

IV DISKUSSION

1 Die Ressourcen

1.1 Das YAC-Contig des EDC

Um den EDC möglichst vollständig analysieren zu können, wurde ein YAC-Contig der chromosomalen Region 1q21 konstruiert, das nicht nur die bekannten, innerhalb von 2 Mb kartierenden EDC-Gene [Gibbs *et al.*, 1993; Volz *et al.*, 1993; Schafer *et al.*, 1995; Mischke *et al.*, 1996] abdeckte, sondern auch die angrenzenden Regionen, ca. 3 Mb proximal und 1 Mb distal des etablierten Genkomplexes. Es stellte eine Hauptgrundlage für das von der Europäischen Union (EU) geförderte Biomed-Projekt „*Integrated analysis of expression and chromosomal organisation of genes localised on human chromosome 1q21: Implications for human disease and cancer*“ dar. Insgesamt besteht das Contig aus 24 überlappenden YACs und elf weiteren Klonen, die aus der instabilen YAC-Kultur 950_e_2 isoliert wurden. Im Laufe der vorliegenden Arbeit wurden 43 Genmarker, die unter Berücksichtigung aller Mitglieder der *SPRR*- und *S100*-Genfamilien [Gibbs *et al.*, 1993; Schafer *et al.*, 1995; Wicki *et al.*, 1996a; Wicki *et al.*, 1996b] 51 Gene repräsentieren, 20 1q21-spezifische Hybridisierungsmarker und sieben polymorphe STS-Marker auf den YACs kartiert. Einschließlich des *ZIRTL* (*zink-iron regulated transporter-like*)-Gens, das ein Metallionen-Transportprotein codiert und von einer kooperierenden Arbeitsgruppe einzelnen YACs des Contigs zugeordnet wurde [Lioumi *et al.*, 1999], beträgt der durchschnittliche Abstand zweier Loci innerhalb des 6 Mb-Contigs nur ca. 70 kb.

Durch die Kartierung der Loci auf einzelnen *SalI*-Restriktionsfragmenten der YACs wurde die Auflösung des Contigs erhöht. Die ermittelten Fragmentgrößen stimmen sowohl mit der *SalI*-Restriktionskarte des PAC/BAC/Cosmid-Contigs des EDC überein, das sich von *NICE6* bis *S100A1* erstreckt [South *et al.*, 1999], als auch mit den *SalI*-Fragmenten von YAC 100_f_3 [Schafer *et al.*, 1995]. Im centromeren Bereich ist die Markerdichte des Contigs allerdings zu gering, so daß zwischen *NICE6* und *IVL* nicht alle Fragmente nachgewiesen werden konnten. Dabei entsprechen die Lücken auf der YAC-Restriktionskarte den Fragmentgrößen, die aus dem PAC/BAC/Cosmid-Contig übernommen wurden [South *et al.*, 1999]. Zusätzliche Fragmente wurden distal von *S100A1* und proximal von *NICE6* identifiziert. In der telomeren Region konnten die *SalI*-Fragmente, abgesehen von dem in YACs instabilen Bereich zwischen *S100A1* und 24f11 (s.u.), lückenlos positioniert werden. Dagegen sind die *SalI*-Fragmente in der centromeren Region aufgrund der wenigen zur Verfügung stehenden Marker unvollständig. Sie decken mehr als die Hälfte der in den YACs enthaltenen 3,3 Mb proximal von *NICE6* ab, wobei die *SalI*-Fragmente des 2,2 Mb großen Bereichs bis einschließlich YAC 789_f_2 zu 85% identifiziert werden konnten.

Zwar sind nahezu 90% der Loci über mindestens zwei YACs miteinander verbunden (*doubly linked*) und auf drei YACs vorhanden, so daß die Größen der intakten genomischen *SalI*-Fragmente bestimmt werden konnten – wie bei der Erstellung des Contigs bereits spekuliert wurde, sind aber auch zwei Abschnitte zu erkennen, die maximal von einem YAC abgedeckt sind und deren *SalI*-Fragmentgrößen somit nicht gesichert sind. *ANXA9* und *NICE8* wurden einzig auf YAC 764_a_1 nachgewiesen, eine Verbindung zum proximal benachbarten Marker war mit Hilfe der YACs des Contigs nicht herzustellen. Dieses Ergebnis deckt sich mit den Beobachtungen, die bei der Erstellung einer ebenfalls hochauflösenden Karte der Region, bestehend aus zwölf fragmentierten YACs des EDC-Contigs, gemacht wurden. Deren Charakterisierung ergab eine zuvor nicht erkennbare Lücke von ca. 150 kb zwischen den YACs 764_a_1 und 954_e_4 [Lioumi *et al.*, 1998]. Da YAC 929_b_2, das ebenfalls im centromeren Bereich des Contigs kartiert, nicht für die Fragmentierung genutzt wurde, könnte eine noch nachzuweisende Überlappung mit YAC 764_a_1 diese Lücke schließen. Im telomeren Bereich des Contigs konnte die Verbindung zwischen *S100A1* und 24f11 nur auf YAC 950_e_2 nachgewiesen werden, das obendrein eine Deletion zwischen diesen Loci aufweist. Unter der Voraussetzung, daß die angrenzenden YACs 100_f_3 (mit *S100A1*) und 713_h_12 (mit 24f11) nicht rearrangiert sind und überlappen, würde das Contig hier dennoch die menschliche DNA kontinuierlich abdecken (Abb. 24B).

Interessanterweise kartiert *TPM3*, das an diversen tumorassoziierten Umlagerungen im Genom des Menschen beteiligt ist [Butti *et al.*, 1995; Lamant *et al.*, 1999], in der Nähe der in den YACs eliminierten DNA-Sequenz zwischen *S100A1* und 24f11. In der Mehrzahl der YACs ist *TPM3* sogar Bestandteil der Deletion, was darauf hinweisen könnte, daß die Instabilität der YACs mit erhöhten Umlagerungsaktivitäten im menschlichen Genom zusammenfällt. YACs, welche die intakte Region um *TPM3* enthalten, könnten somit helfen, die Ursachen dieser Umlagerungen zu erkennen und darüber hinaus zur Aufklärung einer weiteren Krankheit beitragen, die mit dem *TPM3*-Gen assoziiert ist, der autosomal-dominanten Nemanin-Myopathie 1 (NEM-1) [Laing *et al.*, 1995].

Der Vergleich mit der genomischen Restriktionskarte bestätigte die Ausdehnung des Contigs von ca. 6 Mb. Er offenbarte aber auch rearrangierte YACs, erkennbar an Diskrepanzen zwischen den Abständen auf der genomischen und denen auf der YAC-Restriktionskarte. Weitere Umlagerungen sind durch – vom vorherrschenden *SalI*-Fragment – abweichende Fragmentgrößen oder durch einen widersprüchlichen Markergehalt der YACs gekennzeichnet. Auf diese Weise wurden in fast 60% der YACs des Contigs Umlagerungen nachgewiesen. Einschließlich der Umlagerungen, die vom STS-Gehalt der YACs abgeleitet wurden (Tab. 14), sind etwa zwei Drittel der YACs des Contigs rearrangiert. Dieser Wert liegt über den für YAC-Bibliotheken ermittelten bis zu 50% chimären Klonen [Green *et al.*, 1991; Cohen *et al.*, 1993]. Er ist wahrscheinlich auf die ausgeprägte Instabilität dieser Region nicht nur in YACs, sondern auch im menschlichen Genom zurückzuführen.

Die hohe Qualität des YAC-Contigs kam nicht nur den Zielen dieser Arbeit zugute, sondern wurde auch von anderen Arbeitsgruppen bereits ausgiebig genutzt. So waren die YACs, die wenigstens zur Abdeckung des Contigs notwendig sind (*minimal tiling path*), Ausgangspunkt für die Erstellung einer

Karte aus fragmentierten YACs der Region 1q21 [Lioumi *et al.*, 1998]. Ausgewählte YACs des Contigs und DNA-Sonden aus subklonierter YAC-DNA dienten außerdem der Identifizierung von PAC-, BAC- und Cosmid-Klonen zur Erstellung eines weiteren hochauflösenden Contigs des EDC [South *et al.*, 1999]. Mittlerweile werden diese Klone vom *Sanger Center* in Hinxton eingesetzt, um den gesamten EDC zu sequenzieren. Ein weiteres Projekt beschäftigt sich mit der Identifizierung von Amplifikationen im menschlichen Genom, die u.a. für die Entstehung von Tumoren verantwortlich gemacht werden [Forus *et al.*, 1998]. Hierbei werden die YACs des EDC als FISH-Sonden eingesetzt. Anhand der Intensität des Signals auf dem Chromosom kann anschließend festgestellt werden, ob die entsprechenden Sequenzen der chromosomalen Bande 1q21 im Genom des Menschen amplifiziert sind. Über eine cDNA-Selektion unter Verwendung der YACs werden anschließend die vervielfältigten Gene bestimmt.

1.2 Die feingerasterte Keratinozyten-cDNA-Bibliothek

Für die Analyse der Gene, die während der Differenzierung der Epidermis exprimiert werden, wurde am Ressourcenzentrum in Berlin eine feingerasterte cDNA-Bibliothek aus primären Keratinozyten des Menschen hergestellt, die aufgrund unterschiedlicher Kultivierungsbedingungen mehrere Differenzierungsstadien abdeckten [Fischer *et al.*, 1996]. Die hohe Komplexität der cDNA-Bibliothek von ca. 4×10^6 Klonen hat zur Folge, daß bei einem geschätzten mRNA-Gehalt menschlicher Zellen von 300000 Molekülen eine nur einmal vorkommende Sequenz mit einer Wahrscheinlichkeit von nahezu 100% in der Bibliothek vorhanden ist [Clarke & Carbon, 1976]. Von der Gesamtpopulation wurden 184320 einzelne cDNA-Klone in Mikrotiterplatten übertragen und anschließend jeweils doppelt, nach einem bestimmten Muster, in hoher Dichte auf Nylonmembranen angeordnet. Damit liegt die Wahrscheinlichkeit, eine einmal pro Zelle vorkommende mRNA auf den Filtern wiederzufinden, bei fast 50%. Insgesamt wurden zehn identische Filtersets, bestehend aus zehn Filtern mit jeweils 18432 doppelt aufgetragenen cDNA-Klonen, für die Arbeitsgruppen des Biomed-Projekts hergestellt.

Im Gegensatz zu einer herkömmlichen cDNA-Bibliothek, die als komplexe Lösung vorliegt und nach dem Ausbreiten und Aufziehen der Klone auf Agarplatten durch kurzes Auflegen eines Filters ungeordnet auf diesen übertragen wird (*colony lifting*) [Benton & Davis, 1977], ermöglicht die feingerasterte cDNA-Bibliothek, bei der jedem Klon eine feste Koordinate zugeordnet wird, eine wesentlich effizientere Arbeitsweise. Die Aufbewahrung in Mikrotiterplatten gewährleistet einen schnellen und gezielten Zugriff auf die einzelnen Klone. Die Herstellung beliebig vieler, identischer Kopien der cDNA-Bibliothek bzw. der Filter ermöglicht nicht nur die gleichzeitige Nutzung der Bibliothek durch mehrere Arbeitsgruppen, die Ergebnisse verschiedener Laboratorien sind auch direkt miteinander vergleichbar und integrierbar, wodurch aufwendige Mehrfachuntersuchungen entfallen. Die dichte und systematische Anordnung der Klone auf den Filtern erlaubt einen hohen Umsatz bei gleichzeitig erhöhter Sensitivität, weil bei der Autoradiographie aufgrund des definierten Musters der doppelt aufge-

tragenen Klone auch sehr schwache Hybridisierungssignale eindeutig zu erkennen und trotz langer Expositionszeit vom Hintergrund abzugrenzen sind.

Ein wichtiges Qualitätsmerkmal einer feingerasterten DNA-Bibliothek ist ihr Gehalt an kontaminierten Klonen, denn die zahlreichen Herstellungsschritte, die genutzten Arbeitsplätze in diversen Räumen, die verwendeten Geräte und Reagenzien stellen gleichzeitig potentielle Kontaminationsquellen dar. Durch strikte Einhaltung steriler bzw. keimarmer Bedingungen, ständige Sterilitätsprüfungen und Qualitätskontrollen konnte der Anteil der mit Pilzen kontaminierten Klone unter 0,01% gehalten werden (visuelle Kontrolle der Filter auf ungewöhnlich wachsende Kolonien). Alle Auffälligkeiten der Filter wurden notiert und konnten so bei den nachfolgenden Experimenten berücksichtigt werden. Herstellungsbedingt waren ca. 5% der Klone mit einem zweiten kontaminiert. Dieser Wert lag aufgrund der großzügigen Justierung des Laborautomaten, der sich im Probetrieb befand, etwas über dem Durchschnitt.

Für die analysierten Klone der Keratinozyten-cDNA-Bibliothek wurde eine durchschnittliche Insertgröße von ca.1500 bp ermittelt. Dabei ist der überproportionale Anteil kleiner cDNA-Sequenzen (<1000 kb) unter den identifizierten Klonen zu beachten, der auf die im EDC kartierenden *SPRR*- und *S100*-Gene zurückzuführen ist – die durchschnittliche Insertgröße der gesamten Bibliothek dürfte daher über dem obengenannten Wert liegen. Internes Ansetzen des Oligo(dT)-Primers wurde nicht beobachtet. Etwa 5% der Klone waren chimär, ebensoviele waren fehlerhaft transformiert und nicht gegen Ampicillin resistent. Mehr als 60% der analysierten Klone enthielten vollständige cDNA-Sequenzen, die Größen von mehr als 4 kb erreichten.

Mit der bisher einzigen feingerasterten cDNA-Bibliothek aus differenzierenden primären Keratinozyten des Menschen ist eine Analyse nahezu aller in der Epidermis transkribierten Gene möglich. Es existieren zwar noch zwei weitere feingerasterte Keratinozyten-cDNA-Bibliotheken (Nr. 550, KER1 und Nr. 441, HKer1 cDNA am Ressourcenzentrum, Berlin); diese sind mit jeweils 27648 enthaltenen Klonen jedoch äußerst unvollständig und für Expressionsanalysen nur bedingt geeignet. Dagegen konnte die hohe Qualität der eingesetzten cDNA-Bibliothek durch die Ergebnisse der subtraktiven Hybridisierung eindrucksvoll bestätigt werden. Als charakteristische epidermale Differenzierungsmarker wurden cDNA-Klone diverser *SPRRs*, von Involucrin und von Trichohyalin identifiziert, deren Expression überwiegend im *Stratum spinosum* differenzierter Keratinozyten beginnt [Banks-Schlegel & Green, 1981; Kartasova *et al.*, 1988; Hamilton *et al.*, 1991; Hohl *et al.*, 1995]. Von den Transkripten der in die Untersuchungen eingeschlossenen etablierten EDC-Gene, die nachweislich in Keratinozyten exprimiert werden, wurde nur die Profilaggrin-cDNA nicht wiedergefunden. Die Gründe hierfür liegen vermutlich in der sehr späten Expression Profilaggrins, beim Übergang der Keratinozyten in das *Stratum granulosum* [Dale *et al.*, 1985b; McKinley-Grant *et al.*, 1989], sowie in der auffallenden Instabilität der mRNA [McKinley-Grant *et al.*, 1989]. In Zellkultur findet die Expression darüberhinaus nur in konfluenten, gut geschichteten Keratinozyten statt [Fleckmann *et al.*, 1985]. Das gleiche gilt möglicherweise für Repetin, das strukturell mit Filaggrin und Trichohyalin verwandt ist [Krieg *et al.*,

1997] und dessen cDNA in der Bibliothek ebenfalls noch nicht nachgewiesen werden konnte. Auch die limitierte Anzahl analysierter Klone kann dafür verantwortlich sein, daß cDNA-Sequenzen selten vorkommender Transkripte nicht immer erfaßt werden.

Abgesehen von der Verwendung bei der subtraktiven Hybridisierung wurde die feingerasterte Keratinozyten-cDNA-Bibliothek zur Identifizierung weiterer Gene der Region eingesetzt. Durch Hybridisierungen mit PACs aus der Region wurden diverse cDNA-Sequenzen isoliert, die u.a. zur Identifizierung des *ZIRTL*-Gens führten [Lioumi *et al.*, 1999]. Andere cDNA-Klone wurden zur Erhöhung der Markerdichte des PAC/BAC/Cosmid-Contigs des EDC verwendet [South *et al.*, 1999]. Die cDNA-Bibliothek bildete außerdem die Grundlage für die Charakterisierung der *SPRR3*-Transkripte einschließlich der Identifizierung von Transkriptionsfaktoren [Fischer *et al.*, 1999].

2 Die Hybridisierungsmethode

Die in dieser Arbeit erstmals beschriebene subtraktive Hybridisierung einer feingerasterten cDNA-Bibliothek mit YACs stellte sich als äußerst effiziente Methode zur Identifizierung exprimierter Gene einer definierten chromosomalen Region heraus. Dabei hatten frühere Untersuchungen vor allem die Probleme aufgezeigt, die von hochkomplexen DNA-Sonden, wie YACs, bei der Hybridisierung verursacht werden. Aufgrund der unspezifischen Reaktion mit repetitiven oder ribosomalen Sequenzen erreichte der Anteil positiver Klone nicht einmal 15% [Elvin *et al.*, 1990; Geraghty *et al.*, 1993; Boulwood *et al.*, 1997]. Ein weiterer Nachteil war der hohe Anteil chimärer YACs in den verfügbaren YAC-Bibliotheken, der sich auf bis zu 50% beläuft und ebenfalls falsch positive Ergebnisse begünstigt [Green *et al.*, 1991; Cohen *et al.*, 1993]. Schließlich waren die eingesetzten YAC-Sonden auf Molekülgrößen von ca. 500 kb limitiert [Elvin *et al.*, 1990; Geraghty *et al.*, 1993; Boulwood *et al.*, 1997], da die für die DNA-Isolierung verwendeten Elektrophoresemethoden in den darüber liegenden Bereichen keine ausreichende Trennung von den Hefechromosomen mehr zuließen. Trotz dieser Nachteile ergab ein Vergleich der am weitesten verbreiteten Methoden der Genidentifizierung, daß die Verwendung von YACs (zur cDNA-Selektion) immer noch erfolversprechender war als die Identifizierung von Exons durch *exon trapping* oder die computergestützte Vorhersage von Genen aus der genomischen DNA-Sequenz [Hisama *et al.*, 1998].

Für die Identifizierung bisher unbekannter Transkripte des EDC wurde deshalb eine Hybridisierungsmethode etabliert, die den Vorteil der YACs, das Abdecken großer DNA-Bereiche von mehreren Megabasen mit nur wenigen Klonen, nutzte und gleichzeitig die oben beschriebenen, von YACs als DNA-Sonden ausgehenden Probleme fast vollständig beseitigte. Eine feingerasterte Keratinozyten-cDNA-Bibliothek, die einen direkten Vergleich mehrerer Experimente ermöglicht, wurde nacheinander mit mehreren YACs, die nahezu das gesamte Contig des EDC abdeckten, hybridisiert. Durch anschließende subtraktive Auswertung der Hybridisierungsergebnisse jeweils nicht überlappender YACs

– gemeinsam identifizierte cDNA-Klone wurden auf unspezifische Hybridisierung mit in beiden YACs enthaltenen repetitiven Elementen zurückgeführt und nicht berücksichtigt – konnte der Anteil falsch positiver Klone erheblich gesenkt werden.

Neben der feingerasterten cDNA-Bibliothek war das gut charakterisierte YAC-Contig eine wichtige Voraussetzung dieser subtraktiven Vorgehensweise. Es ermöglichte eine zweckmäßige Auswahl der YAC-Sonden hinsichtlich der von ihnen abgedeckten DNA-Region, so daß Überlappungen bei der Auswertung ausgeschlossen waren und chimäre YACs möglichst nicht eingesetzt wurden. Zudem sorgte die auch in hochmolekularen Bereichen hochauflösende ROFE [Ziegler & Volz, 1992] dafür, daß bei der DNA-Trennung eine Verunreinigung der YAC-Sonden mit Hefe-DNA – als Ursache für die Hybridisierung mit nicht in der Region kartierenden, ribosomalen cDNA-Sequenzen – weitgehend vermieden werden konnte. So waren für die Transkriptionsanalyse von ca. 5 Mb der Region 1q21 nur vier bis zu 1440 kb große YACs notwendig.

Insgesamt identifizierte die subtraktive Hybridisierung 40 EDC-spezifische Transkripte aus Keratinozyten, die von 33 verschiedenen Genen der Region 1q21 exprimiert wurden und von denen 21 erstmals dem EDC zugeordnet werden konnten. Die für die YACs 986_e_10 und 950_e_2 ausführlich durchgeführte Auswertung ergab, daß, obwohl nur 50 cDNA-Klone pro YAC analysiert wurden, die dabei isolierten 1q21-spezifischen cDNA-Sequenzen in 85% (YAC 986_e_10) bzw. 92% (YAC 950_e_2) derjenigen Klone vorhanden waren, die ausschließlich von einem der beiden YACs nachgewiesen worden waren – im Vergleich dazu betrug der Anteil positiver Klone bei der nicht subtraktiven Auswertung der Hybridisierung mit YAC 986_e_10 ganze 15%. Von den bekannten EDC-Genen, deren Expression in Keratinozyten nachgewiesen ist, wurden bis auf die von den eingesetzten YACs nicht abgedeckten Gene von *LOR* bis *S100A8* lediglich die cDNA-Sequenzen von *FLG* und *RPTN* noch nicht aufgefunden. Die Fortsetzung der Auswertung der subtraktiv ausgewählten cDNA-Klone könnte somit weitere Transkripte, auch von selten exprimierten Genen des EDC, hervorbringen und die Erfolgsrate noch erhöhen.

Letztendlich bleibt ein grundsätzlicher Konflikt bestehen, der in der Natur der subtraktiven Hybridisierung liegt: Einerseits wird durch die Subtraktion eine hohe Effizienz erreicht, andererseits gehen aber Transkripte verloren, die repetitive Elemente enthalten und gleichzeitig in der zu analysierenden Region kartieren. Geht man davon aus, daß die am häufigsten im menschlichen Genom vorkommenden repetitiven Elemente sowohl in YAC 986_e_10 als auch in YAC 950_e_2 vorhanden sind, so stellen die ca. 10000 Klone, die mit beiden YACs hybridisierten, bereits den Großteil aller in der Keratinozyten-cDNA-Bibliothek enthaltenen Klone mit repetitiven Elementen dar. Dementsprechend würden nur etwas mehr als 5% der 184320 cDNA-Klone bei der subtraktiven Hybridisierung unberücksichtigt bleiben – bei diesem Ergebnis ein vertretbarer Wert.

Da die von weltweiten Projekten zur Sequenzierung des menschlichen Genoms und zur Genkartierung mit Hilfe von ESTs produzierten Datenmengen rapide zunehmen, wurden die Ergebnisse der subtraktiven Hybridisierung mit den in frei zugänglichen Datenbanken verfügbaren Daten verglichen. Als

Grundlage dienten die mit YAC 986_e_10 und YAC 950_e_2 neu identifizierten Gene *NICE1* bis *NICE5*. Kein Datenbankeintrag existierte für die *NICE2*-Sequenz; einzelne Spleißprodukte der Gene *NICE3* und *NICE4* konnten auf diese Weise ebenfalls nicht gefunden werden; und für *NICE1* war nur eine partielle Sequenz vorhanden. Folglich war *NICE2* auch nicht kartiert. EST-Sequenzen, die von den anderen vier *NICE*-Genen stammten, waren zwar auf der RH-Karte des Menschen lokalisiert (GenMap99) – allerdings in einem 13 cM-Intervall, zwischen den Markern D1S514 in der Region 1p13 und D1S2635 in 1q22 (Tab. 19). Eine derart grobe Kartierung ist für die Zuordnung der Gene zum EDC wie auch für die meisten anderen Anwendungen zu ungenau. Um eine überschaubare Anzahl an Kandidatengenen zu erhalten, muß z.B. bei der positionellen Klonierung der zu analysierende Bereich auf wenige CentiMorgan bzw. Megabasen eingegrenzt werden. Die subtraktive Hybridisierung mit YACs ist somit die Methode der Wahl, um die Gene einer Region gezielt und mit vertretbarem Aufwand zu identifizieren.

3 Die integrierte Karte der Region 1q21

Die Vereinigung der genomischen und der YAC-Restriktionskarte des EDC ergab eine hochauflösende physikalische Karte, die sich über ca. 6 Mb oder 2% von Chromosom 1 erstreckt. Über einen Bereich von 4,5 Mb konnten die Fragmente der genomischen Restriktionskarte kontinuierlich nachgewiesen werden. Durch FISH zweier YACs wurde darüberhinaus der Bereich von *ANXA9* bis *ADARI* der Bande 1q21.3 auf der cytogenetischen Karte zugeordnet. Dabei ist jedoch zu beachten, daß die YACs aufgrund der relativ geringen Auflösung der *in situ*-Hybridisierung auch in die benachbarten Regionen 1q21.2 bzw. 1q22 hineinragen könnten.

Von den 21 durch subtraktive Hybridisierung identifizierten und auf den Restriktionsfragmenten lokalisierten neuen EDC-Genen waren 17 Gene zuvor nicht auf diesem Chromosomenabschnitt kartiert oder lediglich als EST-Sequenzen der Region 1p13-1q22 zugeordnet (Tab. 19). Für *ADARI*, *ANXA9*, *PIAS3* und *PSMB4* verfeinerte die Restriktionskartierung die auf cytogenetischer bzw. RH-Kartierung basierende Zuordnung zu den chromosomalen Banden 1q21 (*ANXA9* [Morgan *et*

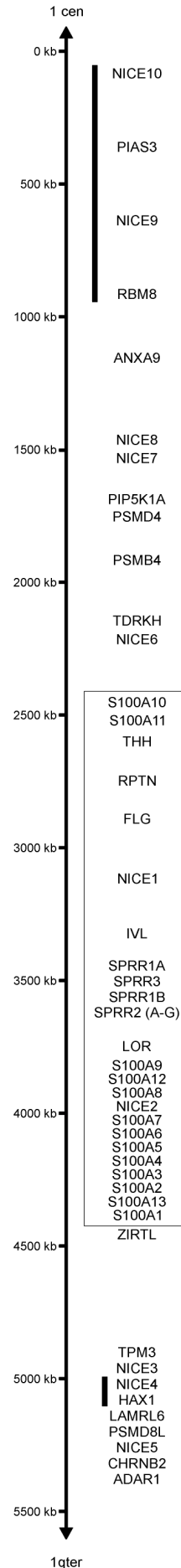


Abbildung 33: Die Gene des erweiterten EDC. Die vereinfacht dargestellte Karte enthält alle Gene der integrierten Karte sowie das *ZIRTL*-Gen in der entsprechenden Reihenfolge. Ein vertikaler Balken zeigt an, daß die Reihenfolge der rechts stehenden Gene nicht aufgelöst werden konnte. Die Region des ursprünglichen EDC ist eingrahmt.

al., 1999], *PIAS3* [Ueki *et al.*, 1999] und *PSMB4* [McCusker *et al.*, 1997]) bzw. 1q21.1-q21.2 (*ADARI* [Weier *et al.*, 1995]). *TPM3* und *PIP5K1A* wurden durch FISH in den Regionen 1q22-q23 (*TPM3* [Wilton *et al.*, 1995]) bzw. 1q22-q24 [Xie *et al.*, 2000] nachgewiesen; ihre Positionen konnten durch die Restriktionskartierung weitaus genauer bestimmt werden. *RBM8* wurde sowohl auf Chromosom 14 [Conklin *et al.*, 2000] als auch durch RH-Analyse in der Region 1q12 [Zhao *et al.*, 2000] positioniert; hierbei könnte es sich um mehrere Mitglieder einer Genfamilie oder um ein Pseudogen handeln. Das neue Gen für einen Lamininrezeptor (*LAMRL6*) in der Region 1q21 muß noch auf seine transkriptionelle Aktivität hin untersucht werden, da neben dem Lamininrezeptor-Gen (*LAMRI*) auf Chromosome 3 [Jackers *et al.*, 1996a] zahlreiche Pseudogene im menschlichen Genom vorhanden sind, die auf Retropositionen zurückgeführt werden [Bignon *et al.*, 1991; Jackers *et al.*, 1996b]. Auch die funktionelle Bedeutung des Locus für die Proteasom-Untereinheit nicht-ATPase 8 (*PSMD8L*) ist noch zu klären; in diesem Fall ist ein komplettes *PSMD8*-Gen (GenBank Acc AC005789) in einem Cosmid enthalten, das auf Chromosom 19 kartiert. Auf Chromosom 1 könnte es sich um ein noch nicht beschriebenes verwandtes Gen handeln, aber auch das Vorhandensein eines Pseudogens muß in Betracht gezogen werden. Ähnliches gilt für das *PSMD4*-Gen auf Chromosom 1, wobei in diesem Fall ein Pseudogen (*PSMD4P*) auf Chromosom 21 liegt [The chromosome 21 mapping and sequencing consortium, 2000].

Die Positionierung sieben polymorpher STS-Marker innerhalb des Contigs verknüpfte schließlich die genetischen und physikalischen Daten in der integrierten Karte der Region 1q21. Dabei stimmt die Reihenfolge 1cen – D1S442 – D1S498/D1S2347 – D1S2345 – D1S1664 – D1S2346 – D1S305 – 1qter sowohl mit den genetischen Karten [Dib *et al.*, 1996; Broman *et al.*, 1998] als auch mit der auf YACs basierenden physikalischen Karte des menschlichen Genoms [Hudson *et al.*, 1995] überein. Der durchschnittliche Abstand der STS-Marker beträgt ca. 1 Mb – aufgrund der relativ gleichmäßigen Verteilung ist auch der maximale Abstand zweier Marker nur unwesentlich größer. Mit den 53 umliegenden Genen, deren Entfernungen zu den STS-Markern bestimmt wurden, ist die hier veröffentlichte integrierte Karte ein wertvolles Hilfsmittel für genetische Kopplungsanalysen zur Identifizierung von Krankheitsgenen innerhalb des EDC. Mit ihrer Hilfe können die Informationen der vier wichtigsten Kartierungsarten – der cytogenetischen und der genetischen Kartierung, der Restriktionskartierung sowie der Kartierung auf DNA-Klonen – kombiniert ausgewertet werden.

4 Die neuen EDC-Gene

Die zu Beginn des Projekts bekannten 24 Gene des EDC zeichneten sich durch evolutionär bedingte Gemeinsamkeiten sowohl in struktureller als auch in funktioneller Hinsicht aus [Mischke *et al.*, 1996]. Einerseits gab es die für Strukturproteine charakteristischen Sequenzwiederholungen der zentralen Gene, zu denen *LOR*, zehn *SPRR*-Gene und *IVL* zählten, andererseits die Ca²⁺-bindenden Motive vom

EF-Hand-Typ der außen liegenden *S100*-Gene *S100A1* bis *S100A10*, denen eine Vielzahl regulatorischer Funktionen in der Zelle zugeschrieben wird. *THH* und *FLG* wiederum vereinten als sogenannte fusionierte Gene beide dieser Strukturmerkmale. Von den neu kartierten EDC-Genen reißen sich das Repetin-Gen (*RPTN*) und *S100A11* in die bekannten Genfamilien ein, wobei *S100A11* im Gegensatz zu *RPTN* auch in der feingerasterten Keratinozyten-cDNA-Bibliothek nachgewiesen werden konnte. Außerdem wurden *NICE1* und *NICE2* innerhalb des ursprünglichen EDC lokalisiert (Abb. 33). Alle anderen Gene liegen proximal oder distal von dieser Kernregion und zeigen, abgesehen von ihrer Expression in Keratinozyten, weder Gemeinsamkeiten in der Sequenz ihrer Transkripte noch in den bisher bekannten Funktionen der entsprechenden Proteine. Von den Genen des erweiterten EDC wurden *CHRN2* und *TDRKH* ebenfalls nicht durch subtraktive Hybridisierung identifiziert.

S100A11 (Calgizzarin oder S100C) gehört zur Familie der S100-Proteine mit zwei Ca^{2+} -bindenden EF-Hand-Motiven, die in der Regel Ca^{2+} -abhängig ihre Konformation ändern und mit spezifischen Zielproteinen interagieren. Mit Annexin I bildet es das Heterotetramer Annexin I₂-S100A11₂ [Seemann *et al.*, 1996], analog der für S100A10 beschriebenen Komplexbildung mit Annexin II zum Calpactin I [Gerke & Weber, 1985]. Beide Komplexe sind offenbar an Membranfusionen beteiligt [Nakata *et al.*, 1990; Seemann *et al.*, 1996]; außerdem wurden sowohl S100A11 als auch S100A10 als Bestandteile des CE identifiziert [Robinson *et al.*, 1998]. Möglicherweise besteht die Aufgabe der Annexine darin, durch Bindung an Membranphospholipide S100A11 und S100A10 zur Zellmembran zu geleiten, an deren Innenseite die Proteine dann mit dem CE vernetzt werden. Es ist in diesem Zusammenhang interessant, daß auch ein Annexin-Gen, *ANXA9*, als Bestandteil des erweiterten EDC identifiziert wurde (s.u.).

Eine mRNA, die in murinen Hauttumoren entdeckt wurde, war der Ausgangspunkt für die Isolierung des Repetin-Gens (*Rptn*) der Maus [Krieg *et al.*, 1997]. Es besitzt die gleiche Exon/Intron-Verteilung wie die Gene der mit den IF assoziierten Proteine Filaggrin und Trichohyalin, und auch die abgeleitete Proteinsequenz weist auf eine Verwandtschaft mit diesen Proteinen hin: Am N-Terminus ist eine starke Homologie zu zwei Ca^{2+} -bindenden EF-Hand-Motiven zu erkennen und im zentralen Bereich ein sich 49-fach wiederholender Sequenzabschnitt von zwölf Aminosäuren. Ein hoher Gehalt an Glutamin (mehr als 20%) macht Repetin zu einem möglichen Substrat für TGasen [Krieg *et al.*, 1997], die auch für die Vernetzung des CE verantwortlich sind. Erste Analysen des menschlichen Repetin-Gens (*RPTN*) führten zu vergleichbaren Ergebnissen, so daß *RPTN* aller Voraussicht nach zur Familie der fusionierten Gene *FLG* und *THH*, welche Keratinmatrixproteine codieren, gezählt werden kann.

Der *SPRR*-Familie konnte zwar kein neues Gen zugeordnet werden, aber ein zweites, zuvor unentdecktes Transkript, das in der cDNA-Bibliothek identifiziert wurde. Entgegen der bekannten Sequenz, die im zentralen Bereich 14 sich wiederholende Sequenzabschnitte aus 24 Nucleotiden aufweist und ein Protein mit 169 Aminosäuren codiert [Gibbs *et al.*, 1993], enthalten mehrere cDNA-Klone nur 13 Sequenzwiederholungen, an zehn weiteren Positionen sind einzelne Nucleotide ausgetauscht. Das hieraus abgeleitete Protein besteht aus 161 Aminosäuren, von denen eine zusätzlich substituiert ist.

Offenbar handelt es sich hierbei um einen Polymorphismus des *SPRR3*-Gens, vergleichbar mit der variablen Anzahl Sequenzwiederholungen, die für das Involucrin-Gen des Menschen beschrieben wurde [Simon *et al.*, 1991]. Gemäß der angenommenen Evolution der *SPRR*-Gene, nach der die zentral gelegenen Sequenzwiederholungen entwicklungsgeschichtlich später entstanden sind [Gibbs *et al.*, 1993], weist das Allel mit 13 Wiederholungen auf einen früheren Ursprung hin.

Von den Genen, die durch subtraktive Hybridisierung erstmals dem EDC zugeordnet werden konnten, kartieren nur zwei, *NICE1* und *NICE2*, innerhalb des ursprünglichen Genkomplexes. Obwohl es dabei keinen Hinweis auf eine Verwandtschaft mit den bekannten EDC-Genen gibt, konnten für *NICE1* einige interessante Details ermittelt werden, die für eine nicht unbedeutende Rolle bei der Keratinozytendifferenzierung sprechen. Abgesehen von der Lokalisierung inmitten der etablierten Strukturprotein-Gene, zwischen *FLG* und *IVL*, gilt dies vor allem für die stark erhöhte Expression in differenzierten Keratinozyten, vergleichbar der des *SPRR2*-Gens. Außerdem sprechen mehrere Glutamin- und Lysin-Reste in der vorhergesagten Proteinsequenz für ein potientielles Transglutaminasesubstrat, als Baustein des CE. Darüberhinaus ist *NICE1* zu einem anderen vorhergesagten Protein homolog, dessen cDNA kürzlich aus menschlicher Haut synthetisiert wurde und das von dem Gen *xp5* abstammt [Zhao & Elder, 1997]. *Xp5* wird in normaler und in psoriatischer Haut exprimiert, aber nicht in undifferenzierten Keratinozyten. Es wurde ebenfalls nahe *IVL* und den *SPRR*-Genen lokalisiert [Zhao & Elder, 1997]. Möglicherweise handelt es sich bei *NICE1* und *xp5* um eine neue Genfamilie des EDC.

TPM3, eines von vier Tropomyosin-Genen des Menschen, exprimiert zwei alternativ gespleißte Isoformen, das Muskel-spezifische α -Tropomyosin 3 und die *non muscle*-Isoform TM30nm [Clayton *et al.*, 1988]. In Muskelzellen bilden die an Aktin gebundenen Tropomyosine zusammen mit dem Troponin-Komplex die dünnen Filamente und sind für die Ca^{2+} -abhängige Kontraktion verantwortlich [Übersicht in Gordon *et al.*, 2000]. Das differenzierungs- und gewebespezifische Vorkommen bestimmter Tropomyosin-Isoformen in Nichtmuskelzellen spricht für die Ausübung spezialisierter Funktionen in den einzelnen Zellpopulationen [Übersicht in Lin *et al.*, 1997]. Über die Stabilisierung der Aktin-Mikrofilamente werden offenbar Beweglichkeit und Form der Zellen beeinflusst. Die Isoform TM30nm wurde bereits in menschlichen Keratinozyten nachgewiesen [Olsen *et al.*, 1995]. Für das Maus homolog konnte gezeigt werden, daß eine verminderte Proteinsynthese die Motilität und somit die Metastasierungsfähigkeit von Melanomzellen reduziert [Miyado *et al.*, 1996]. Translokationen unter Beteiligung des *TPM3*-Gens sind an der Entstehung zweier Onkogene beteiligt [Butti *et al.*, 1995; Lamant *et al.*, 1999], eine Mutation wird mit der Nemalin-Myopathie 1 (NEM-1) in Verbindung gebracht [Laing *et al.*, 1995].

Als Interaktionspartner von **Hax-1** (HS1-assoziiertes oder HS1-bindendes Protein X-1) wurden die mit dem Aktin-Zytoskelett assoziierten Proteine HS1 (*hematopoietic lineage cell-specific protein 1*) [Suzuki *et al.*, 1997] und Cortactin [Gallagher *et al.*, 2000] identifiziert. Während Cortactin in unterschiedlichen Geweben ein potentieller Signalüberträger zwischen Zellmatrix und Zytoskelett ist [Schuuring *et al.*, 1998], kommt HS1 ausschließlich in hämatopoietischen Zellen vor [Kitamura *et al.*,

1989], wo es an einer Antigenrezeptor-vermittelten Signalkaskade beteiligt ist, die Proliferation und Differenzierung bestimmter B-Lymphozytenklone bewirkt bzw. deren Wachstum hemmt [Yamanashi *et al.*, 1993].

Die dem **LAMRL6**-Gen entsprechende cDNA-Sequenz ist identisch mit dem Transkript des Lamininrezeptor-Gens **LAMR1** [Jackers *et al.*, 1996a]. Sie codiert ein 37 kD-Vorläuferprotein (37LRP, 37 kD laminin receptor precursor), aus dem durch Acylierung der 67 kD-Lamininrezeptor (67LR) entsteht [Übersicht in Menard *et al.*, 1997; Buto *et al.*, 1998]. 67LR zählt zu den membranständigen Zelladhäsionsrezeptoren der Integrin-Familie [Gehlsen *et al.*, 1988], die mit Proteinen der extrazellulären Matrix interagieren und so Signale in die Zelle übertragen. Als Ligand fungiert das u.a. in der Basalmembran der Haut vorkommende Laminin, welches die Zelladhäsion, -proliferation und -motilität beeinflusst. In Tumorzellen konnte gezeigt werden, daß eine erhöhte Zahl an 67 kD-Lamininrezeptoren ein invasives und metastasierendes Wachstum der Zellen fördert [Übersicht in Castronovo, 1993].

Die nicotineren Acetylcholinrezeptoren (nAChRs) zählen zur Superfamilie der Liganden-abhängigen Ionenkanäle, die eine schnelle Signalübertragung an Synapsen vermitteln. Sie setzen sich aus mehreren Untereinheiten zusammen, die sowohl Hetero- als auch Homooligomere bilden können [Übersicht in Utkin *et al.*, 2000]. In Keratinozyten wurde das $\beta 2$ -Polypeptid des nicotineren Acetylcholinrezeptors **CHRN2** (*cholinergic receptor, neuronal nicotinic, β -polypeptide*) zusammen mit den Untereinheiten $\alpha 3$, $\alpha 5$, $\alpha 7$ und $\beta 4$ gefunden, was auf die Existenz sowohl heteromerer Rezeptoren, die aus den Untereinheiten $\alpha 3$, $\alpha 5$, $\beta 2$ und $\beta 4$ bestehen, als auch homomerer $\alpha 7$ -Rezeptoren schließen läßt [Grando *et al.*, 1996]. Durch Ligandenbindung an die Nicotinrezeptoren der Epidermis, deren Zahl sich in den suprabasalen Schichten mit dem Differenzierungsgrad der Keratinozyten erhöht, werden Zelldifferenzierung, -migration und -adhäsion beeinflusst [Grando *et al.*, 1995; Grando *et al.*, 1996]. Da Nicotinrezeptoragonisten, wie Acetylcholin, den Ca^{2+} -Transport in die Zelle auslösen und die intrazelluläre Ca^{2+} -Konzentration Einfluß auf diverse Differenzierungsvorgänge in Keratinozyten hat, dienen Calciumionen möglicherweise als sekundäre Botenstoffe (*second messenger*) dieser Signalkaskade [Grando *et al.*, 1996].

Die für doppelsträngige RNA spezifische Adenosindesaminase **ADAR1** (DRADA oder DSRAD) ist ein RNA modifizierendes Enzym, das in prä-mRNA Adenosin hydrolytisch zu Inosin desaminiert [Kim *et al.*, 1994], welches bei der anschließenden Translation dem Guanosin gleichgestellt ist [Basilio *et al.*, 1962]. Bei den bisher identifizierten ADAR1-Substraten handelt es sich um die prä-mRNA neuraler Glutamatrezeptoren [Wang *et al.*, 1995] sowie des Serotonin 2C Rezeptors im Gehirn [Burns *et al.*, 1997], deren Aktivität auf diese Weise modifiziert wird.

Das **TDRKH**-Gen wurde nach einer in der vorhergesagten Aminosäuresequenz enthaltenen Tudor- und zwei K-Homologie (KH)-Domänen benannt [Lamb *et al.*, 2000]. Beide Sequenzmotive ermöglichen eine Proteinbindung an RNA bzw. einzelsträngige DNA und werden mit RNA-Metabolismus und -Transport in Verbindung gebracht [Dejgaard & Leffers, 1996; Übersicht in Ponting, 1997].

TDRKH wird in menschlichen Geweben ubiquitär exprimiert; mehrere alternativ gespleißte bzw. polyadenylierte Genprodukte konnten nachgewiesen werden [Lamb *et al.*, 2000].

In eukaryontischen Zellen ist das Ubiquitin- und ATP-abhängige 26S-Proteasom für den Abbau von 80-90% aller Proteine verantwortlich. Sein proteolytischer Kern, das 20S-Proteasom, besteht aus einem aus zweimal 14 Untereinheiten aufgebauten Dimer. Es ist von dem regulatorischen 19S-Komplex umgeben, dessen 17 bekannte Untereinheiten für die Erkennung der ubiquitinylierten Substratproteine, deren Desubiquitinylierung, Entfaltung und Transport in das Innere des Komplexes zuständig sind [Übersicht in Voges *et al.*, 1999]. In Keratinozyten konnte nachgewiesen werden, daß z.B. die Zahl der nucleären Retinoid- und Vitamin D₃-Rezeptoren über den Ubiquitin/Proteasom-vermittelten Abbau gesteuert wird [Li *et al.*, 1999; Boudjelal *et al.*, 2000]. Mit *PSMB4* (Proteasom-Untereinheit β 4, HSBpros26 oder HsN3), *PSMD4* (Proteasom-Untereinheit nicht-ATPase 4, antisekretorischer Faktor 1 oder S5a) sowie dem zu *PSMD8* (Proteasom-Untereinheit nicht-ATPase 8, p31 oder S14) homologen *PSMD8L* konnten die Loci gleich dreier Untereinheiten des 26S-Proteasoms innerhalb des EDC kartiert werden. Die Untereinheit β 4 ist Bestandteil des 20S-Proteasoms [Northwang *et al.*, 1994], für das Hefehomolog wurde eine Beteiligung an proteolytischen Prozessen nachgewiesen [Gerards *et al.*, 1994]. Als einzige Ubiquitin-bindende Komponente des 19S-Komplexes wurde die Untereinheit S5a identifiziert [Deveraux *et al.*, 1994]; sie wird mit der Selektion und dem Transport der Substrate in Verbindung gebracht, aber auch ein davon unabhängiger Einfluß auf den Proteinabbau wurde nachgewiesen [Fu *et al.*, 1998]. Im Gegensatz zu den anderen Untereinheiten des 19S-Komplexes kommt S5a auch in freier Form vor, wobei in Pflanzenzellen hohe Konzentrationen den Abbau ubiquitinylierter Proteine durch das 26S-Proteasom inhibieren können [Deveraux *et al.*, 1995]. S14 ist ebenfalls Bestandteil des 19S-Komplexes, in verschiedenen Hefestämmen hat eine Mutation des zu S14 homologen Proteins eine Störung des Zellzyklus und ein vermehrtes Auftreten von Ubiquitin-Konjugaten zur Folge [Kominami *et al.*, 1995].

Das *PIPK1A*-Gen codiert eine von drei bekannten Typ I Phosphatidylinosit-4-phosphat-5-kinasen. PIP5KI α , PIP5KI β [Ishihara *et al.*, 1996; Loijens & Anderson, 1996] und PIP5KI γ [Ishihara *et al.*, 1998] nehmen eine Schlüsselposition bei der Synthese diverser Phosphatidylinositphosphate ein [Zhang *et al.*, 1997; Toliás *et al.*, 1998], die als wichtige *second messenger* an vielfältigen zellulären Prozessen beteiligt sind, wie der Regulierung der intrazellulären Ca²⁺-Konzentration und der Aktivierung der Proteinkinase C [Übersicht in Rana & Hokin, 1990], der Organisation des Aktin-Zytoskeletts [Übersicht in Janmey, 1994] und der Exozytose von Neurotransmitter-Vesikeln [Hay *et al.*, 1995]. PIP5KI α wird u.a. in einer vom epidermalen Wachstumsfaktor (EGF, *epidermal growth factor*)-Rezeptor ausgelösten Signalkaskade aktiviert, die für die Polymerisation des Aktin-Zytoskeletts und die Kräuselung der Membran verantwortlich ist [Honda *et al.*, 1999].

Das vom *ANXA9*-Gen abgeleitete Protein Annexin A9 (Annexin 31) [Morgan & Fernandez, 1998] gehört zur wachsenden Familie der Annexine (Lipocortine), deren Aminosäuresequenzen im zentralen

Bereich vier charakteristische Wiederholungen aufweisen. Jede dieser konservierten Proteindomänen enthält eine Typ II Ca^{2+} -Bindungsstelle, die für die Aggregatbildung der Annexine sowie die kooperative Bindung von Phospholipiden verantwortlich sind [Übersicht in Kourie & Wood, 2000]. Außerdem wird der Kernregion die Rolle eines Calciumionenkanals zugeschrieben [Huber *et al.*, 1992]. Weitere mögliche Funktionen der Annexine betreffen Membranfusionen und Vesikeltransport sowie Zellwachstum und -differenzierung [Übersicht in Raynal & Pollard, 1994]. *ANXA9* wird vor allem in differenzierenden Geweben exprimiert [Morgan & Fernandez, 1998].

Nur wenig ist über das kürzlich entdeckte RNA-bindende Protein **RBM8** (*RNA-binding motif protein 8*) und das dazugehörige Gen bekannt [Zhao *et al.*, 2000]. RBM8, das Sequenzhomologien zu mehreren anderen RNA-bindenden Proteinen aufweist, wurde als Bindungspartner des Proteins MAGOH (*mago nashi homologue*) identifiziert, das mit der Zellproliferation in Verbindung gebracht wird [Zhao *et al.*, 1998]. Da die entsprechenden Gene in menschlichen Geweben gemeinsam exprimiert werden [Zhao *et al.*, 2000], könnte RBM8 eine funktionelle Verbindung zwischen MAGOH und der RNA herstellen.

Pias (Proteininhibitor des aktivierten Stat)-Proteine wurden als Inhibitoren der Stat (Signaltransduktoren und Aktivatoren der Transkription)-Signalkaskaden identifiziert [Übersicht in Shuai, 2000]. Die Proteine der Stat-Familie sind Transkriptionsfaktoren, die von membranständigen Zytokin- und Wachstumsfaktor-Rezeptoren aktiviert werden und nach Translokation in den Zellkern die Genexpression auslösen [Übersicht in Ihle, 1996]. Das von *PIAS3*-Gen codierte Protein interagiert spezifisch mit dem phosphorylierten Stat3 [Chung *et al.*, 1997], welches Bestandteil einer Signalkaskade ist, die u.a. von Interleukin-6 und vom EGF aktiviert wird und die Auswirkungen auf Zellwachstum, -beweglichkeit und Apoptose hat [Übersicht in Akira, 2000]. In der Haut einer *Knock-out*-Maus beeinträchtigte das gezielte Ausschalten des *STAT3*-Gens die Wundheilung und den Haarzyklus, was offenbar auf eine gestörte Keratinozytenmigration bei normaler Proliferationsrate zurückzuführen ist [Sano *et al.*, 1999].

Die *NICE9*-Sequenz wies die größte Übereinstimmung mit dem *NOTCH2*-Transkript auf. Sie stammt offenbar von einem bisher unentdeckten *NOTCH*-Gen des Menschen, da die bekannten Mitglieder dieser Genfamilie, *NOTCH1*, 2, 3 und 4, außerhalb der chromosomalen Region 1q21 kartieren [Sugaya *et al.*, 1994; Larsson *et al.*, 1994]. Notch-Proteine sind Transmembranproteine, die als Rezeptoren für Zelloberflächenproteine benachbarter Zellen vor allem während der Embryogenese die Zellentwicklung steuern [Übersicht in Artavanis-Tsakonas *et al.*, 1995]. Dem aktivierten Notch kommt dabei hauptsächlich die Rolle eines Inhibitors der Zelldifferenzierung zu, es unterdrückt z.B. die Entwicklung sensorischer Borsten aus Vorläuferzellen der Epidermis von *Drosophila* [Heitzler & Simpson, 1991]. In epidermalen Stammzellen der Maus scheint Notch1 dagegen die Differenzierung zu stimulieren [Lowell *et al.*, 2000]. Eine auffallende Konservierung der Notch-Signalübertragung bei *Drosophila* und Mensch [Übersicht in Artavanis-Tsakonas *et al.*, 1995] sowie die Verbreitung der Notch-

Proteine in den unterschiedlichsten Zellarten [Williams *et al.*, 1995; Lindsell *et al.*, 1996] weisen auf die essentielle Bedeutung dieses Signalübertragungsweges hin.

Diese kurze Beschreibung der neu identifizierten Gene des erweiterten EDC zeigt, daß die strukturelle und funktionelle Vielfalt der in Keratinozyten exprimierten Gene dieser Region weitaus größer ist als bisher angenommen. Die codierten Proteine werden mit Signalkaskaden in Verbindung gebracht, die Zellproliferation und -differenzierung steuern oder sie könnten als RNA- bzw. DNA-bindende Proteine Einfluß auf die Proteinsynthese ausüben (Tab. 20). Darüberhinaus weisen häufige Interaktionen mit Proteinen des Zytoskeletts auf eine mögliche Beteiligung an der Ausbildung zellulärer Strukturen hin. Die ubiquitäre Expression der außerhalb der Kernregion des EDC liegenden Gene macht deutlich, daß die codierten Proteine, ähnlich den S100-Proteinen, ihre Funktionen auch in anderen Geweben ausüben. Obwohl bisher nur wenige Untersuchungen zu den neuen Genen in epidermalen Zellen durchgeführt wurden, konnte bei einigen von ihnen, wie bei *CHRNA2* und *PIAS3*, bereits ein Bezug zur Keratinozytenreifung hergestellt werden, so daß weitergehende Analysen noch interessante Einblicke in epidermale Differenzierungsvorgänge erwarten lassen. Dabei ist aber auch, wie bereits bei der Genkartierung bemerkt wurde, nicht auszuschließen, daß sich einige wenige der neuen EDC-Loci als funktionslose Pseudogene erweisen werden. Insgesamt stehen die Funktionsanalysen der meisten Genprodukte noch am Anfang; sie werden nach der Sequenzierung des Genoms die Hauptaufgabe der nächsten Jahre sein.

Tabelle 20: Mögliche Funktionen und Interaktionen der von den neuen EDC-Genen codierten Proteine

Gen	Interaktionen	Funktionen
<i>ADAR1</i>	RNA von Glutamat-, Serotoninrezeptor	RNA-Modifikationen
<i>ANXA9</i>		Zelldifferenzierung, -wachstum
<i>CHRNA2</i>		Zelldifferenzierung, -beweglichkeit
<i>HAX1</i>	HS1, Cortactin (Zytoskelett-assoziierte Proteine)	
<i>LAMRL6</i>	Laminin	Zelladhäsion, -proliferation
<i>NICE1</i>		Bestandteil des CE
<i>NICE9</i>		Zelldifferenzierung (Notch-Signalübertragung)
<i>PIAS3</i>	Stat3 (Transkriptionsfaktor)	Zellwachstum, -beweglichkeit
<i>PIP5K1A</i>		Zytoskelettaufbau
<i>PSMB4</i>		Proteinabbau (Proteasom-Untereinheit)
<i>PSMD4</i>	Ubiquitin	Proteintransport (Proteasom-Untereinheit)
<i>PSMD8L</i>		Proteasom-Untereinheit
<i>RBM8</i>	RNA-Bindung	
<i>RPTN</i>		IF-assoziiertes Protein
<i>S100A11</i>	Annexin I	Membranfusionen, Bestandteil des CE
<i>TDRKH</i>	RNA-Bindung	
<i>TPM3</i>	Filamente des Zytoskeletts	Tumorentstehung

5 Der Genkomplex

Wie bereits in der Einleitung erwähnt wurde, ist die Definition des Epidermalen Differenzierungskomplexes in der Region 1q21 des menschlichen Genoms ein unmittelbares Ergebnis der vorliegenden Arbeit. Inzwischen ist der Begriff von vielen Autoren wissenschaftlicher Artikel übernommen worden. Dabei zeugt die schnelle Aufnahme des „EDC“ in molekularbiologische [Herrmann & Harris, 1998] und dermatologische Fachbücher [Fritsch, 1998] sowohl von der Bedeutung, die dieser Region beigemessen wird, als auch von der Akzeptanz des Begriffes seit seiner Einführung [Mischke *et al.*, 1996].

5.1 EDC oder „CDC“?

Beschränkte sich der EDC anfangs lediglich auf drei, innerhalb von 2 Mb der Region 1q21 kartierende Genfamilien, die evolutionär bedingt gemeinsame Strukturmerkmale aufweisen und während der Keratinozytendifferenzierung zusammenwirken [Mischke *et al.*, 1996], so kamen im Laufe dieser Arbeit eine Reihe von Genen hinzu, die in menschlichen Geweben ubiquitär exprimiert werden und die keine Gemeinsamkeiten erkennen lassen bis auf die Tatsache, daß ihre Transkripte in Keratinozyten vorkommen. Da diese Gene auf beiden Seiten der ursprünglichen Kernregion liegen, ist mit der funktionellen Erweiterung des Genkomplexes auch eine räumliche Ausdehnung auf ca. 6 Mb verbunden.

Bei der Betrachtung des gesamten Genkomplexes fällt auf, daß nur in der zentralen Region von *THH* bis *LOR* eine spezifische Genexpression hauptsächlich in der Epidermis bzw. in anderen epithelialen Geweben stattfindet [Zhao & Elder, 1997]. Schon für die meisten *S100*-Gene trifft das nicht mehr zu. Zwar ist ihre Expression in der Regel auf wenige Gewebe beschränkt, aber nicht immer zählt die Epidermis dazu [Übersicht in Donato, 1999]. Dennoch wurde in letzter Zeit immer deutlicher, daß *S100*-Proteine auch in der Haut eine wichtige Rolle spielen. Zu den bereits in der Einleitung erwähnten Funktionen diverser *S100*-Proteine in der Epidermis wurde *S100A7* als Bindungspartner des epidermalen Fettsäure-bindenden Proteins (E-FABP, *epidermal-type fatty acid-binding protein*) in Keratinozyten identifiziert [Hagens *et al.*, 1999]; *S100A8* und *S100A9* binden in differenzierenden Keratinozyten als Heterodimer Ca^{2+} -abhängig ungesättigte Fettsäuren [Siegenthaler *et al.*, 1997]; *S100A2* und *S100A10* kommen vermehrt in durch Wachstumsfaktoren stimulierten, proliferierenden Keratinozyten vor [Munz *et al.*, 1997; Stoll *et al.*, 1998]; für *S100A10* und *S100A11* wurde eine direkte Beteiligung am CE nachgewiesen [Robinson *et al.*, 1997]; und *S100A12* wurde aus fötalen epidermalen Keratinozyten isoliert und bisher ausschließlich in der Epidermis nachgewiesen [Hitomi *et al.*, 1996]. Darüberhinaus sind die ungewöhnlichen Konzentrationen bestimmter *S100*-Proteine in fehlerhaft differenzierten Epithelzellen (s. Abschnitt I 6) ein Hinweis auf potentiell differenzierungsspezifische Funktionen auch in der Epidermis. Die weiter distal bzw. proximal kartierenden, neu identifizierten EDC-Gene sind schließlich durch eine weitgehend fehlende Spezifität der Genexpression gekennzeichnet. Ihre Translationsprodukte haben offenbar Funktionen bei der Steuerung des Zellzyklus sowie der Zellproliferation und -differenzierung im allgemeinen inne (s. Abschnitt IV 4). Da sie ihre Aufgaben

auch in Keratinozyten erfüllen, können diese Gene zweifelsfrei zum EDC gerechnet werden. Bei strenger Auslegung dieser Bezeichnung wäre aber eine Beschränkung auf die Strukturprotein-Gene und die diese einrahmenden *S100*-Gene denkbar, der EDC wäre dann möglicherweise Bestandteil eines größeren „Zelldifferenzierungskomplexes“ („CDC“, „*cell differentiation complex*“).

Insgesamt konnte durch die Identifizierung neuer Gene der Region 1q21, die in Keratinozyten exprimiert werden, die Rolle des EDC als Genkomplex gestärkt werden. Mit 52 Genen innerhalb von 6 Mb liegt die Gendichte zwar noch unter der des menschlichen MHC auf Chromosom 6, der mit mehr als 60 Genen einschließlich Pseudogenen pro Megabase bei einer Ausdehnung von 3,6 Mb eine der reichsten Regionen des menschlichen Genoms ist [The MHC sequencing consortium, 1999] – allerdings konnte bisher nur für ca. 40% der exprimierten Gene des MHC eine Beteiligung am Immunsystem nachgewiesen werden. Im Gegensatz dazu werden die meisten Gene, auch des erweiterten EDC, in Keratinozyten exprimiert und haben offenbar eine Funktion in der Epidermis – auch wenn nicht vergessen werden darf, daß das Expressionsmuster in kultivierten Keratinozyten mit dem in der menschlichen Haut nicht exakt übereinstimmt. Es ist zu erwarten, daß die Zahl der Gene innerhalb des EDC noch zunehmen wird, weil die Region zwischen *SPRR2A* und *S100A7* einschließlich *LOR*, *S100A9*, *S100A12* und *S100A8*, die Deletion zwischen *S100A1* und 24f11 in YAC 950_e_2#9 sowie eine mögliche Lücke im centromeren Bereich des Contigs, proximal von *ANXA9*, von den Untersuchungen ausgeschlossen waren. Darüberhinaus sollten durch eine vollständige Analyse der subtraktiv ausgewählten cDNA-Klone weitere, seltener exprimierte Gene zu identifizieren sein. Obwohl die eingesetzte cDNA-Bibliothek verschiedene Differenzierungsstadien der Keratinozyten einschließt, könnte auch eine Änderung der Kultivierungsbedingungen zur Expression weiterer EDC-Gene führen. Andere Gene der Region werden möglicherweise nicht in Keratinozyten exprimiert und sind somit nur unter der Verwendung anderer cDNA-Bibliotheken oder anderer Methoden der Genidentifizierung aufzufinden. Und letztendlich haben die Ergebnisse dieser Arbeit gezeigt, daß die Grenzen des EDC noch nicht abzusehen sind, eine Ausdehnung in benachbarte Regionen nicht unwahrscheinlich ist.

5.2 Kontrolle der Genexpression

Strukturelle Gemeinsamkeiten und eine streng koordinierte Expression der einem Cluster oder Komplex zugehörigen Gene werfen interessante Fragen bezüglich der Evolution und der transkriptionellen Regulation eines solchen Zusammenschlusses auf. In der Regel sind die Mitglieder eines Genclusters durch Genduplikationen und Umlagerungen aus wenigen Vorläufergenen entstanden. Ihr Zusammenhalt könnte dabei auf transkriptionelle Zwänge zurückzuführen sein, die der Expression eines abgespaltenen Gens entgegenstehen und so eine Trennung in der Genomentwicklung verhindert haben. Beim Menschen konnte u.a. für die α -Globin-Gene auf Chromosom 16 eine übergeordnete Kontrollregion, eine sogenannte Locus-Kontrollregion (LCR, *locus control region*), identifiziert werden, die ca. 40 kb strangabwärts des Clusters für eine entwicklungsspezifische Genexpression notwendig ist [Higgs *et al.*, 1990].

Auch für die drei Genfamilien der zentralen EDC-Region, die offenbar auf wenige Urgene zurückgeführt werden können [Backendorf & Hohl 1992; Fietz *et al.*, 1992; Gibbs *et al.*, 1993; Schafer *et al.*, 1995; Ridinger *et al.*, 1998], ist aufgrund ihrer strukturellen Gemeinsamkeiten, der gewebe- und differenzierungsspezifischen Genexpression sowie der funktionellen Kooperation der codierten Proteine die Frage nach einer potentiellen LCR aufgeworfen worden [Hardas *et al.*, 1996; Mischke *et al.*, 1996], insbesondere, da Untersuchungen am Mausgenom gezeigt haben, daß die orthologen Gene der drei Familien auf Mausechromosom 3 während der evolutionären Entwicklung ebenfalls zusammengehalten wurden [Rothnagel *et al.*, 1994; Ridinger *et al.*, 1998; Song *et al.*, 1999]. Es wurde spekuliert, daß die Expressionsmuster der EDC-Gene sich von der telomeren Seite zum Centromer hin schrittweise ändern, von einer wenig differenzierungs- und gewebespezifischen Verteilung der *S100*-Genprodukte zum hochspezifischen Vorkommen von Filaggrin und Trichohyalin in differenzierten Keratinozyten [Zhao & Elder, 1997]. Die für *NICE1* bis *NICE5* durchgeführte Expressionsanalyse unterstützt die Vermutung, daß mit der distalen Entfernung von den Strukturprotein-Genen die Spezifität der Genexpression nachläßt. *NICE1*, welches zwischen *FLG* und *IVL* kartiert, zeigte eine mit der Differenzierung stark zunehmende Expression in Keratinozyten; darüberhinaus wurden Transkripte bisher ausschließlich in einer Herz-cDNA-Bibliothek nachgewiesen. *NICE2* bis *NICE5* dagegen, die inmitten der *S100*-Gene (*NICE2*) bzw. weiter distal (*NICE3*, *NICE4* und *NICE5*) liegen, zeigten eine schwache, differenzierungsunabhängige Expression in Keratinozyten und waren in einer Vielzahl an Zelltypen exprimiert. Die *NICE5*-Sequenz war darüberhinaus bei *Caenorhabditis elegans* und *Drosophila melanogaster* stark konserviert – ein Kennzeichen sogenannter Haushaltsgene, die in nahezu allen Zellen der unterschiedlichsten Organismen exprimiert werden und elementare Funktionen des Zellstoffwechsels innehaben. Das gleiche Nachlassen der Spezifität war aber auch von den Strukturprotein-Genen über *S100A10* und *S100A11* zu den neu identifizierten Genen in der proximalen Region zu beobachten, so daß die Abnahme nicht von der telomeren zur centromeren Seite erfolgt, sondern vom Zentrum des EDC ausgehend zu beiden Seiten (entsprechend der oben gemachten Unterscheidung zwischen EDC und „CDC“). Während die Möglichkeit nicht ausgeschlossen werden kann, daß eine LCR für die koordinierte Aktivierung der zentralen EDC-Gene im Verlauf der terminalen Keratinozytendifferenzierung verantwortlich ist [Zhao & Elder, 1997], legen Promotoruntersuchungen der beteiligten Gene nahe, daß Transkriptionsfaktoren, vor allem der immer wieder involvierten Aktivator-Protein-1 (AP-1)-Familie, die Kontrolle über die Genexpression ausüben [Welter *et al.*, 1995; Fischer *et al.*, 1996; Sark *et al.*, 1998; Fischer *et al.*, 1999].

Da Calciumionen die Keratinozytenproliferation und -differenzierung maßgeblich beeinflussen (s. Abschnitt I 3), ist auch eine Beteiligung der Ca^{2+} -bindenden S100-Proteine an der Transkriptionskontrolle der EDC-Gene bzw. anderer für die Keratinozytendifferenzierung erforderlicher Gene denkbar, z.B. als Bestandteil von Signalkaskaden, an deren Ende die Aktivierung der entsprechenden Gene steht. Für *S100A1* konnte bereits eine Interaktion mit muskelspezifischen Transkriptionsfaktoren nachgewiesen werden, die zur Änderung der DNA-Bindungsaffinität führt und so die Genexpression

beeinflusst [Baudier *et al.*, 1995]. Weitere Kandidaten für die Steuerung der Genexpression in Keratinozyten finden sich unter den neu identifizierten EDC-Genen. Zu ihnen zählen *PIAS3*, das ein Inhibitorprotein des Transkriptionsfaktors Stat3 codiert, und *NICE9*, ein potentiell neues Mitglied der *NOTCH*-Familie, deren Mitglieder entscheidenden Einfluß auf die Zelldifferenzierung ausüben (s. Abschnitt IV 4).

5.3 Orthologie im Mausgenom – der EDC der Maus

Der gemeinsame Vorfahr von Maus und Mensch lebte vor ca. 70 Millionen Jahren. Obwohl sich die beiden Genome seitdem getrennt voneinander entwickelt haben, besitzen sie immer noch eine Reihe von Gemeinsamkeiten, welche die Maus zum bevorzugten Modellorganismus werden ließen [Übersicht in Strachan & Reid, 1999]. Dazu zählt u.a. die sehr ähnliche Embryonalentwicklung mit stark konservierten Expressionsmustern; über die gezielte Manipulation embryonaler Stammzellen können sogenannte *Knock-out*-Mäuse erzeugt werden, bei denen bestimmte Gene inaktiviert werden, um über die physiologischen Defekte die Funktionen der eliminierten Proteine *in vivo* zu analysieren. Auch der Aufbau der genomischen DNA zeigt Übereinstimmungen; beide Genome haben mit 3000 Mb eine vergleichbare Größe, und für beide Organismen wurde eine vergleichbare Anzahl an Genen errechnet, wobei die Sequenzhomologie codierender DNA 70-90% beträgt (Proteine weisen eine Homologie von 75-95% auf). Trotz dieser starken Konservierung unterscheiden sich Genfamilien bei Maus und Mensch häufig in der Zahl ihrer Mitglieder, und für einige Gene des Menschen gibt es kein orthologes Gen bei der Maus. Zurückzuführen sind diese Abweichungen auf Verdoppelungen chromosomaler Abschnitte und Umlagerungen, die nach der Abzweigung beider Linien im Laufe der Evolution stattgefunden haben. Das führte auch dazu, daß die Reihenfolge der Gene wenig konserviert ist. Dennoch finden sich fast 200 meist kleinere chromosomale Abschnitte, in denen die Anordnung orthologer Gene übereinstimmt [Übersicht in DeBry & Seldin, 1996].

Für Chromosom 1 des Menschen wurden acht orthologe Kopplungsgruppen im Mausgenom identifiziert, die auf den Mauschromosomen (mChr) 1, 3, 4, 5, 8 und 13 liegen [Übersicht in DeBry & Seldin, 1996]. Innerhalb dreier dieser Gruppen (1p36-p31 /mChr4, 1p21-q21/mChr3, 1q21-q32/mChr1) ist die Anordnung der Gene über beachtliche Entfernungen der genetischen Karte konserviert, wobei die Reihenfolge von 23 Genen in den orthologen Regionen auf Mauschromosom 1 und in 1q21-q23 beim Menschen auch auf physikalischen Restriktionskarten verifiziert wurde [Oakey *et al.*, 1992]. Auf Chromosom 1 des Menschen schließen die Gene der Region 1p21-q21 das Centromer ein; auf dem orthologen Abschnitt des Mauschromosoms 3 sind die entsprechenden Gene dagegen ohne Unterbrechung angeordnet [Übersicht in DeBry & Seldin, 1996] – ein Hinweis darauf, daß das Einfügen des Centromers ein relativ spätes Ereignis in der Genomentwicklung war.

Zur letztgenannten Kopplungsgruppe zählen auch die EDC-Gene *FLG*, *LOR*, *S100A6* und *S100A4* [Übersicht in DeBry & Seldin, 1996]. Daß die drei Genfamilien der Kernregion des EDC des Menschen auch auf Chromosom 3 der Maus als Genkomplex konserviert sind, haben weitere Untersu-

chungen gezeigt. So konnte mit Hilfe somatischer Zellhybride nachgewiesen werden, daß *Flg* und *Lor* im Mausgenom in einer Entfernung von 1,1-1,5 cM gekoppelt sind [Rothnagel *et al.*, 1994]. Durch zusätzliche Kartierung der *Sprrr*-Gene und von *Ivl* wurde die Reihenfolge *Flg* – *Ivl* – *Sprrr1* (*a,b*)/*Sprrr3* – *Sprrr2* (*a-k*) – *Lor* festgelegt [Song *et al.*, 1999]. Obwohl die Anordnung der zwei *Sprrr1*-Gene und des *Sprrr3*-Gens auf dem Mausechromosom 3 nicht aufgelöst werden konnte, ist der Aufbau des EDC in diesem Bereich offenbar stark konserviert. Abgesehen von dem noch nicht identifizierten Trichohyalin der Maus stimmt nur die Anzahl der *SPRR2*-Loci, elf bei der Maus [Song *et al.*, 1999] und sieben beim Menschen [Gibbs *et al.*, 1993], nicht überein. Anscheinend haben Duplikationen innerhalb der *SPRR*-Genfamilie sowohl vor der Aufspaltung der beiden Linien stattgefunden – sie führten wahrscheinlich zu den drei Untergruppen – als auch danach. Das *Repetin*-Gen, das im EDC des Menschen zwischen *THH* und *FLG* kartiert, wurde durch FISH der Region des Maus-EDC zugeordnet [Krieg *et al.*, 1997]. Schließlich wurden auf einem Maus-YAC acht orthologe *S100*-Gene lokalisiert [Ridinger *et al.*, 1998] sowie das dazwischen liegende *Zirtl*-Gen [Lioumi *et al.*, 1999]. Somit ist auch die Existenz eines *S100*-Genclusters im Maus-EDC gesichert. Die Anordnung *S100A3/A4/A5* – *A6* – *A13* – *Zirtl* – *A9/A8* – *A1* auf dem Maus-YAC [Ridinger *et al.*, 1998; Lioumi *et al.*, 1999] weicht allerdings von der Reihenfolge *S100A9* – *A12* – *A8* – *A7* – *A6* – *A5* – *A4* – *A3* – *A2* – *A13* – *A1* – *ZIRTL* im menschlichen Genom [Schafer *et al.*, 1995; Wicki *et al.*, 1996a; Wicki *et al.*, 1996b; Lioumi *et al.*, 1999] ab. Dies könnte ein Hinweis auf mehrere Umlagerungen, u.a. eine Inversion, sein, die nach der Trennung der beiden Linien stattgefunden haben – oder auf einen rearrangierten YAC. Erwartungsgemäß konnten die Mausgene *S100A10* und *S100A11* nicht auf dem YAC nachgewiesen werden [Ridinger *et al.*, 1998]. Homologie zum EDC des Menschen vorausgesetzt, sollten diese beiden Gene in einiger Entfernung, getrennt von den anderen *S100*-Genen, kartieren. Wie für Trichohyalin wurden bisher keine cDNA-Sequenzen der Maus für *S100A2*, *S100A7* und *S100A12* gefunden. Möglicherweise werden die entsprechenden Gene in den Zellen nur schwach exprimiert oder sie kommen im Mausgenom evolutionär bedingt nicht vor.

Von den Genen des erweiterten EDC wurden bisher die zu *TPM3*, *NICE9* (*NOTCH2*-ähnlich), *CHRN2* und *ADAR1* orthologen Gene *Tpm5* [Gariboldi *et al.*, 1995], *Notch2* [Gao *et al.*, 1998], *Acrb2* [Lueders *et al.*, 1999] und *Adar1* [Weier *et al.*, 2000] auf dem Mausechromosom 3 lokalisiert (Tab. 21, rechte Spalte). Obwohl die physikalische Kartierung des Mausgenoms noch nicht so weit fortgeschritten ist, legen die verfügbaren Daten nahe, daß der gesamte EDC als Bestandteil einer größeren orthologen Region auf dem Mausechromosom 3 konserviert ist – allerdings hat sich im Mausgenom offenbar eine umgekehrte Orientierung herausgebildet, wie für den Bereich von *Acrb2* bis *Flg* gezeigt wurde [Lueders *et al.*, 1999]. Die ausschließlich im *S100*-Gencluster festgestellten Abweichungen in der Reihenfolge der Gene könnten auf eine erhöhte Umlagerungsaktivität in diesem Bereich hinweisen.

Insgesamt ist die Orthologie zwischen Mausechromosom 3 und der Region 1p21-q21 auf Chromosom 1 des Menschen für die Erforschung des EDC eine wertvolle Hilfe, da sich die in einem Organismus

Tabelle 21: Paraloge und Orthologe der neuen EDC-Gene

EDC-Gen (1q21)	paraloge Gene				orthologes Mausgen	Locus
	6p21	9p13-q34	15q	19p13.3-13.1		
<i>NICE9</i> (<i>NOTCH2</i> -ähnlich)	<i>NOTCH4</i>	<i>NOTCH1</i>		<i>NOTCH3</i>	(<i>Notch2</i>)	mChr3
<i>PIAS3</i>			<i>PIAS1</i>			
<i>ANXA9</i>		<i>ANXA1</i>	<i>ANXA2</i>			
<i>PIP5K1A</i>		<i>PIP5K1B</i>		<i>PIP5K1C</i>		
<i>PSMD4</i> , <i>PSMD8L</i>		(<i>PSMD5</i>)		(<i>PSMD8</i>)	<i>Psm�4</i>	
<i>PSMB4</i>	<i>PSMB8</i> , <i>PSMB9</i>	<i>PSMB7</i>			<i>Psmb4</i>	
<i>RPTN</i>					<i>Rptn</i>	mChr3
<i>TPM3</i>		<i>TPM2</i>	<i>TPM1</i>	<i>TPM4</i>	<i>Tpm5</i>	mChr3
<i>CHRNA2</i>			<i>CHRNA4</i>		<i>Acrb2</i>	mChr3
<i>ADAR1</i>					<i>Adar1</i>	mChr3

erzielten Ergebnisse oft direkt auf den anderen übertragen lassen. So führte die Entdeckung des Reptin-Gens auf Chromosom 3 der Maus [Krieg *et al.*, 1997] kurze Zeit später zur Identifizierung des orthologen Gens im EDC des Menschen, und beim *S100*-Gencluster war es genau umgekehrt [Ridinger *et al.*, 1998]. Vor allem aber sind Homologien zwischen dem Genom der Maus und dem des Menschen von großem Nutzen für die Erforschung von Krankheiten, da sich die am Tiermodell gewonnenen Erkenntnisse oftmals direkt auf den Menschen übertragen lassen [Übersicht in DeBry & Seldin, 1996].

5.4 Paralogie im menschlichen Genom

Komplexe Genome sind höchstwahrscheinlich durch DNA-Duplikationen aus genetisch einfacher aufgebauten Organismen hervorgegangen. Die Verdoppelungen betrafen in der Regel nur Teile des Genoms, in seltenen Fällen aber auch ganze Chromosomensätze (Tetraploidisierung) [Übersicht in Lundin, 1993]. Anschließend Mutationen sowie weitere chromosomale Umstrukturierungen führten dann zur Auseinanderentwicklung der duplizierten Gene, wodurch ein neues Mitglied einer Genfamilie, ein Gen mit neuer Funktion oder ein funktionsloses Pseudogen entstehen konnte. Als Folge dieser Ereignisse sind homologe Chromosomenabschnitte sowohl bei verschiedenen Arten (Orthologie) als auch innerhalb des Genoms einer Art (Paralogie) zu beobachten. Da die Übereinstimmung der duplizierten DNA-Sequenzen aufgrund von Mutationen, Umlagerungen oder Deletionen mit der Zeit abnimmt, ist eine Identifizierung der betroffenen chromosomalen Regionen allerdings nicht immer möglich.

Im menschlichen Genom finden sich sowohl Hinweise auf intrachromosomale als auch auf interchromosomale Duplikationen, die zu seiner Entstehung beigetragen haben. Für die den EDC einschließende Region 1q21-q25 auf Chromosom 1 des Menschen wurden drei paraloge Abschnitte auf den Chromosomen 6, 9 und 19 identifiziert, die offensichtlich alle von der gleichen Urregion abstammen

[Katsanis *et al.*, 1996; Kasahara *et al.*, 1996; Übersicht in Kasahara *et al.*, 1997]. Phylogenetische Untersuchungen haben gezeigt, daß die Regionen 1q21-q25 und 9q33-q34 aus der jüngsten Duplikation hervorgegangen sind [Katsanis *et al.*, 1996]. Im Gegensatz dazu konnte die Verwandtschaftsbeziehung der Region 19p13.3-p13.1 zu den drei anderen paralogen Chromosomenabschnitten nicht abschließend geklärt werden.

Zu den Genfamilien, deren Mitglieder durch Duplikationen auf mehrere der vier paralogen Regionen verteilt wurden (Tab. 21), zählen u.a. die Gene dreier Komponenten des an der Immunabwehr beteiligten Komplementsystems (C3, C4 und C5), die Retinoidrezeptor-Genfamilie (*RXRA*, *RXRB* und *RXRG*) sowie eine Familie von Transkriptionsfaktor-Genen, die mit der *pre-B-cell leukemia* in Verbindung gebracht werden (*PBX1*, *PBX2* und *PBX3*) [Katsanis *et al.*, 1996]. Abgesehen von den Genen für die olfaktorischen Rezeptoren – von denen fast 1000 Loci im menschlichen Genom identifiziert wurden und die auf fast allen Chromosomen vorkommen [Glusman *et al.*, 2001] – wurden bisher nur die Mitglieder der *NOTCH*-Familie auf allen vier Chromosomen nachgewiesen [Sugaya *et al.*, 1994; Larsson *et al.*, 1994]. Auf Chromosom 1 liegt der *NOTCH2*-Locus allerdings in der Region 1p13-p11, außerhalb des paralogen Bereichs. Die Kartierung des *NOTCH*-homologen Gens *NICE9* in der Region 1q21 stimmt dagegen mit dem vorhergesagten Locus überein. Möglicherweise handelt es sich hierbei um ein bisher unbekanntes *NOTCH*-Gen, *NOTCH5*. Die auf beiden Seiten des Centromers gelegenen *NOTCH*-Loci unterstützen die Hypothese, daß die centromeren Regionen des langen und des kurzen Arms von Chromosom 1 ebenfalls aus einer Duplikation hervorgegangen sind, in diesem Fall aus einer intrachromosomalen [Wong *et al.*, 1990; Oakey *et al.*, 1992].

Mit dem *PSMB4*-Gen wurde das Mitglied einer weiteren Genfamilie kartiert, die in den paralogen Regionen auf den Chromosomen 6 (*PSMB8* und *PSMB9*) und 9 (*PSMB7*) vertreten ist [McCusker *et al.*, 1997]. Darüberhinaus wurden zwei Loci, die ebenfalls Untereinheiten des 26S-Proteasoms codieren, in der Region 1q21 nachgewiesen: *PSMD4* und *PSMD8L*, eine zu *PSMD8* homologe Gensequenz. Ihnen können verwandte, nicht genauer kartierte Gene auf den Chromosomen 9 (*PSMD5*) [Nomura *et al.*, 1994; Deveraux *et al.*, 1995] und 19 (*PSMD8*; GenBank Acc AC005789) zugeordnet werden. Offenbar haben auch hier, wie bei den *NOTCH*-Loci, intrachromosomale Verdopplungen zur Entstehung der diversen Untereinheiten beigetragen.

Durch die Kartierung des *PIP5K1A*-Gens konnte die Paralogie zwischen den Chromosomen 1, 9 und 19 erhärtet werden, da die anderen *PIP5K1*-Loci in den Regionen 9q13 (*PIP5K1B*) [Pook *et al.*, 1997] und 19p13.3 (*PIP5K1C*) [Nagase *et al.*, 1998; Ishihara *et al.*, 1998] kartieren. Zwar war der paraloge Abschnitt auf Chromosom 9 bisher auf die telomere Region 9q33-q34 beschränkt [Katsanis *et al.*, 1996]; mit *ANXA1* (9q11-q22) [Huebner *et al.*, 1988] und *TPM2* (9p13.2-p13.1) [Tiso *et al.*, 1997] kartieren aber die Mitglieder zweier weiterer Genfamilien, die in der Region 1q21 vertreten sind, nahe des Centromers von Chromosom 9, so daß sich die Paralogie anscheinend bis auf den kurzen Arm in die Region 9p13 hinein erstreckt.

Die vier Tropomyosin-Gene in den paralogen Regionen 1q21 (*TPM3*), 9p13.2-p13.1 (*TPM2*) [Tiso *et al.*, 1997] und 19p13.1 (*TPM4*) [Wilton *et al.*, 1996] sowie in der Region 15q22.1 (*TPM1*) [Tiso *et al.*, 1997] gaben dann den Hinweis auf einen möglichen paralogen Bereich auf Chromosom 15. Insgesamt konnten die Loci von vier Genfamilien in der Region 1q21 nachgewiesen werden, deren paraloge Gene auf dem langen Arm von Chromosom 15 liegen. Neben *TPM3* sind das *ANXA9* mit *ANXA2* in 15q21-q22 [Richard *et al.*, 1994], *CHRNA2* mit *CHRNA4* in 15q24 [Raimondi *et al.*, 1992] und *PIAS3* mit *PIAS1* in 15q (UniGene Cluster Hs.75251). Durch das bereits erwähnte *ANXA1*-Gen wird darüberhinaus die Paralogie zwischen den Regionen 9q und 15q unterstützt. In Übereinstimmung mit diesen Ergebnissen führten Datenbankanalysen und die Kartierung der Tropomodulin-Gene zu weiteren Genen der Region 15q, denen paraloge Loci in den Regionen 1q, 9q oder 19q zugeordnet werden können [Cox & Zoghbi, 2000], so daß die Paralogie der Region 15q als gesichert gelten darf.

Bemerkenswert ist die Tatsache, daß die paralogen Regionen auf den Chromosomen 1 und 6 jeweils einen Genkomplex einschließen, den EDC in 1q21 bzw. den MHC in 6p21.3. Möglicherweise liegt in der Paralogie der Grund für erhöhte Umlagerungsaktivitäten in beiden Regionen, die im Laufe der Evolution zu einer Vielzahl strukturell und funktionell in Beziehung stehender Gene geführt haben. Ob hierbei repetitive Elemente eine Rolle spielen, die sowohl in bestimmten Bereichen des MHC gehäuft vorkommen [Iris *et al.*, 1993; Fukagawa *et al.*, 1995] als auch auf den YACs des EDC, bedarf noch weiterer Untersuchungen.

Neben der vergleichenden Morphologie ist die Molekularphylogenetik für die Aufklärung der Evolutionsgeschichte in den letzten Jahren zu einem wichtigen Hilfsmittel geworden ist. Die Analyse paraloger Gene und Chromosomenabschnitte kann dabei bedeutende Einblicke in den Verlauf der Evolution gewähren. So konnte anhand der *PBX*-Gene (s.o.) gezeigt werden, daß die jüngste Duplikation, die zu den Regionen 1q21-q25 und 9q33-q34 geführt hat, vor der Abzweigung der Knochenfische vor ca. 420 Millionen Jahren stattgefunden hat [Jones *et al.*, 1995]. Die kieferlosen Fische hatten dagegen ihren Ursprung vor ca. 515 Millionen Jahren, vor der ersten interchromosomalen Duplikation dieser Region, die zur Entstehung des MHC auf Chromosom 6 geführt hat [Kasahara *et al.*, 1996].

Eine weiteres Anwendungsgebiet für Paralogien ist die Genkartierung. Wie bei orthologen Regionen unterschiedlicher Arten lassen sich auch auf paralogen Chromosomenabschnitten Gene vorhersagen. Auf diese Weise wurde z.B. ein *PBX*-Locus auf Chromosom 6 sowie ein *NOTCH*-Locus auf Chromosom 1 identifiziert [Katsanis *et al.*, 1996]. Es kann davon ausgegangen werden, daß in den jetzt fünf paralogen Regionen weitere Mitglieder bereits identifizierter Genfamilien vorhanden sind. Denkbar wären z.B. Annexin-Gene in den Regionen 6p und 19p, ein *NOTCH*-Locus in 15q sowie zusätzliche *PIAS*-Gene auf den Chromosomen 6, 9 und 19. Dabei ist aber zu beachten, daß Mutationen, chromosomale Umlagerungen und Deletionen nach einem Duplikationsereignis, das mehrere hundert Millionen Jahre zurückliegt, ein paraloges Gen möglicherweise so stark verändert haben, daß es nicht mehr als solches zu erkennen ist, oder es sogar ganz eliminiert haben.

Schließlich kommen paraloge Gene als Auslöser von Krankheiten in Frage, bei denen mehrere Loci, die mit dem gleichen klinischen Phänotyp in Verbindung gebracht werden, auf paralogen Chromosomenabschnitten kartieren. Für die Psoriasis z.B. wurden mit der Krankheit assoziierte Loci u.a. in den Regionen 1q21 [Capon *et al.*, 1999], 6p21.3 [Balendran *et al.*, 1999] und 19p13 [Lee *et al.*, 2000] kartiert und für zwei Formen der Hämochromatose, bei der es aufgrund einer Funktionsstörung des retikulohistiozytären Systems zu einer allgemeinen Eisenüberladung der Organe kommt, in den Regionen 1q21 [Roetto *et al.*, 1999] und 6p22 [Feder *et al.*, 1996; Totaro *et al.*, 1998].

6 Ausblicke

Obwohl zwei vorläufige Versionen des menschlichen Genoms bereits veröffentlicht wurden [The International Human Genome Mapping Consortium, 2001; Venter *et al.*, 2001], wird es noch einige Zeit in Anspruch nehmen, bis eine lückenlose Sequenz verfügbar sein wird. Der Grund hierfür liegt vor allem in bestimmten Regionen, deren DNA offenbar von den entsprechenden Wirtsorganismen eliminiert wird und somit nur schwer kloniert werden kann [Neil *et al.*, 1990; Kang & Cox, 1996; Bieth *et al.*, 1997; Kouprina *et al.*, 1999]. Im EDC des Menschen konnte das für die Regionen distal von *S100A1* und proximal von *ANXA9* beobachtet werden. Für die Sequenzierung des 2 Mb umfassenden Kernbereichs des EDC wurden mit dieser Arbeit jedoch die Voraussetzungen geschaffen: Die Klone eines PAC/BAC/Cosmid-Contigs [South *et al.*, 1999] werden zur Zeit am Sanger Center sequenziert; mit der endgültigen Sequenz ist in Kürze zu rechnen.

Zwar sind in der DNA-Sequenz eines Organismus nahezu alle seine Proteine codiert – dennoch erkennen computergestützte Sequenzanalysen nur ca. 70% aller Exons [Lopez *et al.*, 1994]. Für das Auffinden aller codierenden Sequenzen müssen daher ergänzend molekularbiologische Methoden eingesetzt werden, wie die Identifizierung transkribierter Sequenzen durch *exon trapping*. Die in dieser Arbeit beschriebene subtraktive Hybridisierung hat gezeigt, daß sie mit relativ wenig Arbeitseinsatz und Zeitaufwand eine extrem hohe Erfolgsrate bietet. Durch Fortführung der Analysen, vor allem der bisher ausgelassenen Abschnitte des YAC-Contigs, oder durch Verwendung anderer cDNA-Bibliotheken könnten weitere Gene aus der EDC-Region identifiziert werden. Auch die gezielte Suche nach Genen, die in kranken Geweben ungewöhnlich exprimiert werden, wird durch die subtraktive Hybridisierung feingerasterter cDNA-Bibliotheken stark vereinfacht. Auf den EDC bezogen wäre die differenzielle Expressionsanalyse in psoriatischem oder ichthyotischem Gewebe von Interesse, da bei beiden Krankheiten ein Bezug zur Region 1q21 diskutiert wird [Sybert *et al.*, 1985; Nirunsuksiri *et al.*, 1995; Hardas *et al.*, 1996; Capon *et al.*, 1999].

Da in der feingerasterten Keratinozyten-cDNA-Bibliothek die 5'-Enden einiger Transkripte des EDC möglicherweise nicht vorhanden sind, muß versucht werden, diese entweder aus der flüssigen Stammkultur oder aus anderen cDNA-Bibliotheken zu isolieren, z.B. durch Amplifizierung von cDNA-Enden

mittels RACE (*rapid amplification of cDNA ends*) [Frohman *et al.*, 1988]. Anschließend kann mit der Charakterisierung der Genprodukte und den Funktionsanalysen der Proteine der interessanteste Teil des ganzen Vorhabens beginnen. Neben weiteren Expressionsstudien zum Vorkommen und zur Verteilung der Transkripte in verschiedenen Geweben, kann der Wirkort der Proteine in der Zelle bzw. im Gewebe erste Hinweise auf eine mögliche Funktion geben. Um diesen herauszufinden, kann z.B. durch Klonierung in dem entsprechenden Expressionsvektor das zu untersuchende Protein an das grün-fluoreszierende Protein (GFP) gekoppelt werden, das anschließend in den entsprechenden Zellkompartimenten nachgewiesen werden kann [Chalfie *et al.*, 1994]. Als Kandidat für dieses Experiment bietet sich das *NICE1*-Gen an, dessen zugehöriges Protein, ein potentieller Bestandteil des CE, sich in differenzierten Keratinozyten in dieser Struktur anhäufen könnte.

S100-Proteine sind in der Regel Bestandteile von Signalkaskaden. Durch Bindung von Metallionen, vorwiegend Ca^{2+} , aber auch Zn^{2+} und Cu^{2+} , können sie ihre Konformation verändern und mit bestimmten Zielproteinen interagieren [Übersicht in Donato, 1999]. Da S100-Proteine u.a. bei Kardiomyopathien, entzündlichen Prozessen und Tumoren (s. Abschnitt I 6) eine Rolle spielen, können spezifische Antikörper wichtige diagnostische Hilfsmittel für diese Erkrankungen liefern. Die Aufklärung der dreidimensionalen Struktur der S100-Proteine durch Kernresonanzspektroskopie (NMR, *nuclear magnetic resonance*) sowie die Identifizierung möglicher Interaktionspartner, z.B. mit Hilfe des Zwei-Hybrid-Systems (*two-hybrid system*) [Chien *et al.*, 1991], könnte Angriffspunkte für therapeutische Konzepte aufzeigen. Erste (antiallergische) Arzneistoffe, die mit S100-Proteinen interagieren, wurden bereits identifiziert [Shishibori *et al.*, 1999].

Mit fortschreitender Kartierung des EDC der Maus bieten sich neue Möglichkeiten für die Generierung transgener Mäuse, um die physiologischen Aufgaben der einzelnen Proteine im Organismus zu untersuchen. Bisher wurde eine Involucrin-defiziente Maus erzeugt, die keinen auffälligen Phänotyp zeigte [Djian *et al.*, 2000], und eine Loricrin-*Knock-out*-Maus, bei der nach kurzer Zeit andere Strukturproteine des CE die Funktion des Loricrins kompensierten [Koch *et al.*, 2000]. Möglicherweise müssen mehrere EDC-Gene mit ähnlicher Funktion eliminiert werden, damit deren Bedeutung für die Entwicklung der Epidermis bzw. auch anderer Gewebe untersucht werden kann. Der Phänotyp dieser Mäuse könnte dann Übereinstimmungen mit Krankheitsbildern beim Menschen aufweisen.

Schließlich wurde begonnen, die evolutionären Beziehungen zwischen den EDC-Genen anhand des Kugelfischs (*Fugu rubripes rubripes*) zu analysieren. Dieser hat im Vergleich zu den Wirbeltieren ein komprimiertes Genom (400 Mb), die repetitive DNA hat nur einen Anteil von 7,4% (gegenüber ca. 40% beim Menschen), und die Introns sind kürzer [Brenner *et al.*, 1993]. Dementsprechend sollte auch der EDC des Kugelfischs eine geringere Ausdehnung haben. Da im Kugelfischgenom auch mit einer geringeren Anzahl an Genen zu rechnen ist, können so möglicherweise Hinweise auf den Ursprung bzw. die Urgene des EDC gefunden werden.