

# Biochemie und Molekulargenetik 2000

**Die vollständige Sequenzierung von Chromosom 21 war einer der Höhepunkte in der Humangenomforschung. Die Genome von weiteren Organismen wurden mit dem immer umfangreicheren Methodenspektrum der RNA-Analytik entschlüsselt. Weitere Forschungsschwerpunkte waren: Wie verläuft die komplexe Regulation beim programmierten Zelltod (Apoptose)? Und: Welche Rolle spielt das Sexualhormon Estrogen beim Schutz vor neurodegenerativen Erkrankungen?**

## Das Humangenomprojekt

Das Jahr 2000 war ein besonderes Jahr für das Humangenomprojekt (HGP), da erstmalig eine Rohfassung („working draft“) der menschlichen Genomsequenz erstellt wurde. Eine endgültige Version soll bis 2003 fertig gestellt werden. Die vollständige Sequenzierung der menschlichen Chromosomen 21 und 22 und die Verfügbarkeit über eine Rohfassung des restlichen menschlichen Genoms ließen erstmalig fundierte Rückschlüsse über die Gesamtanzahl der menschlichen Gene zu. Diese wird nur noch auf ca. 25000 bis 40000 beziffert, im Vergleich zu früheren Schätzungen von ca. 700000 bis 140000 Genen.

### Rohfassung der Humangenomsequenz

1990 startete das HGP. Ursprünglich sollte die komplette Sequenz des menschlichen Genoms bis zum Jahre 2005 entschlüsselt werden. In der neuen Zielsetzung des HGP für die Jahre 1998 bis 2003 rechnet man mit der fertigen Sequenz im Jahre 2003 oder noch früher.<sup>1)</sup> 1998 waren gerade 3 % des Genoms sequenziert gewesen (Abbildung).<sup>2)</sup> Rund 60 % des Genoms sollen von mehreren US-Sequenzierzentren entschlüsselt werden, finanziert vom National Institute of Health (NIH) und dem US-Department of Energy (DoE). Das britische Sanger-Sequenzierzentrum, im wesentlichen finanziert vom Wellcome Trust, Imperial Cancer Research Fund und Medical Research Council (MRC), soll ca. 26 % entschlüsseln. Den Rest teilen sich Sequenzierzentren in Deutschland (finanziert vom Bundesministerium für Bildung und Forschung, BMBF), in Frankreich, Japan und China.<sup>3)</sup>

Die dem internationalen HGP-Konsortium zugehörigen Institute führen eine gezielte Sequenzierung von bereits physikalisch kartierten Klonen, wie BACs (bacterial artificial chromosomes) und PACs (PI-derived artificial chromosomes) durch. Etwa 24000 sich überlappende und geordnete Klone decken das menschliche Genom ab. Ein BAC umfasst im Durchschnitt ca. 100 bis 120 Kb. Diese werden nach dem Zufallsprinzip („Shotgun-Technik“, siehe Kasten, S. 324) fragmentiert, kloniert, sequenziert und die Sequenz anschließend mit Hilfe von computerunterstützten Programmen in der richtigen Reihenfolge zusammengefügt.<sup>4)</sup>

1998 kündete Craig Venter, Direktor von Celera Genomics Systems of Rockville, Maryland/USA, an, dass das private Unternehmen im Alleingang, in nur drei Jahren das komplette menschliche Genom entschlüsseln werde. Venter wählte eine andere Sequenzierstrategie, und zwar das Gesamtgenom-Shotgun-Verfahren. Mitarbeiter des Unternehmens zerlegten das gesamte menschliche Genom in kleine Stücke, klonierten diese und sequenzierten sie von beiden Enden. Die erhaltenen Daten wurden in Hochleistungscomputern gespeichert und assembliert (d.h. in die richtige Reihenfolge gebracht).<sup>3)</sup>

Ein harter Wettkampf zwischen dem Internationalen HGP-Konsortium, geleitet von Francis Collins, und der privaten Celera hatte dem HGP ein atemberaubendes Tempo verliehen. Dazu kam die Verfügbarkeit von Hochdurchsatz-Kapillar-Sequenziergeräten (Typ 3700 von Applied Biosystems). Am 26. Juni 2000 haben die beiden Kontrahenten unter der Schirmherrschaft von Bill Clinton und Tony Blair der Öffentlichkeit mitgeteilt, dass eine Rohfassung des Humangenoms fertig gestellt wurde.<sup>3)</sup> In der Rohfassung muss jedes Basenpaar mindestens fünf mal, in der endgültigen Version mindestens zehn mal sequenziert worden sein. Die Rohfassung ergibt eine Sequenz mit einer Genauigkeit von 90 %, wobei noch nicht alle sequenzierten Stücke korrekt assembliert und lückenlos sind. Die fertige Version soll dagegen eine Richtigkeit von 99,99

% aufweisen, das entspricht maximal einem Fehler pro 10000 Basenpaaren. Celera hatte 99 % des Genoms 4,6 mal sequenziert, das Internationale HGP Konsortium 87 % des Genoms 7 mal, wobei 40 % davon bereits korrekt physikalisch assembliert sind, also in ihrer tatsächlichen Reihenfolge vorliegen.<sup>5)</sup>

Der wesentliche Unterschied zwischen den beiden Sequenzier-Parteien war, dass das Internationale HGP-Konsortium aufgrund des „Bermuda“-Abkommens die produzierten Daten innerhalb von 24 Stunden der Öffentlichkeit zur Verfügung stellen musste, während die Daten von Celera nicht öffentlich zugänglich sind.<sup>6)</sup> Das „Bermuda“-Abkommen hat vielen Wissenschaftlern eine immense Erleichterung und hohe finanzielle Ersparnisse bei der Suche nach Genen gebracht.

Das Humangenom umfasst zu ca. 40 % hochrepetitive Sequenzen, die eine korrekte Zusammenstellung von Shotgun-Klonen sehr erschweren.<sup>7,8)</sup> Celera hatte bereits im Frühling 2000 mit Hilfe des Gesamtgenom-Shotgun-Verfahrens das eukaryontische Genom von *Drosophila melanogaster* (120 Kb) sequenziert.<sup>9,10)</sup> Um allerdings eine möglichst lückenlose und korrekte Assemblierung der Fragmente zu erzielen, war eine Zusammenarbeit mit dem Berkeley-*Drosophila*-Genome-Project notwendig, das eine geordnete Sequenzierstrategie wie auch das HGP-Konsortium eingeschlagen hatte.

Für die fertige Version des Humangenoms wird voraussichtlich, ähnlich wie für *Drosophila melanogaster*, eine Zusammenarbeit zwischen dem Internationalen HGP-Konsortium und Celera unabdingbar sein.<sup>11)</sup>

## Sequenzierung der Chromosomen 21 und 22

Im Dezember 1999 war das menschliche Chromosom 22, das zweitkleinste Chromosom im Genom vollständig sequenziert.<sup>7,12)</sup> Im Mai 2000 wurde die komplette Sequenzierung des kleinsten menschlichen Chromosoms, des Chromosoms 21, bekannt gegeben.<sup>8,13)</sup> Die Tabelle zeigt eine Übersicht der wichtigsten Daten, die bei der Sequenzierung der beiden Chromosomen erhalten wurden.

Die Chromosomen 21 und 22 machen ungefähr 2 % des gesamten Genoms aus. Während das Chromosom 22 genreich ist, ist das Chromosom 21 genarm. Zusammen enthalten sie ca. 770 Gene (siehe Tabelle). Vorausgesetzt, dass diese Anzahl einem durchschnittlichen Wert an Genen entspricht, würde das menschliche Genom ungefähr 40000 Gene enthalten.<sup>7,8)</sup> Dies würde den bisherigen Schätzungen von 70000 bis 140000 Genen drastisch widersprechen.<sup>14)</sup> Andere Genanzahl-Vorhersageprogramme, basierend entweder auf der gesamten Rohfassung des humanen Genoms, der vergleichenden Genomanalyse oder auf den ESTs (expressed sequenced tags) schätzten die Anzahl der Gene auf nur 27 700 bis 35000.<sup>15)</sup>

## Sequenzierung weiterer Genome

In der Zwischenzeit wurde klar, dass die Vorhersageprogramme für Gene unzureichend sind. Ca 20 % der Exons (codierender DNA-Bereich) werden nicht erkannt, und 30 % der vorhergesagten Exons können experimentell nicht nachgewiesen werden.<sup>7-9)</sup> Eine wesentliche Hilfe bei der Identifizierung von Genen könnte allerdings die vergleichende Analyse von unterschiedlichen Genomen (comparative genomics) sein.<sup>16)</sup>

Bisher wurden drei eukaryontische Organismen komplett sequenziert, und zwar die Hefe (*Saccharomyces cerevisiae*),<sup>17)</sup> der Fadenwurm (*Caenorhabditis elegans*)<sup>18)</sup> und die Taufliege (*Drosophila melanogaster*).<sup>9,10)</sup> Die Sequenz weiterer Eukaryonten soll in den nächsten ein bis drei Jahren entschlüsselt werden und zwar von Maus, Ratte, Schwein, Pufferfisch, Zebrafisch und Reis. Diese Sequenzierdaten werden bei der Identifizierung von bisher unbekanntem Genen und regulatorischen Sequenzen eine entscheidende Rolle spielen.<sup>16)</sup>

## Human Genome Diversity Project

Natürlich vorkommende Sequenzvariabilität ist eine essentielle Eigenschaft eines jeden Genoms. Zwei haploide humane Genome unterscheiden sich im Durchschnitt in mindestens einem von 1000 Basenpaaren.<sup>19)</sup> Zeigt der Unterschied eine Häufigkeit von mehr als 1% in der Bevölkerung, spricht man von einem Polymorphismus. Neben Einzelbasenaustauschen (single nucleotide polymorphism, SNPs, [Nachr. Chem. 2000, 48, 1342]) gibt es viele andere Sequenzunterschiede im Genom, die beispielsweise durch Insertionen, Duplikationen, Deletionen, Rearrangements entstehen, aber viel seltener vorkommen.

Befindet sich ein SNP in einem kodierenden oder regulatorischen Bereich, kann dieser zu einer phänotypischen Variabilität führen. Viele der SNPs liegen allerdings in irrelevanten genomischen Bereichen, können jedoch aufgrund der Nähe zu einem bestimmten Gen als Marker bei der Aufdeckung von Suszeptibilitätsloci dienen. In den nächsten drei Jahren soll eine dichte Karte mit SNPs erstellt werden. Das ursprüngliche Ziel des SNP-Konsortiums, das aus dem Wellcome Trust und mehreren Pharma-Firmen besteht, war es, 1 SNP pro 10 Kb zu identifizieren. Demnach müssten ca. 300000 SNPs gefunden werden.<sup>20)</sup> Neueste Assoziationsstudien, basierend auf SNP-Analysen, wie bei Morbus Alzheimer zeigten, dass unter Umständen sehr viel mehr SNPs erforderlich sind.<sup>21)</sup> Anfang September 2000 meldete das SNP-Konsortium die Identifizierung von über 800000 SNPs.<sup>22)</sup> Das neue Ziel ist nun eine Karte mit 1 SNP pro 3 bis 5 Kb.<sup>20,22)</sup> Kurz darauf meldete Celera die Identifizierung von 2,5 Mio. SNPs durch Sequenzierung von fünf Personen mit ethnisch verschiedener Abstammung.<sup>23)</sup>

Zwei Methoden stellten sich für die Identifizierung von SNPs als sehr effizient heraus:

- die „Reduced representation shotgun(RRS)“-Strategie und
- die „Genomic-alignment“-Strategie.

Bei der RRS-Strategie verwendet man die DNA von mehreren nicht-verwandten Personen, schneidet sie mit einem bestimmten Restriktionsenzym (Hind III oder Bgl II), trennt sie gelelektrophoretisch auf, eluiert die Fragmente einer bestimmten Größe (z. B. 500 bp), erstellt eine „reduzierte“ Genbank und sequenziert die entsprechenden Klone. In der Regel kommt jeder Klon mindestens zwei mal vor. Unterscheiden sich die Sequenzen in einem Basenpaar, in einer ansonsten eindeutig lesbaren Sequenz, entspricht dies in ca. 95 % der Fälle einem SNP.<sup>24,25)</sup>

Die zweite Methode ist viel einfacher, setzt aber eine endgültige Sequenz des menschlichen Genoms voraus. Diese Methode wurde beispielhaft für die Identifizierung von SNPs vom Chromosom 22 angewandt. Eine einmalige Sequenzierung von Chromosom 22-spezifischen Shotgun-Klonen und der Vergleich mit der vollständig vorhandenen Sequenz des Chromosoms 22 machte die Aufdeckung von 1 SNP pro 12 Kb möglich.<sup>24)</sup>

Die Variabilität im menschlichen Genom wird demnächst mehr und mehr bei der Eingrenzung vorwiegend von genetisch komplexen Erkrankungen Anwendung finden. Weiterhin geht man davon aus, dass die SNPs in den kodierenden Bereichen, Promotoren oder Spleißerkennungssequenzen essentiell für die Pharmakogenetik und Entwicklung von genotypspezifischen Therapien sein werden.

*Brunhilde Wirth, Institut für Humangenetik, Universität Bonn E-Mail bwirth@uni-bonn.de*

- 1) F. C. Collins, Science 1999, 282, 682.
- 2) <http://www.ebi.ac.uk/~sterk/genome-MOT/index.html>
- 3) E. Marshall, Science 1999, 284, 1906
- 4) <http://www.sanger.ac.uk/HGP/draft2000/>
- 5) C. Macilwain, Nature, 2000, 405, 983.
- 6) D. R. Bentley, Science 1996, 274, 533.
- 7) I. Dunham et al., Nature 1999, 402, 489.
- 8) M. Hattori et al., Nature 2000, 405, 311.
- 9) M. A. Adams et al., Science 2000, 287, 2185.
- 10) E. W. Myers et al., Science 2000, 287, 2196.
- 11) P. Smaglik, D. Butler, Nature 2000, 404, 119.
- 12) <http://www.sanger.ac.uk/HGP/Chr22>; <http://www.genome.ou.edu/Chr22.html>
- 13) <http://hgp.gsc.riken.go.jp/chr21/index.html>; <http://chr21.rz-berlin.mpg.de/>
- 14) C. Fields, Nature Genet. 1994, 7, 345.
- 15) E. Pennisi, Science 2000, 288, 1146.
- 16) G. M. Rubin et al., Science 2000, 287, 2204.
- 17) The C. elegans Sequencing consortium, Science 1998, 282, 2013.
- 18) A. Goffeau et al., Science 1996, 274, 546.
- 19) D. G. Wang, Science 1998, 280, 1077.
- 20) <http://snp.cshl.org/>
- 21) E. R. Martin et al., Am. J. Hum. Genet. 2000, 67, 383.
- 22) L. Roberts Science 2000, 287, 1898.
- 23) <http://www.celera.org/>
- 24) J. C. Mullikin et al., Nature 2000, 407, 516.
- 25) D. Altshuler et al., Nature, 2000, 407, 513.
- 26) <http://www.ncbi.nlm.nih.gov/omim/>