

Genomforschung

Das größte und wichtigste Projekt der Biomedizin in unserer Zeit ist wohl das Humangenomprojekt (HGP). Das erst 1990 begonnene Projekt, das weltweit vorwiegend akademische Institutionen durchführen und international koordinieren, hat bereits zu einer Vielzahl an neuen Erkenntnissen geführt. Das HGP wird in Zukunft sowohl die Biologie als auch die Medizin tiefgreifend verändern. Darüber hinaus initiiert das Projekt neue Technologien und neue Ideen. Ein ganzes neues Fachgebiet ist entstanden: Genomics.

Das Humangenomprojekt

Hauptziel des Humangenomprojekts war ursprünglich die vollständige Sequenzierung des menschlichen Genoms bis zum Jahre 2005. Da die Entzifferung des Genoms schneller als erwartet voranging und die vorläufige Fünfjahresplanung bereits im Jahr 1998 erfüllt wurde, erwartet man das Ende des Projekts zwei Jahre früher, nämlich im Jahr 2003,¹⁾ passend zur 50-jährigen Feier der Entdeckung der DNA durch James D. Watson und Francis H. Crick.²⁾ Um dennoch die Erwartungen aller Wissenschaftler zu erfüllen, soll die Qualität der Daten (1 Fehler je 10 000 Basen) entsprechend hoch gehalten werden.

Noch im Jahr 1999 wurde die Planung des HGP erneut ergänzt:³⁾ so soll eine „Arbeitsfassung“ des Genoms schon im Frühjahr 2000 geliefert werden. Diese „Rohfassung“ soll bereits das Gerüst des Humangenoms enthalten. Dies bedeutet eine 4- bis 5-fache redundante Sequenzierung von 90 % des Genoms, die noch viele Lücken enthält; dennoch wird diese Rohfassung sehr hilfreich sein, um nach bestimmten Sequenzmotiven zu suchen und um Gene zu identifizieren. Weiter wird sie erstmals die Kartierung von Polymorphismen ermöglichen. Neue Methoden (Kapillarelektrophorese) und die Automatisierung verschiedener Schritte der Sequenzierung ermöglichten diese Beschleunigung des HGP.

Inzwischen wurden zusätzlich zu den ursprünglichen Zielen des Humangenomprojekts in den neuen Plänen auch weiterführende Ziele definiert (Tabelle).

Parallel zu dem Humangenomprojekt gibt es Initiativen von privaten Unternehmen (z. B. Celera Genomics/ USA), die ebenfalls mit Sequenzierungsprojekten eine Fülle von Informationen gewinnen, die die öffentlichen Daten in Zukunft ergänzen werden.

Zu beobachten ist eine enorme Zunahme der Informationen: So wurde bereits Anfang Dezember 1999 die komplette Sequenzierung des ersten menschlichen Chromosoms (Chromosom 22) veröffentlicht.⁴⁾ Einige Wochen vorher hatte man in Washington D.C. die Sequenzierung der milliardsten Base gefeiert.

Der Stand des HGP Ende November:^{5,6)} Von 14,7 % des humanen Genoms war die komplette Sequenz (473 740 kb) bekannt, von 21,8 % lagen die „Rohfassungssequenzen“ vor (697 540 kb), also waren insgesamt 36,5 % des humanen Genoms sequenziert.

SNPs

Jedes Genom unterscheidet sich von den anderen: die häufigsten Variationen sind Unterschiede in einzelnen Basen, die SNPs (ausgesprochen „snips“ – „single nucleotide polymorphisms“).⁷⁾ Im Durchschnitt gibt es zwischen zwei (haploiden) Genomen ein SNP pro 1000 Basen.

Die Verteilung der SNPs in Populationen und ihr Vergleich mit dem Auftreten von Krankheiten kann in Zukunft essentielle Informationen über die genetische Ursache der Erkrankungen liefern. Je häufiger solche Marker und ihre Positionen bekannt sind, desto leichter können Korrelationen mit genetischem Material aus Menschen hergestellt werden.⁸⁾

SNPs können auch bei der Identifizierung der genetischen Unterschiede helfen, die einige Menschen anfälliger für bestimmte Krankheiten machen als andere oder unterschiedliche Reaktionen auf Medikamente hervorrufen.

Man erkannte auch die Bedeutung von SNPs für die Entwicklung sicherer und effektiverer Medikamente. So haben zehn größere Pharmaunternehmen mit dem „Wellcome Trust“ ein internationales Konsortium gegründet,⁹⁾ das sich zum Ziel gesetzt hat, mehrere bekannte Laboratorien weltweit zu unterstützen, um 300 000 SNPs zu identifizieren und zu kartieren. Bis Mitte 2001 plant man die Kartierung von mindestens 150 000 SNPs. Diese Information wird in einer öffentlichen Datenbank hinterlegt (dbSNP),¹⁰⁾ die bereits bis heute 2500 kartierte SNPs enthält – 15 000 weitere Kandidaten-SNPs sind bisher vom diesem Konsortium identifiziert. Die dafür verwendete DNA stammt von 24 Personen, die unterschiedlichen ethnischen Gruppen angehören.

Die Techniken, die bei Identifizierung und Bewertung der SNPs Anwendung finden, entwickeln sich kontinuierlich weiter. Jedoch ist es wichtig, parallel dazu Methoden auszuarbeiten, um solche SNP-Karten (Genotyp) mit dem Erscheinungsbild (Phänotyp) der verschiedenen Individuen vergleichen zu können.

Funktionsanalyse

Die neuen Daten über die Zusammensetzung der Gene stellen uns die Frage nach deren Funktion. Bisher waren es Wissenschaftler gewöhnt, aus der betrachteten Funktion nach der Ursache – dem Protein oder dem Gen – zu suchen. Nun ist der Ansatz umgekehrt und erfordert nicht nur eine gedankliche, sondern auch eine methodische Umstellung. Um dies in einem „genomischen“ Maßstab zu erreichen, kann man folgende Vorgehensweisen verfolgen:

- Sequenzvergleiche (Analyse von Mustern und Homologien mit bekannten Proteinen und deren Funktionen),
- Analyse der Genexpression in hohem Durchsatz (auf RNA- oder Proteinebene),
- Verschiedene Ansätze der Ausschaltung der Gene.

Für die Analyse der Expression auf RNA- und DNA-Ebene hat in den letzten Jahren die Entwicklung der Chiptechnologie eine große Rolle gespielt.¹¹⁾ DNA-Chips sind miniaturisierte Vorrichtungen, die in einer matrix-ähnlichen Anordnung synthetische Oligonukleotide oder cDNA-Fragmente enthalten. Dies ermöglicht es, eine große Anzahl von Hybridisierungsexperimenten parallel durchzuführen und Genexpression zu untersuchen. Dabei wird eine fluoreszenz- oder radioaktivmarkierte Sonde aus biologischem Material, beispielsweise aus exprimierter mRNA, auf dem Chip getestet: die komplementären Stränge hybridisieren auf dem Chip und liefern ein auswertbares Signal.

Zur Zeit überwiegen zwei Techniken: die eine basiert auf In-situ-Synthese von Oligonukleotiden auf einem Chip durch photolithographische Verfahren und die andere auf dem automatisierten „Spotting“ von cDNA-Fragmenten auf der Chipoberfläche. Weitere Chiptechniken sind in der Entwicklung. Die wichtige Rolle der DNA-Chips für die unterschiedlichen Anwendungen (z. B. auch für SNP-Analysen) in der Zukunft ist dabei unumstritten.

Die optimale Auswertung solcher Expressionsexperimente mit DNA-Microarrays stellt ebenfalls eine Herausforderung dar; sie soll es ermöglichen, daß man aus der Fülle der Daten Informationen über die Funktionen der Gene gewinnt.¹²⁾ Beispielhaft für einen solchen Ansatz war die Sequenzierung des Hefegenoms und die Analyse von dessen Expression.^{13,14)}

Die Komplexität dieser Ansätze ist bei höheren Organismen viel größer. Einige Untersuchungen haben sich daher auf bestimmte Zelltypen und Gewebe konzentriert. So könnten interessante Erkenntnisse über wichtige Transduktionswege etwa in Brusttumoren¹⁵⁾ oder beim Darmkrebs¹⁶⁾ gewonnen werden.

Der nächste Schritt, die Analyse der Expression auf Proteinebene, hat in der letzten Zeit eine rasche Entwicklung erfahren. Ein neuer Begriff ist auch dafür entstanden: Proteomics.¹⁷⁾ Aufbauend auf traditionellen Methoden der Proteinchemie werden nach zweidimensionaler Gelelektrophorese Spots aus den Gelen ausgeschnitten und in hohem Durchsatz durch Massenspektrometrie analysiert. Auch automatisierte Assays zu Protein-Protein-Wechselwirkungen (yeast two hybrid screens) liefern zusätzliche Informationen über das aktuelle Proteinexpressionsprofil einer Zelle.

Modellorganismen

In der molekularen Genetik gewinnt man Erkenntnisse über neue Gene und deren Funktion durch den Vergleich mit ähnlichen Genen in anderen Spezies. Die gemeinsame Evolution aller Organismen ermöglicht es, dieses Wissen über einen Organismus auch auf andere zu übertragen. Die vollständige Sequenzierung der ersten Genome von Modellorganismen, *Escherichia coli*, *Saccharomyces cerevisiae*,^{18,19)} und *Caenorhabditis elegans*,²⁰⁾ erzeugte eine Fülle von neuen Daten. Die bisher durchgeführten Vergleiche ließen erkennen, daß obwohl nicht unbedingt ganze Proteine konserviert sind, funktionelle Bereiche (Domänen) durchaus in der Evolution erhalten sind. Dabei ist die Anzahl und die Anordnung solcher Domänen zwischen den Spezies nicht konserviert.^{21,22)} Je komplexer ein Organismus ist, desto vielfältiger ist auch die Anzahl und Lokalisierung solcher Domänen und desto größer ist auch die Gesamtzahl der komplexeren Proteine.

Vergleichende Studien mit Modellorganismen sollen uns helfen, auch menschliche Krankheiten besser zu verstehen. Es stellt sich dabei die Frage, wie relevant die Hefe oder der Fadenwurm sind, um Krankheiten wie Blindheit oder Gehirnerkrankungen zu erforschen. Obwohl das volle Bild solcher Krankheiten auf diese Organismen nicht übertragbar sein wird, kann man durchaus Wissen über Mechanismen auf zellulärer Ebene gewinnen. Dafür gibt es schon zahlreiche Beispiele: ein Gen assoziiert mit einer bestimmten Form von Taubheit (Mohr-Tranebjaerg-Syndrom) hat Ähnlichkeit mit kleinen Hefeproteinen (Tims), die für den Proteinimport in die Mitochondrien eine Rolle spielen;²³⁾ oder das mit Batten-Syndrom (neurodegenerative Krankheit) assoziierte Gen (CLN3) hat ein homologes Gen in der Hefe, das die Expression von anderen Proteinen beeinflusst, die wiederum für die pH-Kontrolle in der Vakuole (und beim Menschen vielleicht in Lysosomen) wichtig sind.²⁴⁾

Dabei sollte man bei solchen Beispielen beachten, inwieweit diese Vergleiche eine biologische Relevanz haben und welche Rolle andere Faktoren wie zelluläre Lokalisation und Protein-Protein-Wechselwirkungen spielen.

Auch Drosophila-Genom fast vollständig aufgeklärt

Das nächste Genom, dessen Sequenzierung fast vollständig ist, ist das der Fruchtfliege, *Drosophila melanogaster*.²⁵⁾ Genetische Experimente mit *Drosophila* werden schon seit 1906²⁶⁾ durchgeführt, und viele grundlegende genetischen Erkenntnisse wurden mit Hilfe dieser etwa 3 mm großen Fliege gewonnen. An *Drosophila* sind Hunderte von unterschiedlichen Mutationen untersucht und verschiedene Mechanismen in der Entwicklungsbiologie aufgeklärt worden. Mehrere Genomzentren und die Celera Genomics arbeiten an der Sequenzierung des *Drosophila*-Genoms in einem gemeinsamen Projekt.²⁷⁾ Die Publikation der Sequenz von vier Chromosomenpaaren (180 Millionen Basen und ca. 12 000 Gene) soll Anfang 2000 erfolgen. Die Wissenschaft wartet gespannt auf diese Ergebnisse, die zusammen mit dem enormen Wissen über diesen Modellorganismus zu neuen Einblicken in die Biologie führen sollen.

Der vollständige Satz der menschlichen Gene, wie ein „biologisches Periodensystem“, soll es erlauben, deren Funktion und Zusammenspiel zu verstehen. Die sich rasch entwickelnden Technologien können es ermöglichen, mehrere Gene in einem einzigen Experiment zu analysieren. Dies wird den Wissenschaftlern helfen, Krankheitsmechanismen besser zu durchschauen und die Medizin auf neue Wege der Prävention und Behandlung von Krankheiten zu führen.

Evangelia Vakalopoulou, Schering, Berlin, E-Mail: Evangelia.Vakalopoulou@schering.de

- 1) F. S. Collins et al., *Science* 1998, 282, 682.
- 2) J. D. Watson, F. H. C. Crick, *Nature* 1953, 171, 737.
- 3) E. Pennisi, *Science* 1999, 283, 1822.
- 4) I. Dunham et al., *Nature* 1999, 402, 489.
- 5) W. Jang et al., *Trends Genet.* 1999, 15, 284.
- 6) www.ncbi.nlm.nih.gov/genome/seq/
- 7) www.nhgri.nih.gov/About_NHGRI/Der/variati.htm
- 8) E. Zubritzky, *Anal. Chem. News Feat.* 1999, 683A.
- 9) www.snp.cshl.org
- 10) www.ncbi.nlm.nih.gov/SNP
- 11) The Chipping Forecast, *Nat. Genet.* 1999, 21 Supplement, 1.
- 12) M. Q. Zhang, *Genome Res.* 1999, 9, 681.
- 13) D. J. Lockhart et al., *Nat. Biotechnol.* 1996, 14, 1675.
- 14) L. D. Stein, *Current Protocols in Human Genetics* 1998, pp.7.9.1.-7.9.8. Wiley & Sons.
- 15) C. M. Perou et al., *Proc. Natl. Acad. Sci. USA* 1999, 96, 9212.
- 16) U. Alon et al., *Proc. Natl. Acad. Sci. USA* 1999, 96, 6745.
- 17) A. Dove, *Nature Biotechnol.* 1999, 17, 233.
- 18) A. B. Goffeau et al., *Science* 1996, 274, 546.
- 19) genome-www.stanford.edu/Saccharomyces/yeast_info.html
- 20) The C. elegans Sequencing Consortium, *Science* 1998, 282, 2013.
- 21) L. Aravind, E. V. Koonin, *Nucleic Acids Res.* 1999, 27, 1609.
- 22) S. A. Chervitz et al., *Science* 1998, 282, 2022.
- 23) C. M. Koehler et al., *Proc. Natl. Acad. Sci. USA* 1999, 96, 2141.
- 24) D. A. Pearce et al., *Nat. Genet.* 1999, 22, 55.
- 25) flybase.bio.indiana.edu/
- 26) W. E. Castle et al., *Proc. Am. Acad. Arts Sci.* 1906, 41, 729.
- 27) D. Butler, *Nature* 1999, 401, 729.

Tabelle. **Zusätzlich zu den ursprünglichen Zielen des Humangenomprojekts wurden nun auch weiterführende Ziele in den neuen Plänen definiert.**