

Das menschliche Genom: Grundsätzliches zur DNA

E. G. Berger, T. Hennet

Einführung

Wer ist HUGO? Dieses einprägsame Akronym steht für die **H**uman **G**enome **O**rganisation, welche im Februar mit der Publikation der Struktur des menschlichen Genoms an die Öffentlichkeit getreten ist. Während viele Schlagzeilen aus Forschung und Technik, so spektakulär sie manchmal auch klingen, mehr Seifenblasen entsprechen, ist das, was HUGO der Weltöffentlichkeit und der Ärzteschaft mitzuteilen hatte, eine echte Revolution, vergleichbar der anatomischen Exploration des menschlichen Körpers durch Vesal 1543 oder der Etablierung des periodischen Systems in der Chemie. Sicher wird sich in einigen Jahrzehnten ein Rückblick auf den Februar 2001 lohnen, mit der Frage, was HUGO damals ausgelöst hat. Auch dürfte es spannend sein, Parallelen im medizinischen Fortschritt nach der ersten anatomischen Sektion und der Strukturaufklärung des Genoms aufzuzeigen. Beiden gemeinsam ist die Schaffung einer neuen Sprache, damals lateinisch, jetzt englisch, den sogenannten «genome speak». Unterschiedlich zwischen damals und heute ist die Geschwindigkeit in der Umsetzung des neuen Wissens. Bereits jetzt ist der biomedizinischen Forschungsgemeinschaft klar, dass die Genomdatenbanken eine unerschöpfliche Quelle von neuem Wissen darstellen. Wir glauben, dass unsere in der Praxis tätigen Kolleginnen und Kollegen daran teilhaben sollten, sei es nur um diese kulturelle Leistung zu würdigen, oder sei es, um einen Zugang zur modernen Molekular- und Zellbiologie zu finden, die zunehmend den medizinischen Alltag prägt. In der Ausbildung unserer jungen Studierenden beabsichtigen wir, dieser Umwälzung Rechnung zu tragen, indem wir in Bälde im ersten Studienjahr einen integrierten Ausbildungsgang in Zell- und Molekularbiologie anbieten werden, der auf einem systematisch aufgebauten interdisziplinären Vorlesungszyklus mit Praktika und projektorientierten Tutoraten in molekularer Medizin und einem Internet-basierten Multimedia Auftritt beruhen wird. Diese Grundlage tritt an die Stelle der Zoologie, der vergleichenden Mor-

phologie der Wirbeltiere und der Botanik, wie sie von Gesetzes wegen aber noch verlangt werden.

In den in loser Reihenfolge erscheinenden Kurzbeiträgen über das menschliche Genom geben wir zuerst die Grundlagen und rufen einige elementare Eigenschaften des Trägers der genetischen Information, der DNA, in Erinnerung. Im folgenden werden wir versuchen, das medizinische Potential der Informationsfülle, welche nun in vielen Datenbanken zugänglich gemacht worden ist*, zu beschreiben. Dabei steht uns auch eine Serie von Beiträgen, die vor einigen Jahren im *New England Journal of Medicine* erschienen sind und ein gutes Echo fanden, Pate.

Warum die Veröffentlichung des Humangenoms jetzt?

Die Angabe einiger Meilensteine in der modernen Genforschung (siehe Kästchen) enthält auch die mit dem Nobelpreis ausgezeichnete Methode der Sequenzierung der DNA durch Sanger. Diese Methode erwies sich als so robust, dass eine Automatisierung möglich wurde. Die Machbarkeit der Strukturaufklärung des menschlichen Genoms zeichnete sich schon Mitte der achtziger Jahre ab: im September 1988 wurde in Montreux unter der Führung von V. McKusick ein von öffentlichen Geldern unterstütztes Konsortium konstituiert, namens HUGO (www.gene.ucl.ac.uk/hugo/history.htm.) Die Organisation ist in Genf eingetragen und umfasst 220 Mitglieder aus 23 Ländern. Die Organisation setzte das klare Ziel der Aufklärung des Humangenoms und koordinierte die Anstrengungen unter den beteiligten Laboratorien. Vor vier Jahren erwuchs den eher gemächlich voranschreitenden Bemühungen von HUGO eine mächtige Konkurrenz, weil eine Firma (Celera) unter der Führung von Craig Venter mit einer andern Strategie das gleiche Ziel erklärte, aber mit der Ankündigung des patentrechtlichen Schutzes. Es bestand die Gefahr der Abschottung der privat gewonnenen Sequenzinformation, welche erst im 2000 durch eine gemeinsame Erklärung von Clinton und Blair, den Chefs der beiden führenden Nationen auf diesem Gebiet, abgewendet werden konnte. Der Druck, den Celera ausübte, beschleunigte die Strukturaufklärung erheblich, so dass nun 94% des Humangenoms durchsequenziert sind. Die Zeitschriften *Nature* und *Science* haben die verschiedenen Strategien aufgeteilt: *Nature* berichtet über das öffentliche, mit Steuergeldern finanzierte Vorgehen, *Science* über die Celera-Methode. Der Zugang zu dieser Information ist nun allgemein zugänglich, ihre Umsetzung stellt aber die grosse Herausforderung dar: Ein Koch im Besitze

Physiologisches Institut
der Universität Zürich
Dr. med. E. G. Berger
Winterthurerstr. 190
CH-8057 Zürich

*www.gdb.org/hugo
www.nhgri.nih.gov
www.ornl.gov/hgmis
www.hhmi.org
www.nature.com/genomics
www.scienceonline.org
www.faseb.org/genetics/ashg/ashgmenu.htm
www.nhgri.nih.gov/educationkit/

egberger@physiol.unizh.ch

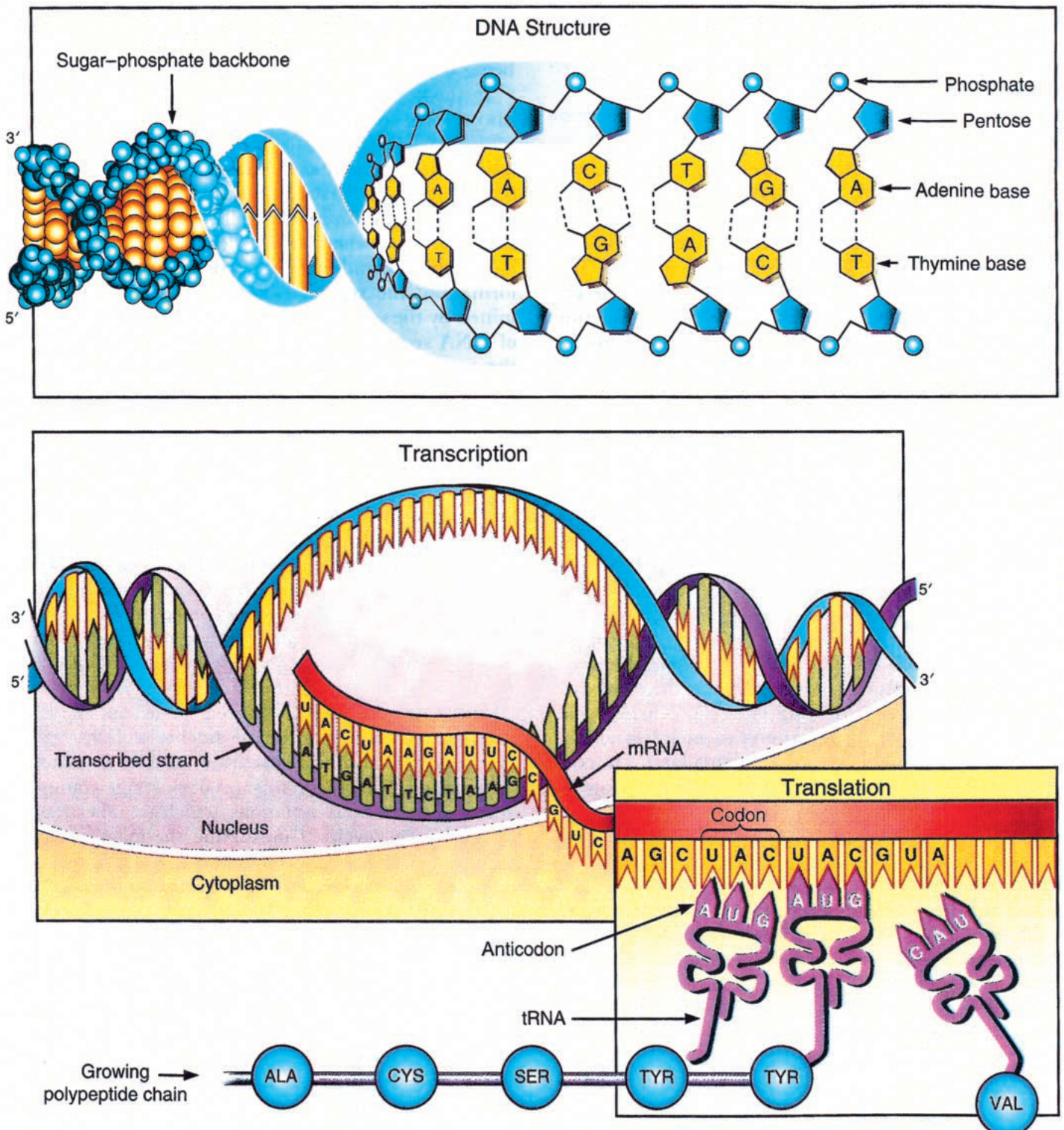
Abbildung 1

Oben: Darstellung der Doppelhelix-Struktur der DNA. Die Paarung beider Einzelstränge erfolgt über Wasserstoffbrücken (gestrichelte Linien).

Mitte: Transkription: Die Verwendung der genetischen Information durch die Zelle erfolgt über die Kopie der abzulesenden Stelle. Die Kopie (RNA) ist chemisch leicht verschieden (siehe Text), was eine stabile Paarung verhindert.

Unten: Die mRNA wandert vom Kern in das Zytoplasma, wo die Information in eine kolineare Aminosäuresequenz übersetzt wird. Als Adaptoren dient die transfer RNA, welche einerseits ein komplementäres Basentriplett, andererseits die dem genetischen Code entsprechende Aminosäure trägt. Die Verknüpfung der Aminosäuren zum Protein erfolgt im Ribosom, welches aus sogenannter ribosomaler RNA und einer Vielzahl von Proteinen besteht.

Modifiziert nach Rosenthal N, N Engl J Med 1994;331(1):39-41. Mit der freundlichen Genehmigung des Verlags.



Meilensteine der Erforschung der DNA.

1869	Miescher entdeckt Nukleinsäuren in Leukozyten
1944	Nachweis der DNA als Erbsubstanz durch Avery
1953	Aufklärung der Doppelhelix-Struktur der DNA durch Watson und Crick
1961	Genetischer Code aufgeklärt durch Nirenberg und Ochoa . Der Code bestimmt die Umsetzung der genetischen Information in Proteine.
1962	Entdeckung der Restriktionsendonukleasen durch Dussoix und Arber
1973	Erste Klonierung eines bakteriellen Plasmids durch Boyer und Cohen
1975	Sequenzierungsmethode durch Sanger und Barrell
1977	Entdeckung der Introns durch Roberts und Sharp
2000	Beendigung der Kartierung und der provisorischen Durchsequenzierung des menschlichen Genoms
2001	Die Zeitschriften Nature und Science publizieren eine Genkarte des Human-genoms, welche zu 94% durchsequenziert ist.

eines Rezeptes, das er lesen kann, hat noch keine Gaumenfreuden erzeugt.

Die DNA als Trägerin der genetischen Information

Die DNA (Deoxyribonucleic Acid, deutsch: DNS, Desoxyribonukleinsäure) ist die chemische Grundlage der genetischen Information von jedem Lebewesen (Mikroorganismen, Pflanzen, Tiere und Menschen). Diese ist in der unterschiedlichen Reihenfolge der Basen A, G, C, T (Abb. 1) festgeschrieben. Die daraus entstehende «genetische Sprache» mit vier Buchstaben bedingt letztlich die Diversität der Schöpfung. Zwischen einer pflanzlichen, tierischen oder menschlichen DNA gibt es keine fundamentalen Unterschiede des chemischen Verhaltens. Deshalb ist es ohne weiteres möglich, beispielsweise menschliche DNA-Fragmente in eine pflanzliche einzubauen oder umgekehrt. In menschlichen Zellen ist die DNA einerseits im Kern und ein kleiner Anteil in den Mitochondrien verpackt. Ein grosses menschliches Chromosom, also ein linearer DNA-Faden, ist in seiner gestreckten Form fast 10 cm lang. Wie können 46 Chromosomen in einen Zellkern verpackt werden, dessen Durchmesser ungefähr 5–10 µm beträgt? Die Antwort der Natur ist eine eindruckliche Aufknäuelung in sogenannte Nucleosomen, die eine Art Mini-Perlenkette bilden, die ihrerseits auch wieder kompakt aufgeknäuel ist. Der in Mitochondrien vorkommende Teil ist seit etwa 25 Jahren vollständig bekannt und enthält die Information für einen Teil der mitochondrialen Proteine und ihrer Biosynthese. Für die Biologie und Medizin ist die mitochondriale DNA nicht nur wegen ihres Informationsgehaltes für mitochondriale Proteine wichtig, sondern auch wegen ihrer

mütterlichen Vererbungsweise, ihrer Bedeutung bei ererbten Störungen der oxidativen Phosphorylierung (Zellatmung), ihrer «Alterung» und ihrer Bedeutung als Marker für Evolution und Völkerwanderung.

Für die **Praxis** wichtig ist die Tatsache, dass das vollständige Genom in jeder Zelle repräsentiert ist. Es ist also möglich, Genanalysen in Blutproben, Schleimhautabstrichen, Haarwurzeln usw. durchzuführen. Die DNA differiert von Mensch zu Mensch geringfügig, und bildet damit auch die Grundlage der biochemischen Individualität. Damit die Sequenzdaten des Humangenoms, wie sie nun von den Datenbanken abgerufen werden können, repräsentativ sind, wurde bei Celera die DNA von 21 zufällig gewählten Spendern unterschiedlichen Geschlechts, Alters und ethnischer Zugehörigkeit verwendet. Ähnlich ging das HUGO-Konsortium vor. Sämtliche persönlichen Daten der DNA-Spender wurden vernichtet.

Die folgenden Beiträge zur molekularen Medizin werden die interindividuellen Unterschiede der DNA im Rahmen von Polymorphismen (das sind Mutationen ohne offensichtliche pathophysiologische Folgen) und Erbkrankheiten (Genveränderungen mit erwiesenen krankmachenden Eigenschaften) thematisieren.

Funktionen der DNA

Als Informationsspeicher dient die DNA nur dann, wenn die Information einerseits weitergegeben, andererseits aber auch genutzt werden kann. Die Weitergabe an die Tochterzellen erfolgt durch **Replikation**: Bei der Entdeckung der Doppelhelix haben bereits Watson und Crick darauf hingewiesen, dass sich die Komplementarität der Basenpaare ausgezeichnet zur Verdoppelung eignet: Im Originalzitat heisst es mit einem Hauch von «Understatement»: *«It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.»* Der Vorgang der Replikation erfolgt nicht fehlerfrei. Ausgeklügelte Reparaturmechanismen verhindern eine Akkumulation von Fehlkopien (also Mutationen). Man rechnet eine Mutation auf drei Zellteilungen; ein Gen durchschnittlicher Grösse wird in 200 000 Jahren von einer Mutation betroffen. Damit sich eine Mutation auf die Nachkommen auswirkt, muss sie in der Keimbahn stattfinden. Diese Mutationshäufigkeit bildet die Grundlage der Evolution. Mit Erstaunen stellen wir fest, wie die systematischen Sequenzvergleiche der Eiweisse mit der gleichen Funktion (sogenannte orthologe Eiweisse) in der ganzen belebten Natur zu neuen entwicklungsgeschichtlichen Zusammenhängen führen. Die ältere Ärztegeneration erhielt einen früher noch

im ersten propädeutischen Examen geprüften Einblick in die Evolution durch das Studium der vergleichenden Morphologie der Wirbeltiere (Portmann); diese auf Morphologie begründete Entwicklungsgeschichte wird nun durch DNA-Sequenzvergleiche abgelöst.

Das Ablesen der Information erfolgt im Organismus zeitlich und räumlich hochspezifisch: Der Vorgang heisst **Transkription** (Abb. 1) und führt zu einem komplementären Abbild des kopierten DNA-Segmentes in Form einer einzelsträngigen mRNA (messenger RNA, deutsch: Boten-RNS). Die RNA enthält an Stelle des Thymins Uracil und anstelle der Deoxyribose Ribose. Dies verhindert die Bildung einer stabilen Doppelhelix und führt zu ziemlich kurzlebigen Molekülen mit einem hohen Umsatz. Dies erlaubt einen raschen Wechsel der an die Zelle weitergegebenen Information, und somit eine rasche Anpassungsfähigkeit an Signale von aussen.

Schliesslich wird die Information der RNA anhand des genetischen Codes in eine kolineare Sequenz von Aminosäuren übersetzt. Die Übersetzung der als Sequenz von Basenpaaren gespeicherten Information in die entsprechende

Aminosäuresequenz heisst **Translation**. Je drei Buchstaben der DNA (Triplet) codieren für eine der zwanzig Aminosäuren. Da vier Buchstaben zur Verfügung stehen, können 64 verschiedene Triplets gebildet werden. Wie die Abbildung zeigt, verfügt die Zelle über Verbindungsstücke zwischen der mRNA und der einzubauenden Aminosäure. Diese werden als transfer RNA (tRNA) bezeichnet, für welche es auch spezifische Gene gibt. Es gibt Aminosäuren, die von einem einzigen Triplet, wieder andere, die von sechs verschiedenen Triplets codiert werden; zudem gibt es eine Art Interpunktion: Das Triplet AUG bestimmt den Beginn, die Triplets UAG, UAA und UGA das Ende der Translation und damit der Länge eines Eiweissmoleküls. Der genetische Code ist universell, abgesehen von kleinen Ausnahmen bei Mitochondrien und Bakterien. In Lehrbüchern findet man ihn wegen seiner elementaren Wichtigkeit häufig im Einband dargestellt. Die **Bioinformatik** benützt ihn, um neu entschlüsselte DNA-Sequenzen automatisch in die zugehörige Aminosäuresequenz zu übersetzen. Die Primärstruktur eines Proteins (also seine Aminosäuresequenz) wird seit gut zwanzig Jahren praktisch nur noch über die entsprechende Geninformation aufgeklärt. Allerdings führt nur ein geringer Teil der DNA zur Übersetzung in Proteine (siehe Abschnitt «Was ist ein Gen?»). Das Produkt eines Gens ist zunächst ein nacktes Eiweiss, das von der Zelle anschliessend noch weiter bearbeitet wird, bis es seine endgültige Struktur (z.B. als multimeres Glycoprotein) und seinen zellulären oder systemischen Wirkungsort gefunden hat. Die nach der Translation erfolgenden Änderungen der Proteinstruktur nennt man «Posttranslatorische Modifikationen». Diese sind für die Funktionsfähigkeit des Proteins häufig ausschlaggebend (Beispiel: Das Hormon Erythropoetin ist ohne posttranslatorisch übertragene Zuckerreste wirkungslos, analog das Insulin, das nach seiner Bildung aus einer inaktiven Vorstufe zur aktiven Form abgespalten werden muss. Beide Vorgänge, die «Verzuckerung» wie die Spaltung erfolgen nach der Translation, sind also «posttranslatorisch»).

Quantitative Angaben (Steckbrief)

- Jeder Zellkern enthält 46 Chromosomen (Diploider Satz).
- Der haploide Satz enthält 23 Chromosomen, davon 22 Autosomen und ein Geschlechtschromosom; entweder X (weiblich) oder Y (männlich). Im diploiden Satz bedeuten XX (Frau), XY (Mann).
- Chromosom = DNA-Doppelhelix von variabler Länge, gestreckt etwa 1–10 μm .
- Total $2,91 \times 10^9$ Basenpaare (haploid), davon 5% transkribiert
- 30 000–35 000 Gene.

Eigenschaften:

- Aufbau aus Deoxy-Ribose-Ketten (Zucker-molekül mit 5 C-Atomen), die mit einer Phosphodiesterbrücke vom 5' zum 3' verknüpft sind. An das C1 der Ribose ist eine Purin- (Adenin oder Guanin) oder Pyrimidinbase (Cytosin oder Thymin) verknüpft.
- Diese Kette bildet einen Einzelstrang, der durch die Basenpaarung A mit T bzw. C mit G zum Doppelstrang wird. Die Paarung erfolgt über die Bildung von Wasserstoffbrücken, deren Bindungsenergie sich quantifizieren lässt und damit die Grundlage zur Hybridisierung liefert. Hitze bricht die H-Brücken (Denaturierung), Abkühlung erlaubt die Renaturierung, wobei sich komplementäre Abschnitte wieder paaren.

Was ist ein Gen?

In der klassischen Genetik bezeichnete man ein Gen als Erbfaktor, welcher einem Merkmal zugrunde liegt. Heute wird ein Gen als chemisch definierter Abschnitt eines DNA-Strangs bezeichnet, der transkribiert wird, also die Synthese einer RNA bewirkt (codierender Teil), inklusive derjenigen Abschnitte, die die Expression des Gens regulieren (nicht-codierender Teil). Das Transkript führt meistens zur Bildung eines Proteins, in wenigen Fällen aber auch zu

struktureller RNA der Ribosomen. Es wurde dann bald klar, dass sehr viel mehr DNA in den Chromosomen aufgeknäuel ist, als man zur Herstellung von Proteinsequenzen braucht. Merkwürdigerweise enthalten gewisse Pflanzen und Amphibien sogar noch mehr DNA als das menschliche Genom. Die Funktion der nicht direkt für Genprodukte verwendeten DNA, etwa 90%, ist nicht klar. Ein Teil dieser nicht-codierenden DNA ist mitten in die Abschnitte der codierenden DNA eingestreut und werden als Introns bezeichnet, wogegen die dazwischen liegenden (codierenden) Abschnitte als Exons bezeichnet werden. Dazwischen liegt ein Mechanismus des Herausspleissens (Splicing) der Introns, ein Vorgang der nach der Transkription im Kern erfolgt und zur Bildung der für die Translation geeigneten mRNA führt. Das Splicing ist ein Vorgang, der zu einer zusätzlichen Variabilität des Proteinmusters in den Zellen führt und der in gewissen Fällen auch gestört sein kann. Aus der Struktur eines Genprodukts kann also nie auf die Struktur des

Gens selber geschlossen werden: Erstens, weil der genetische Code nicht rückwärts übersetzt werden kann und zweitens weil das Gen selber einen variablen Anteil nicht codierender DNA enthält. Das Insulin-Gen beispielweise enthält nur zwei Introns, wogegen das für die Muskeldystrophie (M. Duchenne) verantwortliche Genprodukt, das Dystrophin, über 50 Introns aufweist.

Auf die einfache Frage, *was ist ein Gen?* gibt es keine einfache Antwort. Der Basensequenz ist die Genstruktur nicht anzusehen, weil es keine für den Anfang oder das Ende der Transkription allgemein gültige Sequenz gibt wie bei der Translation. Man könnte postulieren, dass alles was in RNA übersetzt wird, zu einem Gen gehört. Dieses Postulat ist notwendig, aber nicht hinreichend: Es gibt Hinweise dafür, dass Abschnitte des Genoms transkribiert werden, dass aber die Transkripte nicht zur Verwendung bestimmt sind, sondern abgebaut werden. Eine weitere Schwierigkeit ist die Tatsache, dass Gene mit vielen Introns zu unter-

Glossar

Bioinformatik:	Sammelbegriff für die Computerwissenschaft, welche die Gendatenverarbeitung zum Gegenstand hat.
Chromosom:	Strukturelle Einheit des genetischen Materials, bestehend aus einer linearen DNA-Doppelstranghelix und vielen Proteinen.
DNA:	Deoxyribonucleic Acid, deutsch Deoxyribonukleinsäure: Langes lineares Polymer aus vier verschiedenen Nucleotiden, welches aus zwei antiparallelen Strängen besteht.
Exon:	Kodierender Abschnitt eines Gens.
Genetischer Code:	Drei benachbarte Nucleotide einer Nucleinsäure (Triplet) bestimmen eindeutig eine der zwanzig Aminosäuren. Da es 64 verschiedene Triplets gibt, kodieren zum Teil mehrere für die gleiche Aminosäure.
Genom:	Gesamtheit der genetischen Information eines Organismus.
HUGO:	Human Genome Organization. Eine multinationale, in Genf eingetragene Gesellschaft zur Erforschung des humanen Genoms.
Intron:	Nicht-kodierender Abschnitt eines Gens.
Mitochondrien:	Zellorganell mit eigenem Genom. Energielieferant in Form von ATP durch oxidative Phosphorylierung mit Hilfe der Atmungskette.
Mutation:	In Keimbahn-DNA: dauernde, vererbare Änderung der Nucleotidsequenz eines Gens, mit oder ohne Folgen für die Funktion des Genprodukts. In Körperzelle: sogenannt somatische Mutation, ohne Vererbung, mit oder ohne Folgen für die Funktion des Genprodukts.
Nucleosom:	Aggregat von Histonproteinen, um welches sich die DNA zur Kompaktierung im Kern aufknäuel.
Polymorphismus:	Interindividuelle punktuelle Unterschiede in der DNA-Struktur, ohne Folgen für die Genfunktion.
Replikation:	Vorgang zur Verdoppelung der DNA bei der Zellteilung. Dabei werden die beiden Stränge entwunden, so dass eine komplementäre DNA für jeden Strang gebildet werden kann. Je ein Originalstrang verteilt sich dann zusammen mit einem neugebildeten Strang auf eine Tochterzelle.
Sequenzierung:	Technische Methode zur Bestimmung der Reihenfolge der Nucleotide einer DNA
Spleissen:	Eine der grossen Überraschungen Ende der siebziger Jahre war die Entdeckung von Sequenzen innerhalb von Genen, die nicht ins Protein übersetzt wurden. Diese nicht-kodierenden Abschnitte heissen Introns, wogegen die kodierenden Abschnitte als Exons bezeichnet werden. Bei der Transkription entsteht zunächst ein kolineares komplementäres RNA-Molekül (sogenannt heterogene nukleäre RNA), aus welchen die Intronabschnitte herausgeschnitten (-gespleisst) werden.
Translation:	Herstellung im Ribosom eines neuen Proteins durch Verknüpfung von Aminosäuren entsprechend der von der Basensequenz der mRNA vorgegebenen Reihenfolge.
Transkription:	Ablese der genetischen Information durch Bildung einer komplementären Einzelstrang-Nucleinsäure, der heterogenen nukleären RNA. Diese wird zur funktionellen mRNA zusammengespleisst durch Herausschneiden funktionsloser Sequenzabschnitte (Introns). Die Produkte der Transkription sind mRNA (Boten-RNS) zur Herstellung von Proteinen, rRNA zur Bildung von Ribosomen und tRNA (transfer-RNA) als Brückenglieder zur Verknüpfung der richtigen Aminosäuren zur Proteinsynthese.

schiedlichen Genprodukten führen können, durch sogenanntes alternatives Spleissen. Aus diesen und weiteren Gründen schwanken die Angaben über die Zahl der mutmasslichen Gene beträchtlich. In der Grössenordnung um 30 000 ist man sich jedoch seit der Publikation von 95% der Sequenz des Humangenoms einig. Ebenso besteht Einigkeit darüber, dass die exakte Bestimmung noch einige Jahre brauchen wird.

Schlussbemerkung

Die Pädagogik unserer Zeit betont die Notwendigkeit des lebenslangen Lernens. Praktisch nichts von all dem obenstehenden war vor 40 Jahren bekannt. Die Doppelhelix war zwar publiziert, ihre Bedeutung aber nur einigen Insidern klar. Die dadurch mögliche Entschlüsselung des Humangenoms folgte in erstaunlich kurzer Zeit. Einen ähnlichen Innovationsschub dürfte diese nun für die Medizin auslösen.

Literatur

Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J. Molecular Cell Biology. WH Freeman & Co. 2000, New York.

The human genome, Nature 409 (6822), 2001.
The human genome, Science 291 (5507), 2001.