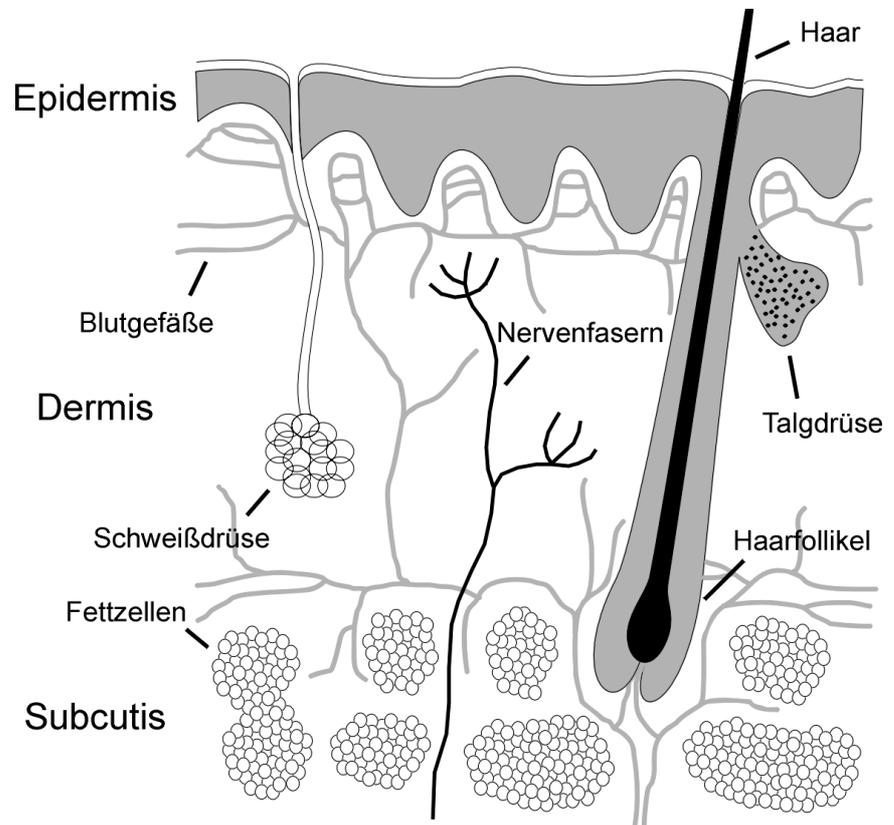


# I EINLEITUNG

## 1 Die Haut – Aufbau und Funktion

Die Haut ist mit einer Oberfläche von ca. 2 m<sup>2</sup> und einem Gewicht von ca. 3 kg (ohne Einrechnung des Fettgewebes) das größte Organ des Menschen. Zusammen mit ihren Anhangsgebilden, den Haaren, Nägeln und Hautdrüsen, bildet sie die Grenzschicht des Organismus zur Umwelt. Obwohl nur 1,5-4 mm dick, hat die Haut äußerst vielfältige und komplexe Funktionen zu erfüllen. Als Barriere verhindert sie sowohl das Austrocknen des Körpers als auch das Eindringen



**Abbildung 1: Der Aufbau der Haut. Schematische Darstellung von Subcutis, Dermis und Epidermis.**

von Wasser und schützt gegen chemische, biologische und physikalische (mechanische, thermische, aktinische) Noxen. Sollte sie dennoch beschädigt werden, besitzt die Haut die Fähigkeit zur Selbstregeneration und zur Abwehr eingedrungener Substanzen und Mikroorganismen. Sie fungiert außerdem als Sinnes- und Ausscheidungsorgan und beeinflusst den Wasserhaushalt und die Körpertemperatur. Nach Aufbau und Funktion lässt sie sich, von innen nach außen, grob in drei Schichten unterteilen (Abb. 1): in Subcutis (Unterhaut), Dermis (Lederhaut) und Epidermis (Oberhaut) [Übersicht in Fritsch, 1998].

Die Subcutis bildet mit ihrem lockeren Bindegewebe den Übergang der Haut zu den darunterliegenden Sehnen und Muskeln. Sie besteht hauptsächlich aus Fettgewebe, das als Energiespeicher, Wärmeisolierung und Schutzpolster gegen mechanische Einwirkungen dient. Mit dem Bindegewebe assoziiert sind Nerven, Blut- und Lymphgefäße, die eine Verbindung zu den oberen Hautschichten herstellen.

Die Hauptkomponente der Dermis ist das aus kollagenen und elastischen Fasern aufgebaute Bindegewebe, das der Haut hohe Reißfestigkeit und Elastizität verleiht. Hier liegen die Gefäße und Nerven zur Versorgung der Haut. Die Blutgefäße verlaufen überwiegend in den Grenzbereichen zur Subcutis und zur Epidermis, die selbst gefäßlos ist und durch Diffusion ernährt wird. Afferente freie Nervenendigungen dienen der Aufnahme und Weiterleitung äußerer Reize, efferente Nerven innervieren die in der Dermis liegenden Blutgefäße, Haarfollikel und Schweißdrüsen. Die Absonderungen der Hautanhangsgebilde, der mit den Haarfollikeln verbundenen Talgdrüsen und der Schweißdrüsen, halten die Haut geschmeidig und belastbar. Gleichzeitig begünstigen sie eine Hautflora, welche die Ansiedlung aggressiver Keime und Pilze auf der Oberfläche verhindert. Der Schweiß trägt außerdem zur Ausscheidung von Stoffwechselprodukten, zur Erzeugung eines individuell spezifischen Geruchs, zur Regulierung des Wasserhaushalts und über die Verdunstungskälte zur Temperaturregelung bei, die aber primär durch Konstriktion und Dilatation der Hautgefäße erreicht wird.

Als äußerste Schicht des Menschen steht die nur ca. 0,1 mm dünne Epidermis in direktem Kontakt zur Umgebung; an ihre Schutzfunktionen werden deshalb höchste Anforderungen gestellt. Sie zeichnet sich durch die Bildung besonders widerstandsfähiger Strukturen aus: Intermediärfilamente (IF) verleihen ihr mechanische Stabilität, und die „Hornzellwand“ (CE, *cornified cell envelope*) verhindert als nahezu undurchlässige Barriere sowohl das Austrocknen des Organismus als auch das Eindringen körperfremder Substanzen. Andere Aufgaben werden von besonders spezialisierten Zellen übernommen: Zum immunologischen Schutz tragen die Langerhans-Zellen bei, für die Absorption von ultravioletter Licht sind die Melanozyten bzw. der von ihnen synthetisierte Farbstoff Melanin verantwortlich, und sensorische Reize von der Hautoberfläche übertragen die Merkel-Zellen.

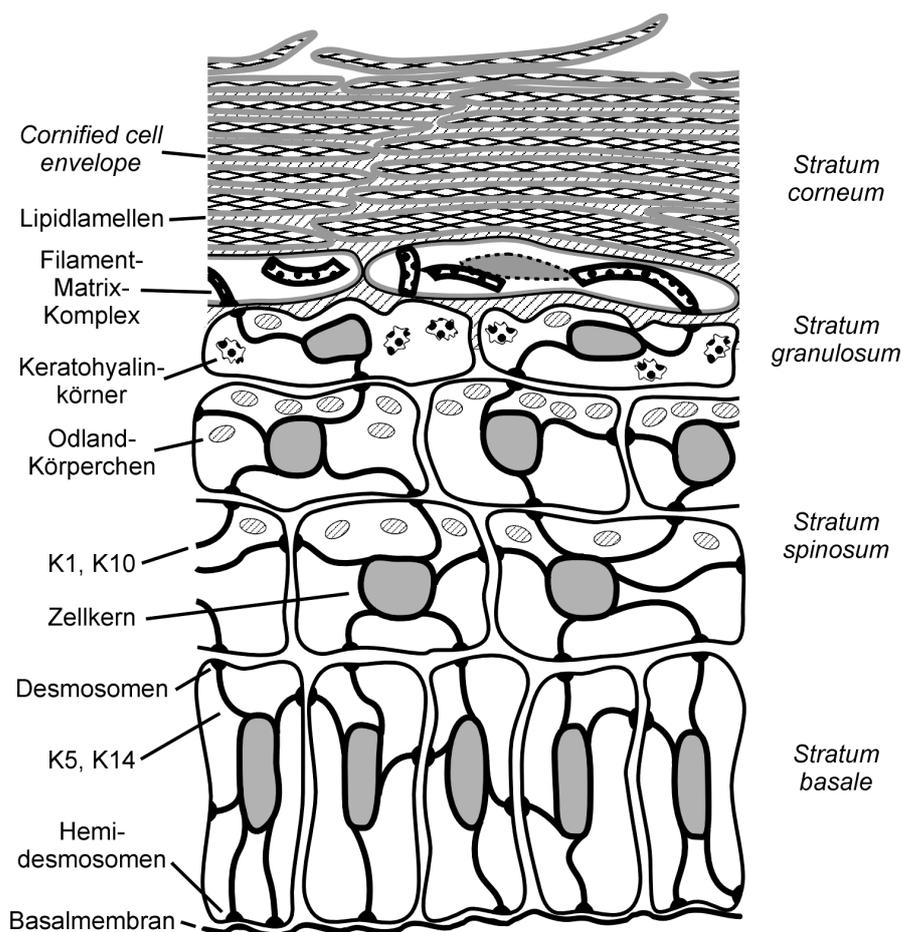
Mit nahezu 90% stellen die Keratinozyten die Hauptpopulation der Epidermiszellen. Neben dem Aufbau der protektiven Zellstrukturen sind sie für eine weitere, für eine Auseinandersetzung mit der Umwelt unabdingbare Funktion verantwortlich: die Regenerationsfähigkeit der Epidermis. Gewährleistet wird diese durch die optimale Anordnung als mehrschichtiges, verhornendes Plattenepithel, für dessen Entstehung und kontinuierliche Erneuerung ein komplexes Differenzierungsprogramm erforderlich ist.

## 2 Die Differenzierung der Epidermis

Im Gegensatz zum geschichteten Aufbau der Haut basiert die Schichtung des Plattenepithels der Epidermis auf unterschiedlichen Differenzierungsstadien einer einzigen Zellpopulation, der Keratinozyten. Nach ihrer Entstehung durch Zellteilung in der Basalschicht (*Stratum basale*) durchlaufen sie in der Stachelzellschicht (*Stratum spinosum*) und der Körnerschicht (*Stratum granulosum*) ansteigende Reifegrade, um schließlich terminal differenziert als tote Keratinozyten von der äußeren Hornschicht (*Stratum corneum*) abgeschilfert zu werden (Abb. 2). Für diesen streng regulierten Differenzierungsgang, der die fortdauernde Erneuerung der Epidermis sicherstellt, benötigen die Keratinozyten unge-

fähr vier Wochen. Jede der vier Zellschichten, die eine hohe Polarität aufweisen, besitzt dabei besondere morphologische und funktionelle Charakteristika [Übersicht in Holbrook, 1994; Übersicht in Fritsch, 1998].

Die unterste Schicht der Epidermis ist das aus zylindrischen Zellen bestehende, einlagige *Stratum basale*. Es ist über die Basalmembran mit der angrenzenden Dermis verbunden und enthält die mitoseaktiven Stammzellen. Von diesen werden vorübergehend teilungsfähige Zellen erzeugt, die nach dem Ablösen von der Basalmembran zur terminalen Differenzierung in suprabasale Schichten aufsteigen.



**Abbildung 2: Die Epidermis. Die Epidermisschichten mit ihren charakteristischen Strukturelementen sind schematisch dargestellt. Die Keratine K5 und K14 bilden die Intermediärfilamente der Basalschicht, K1 und K10 die der suprabasalen Schichten.**

Eingestreut findet man hier die Melanozyten und Merkel-Zellen. Als charakteristische Strukturen der Keratinozyten enthält das *Stratum basale* bereits die aus Keratinen bestehenden IF, die zusammen mit den Aktin-Mikrofilamenten und den röhrenförmigen Mikrotubuli das Zytoskelett bilden. Dessen Fasernetze sind sowohl für Struktur und Beweglichkeit der Zelle als auch für die Kommunikation innerhalb der Zelle und nach außen hin verantwortlich. Adhärenz der Keratinozyten untereinander wird über

die mit dem Zytoskelett verbundenen Desmosomen und Adhärenzkontakte vermittelt, zur Basalmembran über Hemidesmosomen. Spaltbrücken (*gap junctions*) dienen dem interzellulären Stoffaustausch. Nach dem Verlassen der Basalschicht stellen die teilungsfähigen Keratinozyten im 2-5 Zellagen dicken *Stratum spinosum* ihre Proliferation allmählich ein; das Zellvolumen nimmt zu und es findet eine horizontale Umorientierung der Zellachse statt. Im oberen Bereich des *Stratum spinosum* werden lipid- und enzymreiche Odland-Körperchen (*lamellar bodies*) synthetisiert, die im *Stratum granulosum* durch Exozytose in den Interzellularraum abgegeben werden und dort zur Barrierefunktion der oberen

Schichten beitragen. Das *Stratum spinosum* enthält außerdem die immunologisch bedeutsamen Langerhans-Zellen.

Das *Stratum granulosum* umfaßt 1-3 Zellagen und ist durch die zahlreichen basophilen Keratohyalinkörper gekennzeichnet, den Vorläufern der Keratinmatrix. Hier laufen die abschließenden Differenzierungsvorgänge ab: das Abplatten der Zellen durch Dehydratation und der Abbau der meisten Zellorganellen einschließlich der Zellkerne – aber auch der Aufbau des CE an der Innenseite der Plasmamembran sowie die Aggregation und Vernetzung der Keratine.

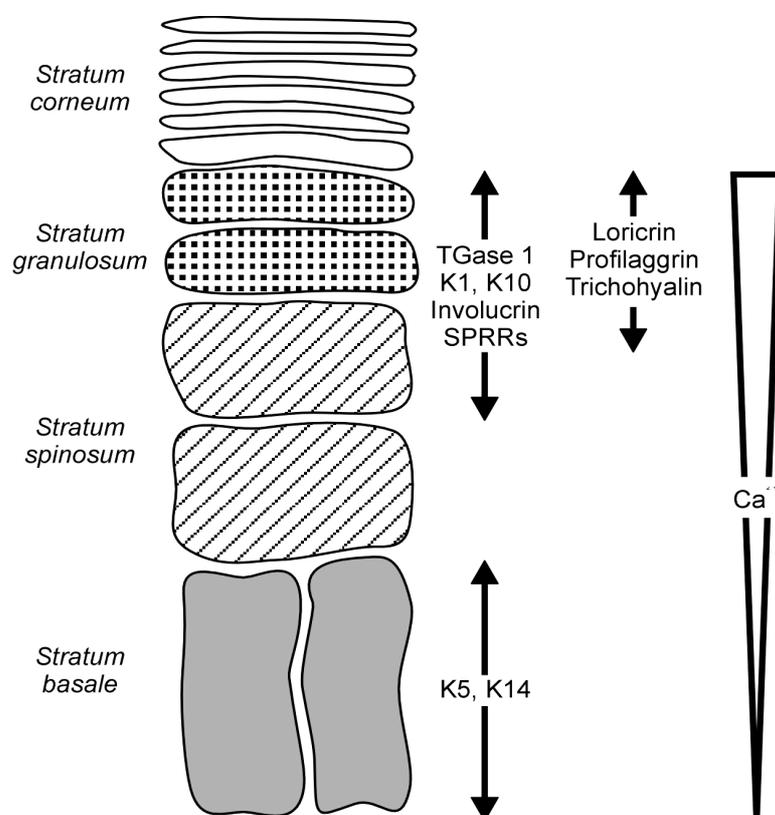
Mit dem Übergang in das *Stratum corneum*, das sich als mechanisch-chemische Barrierschicht der Epidermis über 10-20 Zellagen erstreckt, haben die Keratinozyten, die jetzt als Korneozyten bezeichnet werden, ihre abschließende funktionstragende Form erreicht. Korneozyten sind flache, hexagonale, kernlose Zellen, die mit 30 µm den größten Durchmesser aller Hautzellen haben. Sie bestehen nur noch aus den in eine Proteinmatrix eingebetteten Keratinfilamenten, die von dem starren, nahezu undurchlässigen und chemisch resistenten CE umgeben sind. An dessen Außenseite ist die Plasmamembran vollständig durch eine kovalent gebundene Lipidschicht ersetzt, die zusammen mit den Lipidlamellen des Interzellularraums die Zwischenräume nahezu wasserdicht abschließt. Ein begrenzter Austausch hydrophiler Stoffe über die Haut durch Diffusion ist dennoch möglich, da das *Stratum corneum* aufgrund hygroskopisch wirkender Proteine und Abbauprodukte der Keratinmatrix einen geringen Wasseranteil aufweist. Zusammen mit den Desmosomen ist das Barriere lipid außerdem für die Kohärenz der Zellen im *Stratum corneum* verantwortlich. Der regelmäßige und kompakte Aufbau geht erst in den obersten Zellagen durch den Zerfall der Dichtsubstanz und den Abbau der Keratinmatrix verloren, infolgedessen die Zellen schließlich abgestoßen werden.

### 3 Molekularbiologie der Keratinozyten

Auf ihrem Weg vom *Stratum basale* bis zur Abschilferung im obersten *Stratum corneum* durchlaufen die Keratinozyten unterschiedliche Reifestadien, die durch das Vorhandensein bzw. Nichtvorhandensein spezifischer Zellstrukturen und -organellen gekennzeichnet sind. Dabei verändern sich mit der Morphologie auch die in den Keratinozyten ablaufenden biochemischen Prozesse. Die für die Auf- und Abbauvorgänge erforderliche, zeitlich und räumlich begrenzte Aktivierung bestimmter Gene führt zur Expression charakteristischer Differenzierungsmarker, Proteinen, die in einzelnen Schichten der Epidermis vorkommen und den Differenzierungsgrad der Zelle widerspiegeln (Abb. 3). Dabei handelt es sich sowohl um diverse Strukturproteine als auch um Enzyme, die differenzierungsspezifische Stoffwechselfvorgänge beeinflussen [Übersicht in Eckert *et al.*, 1997].

Zu den vorherrschenden epidermalen Strukturproteinen gehören die Keratine. Sie bilden die vor allem in Keratinozyten ausgeprägten IF, die als intrazelluläres Netzwerk für die mechanische Stabilität der Zellen verantwortlich sind. Zur Zeit sind etwa 30 verschiedene Keratine bekannt, die zwei Familien

zugeordnet werden können: den sauren Typ I-Keratinen, deren Gene in der Region 17q, auf dem langen Arm von Chromosom 17, liegen, sowie den neutralen bis basischen Typ II-Keratinen mit den entsprechenden Loci in der Region 12q [Übersicht in Blumenberg, 1993; Übersicht in Mischke, 1998].



**Abbildung 3: Differenzierungsmarker der Epidermis. Die Doppelpfeile kennzeichnen die Expression der entsprechenden Gene in den jeweiligen Epidermisschichten. Das Dreieck symbolisiert die ansteigende Calciumionenkonzentration.**

Die Vielzahl der vorhandenen Keratingene und die charakteristische Ausbildung von Heterodimeren als Grundbaustein der IF, die aus je einem Typ I- und einem Typ II-Keratin zusammengesetzt sind, führen dabei je nach Gewebetyp und Differenzierungsgrad der Zellen zur Ausprägung spezifischer Expressionsmuster. Während in der Basalschicht der Epidermis ausschließlich die niedermolekularen Keratine K5 und K14 vorkommen, werden diese nach Einsetzen der Differenzierung durch die hochmolekulare Paare K1/K10 und K1/K11 ersetzt. Andere Keratinpaare finden sich z.B. in suprabasalen Zellen nicht verhornender Epithelien (K4/K13), in Hornhautepithelien (K3/K12) so-

wie bei der Wundheilung (K6/K16), wobei die unterschiedlichen Kombinationen Einfluß auf die Interaktionen mit Proteinen und Zellorganellen sowie auf das Aggregationsverhalten der Keratine haben [Übersicht in Morley & Lane, 1994].

Ein charakteristischer Marker für die fortgeschrittene Differenzierung verhornender Plattenepithelien ist Filaggrin, das mit den Keratin-IF assoziiert ist, deren Aggregation und Vernetzung es fördert [Übersicht in Dale *et al.*, 1985a]. Es entsteht aus dem funktionell inaktiven Profilaggrin, dem Hauptbestandteil der Keratohyalinkörner des *Stratum granulosum*. Die Synthese des hochmolekularen Profilaggrins (ca. 400 kD) setzt im obersten *Stratum spinosum* ein [Dale *et al.*, 1985b; McKinley-Grant *et al.*, 1989]. Durch die anschließende Phosphorylierung bilden sich unlösliche Aggregate, die nach Dephosphorylierung und  $\text{Ca}^{2+}$ -abhängiger Proteolyse beim Übergang in das *Stratum corneum* reaktives Filaggrin freisetzen [Übersicht in Dale *et al.*, 1993; Presland *et al.*, 1995]. Dabei gehen aus einem Profilaggrin-Molekül durch proteolytische Spaltung 10-12 Filaggrin-Monomere (35 kD) hervor [Gan *et al.*, 1990], die als Matrix für die Keratinfibrillen dienen. Gleichzeitig wandert der abgespaltene Ami-

no (N)-Terminus des Profilaggrins in den Zellkern, wo er für apoptotische Prozesse verantwortlich gemacht wird [Ishida-Yamamoto *et al.*, 1998]. Ein strukturell und funktionell ähnliches Protein ist Trichohyalin, welches ebenfalls mit den Keratinfilamenten der Epidermis interagiert und hauptsächlich in den Haarfollikeln vorkommt [Rothnagel & Rogers, 1986; Fietz *et al.*, 1990; O'Guin *et al.*, 1992]. Als Bestandteil der Keratohyalingranula wurde es darüberhinaus in spezialisierten Epithelien, u.a. der Zunge und der Vorhaut, gefunden [Manabe & O'Guin, 1994]. Wie Profilaggrin zeigt Trichohyalin  $\text{Ca}^{2+}$ -abhängige posttranslationelle Modifikationen [Tarcza *et al.*, 1997]. Auch konnte für beide Proteine eine Beteiligung am CE nachgewiesen werden [Steinert & Marekov, 1995; Steinert *et al.*, 1998].

Neben den Keratinfilamenten ist der CE eine weitere für terminal differenzierte Keratinozyten der Epidermis spezifische Struktur [Übersicht in Simon, 1994; Steinert, 1995]. Seine Hauptbestandteile sind u.a. Loricrin, Involucrin und die „kleinen Prolin-reichen Proteine“ (SPRRs, *small proline-rich proteins*) [Steven & Steinert, 1994]. Im *Stratum spinosum* beginnt die Synthese der Vorläuferproteine dieser widerstandsfähigen Verhornungsstruktur [Banks-Schlegel & Green, 1981; Kartasova *et al.*, 1988; Hohl *et al.*, 1991b; Hohl *et al.*, 1995]. Anschließend werden sie im *Stratum granulosum* an der Innenseite der Plasmamembran über Glutamyl-Lysyl-Isopeptid-Brückenbindungen miteinander vernetzt [Abernathy *et al.*, 1977; Rice & Green, 1977], katalysiert durch  $\text{Ca}^{2+}$ -abhängige Transglutaminasen (TGasen) [Rothnagel & Rogers, 1984; Thacher & Rice, 1985; Schmidt *et al.*, 1988], die ebenfalls als Differenzierungsmarker dienen. Die SPRRs, die sich aus den Untergruppen SPRR1, SPRR2 und SPRR3 zusammensetzen [Gibbs *et al.*, 1993], fungieren dabei als Brückenbildner zwischen den größeren Proteinen Loricrin und Involucrin [Steinert *et al.*, 1998]. Sie bestimmen über die unterschiedlichen beteiligten Untergruppen und ihren gewebespezifischen Anteil die Eigenschaften des CE. Im *Stratum corneum* ist schließlich die fertige „Hornzellwand“ nachweisbar, zu deren weiteren Bestandteilen u.a. Keratin-IF und Elafin [Steinert & Marekov, 1995], Cystatin A (Keratolinin) [Takahashi *et al.*, 1992], Envoplakin [Ruhrberg *et al.*, 1996], Periplakin [Ruhrberg *et al.*, 1997] und andere desmosomale Proteine [Haftik *et al.*, 1991] zählen. An der Außenseite des CE werden nach dem Abbau der Zellmembran diverse Ceramide und Lipide kovalent gebunden, die für den Zusammenhalt der Korneozyten sorgen und die Zellzwischenräume abdichten [Swartzendruber *et al.*, 1987; Wertz *et al.*, 1989; Marekov & Steinert, 1998].

Proliferation und Differenzierung der Keratinozyten müssen für die Ausbildung einer intakten Epidermis in einem ausgewogenen Verhältnis zueinander stehen. Für die Feinabstimmung dieser Prozesse sind eine Vielzahl endogener und exogener Faktoren verantwortlich, u.a. diverse Zytokine, wie Interleukine und Wachstumsfaktoren, die sowohl von den Keratinozyten selbst als auch von den angrenzenden Zellen der Dermis gebildet werden, und Stoffe wie Phorbol ester, Hydrocortison und das Cholera toxin (Tab. 1) [Übersicht in Eckert *et al.*, 1997]. Zu den am besten untersuchten Einflußfaktoren zählen die aus Vitamin A (Retinol) hervorgehende Retinsäure mit ihren Derivaten, den Retinoiden, und Vitamin D<sub>3</sub> (Cholecalciferol,  $1\alpha,25\text{-(OH)}_2\text{D}_3$ ), deren Auswirkungen über Kernrezeptoren, soge-

nannte Steroidrezeptoren, vermittelt werden, sowie die ubiquitär im Organismus vorkommenden Calciumionen.

*In vitro* wirken Retinoide der Keratinozytendifferenzierung entgegen; es wurde sowohl ein hemmender Einfluß auf die Genexpression und Synthese der suprabasalen Keratine [Fuchs & Green, 1981; Kopan *et al.*, 1987] als auch auf den Aufbau des CE beschrieben [Green & Watt, 1982], wobei letzterer auf eine verminderte Expression des Loricrins [Hohl *et al.*, 1991a] und eine reduzierte TGase-Aktivität zurückgeführt wird [Rubin & Rice, 1986; Floyd & Jetten, 1989]. Involvierte zelleigene Proteine sind u.a. die Retinoid-bindenden Proteine CRABP (*cellular retinoic acid binding protein*)-I und CRABP-II, die dem Transport durch das Zytosol dienen und die die freie Retinoidkonzentration beeinflussen [Übersicht in Siegenthaler, 1996], sowie die nucleären Retinoidrezeptoren RAR (*retinoic acid receptors*)- $\alpha$ , - $\beta$ , - $\gamma$  und RXR (*retinoid X receptors*)- $\alpha$ , - $\beta$ , - $\gamma$ , die als Heterodimere durch Bindung an Regulatorsequenzen der Desoxyribonucleinsäure (DNA, *deoxyribonucleic acid*) die Genexpression beeinflussen [Übersicht in Fisher & Voorhees, 1996]. Vitamin D<sub>3</sub> wird in der Haut unter Einwirkung des Sonnenlichts aus Provitamin D<sub>3</sub> gebildet. Seine Wirkung entfaltet es in Verbindung mit dem Vitamin D<sub>3</sub>-Rezeptor (VDR), der die gleichen regulatorischen Elemente erkennt wie die Retinoidrezeptoren [Schule *et al.*, 1990]. Jedoch sind bei physiologischen Vitamin D<sub>3</sub>-Konzentrationen entgegengesetzte *in vitro*-Effekte zu beobachten: Basalzellen treten vermehrt in die terminale Differenzierung ein [Regnier & Darmon, 1991], und die Synthese der Keratine K1 und K10 sowie die Bildung des CE werden induziert [Hosomi *et al.*, 1983]. Die Wirkungen des Vitamin D<sub>3</sub> sind eng mit der Ca<sup>2+</sup>-Konzentration verknüpft [Su *et al.*, 1994]. Calciumionen weisen einen mit der Differenzierung der Keratinozyten ansteigenden Konzentrationsgradienten in der Epidermis auf [Menon *et al.*, 1985]. Sie beeinflussen die interzelluläre Adhäsion der Keratinozyten [Hennings *et al.*, 1980], sind notwendig für die Expression diverser Keratinozytendifferenzierungsmarker einschließlich Loricrin [Hohl *et al.*, 1991a], Profilaggrin [Yuspa *et al.*, 1989] und TGase [Rubin & Rice, 1986] und sind als Cofaktoren an differenzierungsspezifischen Reaktionen beteiligt, wie der Prozessierung des Profilaggrins [Presland *et al.*, 1995] und des Trichohyalins [Tarcza *et al.*, 1997] sowie der enzymatischen Vernetzung des CE [Rice & Green, 1979].

**Tabelle 1: Einfluß verschiedener Faktoren auf Keratinozytenproliferation und -differenzierung**

	Proliferation	Differenzierung
IL-1	↑	
IL-6	↑	
IGF-I	↑	
EGF	↑	
TGF- $\alpha$	↑	
TGF- $\beta$ 1	↓	↑
KGF	↑	↑
IFN- $\gamma$	↓	↑
Retinoide		↓
Vitamin D <sub>3</sub>	↓ <sup>a</sup>	↑
Ca <sup>2+</sup>	↓	↑
Hydrocortison		↑
Choleratoxin	↑	
Phorbolster	↑	↑

IL, Interleukin; IGF, *insulin-like growth factor*; EGF, *epidermal growth factor*; TGF, *transforming growth factor*; KGF, *keratinocyte growth factor*; IFN, Interferon; <sup>a</sup> bei physiologischen Konzentrationen.

Basalzellen treten vermehrt in die terminale Differenzierung ein [Regnier & Darmon, 1991], und die Synthese der Keratine K1 und K10 sowie die Bildung des CE werden induziert [Hosomi *et al.*, 1983]. Die Wirkungen des Vitamin D<sub>3</sub> sind eng mit der Ca<sup>2+</sup>-Konzentration verknüpft [Su *et al.*, 1994]. Calciumionen weisen einen mit der Differenzierung der Keratinozyten ansteigenden Konzentrationsgradienten in der Epidermis auf [Menon *et al.*, 1985]. Sie beeinflussen die interzelluläre Adhäsion der Keratinozyten [Hennings *et al.*, 1980], sind notwendig für die Expression diverser Keratinozytendifferenzierungsmarker einschließlich Loricrin [Hohl *et al.*, 1991a], Profilaggrin [Yuspa *et al.*,

1989] und TGase [Rubin & Rice, 1986] und sind als Cofaktoren an differenzierungsspezifischen Reaktionen beteiligt, wie der Prozessierung des Profilaggrins [Presland *et al.*, 1995] und des Trichohyalins [Tarcza *et al.*, 1997] sowie der enzymatischen Vernetzung des CE [Rice & Green, 1979].

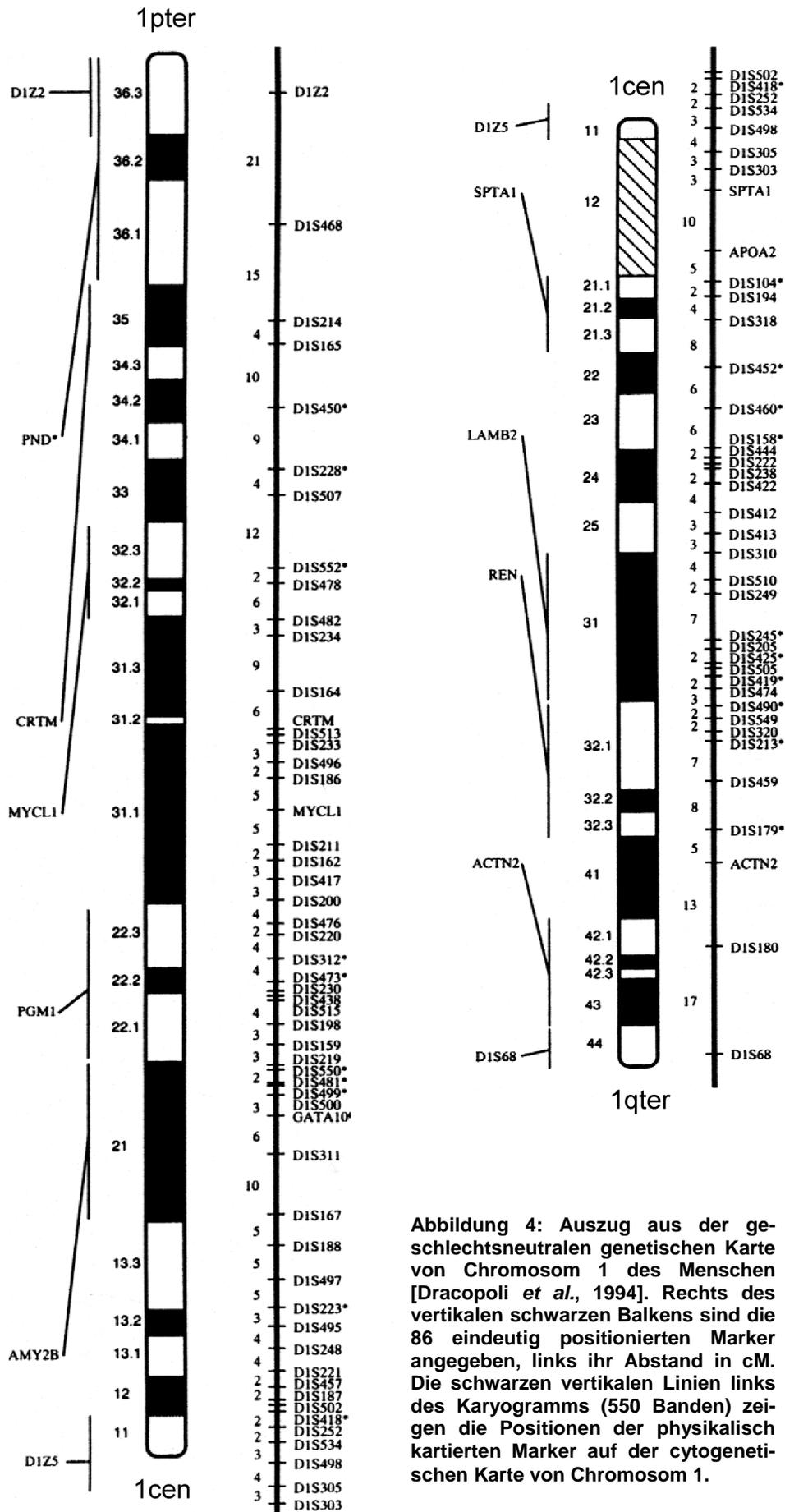
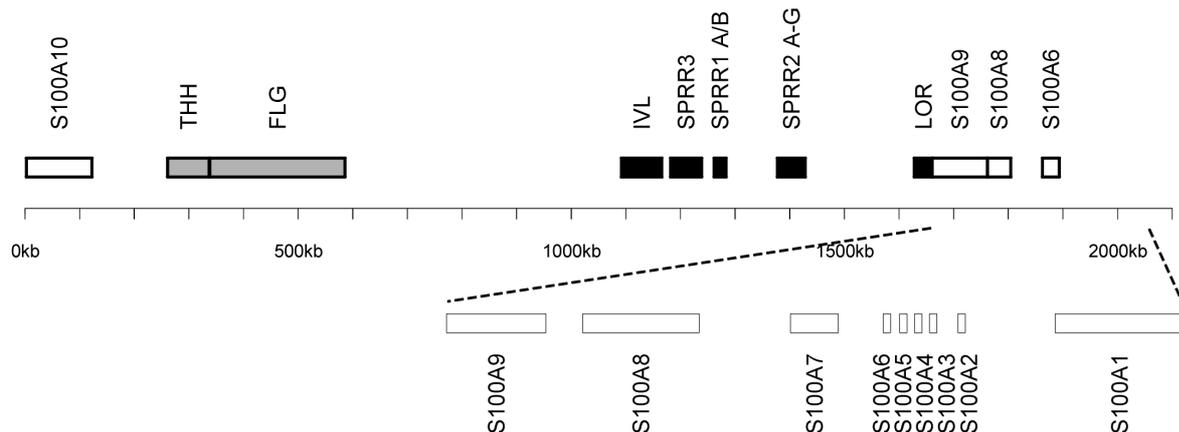


Abbildung 4: Auszug aus der geschlechtsneutralen genetischen Karte von Chromosom 1 des Menschen [Dracopoli *et al.*, 1994]. Rechts des vertikalen schwarzen Balkens sind die 86 eindeutig positionierten Marker angegeben, links ihr Abstand in cM. Die schwarzen vertikalen Linien links des Karyogramms (550 Banden) zeigen die Positionen der physikalisch kartierten Marker auf der cytogenetischen Karte von Chromosom 1.

## 4 Die Bedeutung der Region 1q21 für die Epidermidifferenzierung

Chromosom 1 ist mit einer geschätzten physikalischen Länge von 300 Mb das größte Chromosom des Menschen; es enthält etwa 8% der gesamten genetischen Information. 1994 fand der *First International Workshop on Chromosome 1 Mapping 1994* in Bethesda, Maryland, statt, auf dem die zu diesem Zeitpunkt bekannten Kartierungsergebnisse zusammengetragen wurden, um einen Informationsaustausch und Datenvergleich zwischen den einzelnen Arbeitsgruppen zu ermöglichen. Insgesamt waren 395 Gene (Genome Database, Stand 2/94) und 55 Krankheiten Chromosom 1 zugeordnet, darunter eine Form des Weitwinkelglaukoms, eine neurale Muskelatrophie (Charcot-Marie-Tooth-Krankheit), eine Form des mit Gehörlosigkeit und Retinitis pigmentosa einhergehenden Usher Syndroms und die Nemalin-Myopathie, eine nicht fortschreitende, allgemeine Muskelschwäche. Die auf dem Workshop veröffentlichte genetische Karte (s. Abschnitt I 8) enthielt bei einer geschlechtsneutralen Chromosomenlänge von 367 cM 192 Loci, von denen lediglich 86 genau positioniert waren, was einem durchschnittlichen Abstand von 4,3 cM entspricht (Abb. 4). Die 35 physikalischen Karten mit einer Ausdehnung von 42 kb bis 6 Mb deckten nur kleine Bereiche ab und waren über das gesamte Chromosom verteilt [Dracopoli *et al.*, 1994].

Die chromosomale Bande 1q21 liegt auf dem langen Arm von Chromosom 1 in der Nähe des Centromers. Sie läßt sich bei höherer Auflösung in die Unterbanden q21.1-q21.3 unterteilen (Abb. 4) und entspricht einer ungefähren physikalischen Ausdehnung von 15 Mb. Mit der Kartierung des Profilaggrin-Gens (*FLG*) durch *in situ*-Hybridisierung [McKinley-Grant *et al.*, 1989] wurde ein erster Bezug zwischen der Region 1q21 und der Differenzierung der Epidermis hergestellt. Es folgten mehr oder weniger eng benachbart die Loci für Involucrin (*IVL*; 1q21-q22) [Simon *et al.*, 1989], *SPRR1* (*SPRR1*; 1q22-q23) [An *et al.*, 1992], Loricrin (*LOR*; 1q21) [Yoneda *et al.*, 1992] und Trichohyalin (*THH*; 1q21-q23) [Fietz *et al.*, 1992]. Volz und Mitarbeitern (1993) gelang es schließlich auf einer physikalischen Restriktionskarte (s. Abschnitt I 9) nachzuweisen, daß die Mitglieder dreier Genfamilien innerhalb 2 Mb der Region 1q21 dicht beieinander liegen und dabei eine charakteristische Anordnung aufweisen (Abb. 5). Sowohl *IVL*, *SPRR1* und *LOR*, welche Vorläuferproteine des CE codieren, als auch *FLG* und *THH*, Gene für Proteine der Keratinmatrix, bilden jeweils ein Cluster (von *cluster*, Anhäufung) im menschlichen Genom. Eingerahmt wurden diese beiden Genfamilien von den Loci zweier  $\text{Ca}^{2+}$ -bindender Proteine vom S100-Typ (von *soluble in 100% ammonium sulfate*): *S100A6* (früher *CACY* von Calcyclin), welches zuvor in der Region 1q21-q25 kartiert worden war [Dorin *et al.*, 1990], und *S100A10* (früher *CALIL* von Calpactin I leichte Kette). Kurze Zeit später konnten insgesamt zehn *S100*-Gene (*S100A1* bis *S100A10*) und ebensoviele *SPRR*-Gene, die drei Untergruppen zugeordnet werden (*SPRR1A* und *1B*, *SPRR2A* bis *2G*, *SPRR3*), in direkter Nachbarschaft zu den bereits bekannten Genen lokalisiert werden [Gibbs *et al.*, 1993; Schafer *et al.*, 1995; Mischke *et al.*, 1996] (Abb. 5). Neben ihrer funktionellen Beziehung zur epidermalen Differenzierung zeigen die drei Familien ausgeprägte Überschneidungen hinsichtlich ihrer Protein- und Genstruktur (Abb. 6).



**Abbildung 5: Physikalische Karte der Genfamilien der Region 1q21. Rechtecke zeigen den Bereich an, in dem das entsprechende Gen durch Restriktionskartierung lokalisiert wurde. Schwarze Rechtecke repräsentieren die Gene der Vorläufer des CE, weiße die Gene der  $\text{Ca}^{2+}$ -bindenden S100-Proteine und graue die „fusionierten“ Gene der mit den Intermediärfilamenten assoziierten Proteine. Für die Kartierung der Gene *S100A10*, *THH*, *FLG*, *IVL*, *SPRR3*, *SPRR1B*, *SPRR2A*, *LOR*, *S100A9*, *S100A8* und *S100A6* wurde DNA der menschlichen Zelllinie H2LCL verwendet [Volz *et al.*, 1993; Mischke *et al.*, 1996], für die Kartierung der zehn *SPRR*-Gene DNA aus primären Fibroblasten des Menschen [Gibbs *et al.*, 1993]. *S100A1* bis *S100A9* wurden auf einem YAC kartiert [Schafer *et al.*, 1995].**

Die Bausteine des CE Involucrin, Loricrin und die *SPRRs* enthalten auffällige Wiederholungen kurzer Sequenzeinheiten im zentralen Bereich des Proteins, deren Anzahl von 3-14 für die *SPRRs* [Gibbs *et al.*, 1993] über 20 für Loricrin [Hohl *et al.*, 1991b] bis 39 für Involucrin reicht [Eckert & Green, 1986]. Als Transglutaminasesubstrate werden die Proteine über Glutamin- und Lysinreste miteinander vernetzt, wobei diese bei den brückenbildenden *SPRRs* ausschließlich in den terminalen Regionen liegen [Steinert *et al.*, 1998]. Darüberhinaus weisen sie deutliche Homologien in den Amino- und Carboxy (C)-terminalen Bereichen auf. Weitere Übereinstimmungen sind in ihrer Genstruktur zu erkennen: Alle Gene haben ein kurzes, nicht translatiertes Exon, gefolgt von einem zweiten, das die gesamte codierende Region enthält [Eckert & Green, 1986; Yoneda *et al.*, 1992; Gibbs *et al.*, 1993]. Für die Proteine des CE wird deshalb ein Urgen postuliert, aus dem durch Duplikationen anfangs drei Gene entstanden sind. Diese haben sich dann getrennt entwickelt: *IVL* und *LOR* durch die Amplifizierung interner Sequenzabschnitte und die *SPRR*-Gene zusätzlich durch Genamplifizierung [Backendorf & Hohl, 1992].

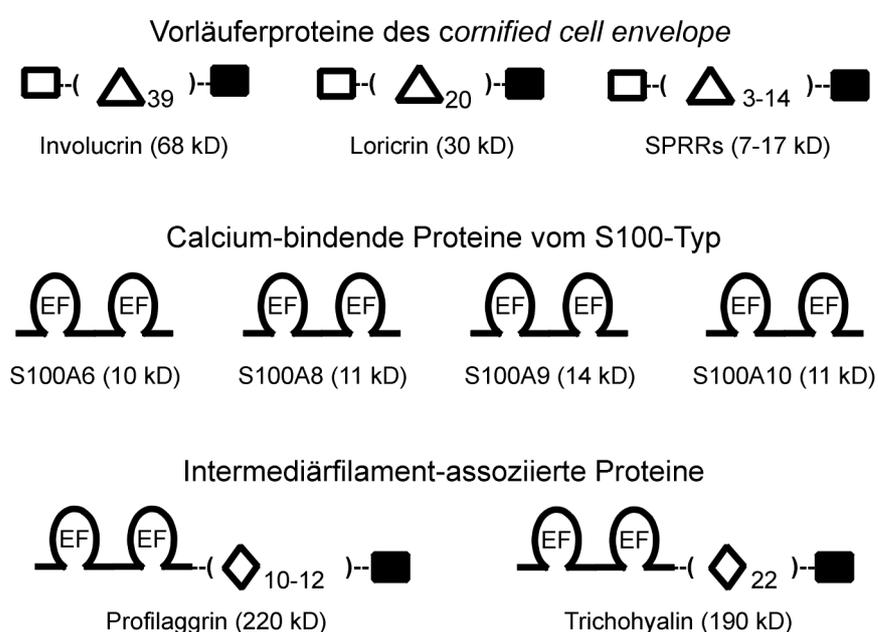
Von ungewöhnlicher Struktur sind die hochmolekularen, mit den IF assoziierten Strukturproteine Profilaggrin und Trichohyalin. Als mögliche Fusionsprodukte vereinen sie die Merkmale der von den benachbarten Genfamilien codierten Proteine: Interne Sequenzwiederholungen, Kennzeichen der Vorläuferproteine des CE, sind am N-Terminus mit den für die S100-Proteine charakteristischen zwei  $\text{Ca}^{2+}$ -bindenden EF-Hand-Motiven (s.u.) verbunden [Presland *et al.*, 1992; Lee *et al.*, 1993]. Letztere sind offenbar für die posttranslationalen Modifikationen verantwortlich, die Peptidwiederholungen für ihre Funktion als Bestandteil der Keratinmatrix und des CE. Auch in den Genstrukturen lassen sich

Gemeinsamkeiten sowohl mit den *S100*-Genen als auch mit *LOR*, *IVL* und den *SPRR*-Genen erkennen: Bei allen ist das erste Exon das kleinste und wird nicht translatiert; wie fast alle *S100*-Gene bestehen *FLG* und *THH* aus drei Exons, wobei die zwei  $\text{Ca}^{2+}$ -bindenden Domänen jeweils von einem Exon kodiert werden; und wie bei *LOR*, *IVL* und den *SPRR*-Genen ist das letzte Exon das größte und enthält die Sequenzwiederholungen [Eckert & Green, 1986; Yoneda *et al.*, 1992; Presland *et al.*, 1992; Gibbs *et al.*, 1993; Lee *et al.*, 1993; Zimmer *et al.*, 1996].

Da die Mechanismen, über die Calciumionen die Differenzierung der Epidermis beeinflussen, noch weitgehend unbekannt sind, war die Identifizierung der Loci  $\text{Ca}^{2+}$ -bindender Proteine der *S100*-Familie zu beiden Seiten der Strukturproteingene in der Region 1q21 äußerst bemerkenswert [Volz *et al.*, 1993; Schafer *et al.*, 1995; Mischke *et al.*, 1996]. Neben dieser Konstellation fällt die strukturelle Gemeinsamkeit zwischen

den Keratinmatrixproteinen und den *S100*-Proteinen auf. Beide enthalten zwei EF-Hand-Motive (benannt nach den E- und F-Helices des Parvalbumins [Übersicht in Kretsinger, 1976]), die jeweils ein zweiwertiges Kation binden können. *S100*-Proteine sind relativ kleine Proteine, die  $\text{Ca}^{2+}$ -abhängig die unterschiedlichsten Zielproteine erkennen. Sie modifizieren Enzyme diverser Signalkaskaden und des Energie-

stoffwechsels, und sie interagieren mit Transkriptionsfaktoren und mit den Filamenten des Zytoskeletts. Auf diese Weise üben sie Einfluß auf Beweglichkeit und Form der Zelle, auf den Zellzyklus und die -differenzierung aus [Übersicht in Schafer & Heizmann, 1996]. Ihr Vorkommen in der menschlichen Haut war ein Hinweis auf mögliche Funktionen bei der Differenzierung der Epidermis: *S100A3* (*S100E*) wurde in der Haarcuticula nachgewiesen [Kizawa *et al.*, 1996], *S100A6* in postmitotischen Zellen der Haarfollikel [Wood *et al.*, 1991], *S100A7* (Psoriasisin) wird verstärkt in psoriatischer Epidermis synthetisiert [Madsen *et al.*, 1991], *S100A8* (MRP8 oder Calgranulin A) und *S100A9* (MRP14 oder Calgranulin B) stehen  $\text{Ca}^{2+}$ -abhängig in Kontakt mit den IF [Goebeler *et al.*, 1995].



**Abbildung 6: Strukturelle Gemeinsamkeiten der von den Genfamilien der Region 1q21 codierten Proteine. Weiße und schwarze Rechtecke repräsentieren die N- bzw. die C-terminalen Peptidsequenzen, Dreiecke und Rauten interne Sequenzwiederholungen, Schleifen die  $\text{Ca}^{2+}$ -bindenden EF-Hand-Motive. Die jeweilige Anzahl der Peptidwiederholungen ist angegeben.**

## 5 Der Epidermale Differenzierungskomplex (EDC)

Genfamilien, deren Mitglieder hohe Sequenzhomologien aufweisen und auf einem Chromosom dicht zusammenliegen, sogenannte Gencluster, kommen im menschlichen Genom recht häufig vor. Als Beispiel seien hier die  $\alpha$ -Globin-Gene, die eine Untereinheit des Hämoglobins codieren, in der Region 16p [Barton *et al.*, 1982] sowie die Typ I- und Typ II-Keratin-Gene in 17q und 12q genannt [Yoon *et al.*, 1994; Ceratto *et al.*, 1997], die jeweils eine hohe Homologie untereinander aufweisen und ähnliche Funktionen in unterschiedlichen Geweben bzw. unterschiedlichen Entwicklungsstadien desselben Gewebes erfüllen. Auch in der Region 1q21 handelt es sich um Genfamilien, deren strukturell und funktionell verwandte Mitglieder jeweils als Cluster vorliegen. Die Besonderheit liegt hier jedoch darin, daß die Gene der drei Familien, obwohl sie unterschiedliche Funktionen erfüllen, alle an einer gemeinsamen, übergeordneten Aufgabe beteiligt sind: der Ausbildung der Epidermis. Eine derartige Anordnung konnte bisher selten nachgewiesen werden und wird im allgemeinen als Genkomplex bezeichnet. Das bekannteste Beispiel eines Genkomplexes ist der Haupthistokompatibilitätskomplex (MHC, *major histocompatibility complex*) des Menschen auf Chromosom 6. Dieser vereint in der Region 6p21.3 mehr als 50 für ein funktionsfähiges Immunsystem erforderliche Gene diverser Familien, die so unterschiedliche Aufgaben wie die Präsentation von Antigenen auf der Zelloberfläche sowie deren Prozessierung und Transport in der Zelle übernehmen [The MHC sequencing consortium, 1999]. Dementsprechend wurden die Genfamilien der Region 1q21, die während der Differenzierung der Epidermis sowohl interagierende Strukturproteine als auch regulatorische Proteine codieren, im Verlauf dieser Arbeit als „Epidermaler Differenzierungskomplex“ (EDC, *epidermal differentiation complex*) definiert [Mischke *et al.*, 1996].

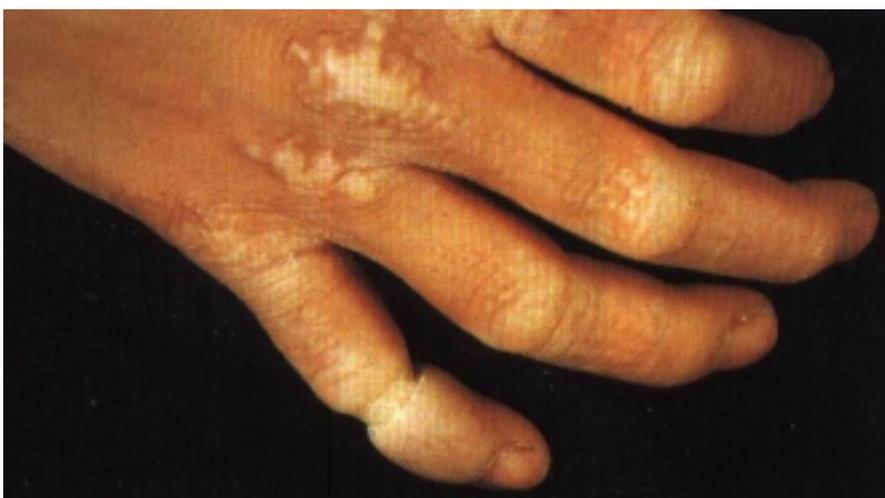
## 6 EDC-assozierte Krankheiten

Bis zum *Sixth International Workshop on Chromosome 1 Mapping 2000* in Iowa City, Iowa, waren 86 Krankheitsgene eindeutig Chromosom 1 zugeordnet [Schutte *et al.*, 2000]. Nur zwei Krankheiten wurden bisher auf eine Mutation in einem EDC-Gen zurückgeführt. Andere zeigten zwar einen Bezug zu abnormal exprimierten Genen des EDC, konnten als polygene oder multifaktorielle Erkrankungen aber nicht eindeutig lokalisiert werden. Neben den mit Verhornungsstörungen einhergehenden Hauterkrankungen fällt die häufige Verbindung der Region 1q21 zu Tumoren auf.

Unterschiedliche Mutationen im Loricrin-Gen sind für die beiden eindeutig mit dem EDC assoziierten Krankheiten verantwortlich. Bei der palmoplantaren Hyperkeratose vom Mutilans-Typ (Vohwinkel Syndrom) [Vohwinkel, 1929] handelt es sich um eine seltene, autosomal-dominante Erbkrankheit, deren diagnostische Charakteristika eine Hyperkeratose der Handflächen und Fußsohlen, hyperkeratose Läsionen und Fingerknöchelpolster auf dem Handrücken sowie Schnürringe an Fingern und

Zehen sind (Abb. 7). Letztere können durch Deformation der darunterliegenden Knochen zu spontanen Autoamputationen führen. Verursacht wird sie durch ein zusätzliches Nucleotid in Codon 231 [Maestrini *et al.*, 1996] oder Codon 209 [Armstrong *et al.*, 1998] des Loricrin-Gens, wodurch es zu einer Verschiebung des Leserahmens kommt, die zu einer Veränderung der C-terminalen Domäne des Proteins führt: Die für die Vernetzung innerhalb des CE verantwortlichen Glutamin- und Lysin-Reste werden durch Arginin ersetzt, und das Protein wird um 22 Aminosäuren verlängert.

Ein ähnlicher Defekt des Loricrin-Gens wurde bei der ebenfalls autosomal-dominant vererbten progressiven symmetrischen Erythrokeratoderma (PSEK) entdeckt [Ishida-Yamamoto *et al.*, 1997]. Cha-



**Abbildung 7: Vohwinkel Syndrom. Erkennbar ist die Hyperkeratose auf dem Handrücken und ein Schnürring am kleinen Finger.**

rakteristische Merkmale der PSEK sind erythematöse Plaques an den Extremitäten in Verbindung mit einer Palmo-plantarkeratose [Darrier, 1911]. Die Insertion liegt hier in Codon 224 und hat vergleichbare Auswirkungen auf die Proteinstruktur [Ishida-Yamamoto *et al.*, 1997]. Bei beiden Krankheiten

konnte als Folge des veränderten Loricrins ein gestörter Aufbau des CE nachgewiesen werden.

Die Bedeutung des CE für eine intakte Epidermis zeigt sich auch in verschiedenen Mutation des TGase 1-Gens (*TGMI*), die sich in einer reduzierten Genexpression bzw. verminderten Aktivität des für die Vernetzung des CE notwendigen Enzyms ausdrückt [Huber *et al.*, 1997]. Eine defekte TGase 1 wurde als Ursache der lamellären Ichthyose identifiziert [Huber *et al.*, 1995], einer seltenen autosomal-rezessiven Erbkrankheit, die durch generalisierte hyperkeratotische Schuppung auf rötlicher, entzündlicher Haut gekennzeichnet ist [Traupe *et al.*, 1984].

Andere Hauterkrankungen, deren Ursachen noch unklar sind, werden mit ungewöhnlichen Konzentrationen bestimmter, von EDC-Genen codierter Proteine assoziiert. Reduzierte bis abwesende Kera-tohyalingranula aufgrund einer verminderten Profilaggrin-Synthese wurden bei der autosomal-dominant vererbten Ichthyosis vulgaris festgestellt [Sybert *et al.*, 1985; Nirunsuksiri *et al.*, 1995], einer milderer Ichthyose mit trockener, schuppiger Haut, aber ohne Anzeichen einer Entzündung [Wells & Kerr, 1965]. Dagegen erfolgt in psoriatischer Haut, die sich durch rezidivierende oder chronische schuppige Exantheme auszeichnet und mit Hyperproliferation sowie immunologisch vermittelten Entzündungsreaktionen einhergeht, eine koordinierte Überexpression der EDC-Gene *S100A7*, *S100A8*,

*S100A9*, *SPRR1* und *SPRR2* [Hardas *et al.*, 1996]. In Übereinstimmung mit diesem Ergebnis konnte ein mit der Psoriasis assoziierter Locus in der Region 1q21 kartiert werden [Capon *et al.*, 1999].

Wie von dem ubiquitären Vorkommen und den vielfältigen Funktionen her zu erwarten ist, werden S100-Proteine mit den unterschiedlichsten Erkrankungen in Verbindung gebracht. Eine verminderte Expression von *S100A1* wurde in Kardiomyopathien nachgewiesen [Remppis *et al.*, 1996], eine erhöhte Coexpression von *S100A8* und *S100A9* bei diversen entzündlichen Prozessen [Bhardwaj *et al.*, 1992; Schmid *et al.*, 1995] und Cystischer Fibrose (Mukoviszidose) [Renaud *et al.*, 1994], einer allgemeinen Sekretionsstörung mit Bindegewebsvermehrung u.a. im Bereich der Bronchien. Ihre Bedeutung für die Zelldifferenzierung unterstreichend kommen S100-Proteine in ungewöhnlich hohen Konzentrationen aber vor allem in verschiedenen Tumorzellen vor: *S100A4* (CAPL) und *S100A7* u.a. beim Mammakarzinom [Pedrocchi *et al.*, 1994; Moog-Lutz *et al.*, 1995; Albertazzi *et al.*, 1998] und *S100A6* beim malignen Melanom [Weterman *et al.*, 1992; Boni *et al.*, 1997; Maelandsmo *et al.*, 1997]. Im Gegensatz zu den anderen Mitgliedern der S100-Familie ist die Synthese von *S100A2* (*S100L*) in entarteten Zellen stark herabgesetzt, was auf eine Tumorsuppressor-Funktion hinweisen könnte [Lee *et al.*, 1992; Maelandsmo *et al.*, 1997; Wicki *et al.*, 1997].

Chromosomale Veränderungen werden oft mit der Entstehung von Tumoren in Verbindung gebracht [Übersicht in Dos Santos & Van Kessel, 1999]. Bei Translokationen werden ganze Abschnitte innerhalb desselben oder zwischen verschiedenen Chromosomen ausgetauscht. Als Folge kann die normale Funktion eines von einem Strangbruch betroffenen Gens verlorengehen, und durch anschließende Fusion eines Bruchstücks mit einem anderen Gen oder dessen Kontrollregion können fehlerhafte Transkripte gebildet bzw. atypisch exprimiert werden. Auch bei Deletionen, die mit dem vollständigen Verlust eines bestimmten DNA-Abschnitts einhergehen, kann es sowohl zum Funktionsausfall als auch zur Bildung eines neuen Gens durch das Zusammenfügen normalerweise getrennter Sequenzen kommen. Duplikationen führen zu Vervielfältigungen bestimmter chromosomaler Bereiche; sie zeigen seltener phänotypische Auswirkungen, können aber zu einer Überexpression einzelner Gene führen. Die Region 1q21 ist relativ häufig an chromosomalen Umlagerungen beteiligt; obwohl sie nur ca. 5% von Chromosom 1 ausmacht, deckt sie mehr als 10% der identifizierten Chromosomenbrüche ab [Mitelman *et al.*, 1997]. Dabei wurden Verbindungen zu einer Vielzahl an Tumoren nachgewiesen, u.a. zu epithelialen Tumoren der Brust und der Niere [Gendler *et al.*, 1990; Bieche *et al.*, 1995; Hoggard *et al.*, 1995; Weterman *et al.*, 1996] und zu Sarkomen der Haut [Forus *et al.*, 1998]. Die von chromosomalen Umlagerungen der Region 1q21 verursachten Tumoren sind möglicherweise ein Anhaltspunkt dafür, daß die dort liegenden Gene die Zelldifferenzierung entscheidend beeinflussen.

## 7 Genomanalyse

Wie die vorhergehenden Abschnitte am Beispiel des EDC gezeigt haben, hat die Molekulargenetik das grundlegende Verständnis biologischer Prozesse in den letzten Jahren entscheidend geprägt und zu einem sprunghaften Anstieg der Erkenntnisse geführt. Das Wissen über das Erbgut ist dabei in vielen Bereichen zu einem unverzichtbaren Faktor geworden. Es ermöglicht neuartige Einblicke in die Entwicklungsgeschichte, lenkt die Aufmerksamkeit auf die noch bestehende Artenvielfalt und dient der Entwicklung neuer diagnostischer und therapeutischer Verfahren. Allerdings dürfen auch die Gefahren nicht übersehen werden, die diese Entwicklung mit sich gebracht hat. Dazu zählen die schwer vorhersehbaren Auswirkungen gentechnischer Eingriffe in die Natur, die vollständige Erfassung eines Individuums anhand seiner Gene und die damit verbundenen Diskriminierungsmöglichkeiten sowie die fortschreitende Kommerzialisierung der natürlichen Ressourcen, die dazu geführt hat, daß auf Gensequenzen, die von der Natur hervorgebracht wurden und seit Millionen von Jahren existieren, Patente vergeben werden.

Obwohl die Anfänge dieser Entwicklung bereits in der Mitte des 19. Jahrhunderts lagen, als der Augustinermönch Gregor Mendel (1866) durch Kreuzungsexperimente an der Gartenerbse *Pisum sativum* Gesetzmäßigkeiten bei der Vererbung verschiedener Merkmale erkannte, wurde erst in den letzten 30 Jahren durch die Entwicklung der grundlegenden molekularbiologischen Techniken der entscheidende Durchbruch in der Genomforschung erzielt. So wurde zwar bereits zu Beginn des 20. Jahrhunderts gefolgert, daß ein Gen auf dem X-Chromosom für die an das männliche Geschlecht gebundene Farbenblindheit verantwortlich ist [Wilson, 1911]; aber erst die Verwendung somatischer Zellhybride, die durch *in vitro*-Fusion von Zellen verschiedener Spezies hergestellt wurden, ermöglichte die erste physikalische Kartierung eines Gens (s. Abschnitt I 9) auf einem Autosom [Migeon & Miller, 1968].

Mit der Entdeckung von Bakterienenzymen, die DNA-Stränge an sequenzspezifischen Positionen schneiden, sogenannten Restriktionsendonucleasen, standen dann erstmals Werkzeuge zur gezielten Manipulation der Erbsubstanz zur Verfügung [Linn & Arber, 1968]. Kurz danach folgte die Identifizierung der Reversen Transkriptase in Retroviren, die entgegen der bei der Proteinsynthese üblichen Richtung Ribonucleinsäure (RNA, *ribonucleic acid*) in DNA umschreibt [Baltimore, 1970; Temin & Mizutani, 1970]. Die dabei entstehende komplementäre DNA (cDNA, *complementary DNA*) ist für die Genidentifizierung und -charakterisierung (s. Abschnitt I 10) unentbehrlich, da RNA aufgrund ihrer Instabilität für die meisten Analysemethoden nicht geeignet ist. Klonierungsvektoren für die Vervielfältigung rekombinierter, eukaryontischer DNA in Bakterien [Cohen *et al.*, 1973] und die DNA-Sequenzierung (s. Abschnitt II 17) [Maxam & Gilbert, 1977; Sanger *et al.*, 1977, 1980] – erste Versuche resultierten in fünf Basen pro Woche – vervollständigten schließlich die wichtigsten gentechnischen Grundlagen, bis 1985 die Einführung der Polymerasekettenreaktion-Methode (PCR, *polymerase chain reaction*) die Molekulargenetik noch einmal revolutionierte [Saiki *et al.*, 1985; Mullis & Faloo-

na *et al.*, 1987; Saiki *et al.*, 1988]: Sie ermöglichte die schnelle Vervielfältigung kleinster DNA-Mengen ohne Vektor und ohne Wirt (s. Abschnitt II 16).

Seitdem bekannt war, daß Gene für die Bildung von Enzymen verantwortlich waren [Beadle & Tatum, 1941], konzentrierte sich die Suche auf defekte Proteine als Auslöser erblich bedingter Krankheiten (z.B. wurde ein verändertes Hämoglobin als Ursache der Sichelzellenanämie identifiziert [Neel, 1949; Pauling *et al.*, 1949]). Aber erst als die entsprechenden Techniken der Genlokalisierung und Klonierung etabliert waren, konnte durch die funktionelle Klonierung ausgehend von einem bekannten Proteindefekt das zugehörige Gen ermittelt werden (z.B. führte die Bluterkrankheit zur Identifizierung des Blutgerinnungsfaktor VIII-Gens [Gitschier *et al.*, 1984]). Da die biochemischen Grundlagen der meisten Erbkrankheiten nicht bekannt sind, entwickelte man mit Hilfe der Molekularbiologie neue Vorgehensweisen zu deren Identifizierung. Ein häufig eingesetztes Verfahren ist die positionelle Klonierung [Collins, 1992], bei der zuerst die chromosomale Lage des gesuchten Gens ermittelt wird. Anschließend wird die DNA dieser Region kloniert, um die dort liegenden Kandidatengene zu isolieren. Über deren Nucleotidsequenzen können Eigenschaften der codierten Proteine abgeleitet und Mutationen identifiziert werden, die als Auslöser einer Krankheit in Frage kommen. Die dabei angewandten Techniken der Genkartierung, DNA-Klonierung und Genidentifizierung werden in den folgenden Kapiteln kurz erläutert.

## 8 Genetische Kartierung

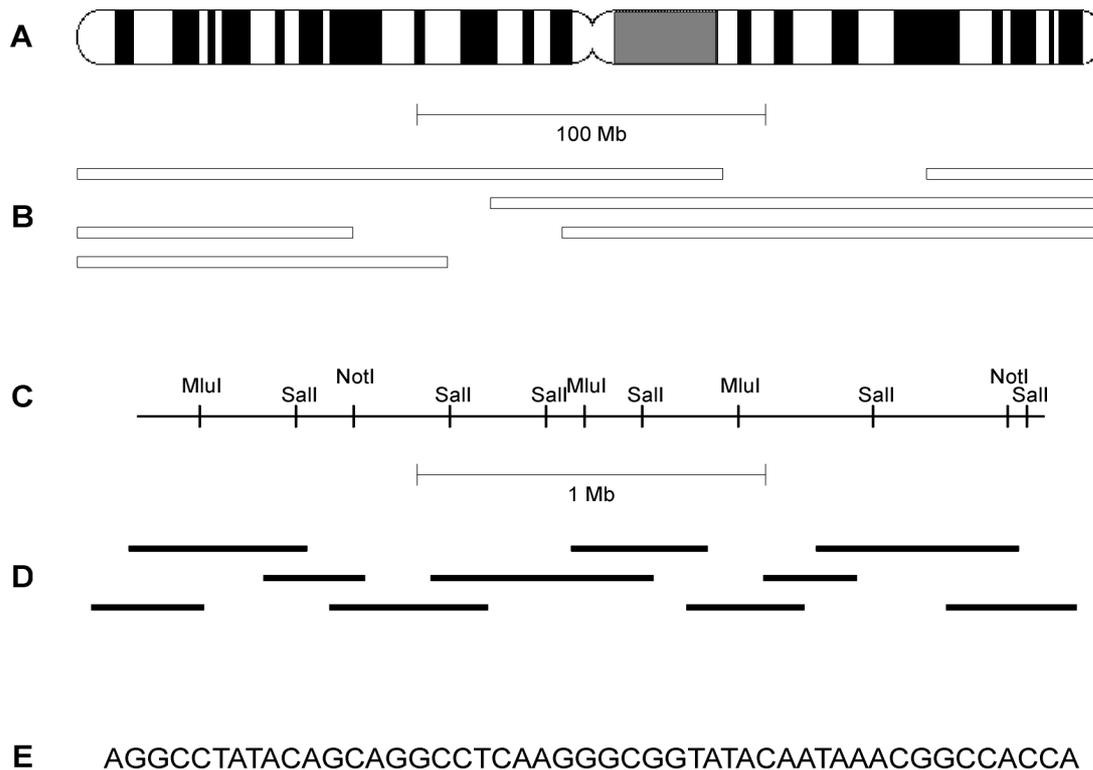
Die genetische Karte (Abb. 4) stellt das wichtigste Hilfsmittel für die Kartierung von Phänotypen dar. Sie gibt die relative Position der auf einem Chromosom liegenden Loci an und basiert auf den Untersuchungen Thomas H. Morgans und seiner Mitarbeiter (1915), die herausfanden, daß durch lineare Anordnung der Gene auf Chromosomen bestimmte Merkmale zusammen vererbt werden, daß es bei der meiotischen Teilung aber auch zu einem Austausch homologer Chromosomenabschnitte durch Überlagerung (*crossing over*) kommen kann. Als Folge werden ursprünglich gemeinsam vererbte (gekoppelte) Merkmale voneinander getrennt, wobei die Wahrscheinlichkeit, daß es zur Trennung zweier Loci kommt, von ihrem Abstand abhängt: Je näher zwei Punkte auf einem Chromosom beieinander liegen, desto unwahrscheinlicher ist eine Rekombination zwischen ihnen. Der Abstand zweier benachbarter Marker wird als Rekombinationshäufigkeit in CentiMorgan (cM) angegeben; 1 cM entspricht dabei einer Rekombinationshäufigkeit von 1%, d.h. bei einer von 100 Meiosen findet eine Neuverteilung statt [Übersicht in White & Lalouel, 1988]. Sie ist nicht nur von der Entfernung abhängig, sondern auch vom Geschlecht und von der chromosomalen Region [Broman *et al.*, 1998]. Nachweisbar sind Rekombinationen aber nur, wenn die zu analysierenden Merkmale unterschiedliche Ausprägungen aufweisen, d.h. wenn die entsprechenden Allele polymorph sind.

Bei den wenigen zur Verfügung stehenden Markern bestand bis Anfang der 80er Jahre ein Problem, da man auf die Analyse der Genprodukte beschränkt war und Polymorphismen in Proteinen seltener vorkommen – nur 25 bis 30 Marker waren damals bekannt, u.a. die Blutgruppen-Antigene [Übersicht in White & Lalouel, 1988]. Dann aber stellte sich heraus, daß die DNA-Sequenz selbst hochpolymorph war – alle 200 bis 500 bp weichen homologe Chromosomen im Durchschnitt voneinander ab – und eine nahezu unbegrenzte Anzahl potentieller Marker enthielt. Erkennen ließ sich eine solche Sequenzabweichung, wenn sie eine spezifische Schnittstelle für ein Restriktionsenzym zerstörte oder wenn durch sie eine neue Schnittstelle entstand, ein sogenannter Restriktionsfragment-Längenpolymorphismus (RFLP) [Übersicht in Botstein *et al.*, 1980]. Die erste veröffentlichte genetische Karte des Menschen basierte auf 403 polymorphen Loci, darunter 393 RFLP-Marker, mit einem durchschnittlichen Abstand von mehr als 10 cM [Donis-Keller *et al.*, 1987]. Weitere Fortschritte konnten durch Mikrosatellitenmarker erzielt werden, ebenfalls häufig vorkommende, kurze, aufeinanderfolgende Sequenzwiederholungen, deren Polymorphismus in der variablen Anzahl ihrer Di-, Tri- oder Tetranucleotidwiederholungen liegt (VNTR, *variable number of tandem repeats*) [Nakamura *et al.*, 1987]. Vor allem in Verbindung mit der PCR sind diese hochpolymorphen DNA-Marker schnell und einfach zu analysieren. Zur Durchführung werden lediglich zwei Oligonucleotid-Primer bekannter Sequenz benötigt, die den Polymorphismus einschließen (s. Abschnitt II 16) – dementsprechend werden die PCR-Marker auch als sequenzmarkierte Stellen (STSs, *sequence-tagged sites*) bezeichnet [Olson *et al.*, 1989].

Die Bedeutung der Kopplungsanalyse liegt vor allem in der Diagnose von Erbkrankheiten und der Kartierung für Krankheiten verantwortlicher Gene. Dabei versucht man, einen Marker zu finden, der mit dem entsprechenden Phänotyp zusammen vererbt wird; die Position des Markers auf der genetischen Karte dient dann als Ausgangspunkt für die physikalische Kartierung und Isolierung des die Krankheit verursachenden Locus.

## 9 Physikalische Kartierung

Mit Hilfe der physikalischen Kartierung läßt sich die exakte Position eines Locus auf einem Chromosom bestimmen; sie ermöglicht je nach Methode eine Auflösung von mehreren Megabasen bis zu einzelnen Basenpaaren. Eine Voraussetzung stellt das Vorhandensein eines spezifischen DNA-Markers (Hybridisierungssonde oder STS) dar; phänotypische Merkmale können somit nicht kartiert werden (ein an ein solches Merkmal gekoppelter Marker kann jedoch zur indirekten Kartierung genutzt werden). Im Gegensatz zur genetischen Karte, bei der 1 cM noch immer einem Abstand von unter 0,1 Mb bis zu mehr als 10 Mb entsprechen kann [Broman *et al.*, 1998], ist durch die physikalische Feinkartierung eine exakte Lokalisierung möglich. Angestrebt werden vor allem integrierte Karten, die durch physikalische Kartierung genetischer Marker die Informationen beider Methoden vereinen, so daß von



**Abbildung 8: Physikalische Kartierungsmethoden.** (A) Das Karyogramm ist Grundlage der cytogenetischen Karte. (B) Die weißen Balken repräsentieren die in den Zellhybriden enthaltenen Chromosomenfragmente des Menschen. (C) Die horizontale Linie symbolisiert einen DNA-Abschnitt, der von den oben angegebenen Restriktionsenzymen an den entsprechenden Stellen geschnitten wird. (D) Die schwarzen Balken stellen ein Contig aus überlappenden DNA-Klonen dar. (E) Eine aus den vier Basen bestehende DNA-Sequenz.

der Kopplungsanalyse direkt auf einen physikalisch begrenzten Chromosomenabschnitt geschlossen werden kann.

Cytogenetische Karte (Abb. 8A): In der Metaphase der Mitose liegen Chromosomen kondensiert vor, sie zeigen den typischen Aufbau zweier am Centromer verbundener Chromatiden. Mit bestimmten Farbstoffen (z.B. Giemsa) entstehen für die einzelnen Chromosomen charakteristische Bandenmuster, wobei sich in der Prometaphase im menschlichen Genom bis zu 850 Banden unterscheiden lassen. Die sich hieraus ergebenden Karyogramme ermöglichen die Identifizierung von Chromosomenbrüchen und sind Voraussetzung der Fluoreszenz-*in situ*-Hybridisierung (FISH). Bei dieser Methode hybridisiert eine an einen Fluoreszenzfarbstoff gekoppelte DNA-Sonde mit denaturierten, kondensierten Metaphase-Chromosomen (550 Banden); unter dem Fluoreszenzmikroskop kann das Signal dann einer bestimmten Bande, die im Durchschnitt etwa 6 Mb entspricht, zugeordnet werden [Fan *et al.*, 1990].

Zellhybridkarte (Abb. 8B): Durch Fusion menschlicher Zellen mit einer Nagetier-Zelllinie entstehen Hybridzellen, die neben dem vollständigen Chromosomensatz des Nagers ganze oder fragmentierte menschliche Chromosomen enthalten (somatische Zellhybride). Nach Kultivierung und Klonierung einzelner Hybridzellen erhält man zahlreiche Zelllinien mit einem definierten, aber unterschiedlichen Gehalt an menschlicher DNA [Weiss & Green, 1967]. Im Idealfall läßt sich so das gesamte Genom abdecken. Die unterschiedlichen Zellhybride werden mit bekannten Markern charakterisiert und kön-

nen anschließend für die Genlokalisierung verwendet werden. Indem man die einzelnen Hybridzelllinien Röntgenstrahlung aussetzt und erneut mit Nagetierzellen fusioniert, entstehen Strahlungshybride (RHs, *radiation hybrids*), deren Anteil an menschlicher DNA weiter herabgesetzt ist, wodurch eine bis auf wenige Megabasen genaue Kartierung möglich wird [Goss & Harris, 1975].

**Restriktionskarte** (Abb. 8C): Für die Restriktionskartierung wird hochmolekulare, genomische DNA in mehreren Ansätzen mit unterschiedlichen Restriktionsenzymen fragmentiert, mit Hilfe einer Gelelektrophorese der Größe nach aufgetrennt und auf eine Membran übertragen (Southern-Blot, s. Abschnitt II 14). Durch Hybridisierungen mit DNA-Sonden können bekannte Marker einzelnen Fragmenten zugeordnet werden. Erkennen zwei Sonden ein gemeinsames Restriktionsfragment, entspricht ihr Abstand im Genom maximal der Größe dieses Fragments. Hybridisieren die Sonden gleichzeitig mit unterschiedlichen, das gemeinsame Fragment überlappenden Fragmenten, die durch Restriktionsspaltung mit einem zweiten Enzym erzeugt wurden, kann die Reihenfolge der Fragmente und der enthaltenen Marker festgelegt werden. Die Auflösung der Restriktionskarte hängt von den verwendeten Enzymen ab; sie liegt zwischen wenigen hundert Basenpaaren für häufig schneidende Restriktionsendonucleasen, wie EcoRI, und mehreren Megabasen für selten schneidende, wie NotI [Übersicht in Evans, 1991]. Die Überbrückung großer Entfernungen ist eng mit der Entwicklung elektrophoretischer Trennverfahren, wie der Pulsfeldgelelektrophorese (PFGE) [Schwartz & Cantor, 1984; Übersicht in Lai *et al.*, 1989] und der Gelelektrophorese im rotierenden Feld (ROFE, *rotating field gel electrophoresis*) [Ziegler & Volz, 1992], verbunden, die eine Trennung von DNA-Fragmenten bis zu einer Größe von ca. 10 Mb ermöglicht.

**Klon-Contig** (Abb. 8D): Eine besonders effektive Charakterisierung einer Region ist durch das Anordnen überlappender DNA-Klone zu einem Contig (von *contiguous*, aneinander grenzend) möglich [Coulson *et al.*, 1988]. Als Ausgangspunkt dient eine DNA-Bibliothek, die durch Restriktionsspaltung

**Tabelle 2: Vor- und Nachteile gebräuchlicher Klonierungssysteme**

	Wirt	Insertgröße in kb	Kopien pro Zelle	DNA-Isolierung	Sequenzierung	Umlagerungen
Plasmid	<i>E. coli</i>	-10	5-500	sehr gut	sehr gut	selten
Bakteriophage $\lambda$	<i>E. coli</i>	5-25	>250	gut	gut	sehr selten
Cosmid	<i>E. coli</i>	35-45	3-50	sehr gut	sehr gut	möglich
PAC	<i>E. coli</i>	70-300	1-2	sehr gut	gut	selten
BAC	<i>E. coli</i>	50-300	1	sehr gut	gut	sehr selten
YAC	<i>S. cerevisiae</i>	50-2000	1	schwierig	schwierig	häufig

PAC, *P1 artificial chromosome*; BAC, *bacterial artificial chromosome*; YAC, *yeast artificial chromosome*.

genomischer DNA und nachfolgende Klonierung hergestellt wird. Durch partielle Spaltung wird sichergestellt, daß dabei auch überlappende Klone entstehen. Anschließend werden Überlappungen durch Hybridisierung zweier Klone mit derselben Sonde nachgewiesen und die identifizierten Klone zu einem Contig, das die zu analysierende genomische Region kontinuierlich abdeckt, zusammengesetzt. Je nach Auswahl des Klonierungsvektors variieren die aufgenommenen DNA-Fragmente, die für die Auflösung des Contigs verantwortlich sind, zwischen 5 und 25 kb für den bakteriellen Vektor

lambda und bis zu 2 Mb für künstliche Hefechromosomen (YACs, *yeast artificial chromosomes*). Durch zusätzliche Restriktionskartierung der Klone kann die Auflösung des Contigs nahezu beliebig erhöht werden. Es muß jedoch beachtet werden, daß bestimmte DNA-Sequenzen des Menschen (z.B. einige repetitive Sequenzen) in Bakterien oder Hefezellen nicht vertreten sind und sich in diesen Organismen nur schwer klonieren lassen [Neil *et al.*, 1990; Kang & Cox, 1996; Bieth *et al.*, 1997; Kouprina *et al.*, 1999]. Bei YACs ist darüberhinaus mit einem hohen Anteil (bis zu 50%) chimärer Klone zu rechnen [Green *et al.*, 1991; Cohen *et al.*, 1993]; die erzielten Ergebnisse müssen somit im menschlichen Genom mittels komplementärer Techniken verifiziert werden. Der wesentliche Vorteil eines Contigs besteht darin, daß im Gegensatz zu den anderen Kartierungsmethoden die DNA der zu analysierenden Region bereits kloniert vorliegt – die Klone eines Contigs finden u.a. Verwendung bei der Herstellung spezifischer Marker, der Genidentifizierung und -sequenzierung. Unterschiedliche Klonierungssysteme mit ihren Anwendungsbereichen, Vor- und Nachteilen sind in Tabelle 2 zusammengefaßt.

Sequenzierung (Abb. 8E): Mit der Nucleotidsequenz wird die genaueste physikalische Kartierung erreicht. Als Ausgangsprodukt dient ein in einem Klonierungsvektor oder durch PCR amplifiziertes DNA-Fragment, dessen Einzelstränge nach Anlagerung je eines spezifischen Primers in getrennten Reaktionen sequenziert werden. Beim Kettenabbruchverfahren nach Sanger [Sanger *et al.*, 1977], das auf der Synthese basenspezifisch endender DNA-Fragmente basiert (s. Abschnitt II 17), können bis zu 1000 Nucleotide pro Reaktion bestimmt werden. Längere DNA-Abschnitte werden schrittweise mit Hilfe wandernder Primer, die an den Enden der vorhergehenden Reaktionen plaziert werden, sequenziert (*primer walking*). Voraussetzung für die Sequenzierung größerer DNA-Abschnitte ist ein Contig, dessen überlappende Klone eine durchgehende Basenfolge ergeben. Gene können anschließend durch computergestützte Analysen der DNA-Sequenz identifiziert werden (s. Abschnitt I 10). Der Nachteil dieser Vorgehensweise ist ihr sehr hoher Arbeitsaufwand.

## 10 Methoden zur Identifizierung neuer Gene

Die zur Identifizierung unbekannter Gene einer bestimmten chromosomalen Region eingesetzten Methoden nutzen zwei für Gene charakteristische Eigenschaften: die Expression von RNA-Produkten sowie die Konservierung codierender und regulatorischer DNA-Sequenzen in der Evolution. Alle diese Methoden sind auf genomische DNA-Klone der zu analysierenden Region angewiesen, die zumeist durch Anlagerung an komplementäre Nucleinsäuren definierter DNA- bzw. Boten-RNA- (mRNA, *messenger RNA*)- oder cDNA-Populationen zur Selektion genspezifischer Sequenzen führen. Für jede der Methoden gelten gewisse Einschränkungen. So können z.B. bei Verwendung von mRNA bzw. cDNA nur die in der zugrundeliegenden Zellpopulation gewebe- und entwicklungsspezifisch exprimierten Gene gefunden werden; und der Einsatz größerer genomischer DNA-Klone, wie künstlicher

Chromosomen des Bakteriophagen P1 (PACs, *P1 artificial chromosomes*), künstlicher Bakterienchromosomen (BACs, *bacterial artificial chromosomes*) und vor allem von YACs, führt aufgrund der unspezifischen Hybridisierung enthaltener repetitiver Sequenzen zu einem hohen Anteil falsch positiver Ergebnisse.

Zoo-Blot-Hybridisierung: Ein DNA-Klon wird unter wenig stringenten Bedingungen, die eine Paarung von Nucleotidsträngen trotz einiger Abweichungen zulassen, mit einem Southern-Blot hybridisiert, auf dem genomische DNA verschiedener Spezies fixiert ist (Zoo-Blot). Positive Hybridisierungssignale weisen darauf hin, daß innerhalb des klonierten Fragments Gensequenzen liegen, die aufgrund ihrer funktionellen Bedeutung während der Evolution in verschiedenen Arten konserviert wurden.

Hybridisierung von mRNA/cDNA: mRNA-Populationen aus verschiedenen Geweben oder Zelllinien werden auf eine Membran übertragen (Northern-Blot) und mit einer genomischen DNA-Sonde hybridisiert [Alwine *et al.*, 1977]. Ein positives Hybridisierungssignal ist ein Anhaltspunkt für codierende Sequenzen in der Sonde. Bei Verwendung einer cDNA-Bibliothek können die von der Sonde identifizierten Klone unmittelbar analysiert werden.

cDNA-Selektion: Bei diesem Verfahren dient die genomische DNA eines Klons als stationäre Phase. Durch Zugabe eines cDNA-Gemisches kommt es bevorzugt zur Hybridisierung von in beiden Populationen übereinstimmenden, genspezifischen Sequenzen. Nachfolgend wird die nicht hybridisierende cDNA abgewaschen; die gebundenen cDNA-Fragmente werden eluiert und mittels PCR amplifiziert. Mehrere Selektionsrunden führen zur Anreicherung der gesuchten cDNA-Sequenzen [Lovett *et al.*, 1991].

Exon-Identifizierung (*exon trapping*): Exons in klonierter DNA lassen sich erkennen, indem man sie in einem geeigneten System zur Expression bringt. Ein Expressionsvektor besteht aus einem Plasmid, das ein komplettes Minigen enthält, mit Replikationsursprung, einem starken Promotor, zwei kleinen Exons und einem Polyadenylierungssignal. Innerhalb des Introns liegt eine multiple Klonierungsstelle, in die ein genomisches Fragment subkloniert wird. Anschließend wird mit diesem Konstrukt eine eukaryontische Zelllinie transfiziert, welche die RNA transkribiert und prozessiert. Über die Größe der gespleißten mRNA kann festgestellt werden, ob sie ein zusätzliches, aus der genomischen DNA stammendes Exon enthält [Duyk *et al.*, 1990].

Identifizierung von CpG-Inseln: In Eukaryonten sind viele Gene vor allem in der Promotorregion am 5'-Ende mit GC-reichen Sequenzen, sogenannten CpG-Inseln, assoziiert. Durch Restriktionsspaltung mit Enzymen, deren Erkennungssequenz ebenfalls GC-reich ist und die in diesen Regionen häufiger schneiden als im übrigen Genom, können CpG-Inseln erkannt werden [Lindsay & Bird, 1987].

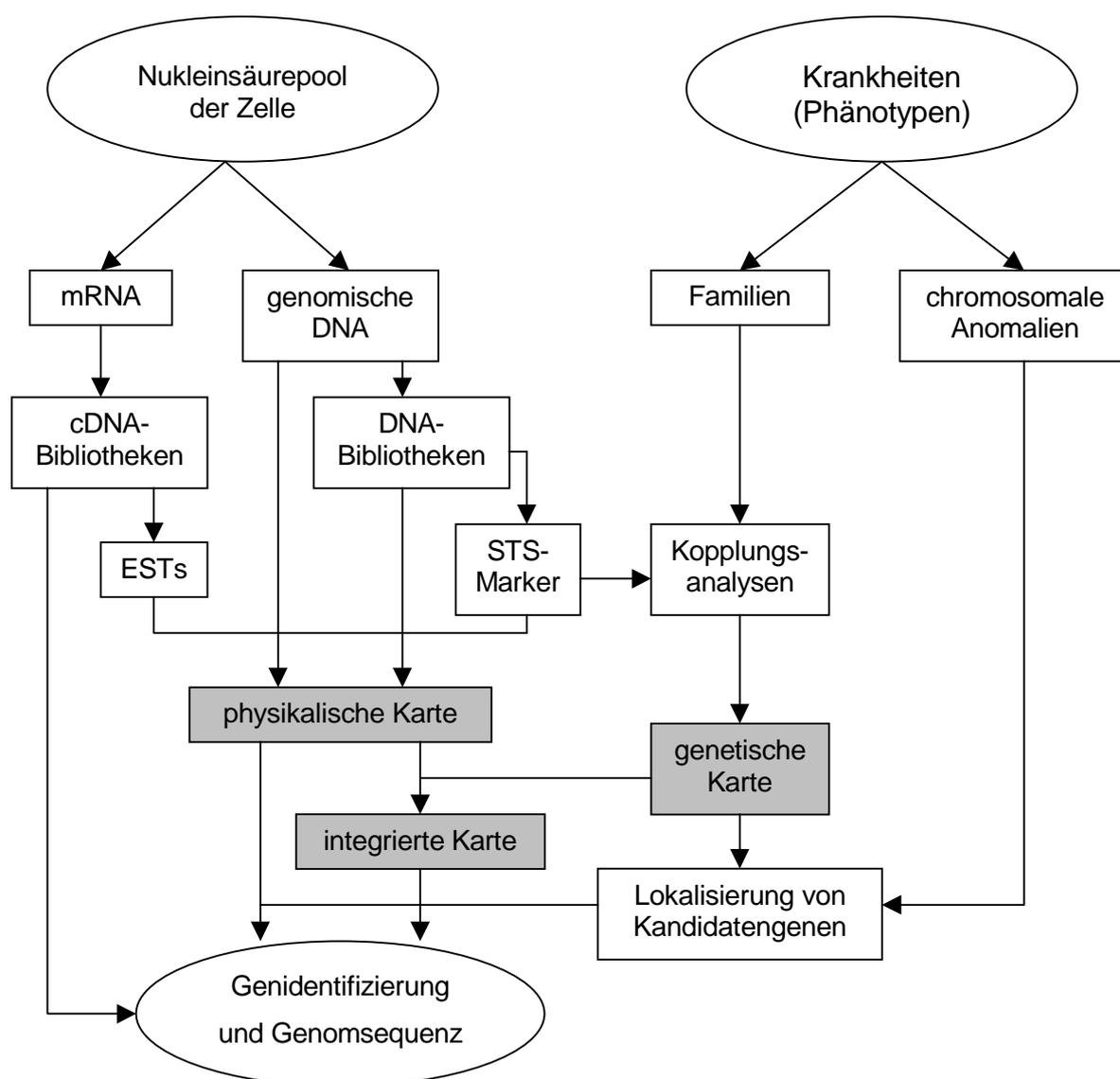
DNA-Sequenzanalyse: Nach Bestimmung der Nucleotidsequenz eines DNA-Klons ermöglichen diverse Computerprogramme ihre Überprüfung auf enthaltene genspezifische Abschnitte. Die Homologiesuche vergleicht die Nucleotidsequenz und davon abgeleitete Aminosäuresequenzen aller sechs Leserahmen (drei für jede Richtung) mit den in den Datenbanken verfügbaren DNA- und Proteinsequenzen.

zen (BLAST, *basic local alignment search tool*) [Übersicht in Altschul *et al.*, 1997]. Eine signifikante Übereinstimmung kann ein Hinweis auf eine enthaltene Gensequenz sein. Eine andere Möglichkeit besteht in der Vorhersage von Exons; die Programme berücksichtigen dabei bestimmte Konsensussequenzen, wie Start- und Stopcodons für die Translation, Exon/Intron-Übergänge und Polyadenylierungssignale sowie die Länge der offenen Leserahmen (ORFs, *open reading frames*) [Überbacher *et al.*, 1996]. Obwohl in der DNA-Sequenz nahezu alle Gene eines Organismus codiert sind, ist ihre Identifizierung mit dieser Methode nicht immer gewährleistet. Die Trefferquote für transkribierte Sequenzen beträgt ungefähr 70%; außerdem muß mit ca. 30% falsch positiven Exons gerechnet werden [Lopez *et al.*, 1994].

## 11 Das Humangenom-Projekt

Angesichts der Größe des menschlichen Genoms – nach neuesten Erkenntnissen geht man von höchstens 40000 Genen aus, die auf etwa 3,3 Milliarden Basenpaare oder 3300 Mb der 23 Chromosomen des einfachen Chromosomensatzes verteilt sind [The International Human Genome Sequencing Consortium, 2001] – ist ein gesuchtes Krankheitsgen oftmals mit der Nadel im Heuhaufen zu vergleichen. Hinzu kommt, daß viele Krankheiten nicht auf ein einzelnes Gen zurückgeführt werden können, sondern polygen oder multifaktoriell bedingt sind. Folglich waren die Bemühungen einzelner Arbeitsgruppen in den 80er Jahren, obgleich erfolgreich, sehr arbeits- und kostenintensiv; z.B. vergingen von der Lokalisierung bis zur Identifizierung des die Cystische Fibrose verursachenden Gens mehr als drei Jahre [Wainwright *et al.*, 1985; Kerem *et al.*, 1989]. Weder die genetischen und physikalischen Karten noch die Techniken an sich waren für eine effiziente Analyse ausreichend entwickelt.

Um ihre Forschungsaktivitäten zu koordinieren, schlossen sich die großen Organisationen der USA 1990 im Humangenom-Projekt (*human genome project*) zusammen, an dem sich nationale Organisationen weltweit beteiligten und an dessen Ende die vollständige Sequenzierung und die Identifizierung aller Gene des menschlichen Genoms stehen soll (Abb. 9) [Collins & Galas, 1993]. Die ersten Etappen wurden mit der Erstellung einer hochauflösenden genetischen Karte mit einem durchschnittlichen Markerabstand von 1,6 cM [Dib *et al.*, 1996] und einer auf YAC-Klonen und STS-Markern basierenden integrierten Karte des Menschen [Chumakov *et al.*, 1995] erreicht. Da YACs für eine verlässliche Analyse einzelner chromosomaler Regionen aufgrund häufiger Umlagerungen jedoch nur bedingt geeignet sind, wurden zusätzlich PAC-, BAC- und Cosmid-Contigs für die abschließende Sequenzierung erstellt. Eine möglichst vollständige Genidentifizierung soll durch die zusätzliche Charakterisierung exprimierter Sequenzen garantiert werden. Zu diesem Zweck wurden Klone diverser gewebespezifischer cDNA-Bibliotheken ansequenziert und als exprimierte sequenzmarkierte Stellen (ESTs, *expressed sequence tags*) auf Strahlungshybriden lokalisiert. Die RH-Transkript-Karte des Menschen enthält etwa 30000 durch EST-Marker repräsentierte Gene [Deloukas *et al.*, 1998].



**Abbildung 9: Die Analyse des menschlichen Genoms. Die wichtigsten Strategien des Humangenom-Projekts und ihre Beziehungen zueinander sind dargestellt.**

Ergänzt wird die Analyse des menschlichen Erbguts durch Genomprojekte verschiedener Modellorganismen. Sie gewähren Einblicke in die evolutionäre Entwicklung der verschiedenen Arten und dienen der Aufklärung grundlegender molekularer Mechanismen, die bei der Genregulation oder bei genetisch bedingten Krankheiten eine Rolle spielen und sich oftmals auch auf den Menschen übertragen lassen. Darüberhinaus konnten die Erfahrungen, die bei der Sequenzierung der Modellorganismen gesammelt wurden, für die Entwicklung neuer Techniken genutzt werden, welche beim Humangenom-Projekt eingesetzt werden. Bisher wurden u.a. vollständig sequenziert: als erster freilebender Organismus das Bakterium *Haemophilus influenzae* (1,8 Mb) [Fleischmann *et al.*, 1995], als erster Eukaryont die Bäckerhefe *Saccharomyces cerevisiae* (14,5 Mb) [Übersicht in Goffeau *et al.*, 1996], das Bakterium *Escherichia coli* (4,6 Mb) [Blattner *et al.*, 1997], als erstes Tier der Fadenwurm *Caenorhabditis elegans* (100 Mb) [Übersicht in The C. elegans Sequencing Consortium, 1998], als erste

Pflanze die Ackerschmalwand *Arabidopsis thaliana* (120 Mb) [The Arabidopsis Genome Initiative, 2000] und als einer der besterforschten Organismen die Fruchtfliege *Drosophila melanogaster* (165 Mb) [Adams *et al.*, 2000]. Obwohl zwei vorläufige Versionen des menschlichen Genoms bereits veröffentlicht wurden [The International Human Genome Sequencing Consortium, 2001; Venter *et al.*, 2001], wird es noch einige Zeit in Anspruch nehmen, bis eine lückenlose Basenfolge vorliegt, da viele Sequenzbruchstücke noch nicht geordnet sind und schwer zu klonierende Regionen größtenteils ausgelassen wurden. Weitere Projekte befassen sich mit dem Kugelfisch *Fugu rubripes rubripes* (400 Mb), der Maus *Mus musculus* (3000 Mb) und der Ratte *Rattus norvegicus* (3000 Mb).

Damit die gewonnenen Informationen jederzeit und weltweit für Recherchen zur Verfügung stehen, werden sie in elektronischen Datenbanken gespeichert, auf die über das Internet, vor allem über das *world wide web* (WWW), zugegriffen werden kann. Die wichtigsten Nucleotidsequenzdatenbanken sind GenBank am *National Center of Biotechnology Information* (NCBI) in den USA, EMBL (*European Molecular Biology Laboratory*) am *European Bioinformatics Institute* (EBI) in Großbritannien und DDBJ (*DNA Data Bank of Japan*) am *National Institute of Genetics* (NIG) in Japan, wobei ein regelmäßiger Datenabgleich der einzelnen Zentren gewährleistet ist. Desweiteren wurden Datenbanken für zahlreiche Themenbereiche eingerichtet, u.a. PIR (*Protein Information Resource*) und SWISS-PROT für Proteinsequenzen, GDB (*Genome Data Base*) für die Genkartierung, OMIM (*Online Mendelian Inheritance in Men*) für genetisch bedingte Krankheiten und Phänotypen, Généthon und CHLC (*Cooperative Human Linkage Center*) für genetische Karten des Menschen, CEPH (*Centre d'Etude du Polymorphisme Humain*) und das *Whitehead Institute/MIT* (*Massachusetts Institute of Technology*) für physikalische Karten auf der Basis von DNA-Klonen. Neben der reinen Archivierung haben die einzelnen Institute zahlreiche Analyseprogramme zur Charakterisierung von Nucleotid- und Proteinsequenzen entwickelt (s. Abschnitt II 18), die im optimalen Fall aus einer genomischen DNA-Sequenz die Funktion des codierten Proteins vorhersagen können.

Wie an der Größe der fertiggestellten Genome (s.o.) zu erkennen ist, haben die zunehmende Verbesserung und Automatisierung der Analysemethoden in den letzten Jahren zu einem rasanten Anstieg der produzierten Daten geführt – mittlerweile verdoppeln sich die Einträge in den Datenbanken alle 18 Monate. Es wird damit gerechnet, daß das menschliche Genom im Jahr 2003 vollständig und lückenlos sequenziert sein wird. Mit der Sequenz werden aber weder alle Gene identifiziert sein noch die genetischen Grundlagen aller Erbkrankheiten – nicht zu vergessen, daß jedes Individuum ein einzigartiges Genom besitzt. Da auch die Funktionen selbst vieler bekannter Gene noch nicht aufgeklärt sind, steht der interessanteste Teil des Projekts noch bevor.

## 12 Ziele der Arbeit

Mit der Definition des Epidermalen Differenzierungskomplexes konnte bereits zu Beginn der Arbeit eine adäquate Bezeichnung gefunden werden, die die Bedeutung der chromosomalen Region 1q21 für die Differenzierung der menschlichen Epidermis widerspiegelt. Durch die vorliegende Arbeit sollte der Stellenwert des EDC auch auf genomischer Ebene weiter untermauert werden. In der ersten Phase sollten die entsprechenden Grundlagen für dieses Vorhaben geschaffen werden. Darauf aufbauend sollte der EDC dann molekulargenetisch analysiert werden, d.h. die Gene der Region sollten möglichst vollständig identifiziert, kartiert und charakterisiert werden. Entsprechend umfangreich waren die von diesem „1q21-Genomprojekt“ verfolgten Ziele:

- 1) Ausgehend von einer genomischen Restriktionskarte der Region 1q21 [Volz *et al.*, 1993], sollten durch Southern-Blot-Hybridisierung mit Sonden bereits kartierter EDC-Gene überlappende DNA-Klone identifiziert und zu einem Contig zusammengesetzt werden, das den EDC möglichst vollständig abdeckt.
- 2) Durch Subklonierung der den EDC begrenzenden DNA-Klone in Plasmide sollten neue Hybridisierungssonden erzeugt werden, mit denen weitere Klone der Region identifiziert werden sollten, um das Contig auszudehnen und den Anschluß an benachbarte physikalische Karten herzustellen.
- 3) Die Orientierung des EDC auf Chromosom 1 und seine genaue Position auf der cytogenetischen Karte sollte durch FISH zweier den EDC begrenzender DNA-Klone auf Metaphase-Chromosomen ermittelt werden.
- 4) Die Kartierung genetischer STS-Marker innerhalb des Contigs sollte die Integration der physikalischen Karte des EDC und der genetischen Karte von Chromosom 1 ermöglichen – als Grundlage für Kopplungsanalysen von Krankheiten, die mit den Genen der Region in Verbindung gebracht werden.
- 5) Eine aus primären Keratinozyten des Menschen gewonnene cDNA-Bibliothek, die innerhalb des Biomed-Projekts „*Integrated analysis of expression and chromosomal organisation of genes localised on human chromosome 1q21: Implications for human disease and cancer*“ zur Verfügung stand, war für die Identifizierung neuer EDC-Gene vorgesehen. Damit sie möglichst effizient und von allen Arbeitsgruppen des Projekts gleichzeitig genutzt werden konnte, sollte die cDNA-Bibliothek mit Unterstützung des Ressourcenzentrums in Berlin in hoher Dichte auf Membranen übertragen werden.
- 6) Die Verwaltung der feingerasterten cDNA-Bibliothek für das gesamte Biomed-Projekt sollte vom Institut für Immunogenetik übernommen werden; für die erwarteten umfangreichen Auswertungen sollte eine Datenbank erstellt werden.
- 7) Aufgrund der unregelmäßigen und relativ große Lücken aufweisenden Anordnung der bekannten EDC-Gene bestand die Aussicht, daß weitere Mitglieder dieses Genkomplexes in der Region lo-

kalisiert sind. Um diese aufzufinden, sollte eine Methode etabliert werden, welche die geschaffenen Ressourcen möglichst effizient nutzt: Durch Hybridisierung der feingerasterten cDNA-Bibliothek mit Sonden aus genomischer DNA sollten in Keratinozyten exprimierte Gene der Region 1q21 identifiziert werden.

- 8) Um die Lage der den cDNA-Klonen entsprechenden Gene im EDC zu verifizieren, sollten sie auf der genomischen Restriktionskarte der Region 1q21 positioniert werden; gleichzeitig sollte die integrierte Karte des EDC ausgedehnt werden.
- 9) Über computergestützte Sequenzanalysen sollten die im EDC kartierenden cDNA-Klone auf charakteristische Sequenzmerkmale hin untersucht werden, wie die  $\text{Ca}^{2+}$ -bindenden EF-Hand-Motive oder die Strukturproteine kennzeichnenden Sequenzwiederholungen der bekannten EDC-Gene. Da die Differenzierung der Epidermis ein äußerst komplexer Vorgang ist, der von den unterschiedlichsten Stoffen beeinflusst wird (s. Abschnitt I 3), kamen auch Gene anderer Proteine, wie diverser Rezeptoren, Mediatoren oder Transkriptionsfaktoren, als Kandidaten für den EDC in Betracht.