



Pressemitteilung

22. Juni 2010

Pilotstudien des 1.000 Genome-Projekts abgeschlossen

Internationales Forschungskonsortium vollendet erste Schritte zur Erstellung einer Datenbank mit 2.500 menschlichen Genomen

Das 1.000 Genome-Projekt ist ein internationales Projekt, in dem Wissenschaftler aus öffentlich finanzierten Forschungseinrichtungen und der Industrie gemeinsam daran arbeiten, eine genaue Karte der genetischen Unterschiede der Menschen zu erstellen. Am Montag, dem 21. Juni 2010 haben die Forscher des 1.000 Genome-Projektes den erfolgreichen Abschluss ihrer drei Pilotstudien bekannt gegeben. Die Daten stehen frei in öffentlichen Datenbanken zur Verfügung. Damit beginnen die Arbeiten an der Hauptphase des Projektes mit dem Ziel, eine öffentliche Datenbank mit genomischen Informationen von 2.500 Individuen aus weltweit 27 verschiedenen Bevölkerungsgruppen zu erstellen.

Anfang 2008 wurde das 1.000 Genome-Projekt mit dem Ziel ins Leben gerufen, eine genaue Karte der genetischen Unterschiede im Erbmateriale verschiedener Individuen zu erstellen. In drei Pilotstudien sollten zunächst Strategien entwickelt werden, um Kataloge derjenigen genetischen Varianten zu erstellen, die mit einer Häufigkeit von über 1% in verschiedenen Volksgruppen auftreten. Solche Kataloge helfen Forschern, den Einfluss bestimmter genetischer Veränderungen auf verschiedene Erkrankungen besser einschätzen zu können. Damit können auch Aussagen über das individuelle Risiko einzelner Patienten für bestimmte Erkrankungen bzw. ihre Reaktion auf bestimmte Medikamente besser vorhergesagt werden. Die Ergebnisse der Pilotstudien sind frei auf der Webseite des 1.000 Genome-Projekts zugänglich. Darüber hinaus können sie aber auch über die Amazon Elastic Compute Cloud (EC2) abgerufen werden. Damit ist dieser beispiellos große Datensatz auch für Wissenschaftler zugänglich, die nicht über genügend große Computerkapazitäten verfügen, um sich die Daten für anschließende Analysen lokal herunterladen zu können.

Das Max-Planck-Institut für molekulare Genetik (MPIMG) in Berlin ist als eines von zwei europäischen Instituten an dem internationalen Schlüsselprojekt der Genomforschung beteiligt. „Obwohl das 1.000 Genome-Projekt noch lange nicht abgeschlossen ist, wird schon jetzt deutlich, dass hier eine universelle Grundlage für alle folgenden, krankheitsorientierten Genomprojekte geschaffen wird. Gegenwärtig befinden sich bereits Projekte in der Startphase, die auf der vergleichenden Genomsequenzierung ganzer Patientenkohorten basieren“, sagt Ralf Sudbrak, Projektleiter des deutschen Beitrages am 1.000 Genome-Projekt. Und Hans Lehrach, Mitglied im Lenkungsgremium des 1000 Genome-Projektes und Direktor der Abteilung des MPIMG, in der die Sequenzierungsarbeiten stattfinden, fügt hinzu: „Die Erfahrungen aus diesem Projekt und die rasante technologische Entwicklung ermöglichen bereits jetzt Initiativen wie das von uns initiierte Treat1000-Projekt, das neue Möglichkeiten für eine personalisierte Medizin schaffen soll. In diesem Projekt sollen in den nächsten Jahren die Genome von tausend Tumorpatienten sowie das veränderte Genmaterial ihrer Tumore sequenziert werden.“



In der jetzt beginnenden, zwei Jahre währenden Hauptphase des 1000 Genome-Projektes werden die genetischen Informationen von 2.500 Individuen aus 27 verschiedenen Bevölkerungsgruppen untersucht. Alle Studienteilnehmer haben dabei einer Veröffentlichung ihrer Daten bereits vorab zugestimmt. Die Forscher hoffen, dass die freie und öffentliche Verfügbarkeit dieser Daten die Entwicklung neuer Methoden und Ansätzen der genetischen Forschung sehr viel schneller vorantreiben wird, als dies bei der Veröffentlichung von ausschliesslich krankheitsbezogenen – und damit wesentlich stärker geschützten - Informationen der Fall wäre. „Wir sind verpflichtet, unsere Daten der Öffentlichkeit zur Verfügung zu stellen. So wollen wir sicher gehen, dass jedes Institut und jeder Wissenschaftler weltweit mit unserem Datensatz arbeiten kann, um Volkskrankheiten besser verstehen zu lernen“, erklärt Sudbrak.

Die durch das 1000 Genome-Projekt produzierte Datenmenge ist bislang einmalig in der biomedizinischen Forschung. Zur Zeit beträgt die Größe des Datensatzes ungefähr 50 Terabyte oder 50.000 Gigabyte. Das entspricht fast 7,3 Billionen (eine Eins mit 12 Nullen) DNA-Buchstaben oder Terabasen an Sequenzdaten. Wissenschaftler haben freien Zugriff auf die Pilotdaten des 1.000 Genome-Projektes durch die Projektwebseite und können sich die Daten über das NCBI (<ftp://ftp-trace.ncbi.nih.gov/1000genomes/>) oder das EBI (<ftp://ftp.1000genomes.ebi.ac.uk/>) herunterladen.

Ermöglicht wird die deutsche Beteiligung am 1.000 Genome-Projekt durch Unterstützung des Bundesministeriums für Bildung und Forschung, welches die Beteiligung der Berliner Forscher an dem internationalen Großprojekt im Rahmen des Programms "Integrierte Verbünde der medizinischen Genomforschung NGFN-Plus" fördert.

Weitere Informationen:

<http://www.1000genomes.org/> - Projektseite des 1.000 Genome-Projekts

<http://www.treat1000.org/> - Projektseite des treat1000-Projekts

<http://www.molgen.mpg.de> – Homepage MPI für molekulare Genetik

Kontakt:

Dr. Ralf Sudbrak

Max-Planck-Institut für molekulare Genetik

Ihnestr. 63-73

14195 Berlin

Tel.: +49 30 8413-1612

Fax: +49 30 8413-1380

Email: sudbrak@molgen.mpg.de

HINTERGRUNDINFORMATION

Die Pilotstudien des 1.000 Genome-Projektes - Abschätzung der grundlegenden Aspekte zur Durchführbarkeit des Projekts

In der **ersten Pilotstudie** wurde die genetische Information von sechs Individuen (zwei sogenannte Kernfamilien mit jeweils beiden Elternteilen und einem Nachkommen) mit einer sehr hohen Abdeckung sequenziert. Dies bedeutet, dass jeder Buchstabe der DNA durchschnittlich 20 bis 60 mal abgelesen wurde, um eine hohe Genauigkeit der resultierenden DNA-Sequenz zu gewährleisten. Die sechs Genome wurden mit den unterschiedlichen zur Verfügung stehenden Technologien von akademischen Instituten in den USA, China, Großbritannien und Deutschland sowie von den drei Firmen 454 Life Sciences/Roche, Applied Biosystems/Applera Corp. Gesellschaft und Illumina Inc. sequenziert.

In der **zweiten Pilotstudie** wurden die Genome von 179 Individuen mit einer wesentlich niedrigeren Abdeckung sequenziert. Im Durchschnitt wurde jeder Buchstabe dreimal gelesen. Obwohl die Sequenzierkosten stetig fallen, ist die vollständige Sequenzierung von hunderten von Genomen immer noch sehr kostspielig. Ein Ansatz zur Bestimmung der genetische Varianten, die mit einer bestimmten Häufigkeit in einer Volksgruppe auftreten, besteht darin, viele Genome mit einer geringen Abdeckung zu analysieren. Anschließend werden die erhaltenen Daten nach genetischen Varianten durchsucht, die bei mehreren Individuen auftreten. Die Ergebnisse dieser Pilotstudie konnte zeigen, dass diese Strategie effektiv und zielführend ist.

In der **dritten Pilotstudie** wurden die kodierenden Regionen („Exone“) von 1.000 Genen in 700 Individuen untersucht. Ziel war es, herauszufinden, wie am besten ein sehr genauer Katalog von genetischen Varianten in den Protein-kodierenden Genen erstellt werden kann. Die Pilotstudie lieferte durch eine beispiellose Probenzahl Muster von sehr seltenen Varianten in der menschlichen Bevölkerung.

Die Ergebnisse aller drei Pilotstudien zeigen, dass die neuen Sequenzierungstechnologien den Zugang auch zu seltenen bzw. weniger häufigen Varianten ermöglichen. Das Sequenzierungsvolumen belief sich auf über 3,8 Terabasen. Dies bedeutet, dass bereits in der Pilotphase eine 1000fache Abdeckung des humanen Genoms erreicht wurde.