

Recherche des IQVIA Institute for Human Data Science:

Genomik-Initiativen im weltweiten Vergleich

Die Genomforschung entwickelt sich durch sinkende Kosten für Sequenzierungen und durch leistungsfähige IT-Infrastrukturen rasant weiter. Bislang war es schwierig, den Überblick zu behalten. Damit sich Entscheider ein besseres Bild machen können, hat das IQVIA Institute for Human Data Science 187 Initiativen in einer Datenbank zusammengestellt und bewertet.

>> Im Jahr 1990 begann das Zeitalter der modernen Genomforschung: 1.000 Wissenschaftler aus 40 Ländern schlossen sich zum Humangenomprojekt zusammen. Ihr Ziel war, das menschliche Genom zu sequenzieren: eine Herausforderung, die bis 2003 erfolgreich bewältigt wurde. Das Gesamtbudget lag bei schätzungsweise 2,7 Milliarden US-Dollar.

Diagnostik und Therapie der Zukunft – mit Genomdaten

Bald darauf sind die Kosten für genomische Sequenzierungen rapide gesunken – von 14 Millionen US-Dollar (2006) auf 1.500 US-Dollar (2016) und weiter auf 300 US-Dollar (2020). Mittlerweile wurden fast 40 Millionen menschlicher Genome bis zu einem gewissen Grad sequenziert oder genotypisiert (Abb. 1). Dieser Trend wird sich weiter fortsetzen. Genaue Prognosen sind schwierig. Keine Zweifel bestehen

jedoch, dass wir uns in einer Phase des exponentiellen Datenwachstums befinden. Davon profitiert vor allem die Medizin.

Genomische Daten kommen bei unterschiedlichen Fragestellungen zum Einsatz. Sie helfen in der Genealogie, um Vorfahren zu identifizieren. Auch die Forensik nutzt molekularbiologische Technologien zur Identifizierung von Individuen. Aber mit Abstand die meisten Forscher setzen Genomdaten ein, um menschliche Krankheiten zu erkennen, zu verstehen und perspektivisch zu behandeln.

Dabei gibt es entscheidende Unterschiede: Über 10.000 Leiden sind monogenetisch; sie werden durch Defekte in einem einzigen Gen verursacht. Bei Krebs, Diabetes, Herz-Kreislauf-Erkrankungen und Asthma ist die Sache weitaus komplizierter. Heute geht man davon aus, dass Gen-Umwelt-Interaktionen zum Ausbruch solcher Leiden führen. Auch bei Infektionskrankheiten wie HIV und Malaria wurden genetische

Komponenten identifiziert, die mit Infektionsrisiken und mit dem Ansprechen von Therapien in Verbindung stehen.

Welches Potenzial in genomischen Daten steckt, zeigt sich schon heute in der Onkologie. Ärzte kennen nicht nur Gene, die mit höheren Krankheitsrisiken assoziiert sind, sondern setzen im Zuge der personalisierten Therapie spezielle Pharmaka ein. Das Organsystem selbst spielt dabei keine Rolle mehr. Etlliche Konsortien arbeiten an diesen Fragestellungen – und es wird immer schwerer, den Überblick zu behalten.

Die IQVIA Genomics Initiative-Datenbank schafft Überblick

Deshalb hat das IQVIA Institute eine eigene Datenbank mit weltweiten Initiativen aus der Humangenomik aufgebaut. Basis sind alle frei zugänglichen Informationen zu den Projekten. Um in die Daten-

bank aufgenommen zu werden, muss eine Initiative genomische Daten zumindest teilweise selbst generiert haben oder im Fall von Biobanken biologisches Material besitzen, das genotypisiert beziehungsweise sequenziert werden kann. Für jede Initiative wurden 26 Deskriptoren erfasst, einschließlich Details zur geografischen Herkunft der Daten, zum Umfang, zur Finanzierung, zum Zweck der Initiative, zur Art der Genomdaten, zu deren Verknüpfung mit weiteren Informationen und zur Zustimmung der Patienten für die weitere Nutzung.

Momentan befinden sich 187 genomische Initiativen in der Datenbank. Sie sind größtenteils aus Nordamerika (93 Projekte) oder Europa (35 Projekte). Weitere 31 Initiativen wurden international konzipiert. Es gibt unterschiedliche Anteile öffentlicher, privat geförderter oder öffentlich-privater Partnerschaften (Abbildung 2).

Öffentlich oder privat finanzierte Genom-Initiativen unter der Lupe

Einige Beispiele zu privatwirtschaftlichen Unternehmen aus der IQVIA-Datenbank: Ancestry.com hat sich auf genealogische Fragestellungen spezialisiert. Dazu gehört auch die genetische Genealogie. Die Kohorte besteht aus 15 Millionen genotypisierten Individuen. Auch 23andMe bietet Genotypisierungen für Privatpersonen an, die etwas über ihre Abstammung und über Krankheitsrisiken wissen wollen. Die Daten fließen in genomweite Assoziationsstudien (GWAS) ein. Genomic Health wiederum hat seinen Schwerpunkt im Bereich somatischer Tumore;

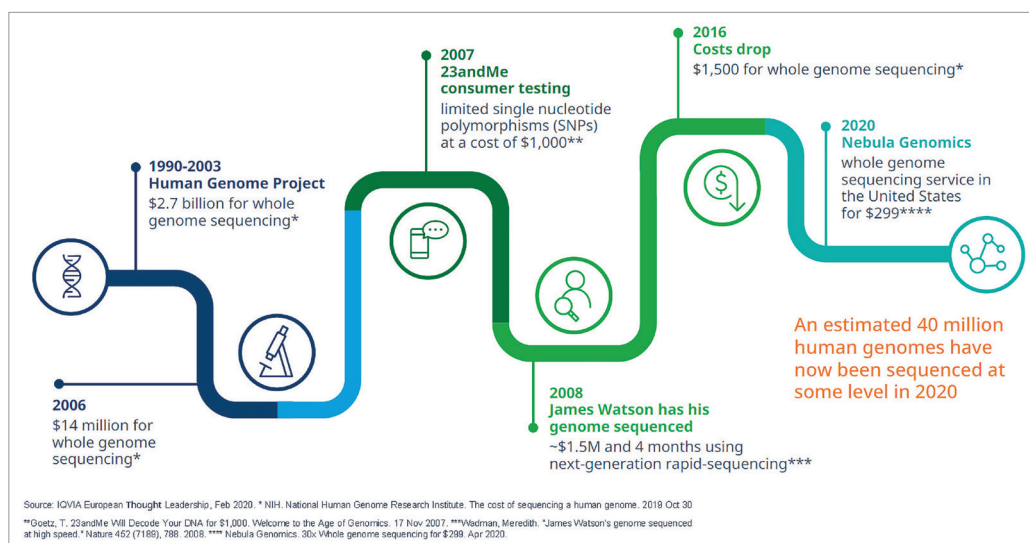


Abbildung 1: Meilensteine und Kosten der Genomsequenzierung in den letzten 30 Jahren; Quelle: IQVIA

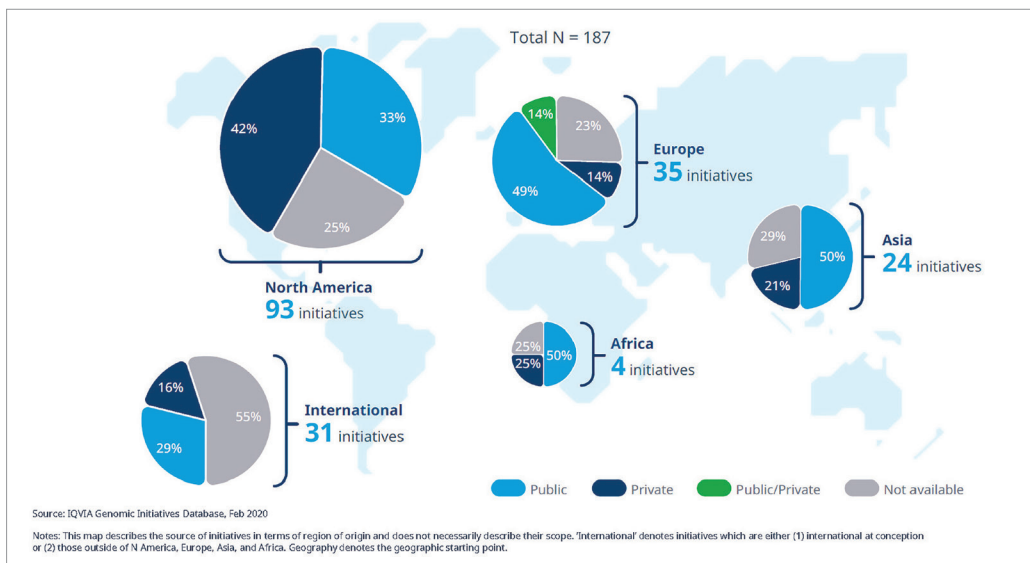


Abbildung 2: Übersicht zur geografischen Verteilung und zur Art der 187 Genom-Initiativen in der IQVIA-Datenbank; Quelle: IQVIA

aktuell umfasst die Kohorte eine Million Datensätze.

Der öffentliche Sektor ist ebenfalls gut aufgestellt. Genomics England, eine Firma im Besitz des britischen Department of Health and Social Care, hat 122.000 Datensätze in ihrer Datenbank. Insgesamt sollen es fünf Millionen werden, davon 500.000 in Form genomischer Sequenzdaten. Ziel der Forscher ist, neue Erkenntnisse über Krebs und seltene Krankheiten zu gewinnen. Aus den USA kommt das All-of-Us-Forschungsprogramm (früher Precision Medicine Initiative Cohort Program). Es wurde im Jahr 2015 während der Amtszeit von Barack Obama ins Leben gerufen, um Fortschritte bei der medizinischen Versorgung zu erzielen. All-of-Us zielt darauf ab, Bioproben, genetische Daten und elektronische Patientenakten von einer Million Menschen zu sammeln. Weitere staatliche Initiativen sind länderspezifisch: Dubai Genomics plant, das Genom aller Einwohner zu sequenzieren, nennt aber keinen zeitlichen Horizont. Das Turkish Genome Project wiederum hat eine Million genomischer Daten zum Ziel – und zwar bis 2023, dem 100. Geburtstag der Türkei. In China setzt man beim Nanjing Project ebenfalls auf eine Million Datensätze.

Viel hilft viel: Techniken der Genomforschung

Welche Daten von Initiative zu Initiative gesammelt wer-

den, variiert ebenfalls stark. Nur die Sequenzierung des gesamten menschlichen Genoms liefert alle Informationen – auch für künftige wissenschaftliche Projekte. Diese Daten sind am besten geeignet, um Krankheitsmechanismen zu verstehen und um klinische Entscheidungen zu treffen. Bleibt als einziger Nachteil, dass es sich um vergleichsweise große Datensätze handelt.

Protein-codierende Regionen, die Exons, machen lediglich 1,5 Prozent davon aus. Sie werden bei der reinen Exom-Sequenzierung erfasst, was zu Lücken führt. Welche Aufgaben Introns als nicht codierende Sequenzen haben, ist teilweise unklar. Diese Bereiche werden zwar nicht in das Alphabet der Proteine übersetzt. Ihnen könnten dennoch wichtige Aufgaben zukommen, wenn es um die Frage geht, wie die Aktivität von Genen reguliert wird oder wie es zu Gen-Umwelt-Interaktionen kommt. Trotzdem reichen sie in vielen Fällen aus, um Personen mit einem höheren Krankheitsrisiko zu identifizieren und individuell zu behandeln.

Noch einfacher sind SNP-Chips. Sie erfassen nur Einzelnukleotid-Polymorphismen (Single Nucleotide Polymorphisms, SNPs). Mehr als 100 Millionen dieser SNPs wurden bisher identifiziert. Ein Chip erfasst normalerweise 500.000 bis 2,5 Millionen SNPs. Er liefert nur simple Ja-Nein-Antworten auf die Frage,

ob an einem bestimmten Ort eine Base im Erbgut ausgetauscht worden ist oder nicht. Solche Untersuchungen eignen sich für gezielte Fragestellungen, aber kaum für die Grundlagenforschung, da zu wenige Informationen generiert werden.

Letztlich führen fallende Kosten für Genomsequenzierung in Kombination mit höheren Rechnerleistungen, günstigen Massenspeichern und mit einem zunehmendem Automatisierungsgrad von Labors dazu, dass nach Möglichkeit immer das gesamte Genom untersucht wird.

Mehr Wissen: Genomische Daten mit Zusatzinformationen verknüpfen

Um den Nutzen zu beurteilen, spielt nicht nur die Art der Sequenzierung eine Rolle. Gibt es keine weiteren Verknüpfungen genomischer Daten mit Informationen zu Probanden, etwa zu deren Alter, Geschlecht, Ethnie, Vorerkrankungen oder Therapien, bietet eine Datenbank wenig Nutzen. Als ideal gelten Links zwischen Sequenzdaten und elektronischen Patientenakten.

Dazu ein paar Zahlen: Nur 42 Prozent aller Quellen in der IQVIA Genomic Initiatives Database haben eine gewisse Verknüpfung von Sequenzdaten mit Angaben zur Demographie oder zur Gesundheit. Und lediglich 28 Prozent der Initiativen stellten Verknüpfungen

ihrer Daten mit elektronischen Patientenakten her.

Das liegt in vielen Fällen an rechtlichen Hürden. Man braucht Einverständniserklärungen und muss die Daten anonymisieren oder pseudonymisieren. Der Aufwand ist hoch, und deshalb entscheiden sich Initiatoren vieler Projekte, ihre Quellen nur geschlossenen Nutzergruppen zugänglich zu machen. Poolt man Daten über mehrere Quellen, werden die Hürden noch höher. Hinzu kommt: Man braucht Standards zur Erfassung genomischer Informationen.

Von solchen Hürden lassen sich Forscher nicht abschrecken. Im Jahr 2018 gab es laut Europäischer Union in 13 Ländern Planungen, Daten zusammenzuführen, nämlich in Bulgarien, Kroatien, Zypern, in der Tschechischen Republik, in Estland, Finnland, Italien, Litauen, Luxemburg, Malta, Portugal, Slowenien, Spanien, Schweden und Großbritannien. Ziel ist, die jeweilige Bevölkerung möglichst repräsentativ abzubilden.

Um dies zu erreichen, brauchen bereits Kliniken eine geeignete Infrastruktur. Sie müssen Daten speichern, analysieren, interpretieren und extern oder intern teilen. Genau im letzten Schritt hapert es mitunter. Denn die meisten genomischen Initiativen arbeiten mit eigenen Protokollen und Systemen für die Handhabung ihrer genomischen Informationen. Dieser Mangel an Standardisierung kann Kompromisse beim Vergleich von Sequenzbefunden aus verschiedenen Labors erforderlich machen, was nicht unbedingt zu mehr Genauigkeit führt. Sowohl bei der Technik selbst als auch bei der Annotation genomischer Daten geht mittelfristig nichts ohne Standardisierung.

Ein Qualitätsindikator für genomische Datenbanken

Damit wird klar: Viele unterschiedliche Parameter bestimmen, welchen Wert Sequenzdaten einer bestimmten Initiative haben. Um Projekte zu bewerten, hat das

IQVIA Institute deshalb ein spezielles Scoring entwickelt.

Jede Initiative wurde je nach Art ihrer Daten mit 1 bis 5 bewertet, wobei die Sequenzierung des gesamten Genoms zur höchsten (5) und die Genotypisierung zur niedrigsten Punktzahl (1) führt. Zwischenwerte von 4, 3 und 2 wurden bei biologischen Proben, Sequenzierungen des gesamten Exoms und gemischten Ansätzen vergeben.

Die Verknüpfung mit Patientendaten ist ein weiteres Kriterium zur Bewertung. Für Links zu klinischen Daten gibt es 4 Punkte und für Verknüpfungen zu Patienten-Fragebögen 2 Punkte. Ohne solche zusätzlichen Informationen wurde kein Punkt vergeben.

Beide Teilskalen führen in Summe zu maximal 10 Punkten – und zwar bei einer Sequenzierung des ganzen Genoms und bei Verknüpfungen zu elektronischen Patientenakten. Ausgewählte Ergebnisse zeigt Abbildung 3.

Fallbeispiel: Genomische Daten und COPD

Genomische Initiativen, die hohen Standards genügen, bringen die biomedizinische Forschung rasch voran, wie Fallbeispiele zeigen.

COPD ist die dritthäufigste Todesursache weltweit. Man kann die Lungenkrankheit momentan nicht heilen, und Prognosen zum Verlauf sind schwierig. Beide Aspekte sollten mit genomweiten Assoziationsstudien (GWAS) untersucht werden. Basis waren 48.943 Datensätze aus der UK Biobank mit 500.000 Personen. Sowohl genomische Daten als auch klinische Spirometrie-Daten lagen vor. Anhand dieser Quelle untersuchten Forscher Assoziationen von 27.624.732 Varianten im Erbgut mit der Lungenfunktion. Weitere Quellen aus China kamen hinzu, um auch außereuropäische Varianten im Genom zu berücksichtigen.

Letztlich fanden Wissenschaftler 81 genetische Vari-

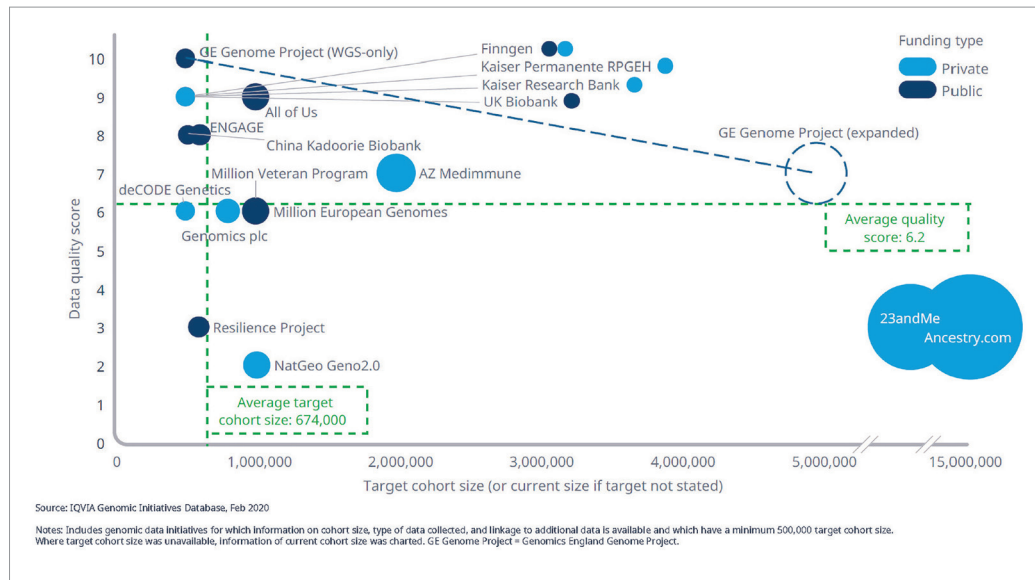


Abbildung 3: Übersicht zu Qualitätsindikatoren bei genomischen Initiativen; Quelle: IQVIA

ten mit Assoziation zur Einsekundenkapazität (FEV1, Forced Expiratory Pressure in 1 Second). FEV1 ist ein diagnostischer Parameter für Obstruktionen der unteren Atemwege.

Die zweite Stufe umfasste 95.375 Datensätze von Personen aus der britischen Biobank, dem SpiroMeta-Konsortium und der UK Households Longitudinal Studie (UKHLS). Ziel war, die bereits identifizierten Varianten hinsichtlich ihrer COPD-Suszeptibilität zu bewerten.

Das Forscherteam fand 43 bislang unbekannte Varianten im Erbgut, welche mit der Lungenfunktion assoziiert sind. Damit stieg die Zahl solcher Signale auf insgesamt 97. Solche Besonderheiten im Erbgut stehen mit einem bis zu vierfach erhöhten COPD-Risiko in Verbindung.

Alle identifizierten Varianten beeinflussen 234 Gene, die für unterschiedliche biochemische Prozesse codieren. Dazu zählen unter anderem Stoffwechselwege im Bereich elastischer Fasern in der Lunge, der Hedgehog-vermittelten Signalübertragung, epigenetische Regulationswege und der Inositolphosphat-Stoffwechselweg. Solche Informationen helfen Arzneimittelherstellern, Proteine als Targets für neue Pharmaka zu entwickeln. Alle derzeit zugelassenen Medikamente und solche in der Entwicklung zielen auf die Gen-

produkte von sieben dieser 234 Gene ab.

Fallbeispiel: Innovative Studiendesigns mit genomischen Daten

Mit ähnlichen Strategien versuchen Onkologen, neue Therapien beim nicht-kleinzelligen Lungenkrebs (NSCLC) zu finden, beispielsweise anhand des Lung Cancer Master Protocol (Lung-MAP). Die Studie wurde im Juni 2014 als öffentlich-private Partnerschaft vom US National Cancer Institute, von den Friends of Cancer Research, von Pfizer, Genentech/Roche, AstraZeneca, Amgen, Bristol-Myers Squibb und Foundation Medicine gestartet. Ziel ist, personalisierte Therapien schneller verfügbar zu machen.

Die Tumore aller Studienteilnehmer wurden mittels Sequenzierung der nächsten Generation auf spezifische genetische Veränderungen untersucht. Wer Biomarker hatte, für die sich experimentelle Therapien eignen könnten, wurde in eine entsprechende Teilstudie aufgenommen. Patienten, deren Tumor nicht zu einem Biomarker-spezifischen Studienmedikament passte, konnten in andere Teilstudien aufgenommen werden. Sie erhielten Immuntherapien oder Medikamente, von denen ein gewisser Effekt über molekulare Subtypen hinweg erwartet wurde.

Bislang haben Forscher 2.800 Patienten aufgenommen, von denen man für 799 experimentelle

Therapien im Rahmen klinischer Studien finden konnte. Zwischen Juni 2014 und Februar 2020 wurden Substudien mit zwölf neuen Therapieprinzipien abgeschlossen. Sieben weitere sollen zusammen mit forschenden Herstellern realisiert werden.

Daten allein sind nur die halbe Miete

Wie geht es weiter? Bis 2025 soll das Genom von 50 bis 100 Millionen Menschen sequenziert worden sein. Diese große Zahl darf nicht darüber hinwegtäuschen, dass es starke Unterschiede in der Qualität und Verfügbarkeit dieser Daten geben wird. Initiativen, die mit genomischen Daten arbeiten, sollten darauf achten, technische Standards einzuhalten, aber auch Verknüpfungen mit weiteren Informationen wie elektronischen Patientenakten herzustellen. Nur so lässt sich der größtmögliche Nutzen aus Datenbanken ziehen. <<

Quelle

IQVIA: Understanding the Global Landscape of Genomics Initiatives. Progress and Promise. Mai 2020. Online unter: <https://www.iqvia.com/insights/the-iqvia-institute/reports/understanding-the-global-landscape-of-genomic-initiatives>

Kontakt: Dr. Gisela Maag, Pressesprecherin/IQVIA
 E-Mail: gisela.maag@iqvia.com