dem Unternehmen SAP für diese Fragestellungen die Plattform CancerLinQ™ (Learning Intelligence Network for Quality), die Informationen über viele Millionen Krebspatienten vorhalten wird. CancerLinQ™ soll Onkologen bei der Therapieentscheidung unterstützen, indem es die Daten des jeweiligen Patienten in Echtzeit mit einer großen Zahl ähnlicher Patienten abgleicht und so Erfahrungen und Behandlungsergebnisse zugänglich macht. Forscher werden die Plattform benutzen können, um die (anonymisierten) Patientendaten

nach Mustern und Korrelationen zu durchsuchen und so Hypothesen für weitere Forschungsarbeiten zu generieren. Bis Ende 2016 wird CancerLinQ™ voraussichtlich 500 000 elektronische Patientenakten aus Tausenden onkologischen Praxen und Behandlungszentren zusammentragen.

Herausforderungen von Big Data



Hohe IT-Anforderungen

Durch das schiere Volumen der Daten werden enorme Herausforderungen an die IT-Infrastruktur gestellt. Die zunehmende Verbreitung immer schnellerer Sequenziergeräte hat die weltweite Kapazität zur Genomsequenzierung in den letzten Jahren sprunghaft ansteigen lassen. Zurzeit verdoppelt sich die weltweite Sequenzierkapazität alle 12 Monate und auch für die nahe Zukunft wird noch eine Verdopplung alle 18 Monate vorausgesagt [9]. Damit ist die Wachstumsgeschwindigkeit der Produktion von Sequenzierdaten deutlich höher als die Fortschritte in der Computertechnologie. Das bedeutet, dass das Sequenzdatenvolumen mittelfristig nicht durch einfache Hochskalierung der Computer-Infrastruktur bewältigt werden kann. Neue Technologien für die Datenspeicherung und Prozessierung werden benötigt.

Zugang erfordert neue Strukturen

Schon jetzt beträgt die Menge in öffentlichen Datenbanken zugänglicher Krebs-Genomdaten mehrere Peta-Byte (1 Peta-Byte entspricht 10^{15} Bytes; man könnte damit den Inhalt von etwa 4 Mrd. Büchern speichern). Allein die Forschungsverbände innerhalb des International Cancer Genome Consortium (ICGC; www.icgc.org) haben in den vergangenen 5 Jahren mehr als 2 Petabyte an Sequenzdaten produziert [10]. Daten im Petabyte-Bereich sind nur sehr eingeschränkt portabel. Das Herunterladen von zentralen Servern würde mit der an Universitäten und Krankenhäusern normalerweise vorhandenen Bandbreite Monate dauern. Daher ist es im Big-Data-Bereich sinnvoller, die Nutzer zu den Daten zu bringen anstatt die Daten zu den Nutzern. Ermöglicht wird das durch Cloudtechnologie. Durch die gemeinsame Nutzung von Hochleistungs-IT sind Clouds zudem kosteneffizienter als getrennte Rechenzentren. Ein Modell für eine öffentliche Cloud, die Wissenschaftlern und Ärzten Zugang zu großen Kollektionen genomischer

Daten sowie zu standardisierten Analyseverfahren geben wird, ist unter www.genome-cloud.de beschrieben.

Auch weiterhin werden großangelegte Sequenzierstudien in internationalen Verbänden wie dem International Cancer Genome Consortium for Medicine (www.icgcm.org) die Entwicklung vorantreiben. Der größere Anteil der Genomdaten wird jedoch schon bald in einer Vielzahl kleinerer Forschungsprojekte und vor allem in der klinischen Anwendung produziert werden. Das vergrößert die Herausforderung, die erzeugten Daten nicht nur für den individuellen Patienten, sondern auch für die breite wissenschaftliche Nutzung verfügbar zu machen. Um das Potenzial der Genomdaten optimal ausnutzen zu können haben sich weltweit über 400 Organisationen zur Global Alliance for Genomics and Health zusammengeschlossen, um ein Rahmenwerk für das Teilen von Genomdaten und klinischen Daten zu entwickeln [11].

Neue Methoden für die Dateninterpretation

Die klinische Interpretation der genomweiten Sequenzdaten ist bisher nur sehr eingeschränkt möglich. Nur für einen Bruchteil der beobachteten Mutationen ist bekannt, ob und wie sie für bestimmte Therapien sensibilisieren bzw. Resistenzen vermitteln. Das liegt zum einen daran, dass nur wenige Mutationen für die Tumorentstehung relevant sind (sogenannte Driver-Mutationen), während die Mehrheit der Mutationen für die Tumorzellen neutral sind (Passenger-Mutationen). Als Zielstrukturen sind jedoch in den meisten Fällen nur Driver-Mutationen geeignet, ihrer Erkennung kommt daher eine große klinische Bedeutung zu.

Genau das ist aber schwierig: Zwar gibt es einige sehr häufig auftretende Driver-Mutationen, die leicht zu erkennen und mittlerweile gut charakterisiert sind. Diesen wenigen häufigen Drivern steht aber eine sehr viel größere Zahl seltener Driver-Mutationen gegenüber, die nur durch die Untersuchung sehr großer Patientengruppen erkannt werden können. Noch komplizierter wird die Analyse, wenn zu den Genomdaten noch weitere Daten (Proteomics oder Metabolomics) hinzukommen. Die gemeinsame Analyse dieser Datentypen wird genauere Vorhersagen zur Tumorsubgruppierung und damit auch

zur optimalen Therapie erlauben. Allerdings ist die Verknüpfung herausfordernd: Während höhere Fallzahlen die Komplexität der Analysen linear ansteigen lassen, führt die Kombination mehrerer Datentypen zu einem exponentiellen Anstieg. Aus großen, heterogenen Datensätzen lassen sich nur mit Techniken des maschinellen Lernens verborgene Information extrahieren.

Tumorheterogenität

Auch die Tumorbiologie setzt den Big-Data-Anwendungen Grenzen, denn grundsätzlich gleicht kein Tumorgenom dem anderen. Weder wird man in zwei Patienten jemals das gleiche Mutationsmuster beobachten, noch wird man in einer größeren Anzahl von Zellen eines Tumors das exakt gleiche Genom vorfinden. Die Heterogenität der Tumoren von verschiedenen Patienten (interpatient heterogeneity) erschwert die Erkennung von Driver-Mutationen und die Patienten-Stratifizierung, und damit letztlich die Therapieempfehlung.

Die Tumor-Heterogenität innerhalb eines Patienten (intrapatient heterogeneity) macht es praktisch unmöglich, den vollständigen Katalog aller in einem Patienten vorkommenden Mutationen zu erfassen. Zwar sind subklonale Varianten, also Varianten, die nur in einem Teil der Tumorzellen vorkommen, normalerweise nicht als therapeutische Targets geeignet. Theoretisch kann jedoch schon eine einzige Krebszelle mit einer Mutation, die eine Resistenz gegenüber einer Therapie bewirkt, zum Versagen dieser Therapie führen. Daher ist die Kenntnis subklonaler Varianten zur Vorhersage von Therapieansprechen und -resistenz von großer Bedeutung.

Zur Abschätzung der Heterogenität eines Tumors und zumindest partieller Erfassung der Subklone können – sofern klinisch machbar – mehrere räumlich getrennte Teile des Tumors sequenziert werden. Durch Sequenzierung an mehreren Zeitpunkten während des Therapieverlaufs kann das Herauswachsen resistenter Subklone festgestellt und das Therapieschema angepasst werden. Eine zunehmende Bedeutung kommt dabei den „liquid biopsies“ zu, die durch die Analyse von freier Tumor-DNA oder zirkulierenden Tumorzellen die nichtinvasive Überwachung des Tumorwachstums und der Tumorevolution ermöglichen.

Datenheterogenität

Zusätzlich erschweren praktische Probleme die Einführung von Big-Data-Analysen in der Onkologie. Die genom-basierte Medizin benötigt die Verknüpfung von Genomdaten mit klinischen Daten, (histo-)pathologischen Parametern und ggf. experimentellen Daten, die mit patientenspezifischen Modellen erzeugt wurden. Diese Daten werden mit verschiedenen, häufig nicht kompatiblen Methoden generiert und in verschiedenen Domänen vorgehalten, zwischen denen eine Verknüpfung nicht ohne weiteres möglich ist. So gibt es öffentliche Datenbanken für Omics-Daten, während klinische Daten häufig in proprietären Formaten auf Servern der Krankenhäuser vorliegen und weitere persönliche Daten auf privaten Servern kommerzieller Anbieter. Klinische und experimentelle Daten werden in den meisten Fällen nicht strukturiert erfasst und im schlechtesten Fall liegen sie nicht elektronisch vor.

Nachweis von Kausalität und Relevanz

Big Data bedeutet eine Verschiebung von der kausalitätsorientierten zur korrelationsbasierten Wissenschaft. Trotz (oder gerade wegen) der Vielzahl der Möglichkeiten, die sehr große Fallzahlen und eine engmaschige Verknüpfung vieler Datentypen bieten, muss bei der Interpretation von Big-Data-Analysen mit Umsicht vorgegangen werden. Systematische Fehler, zum Beispiel durch nicht-standardisierte Datenerfassung und nicht berücksichtigte Störfaktoren, können auch durch größere Fallzahlen nicht kompensiert werden [12]. Aus Korrelationen, wie Big-Data-Analysen sie liefern, kann nicht notwendigerweise auf Kausalität geschlossen werden, genauso wenig wie aus der statistischen Signifikanz, die wegen der großen Fallzahlen verhältnismäßig leicht erreicht wird, direkt eine klinische Relevanz gefolgert werden kann.

Ethische und rechtliche Fragen

Generell steht Big Data in der Medizin im Konfliktfeld zwischen dem Schutz der Privatsphäre des Patienten und Maximierung des Erkenntnisgewinns durch vollständige Ausnutzung der technischen Möglichkeiten. Genomische Daten sind per se nicht anonymisierbar, da das genomische Profil wie ein Fingerabdruck die eindeutige Zuordnung zu einem Individuum ermöglicht. Dabei kann alleine schon

die Re-Identifizierung eines Patienten in einer Krebsgenomdatenbank für den Patienten in vielerlei Hinsicht zum Nachteil werden.

Je mehr Daten miteinander verknüpft werden, desto größer wird das Risiko der Re-Identifizierung, und desto umfassender wird das Bild, das Dritte sich von Patienten machen können. Im Gegensatz dazu stehen die Interessen von Wissenschaftlern und Ärzten. Um das Potenzial von Big-Data-Technologien optimal ausnutzen zu können, sind sowohl eine möglichst breite Verfügbarkeit der Datensätze als auch eine weitgehende Verknüpfung verschiedener Datentypen essenziell. Hier müssen Mechanismen gefunden werden, die einerseits die Forschung nicht stärker als notwendig einschränken, andererseits aber die Wahrung der Rechte und Interessen der Patienten sicherstellen.

Weiterhin muss geklärt werden, wem erhobene Daten gehören. Was sind die Konsequenzen, wenn ein Patient die Einwilligung zur Nutzung seiner Daten widerruft? Wie werden Incidental Findings gehandhabt? Hier gilt es, das Recht auf informationelle Selbstbestimmung, und damit auch das Recht auf Nichtwissen, zu respektieren. Empfehlungen für einen verantwortungsvollen Umgang mit den neuen technischen Möglichkeiten in Klinik und Forschung gibt die Heidelberger Projektgruppe EURAT in ihren „Eckpunkten für eine Heidelberger Praxis der Ganzgenomsequenzierung“ [13].

Ausblick

Die rasante technologische Entwicklung in der Analytik zusammen mit einer zunehmenden Digitalisierung des Gesundheitssystems eröffnen eine Vielzahl neuer Möglichkeiten in der Onkologie. Zusammen haben sie das Potenzial, das Verständnis von Tumorerkrankungen deutlich zu erweitern und Krebsdiagnostik und -therapie grundlegend zu verändern.

Allerdings müssen noch viele ethische Fragen beantwortet und technische Probleme gelöst werden. Darüber hinaus sind im Gesundheitswesen Veränderungen erforderlich, die mehr Zeit benötigen werden als die reine technologische Entwicklung. Die genom-basierte, personalisierte Krebsmedizin braucht entsprechend ausgebildete Ärzte, die neben der Nutzung auch in der Lage sind, den Patienten die

Bedeutung und potenziellen Folgen genomischer Untersuchungen und weitreichender Datenerfassung zu vermitteln. Wie schnell und breit Big-Data-Technologien einen breiten klinischen Nutzen entfalten können, muss sich erst zeigen.

*Dr. Matthias Schlesner
Division Theoretical Bioinformatics
Head of Computational Oncology Group
Deutsches Krebsforschungszentrum
Im Neuenheimer Feld 280
69120 Heidelberg
m.schlesner@dkfz.de*

Interessenkonflikte: Der Autor erklärt, dass keine Interessenkonflikte bestehen.

Literatur

- 1 Weber GM, Mandl KD, Kohne IS. Finding the Missing Link for Big Biomedical Data. *JAMA* 2014; 311: 2479–2480
- 2 Savage N. Mobile data: Made to measure. *Nature* 2015; 527, S12–S13
- 3 Horn S, Figl A, Rachakonda PS et al. TERT promoter mutations in familial and sporadic melanoma. *Science* 2013; 339: 959–961
- 4 Northcott PA, Lee C, Zichner T et al. Enhancer hijacking activates GF11 family oncogenes in medulloblastoma. *Nature* 2014; 511: 428–434
- 5 <http://www.dkfz.de/de/inform/> (letzter Zugriff 02.06.2016)
- 6 Sturm D, Orr BA, Toprak UH et al. New Brain Tumor Entities Emerge from Molecular Classification of CNS-PNETs. *Cell* 2016; 164: 1060–1072
- 7 Muller FL, Aquilanti EA, DePinho RA. Collateral Lethality: A new therapeutic strategy in oncology. *Trends Cancer* 2015; 1: 161–173
- 8 Schork NJ. Personalized medicine: Time for one-person trials. *Nature* 2015; 520: 609–611
- 9 Stephens ZD, Lee SY, Faghri F et al. Big Data: Astronomical or Genomical? *PLoS Biol* 13: e1002195
- 10 Stein LD, Knoppers BM, Campbell P et al. Data analysis: Create a cloud commons. *Nature* 523, 149–151
- 11 genomicsandhealth.org/about-the-global-alliance/key-documents/white-paper-creating-global-alliance-enable-responsible-shar
- 12 Antes G. Big Data und Personalisierte Medizin: Goldene Zukunft oder leere Versprechungen? *Dtsch Arztebl* 2016; 113: A-712 / B-600 / C-592
- 13 https://www.uni-heidelberg.de/md/total-sequenzierung/stellungnahme_2.pdf (letzter Zugriff 02.06.2016)