

## III ERGEBNISSE

Die Ergebnisse dieser Arbeit sind in vier Abschnitten dargestellt. Am Anfang steht die Erstellung eines Contigs der Region 1q21, als Grundlage aller weiteren Arbeiten mit einer kurzen Diskussion. Es folgt die Identifizierung neuer EDC-Gene durch Etablieren einer Hybridisierungsmethode und die Positionierung der Gene auf der integrierten Karte, die sich aus der genomischen Restriktionskarte und dem verfeinerten Contig zusammensetzt. Den Abschluß bildet die Charakterisierung der cDNA-Sequenzen der neuen Gene. Eine ausführliche Diskussion der gesamten Arbeit erfolgt im letzten Kapitel.

### 1 Erstellung eines YAC-Contigs der Region 1q21

Für die molekulargenetische Charakterisierung des EDC mußte ein Contig erstellt werden, das den entsprechenden Abschnitt von Chromosom 1 möglichst vollständig in überlappenden DNA-Klonen abdeckte. Da die zu Beginn des Projekts bekannten EDC-Gene bereits innerhalb von 2 Mb der Region 1q21 kartiert waren [Volz *et al.*, 1993] (Abb. 5), lag es nahe, einen Vektor zu verwenden, der möglichst große Bereiche des menschlichen Genoms aufnehmen kann (Tab. 2). Die entsprechende Entscheidung für das YAC sollte nicht nur die Untersuchung des gesamten EDC mit vertretbarem Aufwand ermöglichen, sondern auch die der angrenzenden Regionen. Als Ausgangspunkt diente die CEPH-Mega-YAC-Bibliothek, die das gesamte Genom des Menschen in YAC-Klonen enthält, welche bereits ansatzweise charakterisiert waren, u.a. durch die Bestimmung der YAC-Größen, die Kartierung von STS-Markern und die Identifizierung überlappender Klone [Cohen *et al.*, 1993]. Dadurch schien eine sinnvolle Auswahl an Klonen im Hinblick auf die zu untersuchende Region möglich zu sein, obwohl die Ergebnisse der Charakterisierungen nicht immer zuverlässig waren und Probleme mit chimären YACs zu erwarten waren [Green *et al.*, 1991; Cohen *et al.*, 1993].

#### 1.1 Auswahl der YAC-Klone

Aufgrund übereinstimmender Daten zur physikalischen Kartierung der einzelnen Gene (s. Abschnitt I 4) war die Lokalisierung des EDC auf der chromosomalen Bande 1q21 gesichert. Die YAC-Klone, die den entsprechenden Abschnitt des menschlichen Genoms enthalten sollten, waren jedoch nur durch wenige genetisch kartierte Marker charakterisiert [Cohen *et al.*, 1993], deren physikalische Positionen noch nicht genau festgelegt waren. Deshalb erfolgte eine Vorauswahl von 26 YACs, die die STS-Marker von D1S189 bis D1S305 der genetischen Karte von Généthon [Gyapay *et al.*, 1994] (Abb. 15,

Tabelle 8A: CEPH-Daten für die ausgewählten YACs [Cohen et al., 1993]

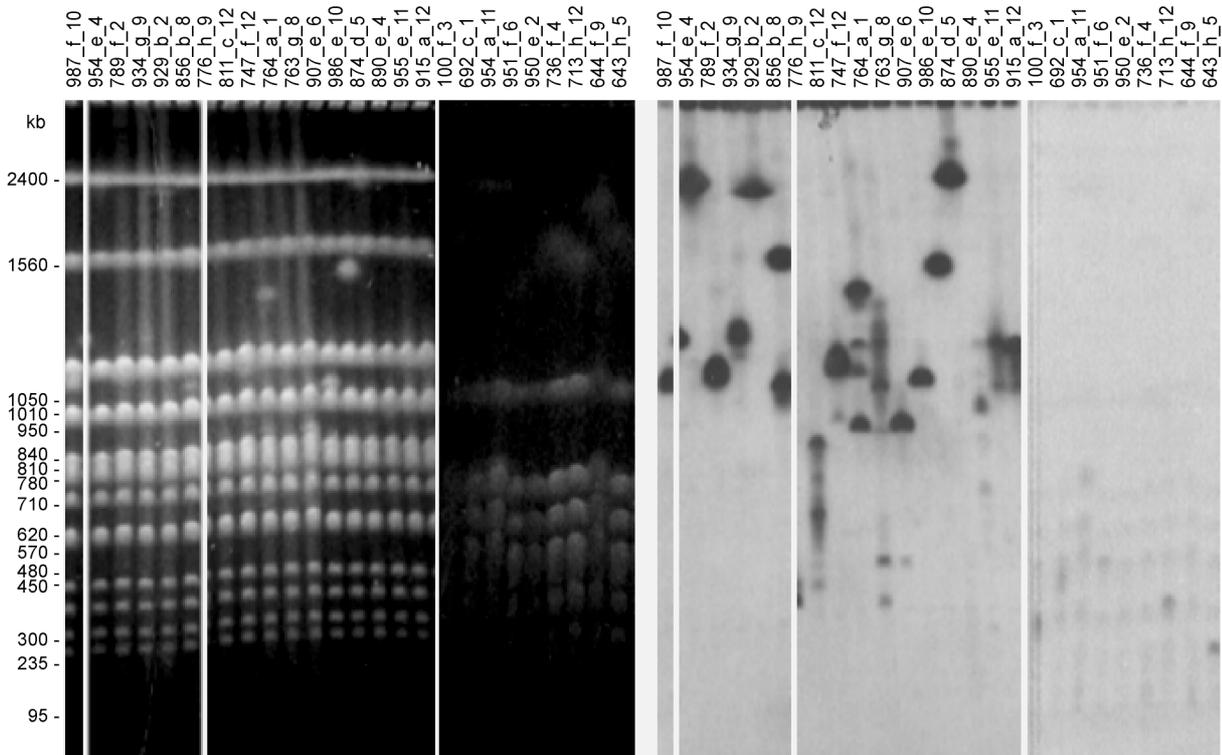
8B: Experimentell ermittelte Daten

YAC	Größe in kb	STS-Marker und Gene	Überlappungen	Größe in kb	STS-Marker
100_f_3 <sup>a</sup>	- (580)	(S100A1-9)		560	
210_h_11	-	CAPL (S100A4)		YAC-Klon ist nicht gewachsen	
643_h_5	430	<b>D1S305</b> , D18S455		450	D1S305
644_f_9	230	<b>D1S305</b>		230	D1S305
692_c_1 <sup>a</sup>	760 (660)	CAPL (S100A4)	<b>955_e_11</b>	740	
713_h_12	450, 660	<b>D1S305</b>		630	D1S305
736_f_4	20	<b>D1S305</b>		260, 380	D1S305
747_f_12	890, 1220, 1760	D8S592, AFMa071yf1	717_c_10, <b>763_g_8</b> , <b>764_a_1</b> , 882_a_2	900, 1360	
763_g_8	940		<b>747_f_12</b> , <b>764_a_1</b> , 811_c_12, <b>874_d_5</b>	890	
764_a_1	-	<b>D1S498</b> , AFM291xh1 (Chr1)	<b>747_f_12</b> , <b>763_g_8</b> , 811_c_12, <b>874_d_5</b>	330-1700	D1S498
776_h_9	860	<b>D1S498</b>	677_a_3, <b>934_g_9</b>	630, 860	D1S498
789_f_2	1300		<b>763_g_8</b> , <b>764_a_1</b>	1120	D1S442, D1S498
811_c_12	1370		<b>959_a_4</b>	1100	
856_b_8	1000	FGFR(Chr8)		1065, 1540	D1S442, D1S498
874_d_5	1450		<b>763_g_8</b> , <b>764_a_1</b> , <b>890_e_4</b> , <b>907_e_6</b> , <b>986_e_10</b>	2250	
890_e_4	1110		<b>874_d_5</b> , <b>955_e_11</b>	950	
907_e_6	1220		<b>874_d_5</b>	1040	
915_a_12	1360	D18S457	898_c_4, 940_c_7, <b>954_a_11</b> , 958_c_1	1160	
929_b_2	80	<b>D1S442</b> , WI-2014		2190	D1S442
934_g_9	1700	<b>D1S442</b>	<b>789_f_2</b>	1240, 1270	D1S442
950_e_2	340, 930, 1290	<b>D1S305</b> , AFM291vh5 (Chr1)		240-860	D1S305
951_f_6	1080	<b>D1S305</b>		850	D1S305
954_a_11	1580	<b>D1S305</b> , D18S457	<b>915_a_12</b>	950, 1170	D1S305
954_e_4	1060, 1080	<b>D1S442</b>	693_d_6, 758_h_9, 764_h_7, 799_c_8, 922_g_3, 929_f_7, <b>987_f_10</b>	2340	D1S442
955_e_11	1310		<b>692_c_1</b> , <b>890_e_4</b>	1010, 1170	
986_e_10	-		745_h_1, <b>874_d_5</b> , 918_h_12	1440	
987_f_10	-		693_d_6, 758_h_9, 764_h_7, 799_c_8, 922_g_3, 929_f_7, <b>954_e_4</b>	1090	

<sup>a</sup> Die in Klammern angegebenen YAC-Größen und der Gengehalt von YAC 100\_f\_3 wurden von Schaefer und Mitarbeitern (1995) bestimmt.

unten) enthielten und innerhalb eines Intervalls von 11 cM lagen, um diejenigen Marker zu bestimmen, die eine Verbindung zum EDC aufwiesen. Durch Southern-Blot-Hybridisierungen mit spezifischen Sonden der Gene *S100A6* und *S100A10*, welche die physikalische Karte des EDC begrenzten, konnte ein einziges YAC identifiziert werden: YAC 950\_e\_2, das den Marker D1S305 enthielt, hybridisierte mit der *S100A6*-Sonde.

Diese Information war der Ausgangspunkt für eine zweite, gezieltere Auswahl von YAC-Klonen für die Konstruktion des Contigs. Eingeschlossen werden sollten alle YACs der CEPH-Mega-YAC-Bibliothek, welche die D1S305 benachbarten Marker der genetischen Karte enthielten, D1S442 und D1S498 auf der einen und D1S303 auf der anderen Seite. Da D1S303 keinem YAC zugeordnet war, beschränkte sich die Auswahl auf die Marker D1S305, D1S442 und D1S498, die einen Bereich von 4 cM der genetischen Karte [Gyapay *et al.*, 1994] begrenzten. Insgesamt konnten zwölf Klone der CEPH-Mega-YAC-Bibliothek identifiziert werden, die je einen der drei Marker enthielten. Weitere zwölf wurden durch direkte Überlappung mit einem der ersten zwölf YACs oder durch weitere Überlappungen ausgewählt (Tab. 8A). YAC 210\_h\_11, dem das *S100A4*-Gen zugeordnet war, zeigte kein Wachstum und konnte daher nicht analysiert werden. Zwei zusätzliche YAC-Klone, 100\_f\_3 und 692\_c\_1, auf denen *S100*-Gene der Region 1q21 kartiert waren [Schafer *et al.*, 1995], wurden von einer kooperierenden Arbeitsgruppe beige-steuert und komplettierten die 26 Kandidaten für das Contig.



**Abbildung 11: Bestimmung der YAC-Größen.** Das mit Ethidiumbromid gefärbte Agarosegel (links) zeigt die durch ROFE aufgetrennten Chromosomen der 26 ausgewählten YAC-Klone. Durchgängig vorhandene Banden stammen von den originären Hefechromosomen, zusätzliche Banden von den künstlichen. Mit Hilfe der Autoradiographie wurden die YACs auf dem entsprechenden Southern-Blot nach Hybridisierung mit  $^{32}\text{P}$ -markierter Gesamt-DNA des Menschen sichtbar gemacht (rechts). Die Größen des Standards stammen von *S. cerevisiae* YPH149. Sie sind nicht auf die neun rechts angeordneten YACs übertragbar, die in einem separaten Gel aufgetrennt wurden.

## 1.2 YAC-Größenbestimmung

Um bei der Erstellung der Contig-Karte die Entfernungen zwischen einzelnen Markern abschätzen zu können und zur Identifizierung instabiler und chimärer Klone, wurden die Größen der YACs bestimmt. Mittels ROFE wurden die künstlichen von den originären Hefechromosomen abgetrennt und durch Southern-Blotting auf Nylonmembranen übertragen. Anschließend wurden sie mit radioaktiv markierter Gesamt-DNA des Menschen hybridisiert und durch Autoradiographie nachgewiesen. So wurde in 17 der vom CEPH erhaltenen YAC-Kulturen ein einzelnes künstliches Hefechromosom identifiziert, sieben Kulturen stellten sich als Gemisch zweier bzw. dreier YACs heraus (736\_f\_4, 747\_f\_12, 776\_h\_9, 856\_b\_8, 934\_g\_9, 954\_a\_11, 955\_e\_11) und zwei Kulturen enthielten eine Vielzahl an YACs (764\_a\_1, 950\_e\_2), die kein eindeutiges Hybridisierungssignal ergaben und an einer breiten, verschmierten Bande zu erkennen waren (Abb. 11). Die YAC-Größen lagen zwischen 230 und 2340 kb (Tab. 8B). Abweichungen zu den vom CEPH ermittelten Werten sind einerseits auf die durch ROFE verbesserte Auflösung großer DNA-Fragmente zurückzuführen, können aber auch Ausdruck der Instabilität und Inhomogenität der einzelnen YAC-Klone bzw. -Kulturen sein.

## 1.3 Charakterisierung instabiler YACs

Eine Charakterisierung der instabilen Klone wurde anhand der YAC-Kulturen 764\_a\_1 und 950\_e\_2 durchgeführt, denen keine eindeutige Größe zugeordnet werden konnte. Nach dem Ausbreiten der

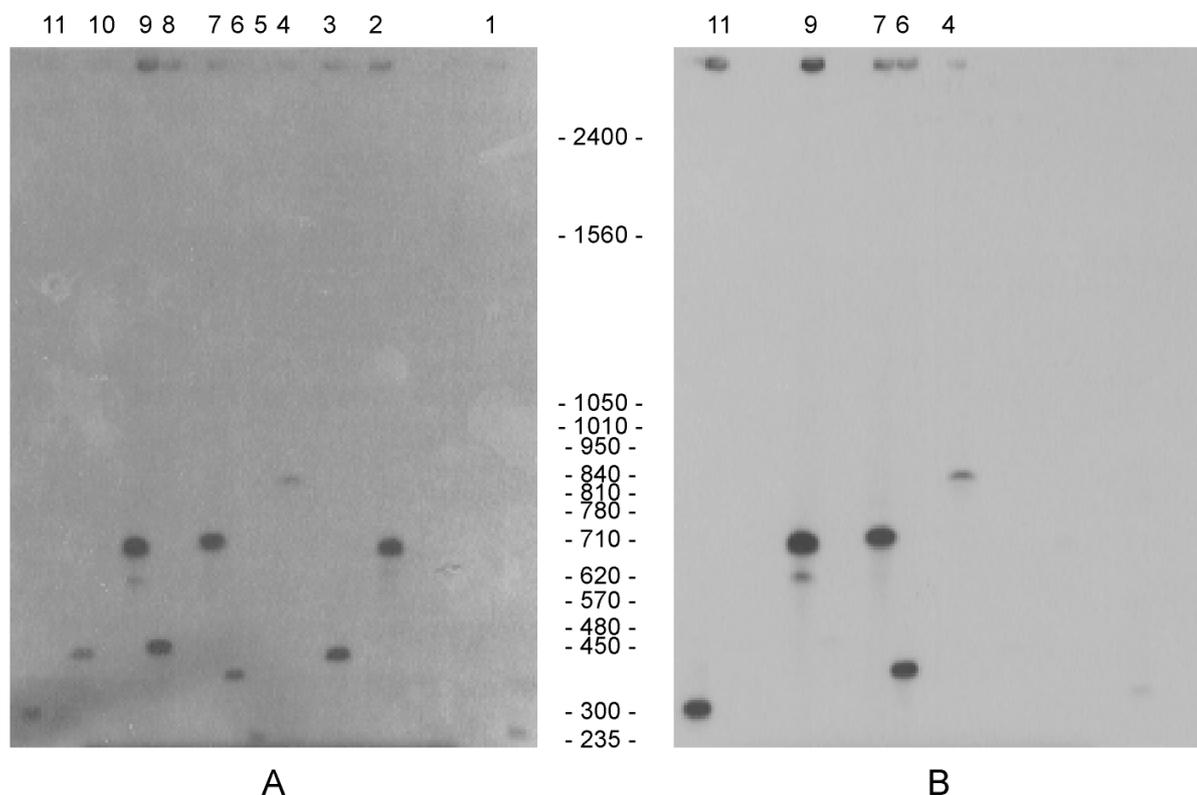


Abbildung 12: Charakterisierung der aus YAC 950\_e\_2 isolierten Kolonien. Die Autoradiogramme zeigen den Southern-Blot, der die durch ROFE aufgetrennten Hefechromosomen der einzelnen YAC-Kolonien enthält, nach Hybridisierung mit  $^{32}\text{P}$ -markierter Gesamt-DNA des Menschen (A) und nach Hybridisierung mit einer spezifischen Sonde für S100A6 (B). Die Größen des Standards stammen von *S. cerevisiae* YPH149. Nur die jeweils positiven Kolonien sind angegeben.

Stammkulturen auf Agarplatten wurden von jeweils 20 isolierten Kolonien die YAC-Größen bestimmt. Für 764\_a\_1 konnten in allen Kolonien YACs nachgewiesen werden (#1 bis #20), die Größen lagen zwischen 330 und 1700 kb. Lediglich elf der aus 950\_e\_2 isolierten Kolonien enthielten YACs (#1 bis #11), ihre Ausdehnung reichte von 240 bis 860 kb (Abb. 12A). Weder für 764\_a\_1 noch für 950\_e\_2 zeigte sich ein bevorzugter Wert (Tab. 9). Weitere Analysen ergaben, daß die YACs die peripheren Bereiche des EDC enthielten. Von den 20 aus 764\_a\_1 isolierten Klone hybridisierten 19 mit spezifischen Sonden der Gene *THH* und *S100A10*, lediglich dem kleinsten YAC, 764\_a\_1#19 (330 kb), fehlte das *S100A10*-Gen (Tab. 9). Zwar hybridisierten nur zehn der elf Klone aus 950\_e\_2 mit der *S100A9*-Sonde und fünf mit der *S100A6*-Sonde (Abb. 12B); durch die Generierung weiterer Marker (s. Abschnitt III 1.6) konnte aber eindeutig nachgewiesen werden, daß auch diese Klone aus der gleichen Region stammten. Offensichtlich waren die unterschiedlichen YACs einer Kultur aus einem Ursprungs-YAC hervorgegangen, das bestimmte DNA-Abschnitte in unterschiedlichem Ausmaß eliminiert hatte. Unter Beachtung dieser Umlagerungen wurden auch die aus mehreren

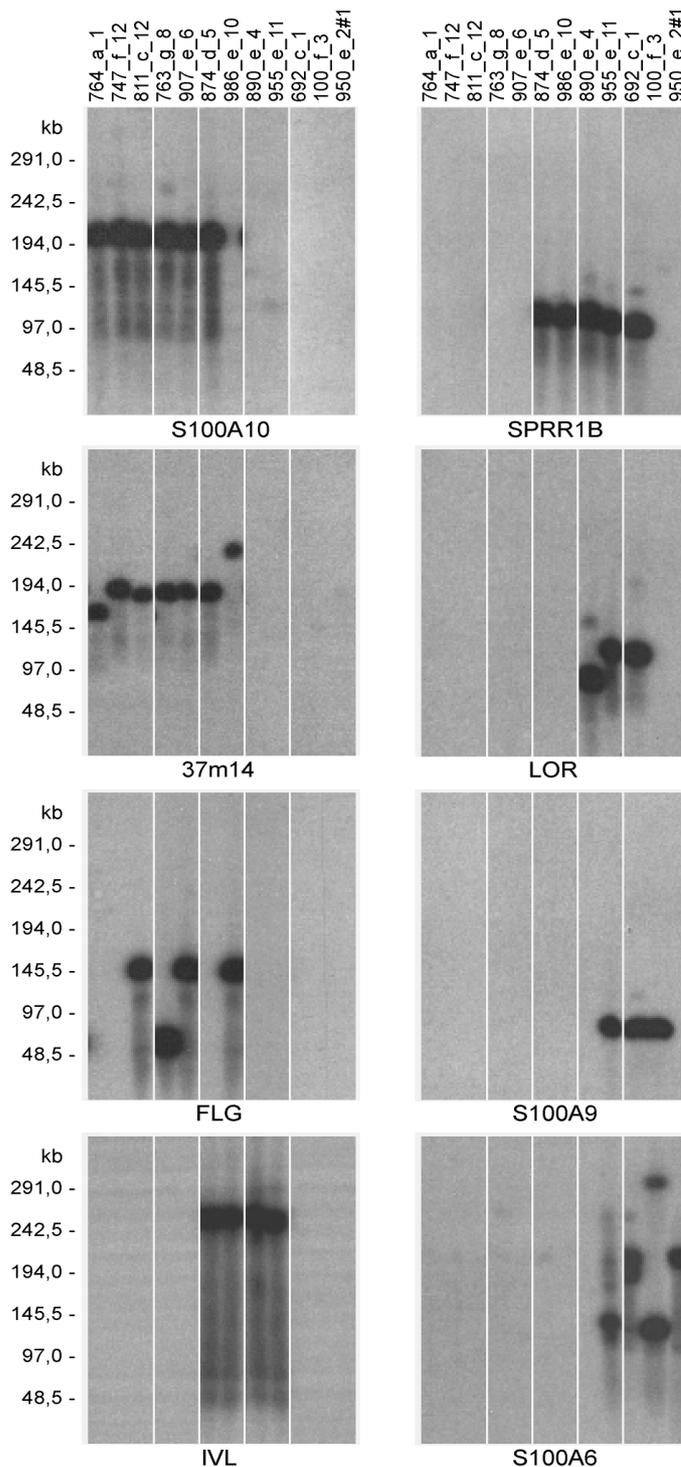
**Tabelle 9: Charakterisierung der instabilen YACs 764\_a\_1 und 950\_e\_2**

YAC	Kolonie	Größe in kb	enthaltene Marker
764_a_1	#1	1320	<i>S100A10, THH</i>
	#2	1300	<i>S100A10, THH</i>
	#3	460	<i>S100A10, THH</i>
	#4	1050	<i>S100A10, THH</i>
	#5	1190	<i>S100A10, THH</i>
	#6	460	<i>S100A10, THH</i>
	#7	1270	<i>S100A10, THH</i>
	#8	1200	<i>S100A10, THH</i>
	#9	1050	<i>S100A10, THH</i>
	#10	1100	<i>S100A10, THH</i>
	#11	1240	<i>S100A10, THH</i>
	#12	1400	<i>S100A10, THH</i>
	#13	1230	<i>S100A10, THH</i>
	#14	450	<i>S100A10, THH</i>
	#15	1700	<i>S100A10, THH</i>
	#16	1180	<i>S100A10, THH</i>
	#17	390	<i>S100A10, THH</i>
	#18	1100	<i>S100A10, THH</i>
	#19	330	<i>THH</i>
	#20	1210	<i>S100A10, THH</i>
950_e_2	#1	250	<i>S100A9</i>
	#2	680	<i>S100A9</i>
	#3	410	-
	#4	860	<i>S100A6, S100A9</i>
	#5	240	<i>S100A9</i>
	#6	340	<i>S100A6, S100A9</i>
	#7	710	<i>S100A6, S100A9</i>
	#8	440	<i>S100A9</i>
	#9	690	<i>S100A6, S100A9</i>
	#10	430	<i>S100A9</i>
	#11	300	<i>S100A6, S100A9</i>

YACs bestehenden Kulturen 764\_a\_1 und 950\_e\_2 in die Arbeit mit eingeschlossen, was vor allem für die spätere Genkartierung von großem Nutzen sein sollte. Zur Vereinfachung werden auch die YAC-Kulturen im folgenden als YAC-Klone bzw. YACs bezeichnet. Nur wenn es auf eine Unterscheidung der einzelnen Kolonien einer YAC-Kultur ankommt, z.B. bei der Genkartierung, wird für diese die YAC-Adresse mit einer nachgestellten Koloniennummer versehen.

#### 1.4 Zusammensetzen eines Contigs des EDC

Um die Überlappungen der einzelnen Klone zuverlässig bestimmen zu können und die physikalische Position der YACs in Bezug auf die genomische Restriktionskarte des EDC festzulegen, wurden die EDC-Gene auf den einzelnen YACs lokalisiert. Zu diesem Zweck wurde die DNA der 26 YAC-Klone

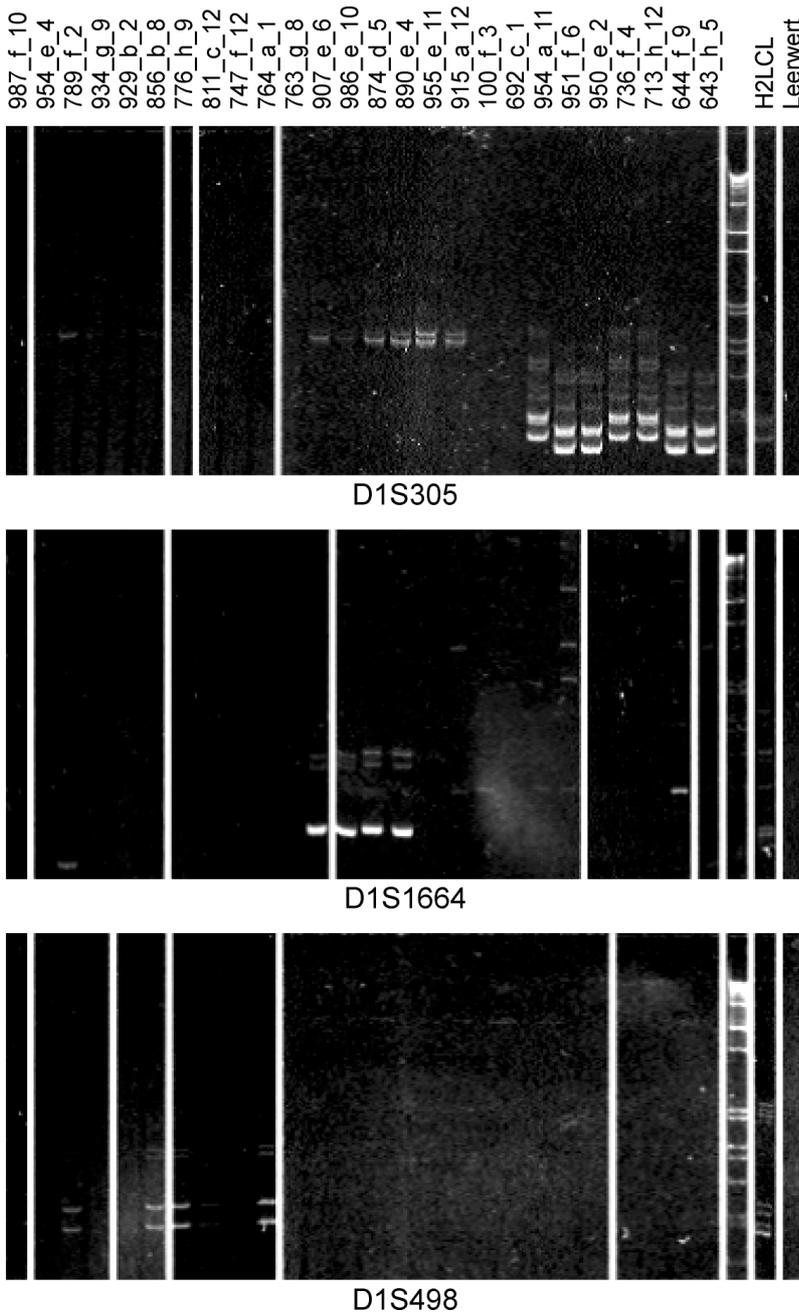


**Abbildung 13:** Nachweis von EDC-Genen auf den YACs des Contigs der Region 1q21. Die Autoradiogramme zeigen die von den  $^{32}\text{P}$ -markierten Sonden der angegebenen Gene und Marker nachgewiesenen *Sall*-Restriktionsfragmente der zwölf überlappenden YACs, die die EDC-Gene enthalten, in der entsprechenden Reihenfolge. Abweichende Fragmentgrößen sind auf YAC-Enden oder chimäre Bereiche zurückzuführen. In YAC 874\_d\_5 ist der *FLG*-Locus deletiert, die angrenzenden Marker 37m14 und *IVL* sind vorhanden. Die Heterogenität der mit der *S100A6*-Sonde hybridisierenden Fragmente weist auf die Instabilität dieser Region in YACs hin. Die YAC-Adressen (oben) und die Größen des  $\lambda$ -Standards (links) sind angegeben.

vor dem Southern-Blotting mit dem Restriktionsenzym *Sall* geschnitten und die DNA-Fragmente durch ROFE aufgetrennt. Die anschließenden Hybridisierungen mit radioaktiv markierten, genspezifischen Sonden identifizierten zwölf YACs, die eines oder mehrere EDC-Gene enthielten (Abb. 13). Dabei konnte jedes Gen mindestens drei unabhängigen YACs zugeordnet werden, womit die gesamten 2 Mb des EDC doppelt mit DNA-Klonen abgedeckt waren (*doubly linked contig*). Die Anordnung der Gene auf den YACs war bis auf eine Deletion des *FLG*-Locus in YAC 874\_d\_5 ohne Widerspruch zur genomischen Restriktionskarte möglich (Abb. 15).

### 1.5 Erweiterung des Contigs durch die Integration genetischer Marker

Als nächstes folgte die Integration der genetischen Marker. Da es sich hierbei um STS-Marker handelte, die durch begrenzte Primersequenzen und das bei der PCR entstehende Produkt definiert sind, wurden PCR-Analysen mit gereinigter, gelfiltrierter DNA der einzelnen YAC-Klone durchgeführt. Zur Anwendung kamen die bereits erwähnten STS-Marker D1S305, D1S442 und D1S498 [Gyapay *et al.*, 1994], D1S2345 sowie D1S1664, D1S2346 und D1S2347, die vom *Whitehead Institute/MIT* einzelnen YACs des Contigs zugeordnet worden waren, und zwar 874\_d\_5 (D1S1664), 955\_e\_11 (D1S2346) bzw. 776\_h\_9 (D1S2347)



**Abbildung 14:** Nachweis von STS-Markern auf den YACs durch PCR. Dargestellt sind die Ethidiumbromid-gefärbten Polyacrylamidgele, in denen die mit spezifischen Primern für D1S305, D1S1664 und D1S498 amplifizierten Fragmente aufgetrennt wurden. DNA aller 26 Kandidaten-YACs wurde eingesetzt. Als Positivkontrolle diente DNA der H2LCL-Zelllinie, als Größenstandard die 1 kb-Leiter (dritte Bahn von rechts).

[Hudson *et al.*, 1995]. In Übereinstimmung mit den CEPH-Daten [Cohen *et al.*, 1993; Chumakov *et al.*, 1995] wurden sowohl D1S305 als auch D1S442 und D1S498 auf den entsprechenden YACs nachgewiesen (Abb. 14, Tab. 8B). Darüberhinaus wurden D1S442 und D1S498 auf drei weiteren YAC-Klonen (789\_f\_2, 811\_c\_12 und 856\_b\_8) gemeinsam identifiziert. Diese Verbindung und die Lokalisierung von D1S498 zusammen mit *S100A10* und *THH* auf YAC 764\_a\_1 legte die Reihenfolge D1S442 – D1S498 – *S100A10* – *THH* fest und fügte dem Contig sechs weitere YACs hinzu. Auf der anderen Seite des EDC stellte YAC 950\_e\_2 die Verbindung zwischen den *S100*-Genen und D1S305 her, wodurch das Contig nochmals um sechs YACs erweitert wurde. Auch die früheren Ergebnisse für D1S1664, D1S2346 und D1S2347 konnten durch die PCR-Analysen bestätigt werden. Darüberhinaus identifizierte D1S1664 drei (Abb. 14),

D1S2346 zwei und D1S2347 fünf weitere Klone des Contigs (Tab. 8B). D1S1664 konnte zwischen den Loci *FLG* und *IVL* kartiert werden und D1S2346 in unmittelbarer Nähe von *LOR*. Die Position von D1S2347 proximal von D1S498 basierte nur auf dem schwach positiven PCR-Ergebnis mit YAC 789\_f\_2. D1S2345 wurde erstmals auf YACs lokalisiert und lag zwischen D1S498 und *S100A10* (Abb. 15). Damit waren alle sieben genetischen Marker in die physikalische Karte integriert. Das Contig umfaßte schließlich 24 YAC-Klone, die ca. 6 Mb der Region 1q21 abdeckten.

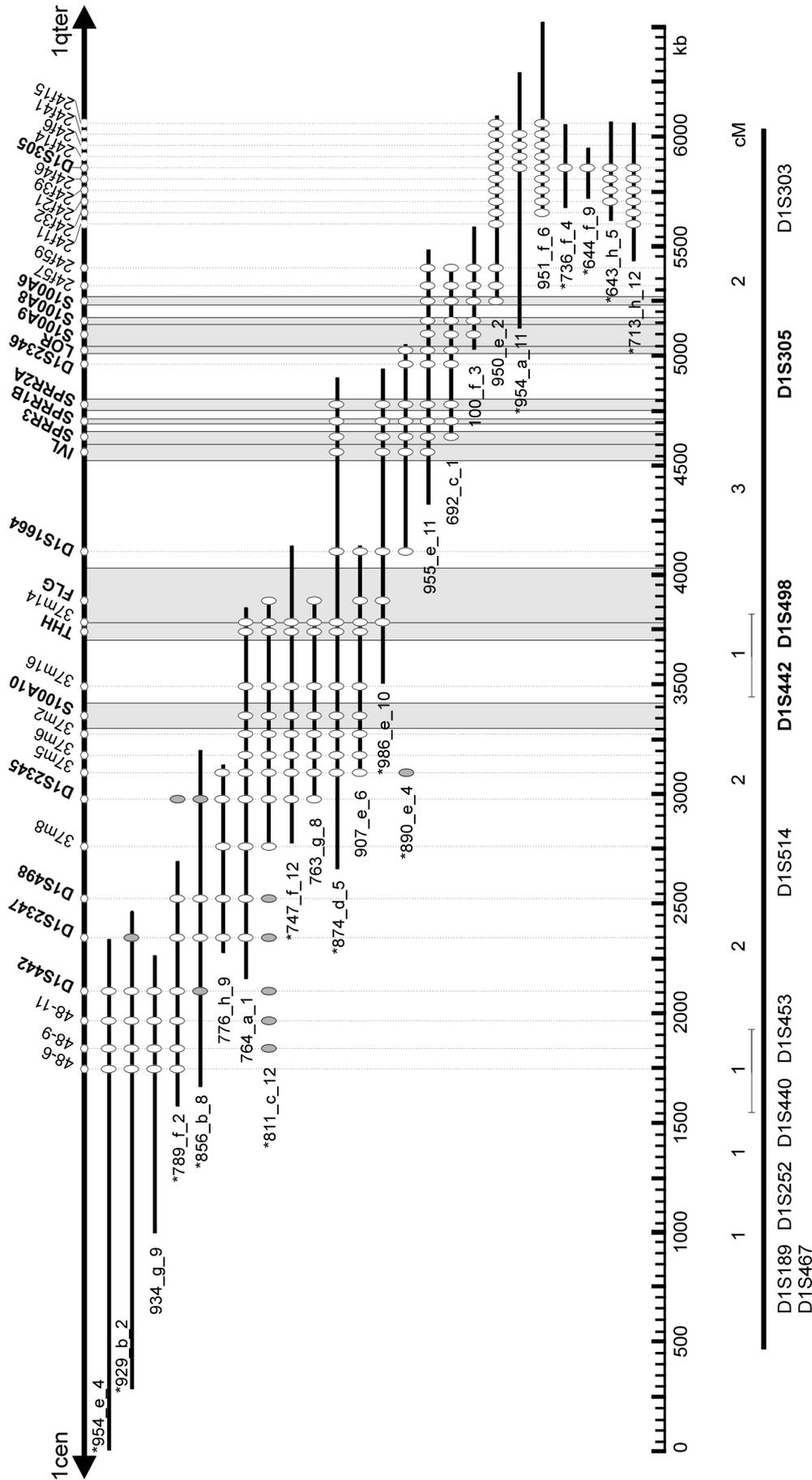
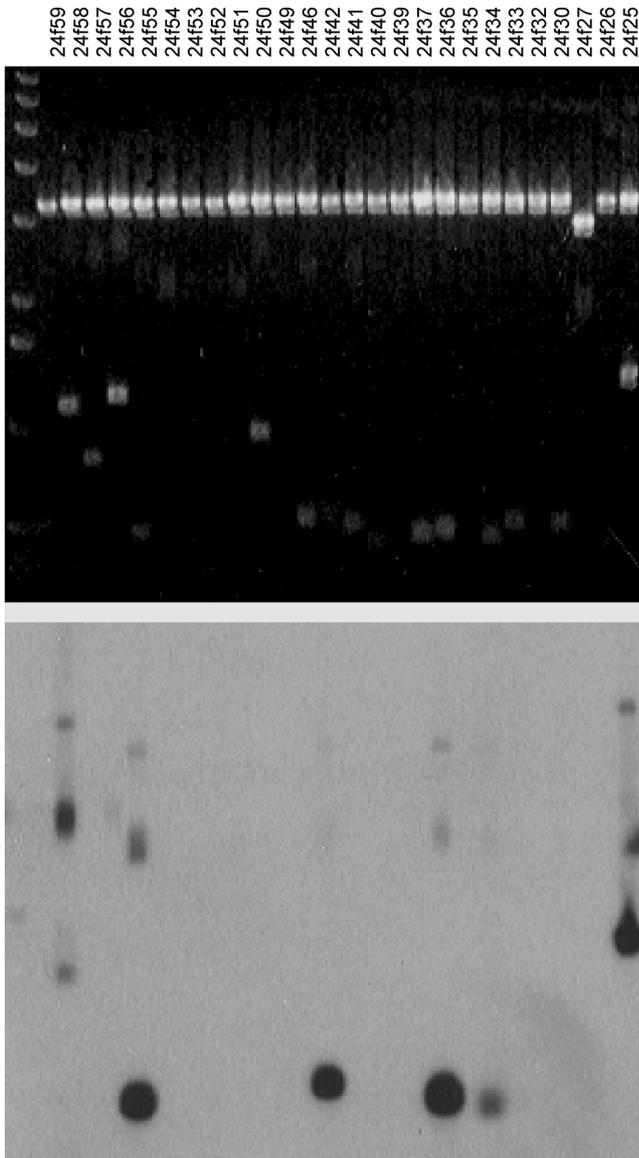


Abbildung 15: YAC-Contig und integrierte Karte der Region 1q21. Die 24 YACs der CEPH-Mega-YAC-Bibliothek, die das Contig der Region 1q21 bilden, sind dargestellt. Die Länge der horizontalen Balken entspricht den durch ROFE ermittelten Größen der YACs (s. Tab. 8B). Enthält eine YAC-Kultur mehrere künstliche Hefechromosomen, wurde das größte berücksichtigt. Graue Flächen geben die Positionen der entsprechenden Gene auf der genomischen Restriktionskarte des EDC [Volz *et al.*, 1993] an, vertikale Linien die relativen Positionen der STS-Marker und der zusätzlich generierten Hybridisierungsmarker. Weiße Ovale kennzeichnen ein starkes Hybridisierungs- bzw. PCR-Signal des entsprechenden Markers mit dem YAC, graue ein schwaches Signal, wahrscheinlich zurückzuführen auf einen zweiten, unterrepräsentierten YAC in der Originalkultur. Rearrangede YACs (gemäß den Daten des CEPH, des *Whitehead Institute/MIT* und den experimentell ermittelten) sind mit einem Sternchen gekennzeichnet. Die Gene *S100A1* bis *S100A9* wurden bereits zuvor auf YAC 100\_f\_3 nachgewiesen [Schafer *et al.*, 1995]. Unterhalb des Contigs ist der entsprechende Abschnitt der genetischen Karte von Chromosom 1 [Gyapay *et al.*, 1994] dargestellt.

### 1.6 Herstellung und Kartierung neuer Sonden zur Erhöhung der Markerdichte

Da die Anzahl der Marker abseits der bekannten EDC-Gene noch relativ gering war, wurden weitere DNA-Sonden für die Hybridisierung des YAC-Contigs generiert: DNA der YACs 764\_a\_1#2, 934\_g\_9 und 950\_e\_2#9, welche die äußeren Bereiche des Contigs abdeckten, wurde über ein ROFE-Gel von den Hefechromosomen abgetrennt, durch Elektroelution isoliert, mit Restriktionsenzymen gespalten und in Plasmidvektoren subkloniert. Nach der Transformation wurden die Plasmide in Bak-



**Abbildung 16: Detektion repetitiver Sequenzen in DNA-Sonden.** Das mit Ethidiumbromid gefärbte Agarosegel zeigt 26 Plasmidinserts, die subklonierete YAC-DNA enthalten, nach Restriktionsspaltung und Gelelektrophorese. Die 3 kb-Bande entspricht der Vektor-DNA. Das Autoradiogramm zeigt das Ergebnis der Hybridisierung des entsprechenden Southern-Blots mit <sup>32</sup>P-markierter Gesamt-DNA des Menschen. Repetitive Inserts sind an einem Hybridisierungssignal zu erkennen. Die äußere linke Bahn enthält als Größenstandard die 1 kb-Leiter.

terien vermehrt und anschließend isoliert. Um repetitive und Hefe-DNA enthaltende Plasmidinserts auszusortieren – letztere sind auf eine Verunreinigung der eingesetzten YAC-DNA zurückzuführen –, wurde der Southern-Blot eines Gels hergestellt, in dem die mit Restriktionsenzymen gespaltenen Plasmide und Inserts aufgetrennt worden waren (Abb. 16, oben). Durch zwei Hybridisierungen mit Gesamt-DNA des Menschen (Abb. 16, unten) und mit Hefe-DNA wurden die DNA-Inserts, die nicht für die Region 1q21 spezifisch waren, identifiziert. Schließlich konnten 20 isolierte Plasmidinserts als Marker für die Region 1q21 verwendet werden. Sie alle hybridisierten mit YACs des Contigs, die bereits durch Gen- oder STS-Marker der Region charakterisiert waren. Da 18 der neuen Hybridisierungsmarker auf drei oder mehr YACs nachgewiesen werden konnten und 24f11 und 24f15 immerhin noch auf zwei YACs (Tab. 10), war ihre Lokalisierung in der Region 1q21 gesichert. Benachbarte Loci des Contigs waren bis auf die Verbindung zwischen 24f59 und dem distal angrenzenden Marker weiterhin auf mindestens zwei YACs vorhanden (Abb. 15). In diesem Bereich war die Anordnung der Marker nicht eindeutig zu bestimmen, die Reihenfolge von 24f11 bis

24f15 konnte unter Beibehaltung der Anzahl rearrangierter YACs umgedreht werden. Ein ähnliches Bild ergab sich in der centromeren Region in der Nähe des Markers 37m8. Auch hier waren vermehrt rearrangierte YACs zu beobachten, so daß die Position des Markers nicht zweifelsfrei festgelegt werden konnte. Mit steigender Zahl der kartierten Loci häuften sich die Marker mit identischen Hybridisierungsmustern, bei denen eine Auflösung der Reihenfolge ebenfalls nicht möglich war.

### 1.7 Identifizierung rearrangierter YACs

Ein der genomischen Restriktionskarte oder mehreren YACs des Contigs widersprechender Markergehalt einzelner YACs sowie der YAC-Größe entgegenstehende Markerabstände ermöglichten die Identifizierung rearrangierter Klone. Bei den YACs 747\_f\_12, 986\_e\_10 und 874\_d\_5, das, wie bereits erwähnt wurde, auch eine Deletion des *FLG*-Locus aufwies, stimmten die YAC-Größen nicht mit den Entfernungen der enthaltenen Marker auf der genomischen Restriktions-

karte überein (Abb. 15) – ein Hinweis auf chimäre YACs. Das gleiche galt für die YACs 643\_h\_5, 713\_h\_12 und 954\_a\_11, deren Markergehalt den übrigen YACs der telomeren Region offenbar widersprach. Die YACs 644\_f\_9, 736\_f\_4 und 856\_b\_8 reagierten lediglich mit den STS-Markern, benachbarte Hybridisierungsmarker konnten nicht nachgewiesen werden. Die YACs 789\_f\_2, 811\_c\_12 und 890\_e\_4 zeigten mit bestimmten Markern schwach positive Ergebnisse, die mit ihrer Größe nicht vereinbar waren, z.B. YAC 811\_c\_12 (1370 kb), das den *FLG*-Locus enthielt, mit 48-6, 48-9 und D1S442, was auf eine YAC-Größe von mehr als 2000 kb schließen ließe (Abb. 15). Diese Resultate wurden auf zwei unterschiedliche Klone in der ursprünglichen YAC-Kultur zurückgeführt, von denen einer stark unterrepräsentiert war.

### 1.8 Chromosomale Orientierung des EDC

Mit der Erstellung des YAC-Contigs der Region 1q21 war bereits der indirekte Nachweis der Orientierung des EDC auf Chromosom 1 gelungen. Die Reihenfolge 1pter – D1S442/D1S498 – D1S305 –

**Tabelle 10: Neue Hybridisierungs sonden aus subklonierter YAC-DNA**

Ursprungs-YAC	Sonde	positive YACs <sup>a</sup>
	24f6	951_f_6, 954_a_11
	24f11	713_h_12
	24f14	951_f_6, 954_a_11
	24f15	951_f_6
	24f21	643_h_5, 713_h_12, 951_f_6
950_e_2#9	24f32	713_h_12, 951_f_6
	24f39	643_h_5, 713_h_12, 951_f_6
	24f41	951_f_6, 954_a_11
	24f46	643_h_5, 713_h_12, 951_f_6
	24f57	100_f_3, 692_c_1, 955_e_11
	24f59	100_f_3, 692_c_1, 955_e_11
	37m2	747_f_12, 763_g_8, 811_c_12, 874_d_5, 907_e_6
	37m5	747_f_12, 763_g_8, 776_h_9, 811_c_12, 874_d_5, 907_e_6
	37m6	747_f_12, 763_g_8, 811_c_12, 874_d_5, 907_e_6
764_a_1#2	37m8	776_h_9, 811_c_12
	37m14	747_f_12, 763_g_8, 811_c_12, 874_d_5, 986_e_10, 907_e_6
	37m16	747_f_12, 763_g_8, 811_c_12, 874_d_5, 907_e_6
	48-6	789_f_2, 929_b_2, 954_e_4
934_g_9	48-9	789_f_2, 929_b_2, 954_e_4
	48-11	789_f_2, 929_b_2, 954_e_4

<sup>a</sup> Die in allen Fällen positiven Ursprungs-YACs und schwach positive YACs wurden nicht berücksichtigt.

1qter der genetischen Karte [Gyapay *et al.*, 1994] zusammen mit der cytogenetischen Kartierung der EDC-Gene auf der chromosomalen Bande 1q21 (s. Abschnitt I 4) führte durch die Verbindung von D1S498 und *S100A10* auf YAC 764\_a\_1 sowie von *S100A6* und D1S305 auf YAC 950\_e\_2 zu der Anordnung 1cen – D1S498 – *S100A10* – *S100A6* – D1S305 – 1qter. Für den direkten Nachweis wurde ein FISH-Experiment mit zwei den EDC begrenzenden YACs durchgeführt. Durch Kopplung der YACs an unterschiedliche Fluoreszenzfarbstoffe und gleichzeitige Hybridisierung der beiden Sonden mit denaturierten Metaphase-Chromosomen (Zwei-Farben-FISH) sollte neben der Zuordnung zu einer

chromosomalen Bande die relative Anordnung der YACs auf Chromosom 1 bestimmt werden. Als Hybridisierungssonden wurden die YACs 764\_a\_1 und 950\_e\_2 eingesetzt, die ca. 1,5 Mb auseinander lagen und die die äußeren Gene des EDC in Verbindung mit den jeweils benachbarten genetischen Markern enthielten (Abb. 15). Beide YACs waren offenbar nicht rearrangiert, zählten aber zu den instabilen Klonen. Deshalb wurden zwei aus den ursprünglichen YAC-Kulturen isolierte



**Abbildung 17: Zwei-Farben-FISH zur Bestimmung der Orientierung des EDC auf Chromosom 1.** Dargestellt ist die *in situ*-Hybridisierung zweier YACs des EDC-Contigs mit Metaphase-Chromosomen unter dem Fluoreszenzmikroskop. YAC 764\_a\_1#2 ist mit Digoxigenin markiert (rot), YAC 950\_e\_2#9 mit Biotin (grün). Zwei von 14 Chromosomen, auf denen YAC 950\_e\_2#9 telomer von YAC 764\_a\_1#2 in der Region 1q21.3 kartiert, sind abgebildet. Angrenzend ist das Heterochromatin der Region 1q12 an einer breiten hellen Bande zu erkennen.

Klone verwendet: 764\_a\_1#2 und 950\_e\_2#9. Das FISH-Ergebnis bestätigte die aus dem YAC-Contig hervorgehende Anordnung: Sowohl 764\_a\_1#2 als auch 950\_e\_2#9 kartierten ausschließlich in der Region 1q21.3 (Abb. 17); auf zwölf von 14 Chromosomen hybridisierte 764\_a\_1#2 (mit D1S498, *S100A10* und *THH*) proximal von 950\_e\_2#9 (mit *S100A6* und D1S305); nur zwei Chromosomen zeigten keine eindeutige Auflösung der Hybridisierungssignale; nicht ein Chromosom mit umgekehrter Anordnung war zu erkennen.

## 1.9 Diskussion

Das in diesem Abschnitt beschriebene YAC-Contig wurde 1996 erstellt und enthielt nicht nur alle 24 zum damaligen Zeitpunkt kartierten Gene des EDC in überlappenden DNA-Klonen, sondern auch ca. 6 Mb der chromosomalen Bande 1q21, die durch sieben STS- und 20 neue Hybridisierungsmarker charakterisiert waren – eine Version mit vier STS- und zehn 1q21-spezifischen Hybridisierungsmarkern wurde im selben Jahr veröffentlicht. Alle Gene waren mindestens dreifach abgedeckt, so daß eine verlässliche Analyse der Region zu erwarten war.

Frühere Untersuchungen hatten auf wenigen separaten YACs basiert, die aus der Region 1q21 bekannt waren und die entweder *SPRR*- oder *S100*-Loci enthielten [Hardas & Elder, 1992; Hardas *et al.*, 1994; Schafer *et al.*, 1995]. Dabei hatte die Gefahr bestanden, daß chimäre YACs unentdeckt blieben und irreführende Resultate lieferten, was eine Veröffentlichung bestätigte, in der die Gene *S100A6* und *CRP* (C-reaktives Protein-Gen) in einem Abstand von 450 kb auf Chromosom 1q21 lokalisiert wurden [Hardas *et al.*, 1994]. Diese Entfernung war offensichtlich das Ergebnis eines rearrangierten YACs, da sie weder durch die genomische Restriktionskarte des EDC [Volz *et al.*, 1993] noch mit Hilfe der YACs des Contigs, die beide ca. 900 kb distal von *S100A6* abdeckten, verifiziert werden konnte – und auch nicht durch die erweiterte integrierte Karte des EDC (s. Abschnitt III 3). Eine größerer Distanz der beiden Gene ergab sich auch aus der Anordnung 1cen – *S100A6* – *D1S305* – 1qter im Contig in Verbindung mit der genetischen Karte des CHLC [Murray *et al.*, 1994], auf der *CRP* 9 cM distal von *D1S305* lokalisiert war, was einer ungefähren physikalischen Entfernung von 9 Mb entspricht.

Dagegen ermöglichte das Contig die Identifizierung chimärer YACs, deren Status bereits teilweise verifiziert werden konnte [Lioumi *et al.*, 1998]. Unter Berücksichtigung der Umlagerungen zeigte die Positionierung der YACs mit EDC-Genen keinen Widerspruch zur genomischen Restriktionskarte. Zwei Abschnitte des Contigs, welche die bekannten Gene flankierten, stellten sich allerdings als instabil heraus, erkennbar an einer Häufung rearrangierter YAC-Klone in der Umgebung von 37m8 sowie an der auf einem einzigen YAC nachgewiesenen Verbindung zwischen 24f59 und dem benachbarten distalen Marker. Bei der späteren Genkartierung konnten diese Bereiche, die vermutlich das Wachstum der Hefen beeinträchtigten und deshalb eliminiert wurden, noch eingengt und genauer charakterisiert werden (s. Abschnitt III 3).

Durch die physikalische Kartierung der STS-Marker *D1S305*, *D1S442*, *D1S498* und *D1S1664* stellte das YAC-Contig die erste integrierte Karte der Region 1q21 dar. Die Anordnung der Loci auf der genetischen Karte 1cen – *D1S442/D1S498* – *D1S305* – 1qter [Gyapay *et al.*, 1994] wurde durch Hybridisierung zweier die entsprechenden Marker enthaltenden YACs *in situ* nachgewiesen. Außerdem konnte die Reihenfolge 1cen – *D1S442* – *D1S498* – 1qter mit Hilfe des Contigs festgelegt werden, in Übereinstimmung mit kurz zuvor veröffentlichten Daten [Hoggard *et al.*, 1995], und mit der Lokalisierung von *D1S1664* konnte eine Lücke von 3 cM zwischen *D1S498* und *D1S305* geschlossen werden. Zusätzlich wurden *D1S2345*, *D1S2346* und *D1S2347* innerhalb des Contigs kartiert, so daß sieben STS-Marker für Kopplungsanalysen der Region zur Verfügung standen, deren physikalische Positio-

nen und deren benachbarte Gene bekannt waren. Ein Vergleich genetischer und physikalischer Daten zeigte zwar eine gute Übereinstimmung (3 cM zwischen D1S305 und D1S498 entsprachen ca. 3 Mb, 1 cM zwischen D1S498 und D1S442 ca. 1 Mb), erforderte jedoch aufgrund der instabilen YACs noch eine Verifizierung. Eine weitere Eingrenzung der STS-Loci wurde dann durch die Lokalisierung zusätzlicher Gene auf der erweiterten integrierten Karte des EDC erreicht (s. Abschnitt III 3).

Mit der Fluoreszenz-*in situ*-Hybridisierung wurde neben der Anordnung der STS-Marker die chromosomale Orientierung des EDC direkt nachgewiesen, da die eingesetzten YACs auch EDC-Gene enthielten. Eine frühere Arbeit, die zu dem gleichen Ergebnis gelangt war, hatte dieses aus einem chimären YAC hergeleitet und war daher nicht aussagekräftig [Hardas *et al.*, 1994] (s.o.).

Schließlich stellte das YAC-Contig eine Verbindung zwischen drei Contigs der physikalischen Karte des *Whitehead Institute/MIT* von Chromosom 1 [Hudson *et al.*, 1995] (*Data Release 8*, 1995) her. Diese beinhalteten die STS-Marker von D1S250 bis D1S442 (WC1.26), D1S498 und D1S1664 (WC1.27) sowie D1S305 (WC1.28). Damit war ein Contig etabliert, das 11 cM zwischen D1S250 auf Chromosom 1p und D1S305 in der Region 1q21 einschließlich des Centromers umfaßte. Mittlerweile wurde die Anordnung bestätigt: Das Contig WC1.16 (*singly linked*) der abschließenden Version des *Whitehead Institute/MIT* (*Data Release 12*, 1997), das sich über 66 cM der genetischen Karte [Dib *et al.*, 1996], von D1S239 bis D1S422, erstreckt, enthält die Marker D1S442, D1S498, D1S2347, D1S1664, D1S2346 und D1S305 in der angegebenen Reihenfolge, wobei D1S498 und D1S2347 auch hier nicht eindeutig angeordnet werden konnten.

## 2 Identifizierung neuer EDC-Gene durch subtraktive Hybridisierung

Da genomische DNA-Klone für nahezu alle molekulargenetischen Untersuchungen unverzichtbar sind, wie z.B. für die Identifizierung und Kartierung neuer Gene, für die Aufklärung der Genstruktur sowie für die Auswertung genetischer Kopplungsanalysen auf molekularer Ebene, war mit dem YAC-Contig die entscheidende Grundlage für die weitere Charakterisierung des EDC geschaffen. Eine zweite Ressource stand in Form einer Keratinozyten-cDNA-Bibliothek zur Verfügung (C. Backendorf, Leiden), mit deren Hilfe die in der Epidermis exprimierten Gene identifiziert und die dazugehörigen Transkripte analysiert werden sollten. Die zugrundeliegende RNA wurde aus primären Keratinozyten des Menschen isoliert, die *in vitro* bei unterschiedlichen  $\text{Ca}^{2+}$ -Konzentrationen kultiviert wurden, um verschiedene Differenzierungsstadien der Zellen zu erfassen [Fischer *et al.*, 1996]. Als Startpunkt für die Synthese des ersten cDNA-Strangs diente ein Oligo(dT)-Primer; die Komplexität der gesamten cDNA-Bibliothek betrug ca.  $4 \times 10^6$  Klone. Um sie optimal zu nutzen, wurde aus 184320 Klonen eine feingerasterte cDNA-Bibliothek (*gridded cDNA library*) hergestellt (s. Abschnitt II 19), die gitterförmig, in hoher Dichte auf zehn Nylonmembranen übertragen wurde (*high density filters*). Das Filter-Set

konnte anschließend für Hybridisierungen verwendet werden; die den einzelnen Klonen zugeordneten Koordinaten ermöglichten den direkten Vergleich mehrerer Experimente.

## 2.1 Eine direkte Methode zur Identifizierung neuer Gene

Mit den YACs und der feingerasterten cDNA-Bibliothek waren sowohl Klone genomischer DNA der EDC-Region verfügbar als auch cDNA-Klone, welche die in differenzierten Keratinozyten transkribierten Sequenzen enthielten. Somit konnten bis auf die Zoo-Blot-Hybridisierung alle in der Einleitung beschriebenen Verfahren zur Identifizierung neuer Gene (s. Abschnitt I 10) angewandt werden. Da das primäre Ziel in der Identifizierung neuer EDC-Gene bestand, d.h. von Genen, die während der Differenzierung der Epidermis in Keratinozyten exprimiert werden und gleichzeitig innerhalb des EDC kartieren, wurde eine Methode gewählt, die beide Ressourcen kombinierte: Die Hybridisierung der feingerasterten Keratinozyten-cDNA-Bibliothek mit einer YAC-Sonde wurde als die Methode erachtet, welche die vorhandenen Mittel am konsequentesten nutzte.

Allerdings gab es Bedenken gegen eine effiziente Durchführung dieser Methode. So hatten frühere Hybridisierungen mit YACs große Probleme mit falsch positiven Ergebnissen offenbart, die auf chimäre YACs und auf repetitive Sequenzen in den hochkomplexen Sonden zurückzuführen waren [Elvin *et al.*, 1990; Geraghty *et al.*, 1993; Boultonwood *et al.*, 1997]. Außerdem bestand bei größeren Sonden die Gefahr, daß die spezifische Aktivität eines einzelnen Gens im Verhältnis zur Aktivität der gesamten markierten DNA nicht ausreichte, um ein eindeutiges Hybridisierungssignal zu erzeugen. Aufgrund der Vorarbeiten wurde das Experiment dennoch mit berechtigter Aussicht auf Erfolg angegangen. Die YACs waren durch das Contig bereits ausgiebig charakterisiert – chimäre Klone sollten demnach erkennbar sein und nicht als Sonde verwendet werden. Die feingerasterte cDNA-Bibliothek versprach eine erhöhte Sensitivität bei der Autoradiographie, da die Filter auch bei langer Expositionszeit der Filme eine eindeutige Abgrenzung der gitterförmig angeordneten, doppelt aufgetragenen Klone von unspezifischen Hintergrundsignalen ermöglichen sollten. Schließlich sollten repetitive Sequenzen durch kompetitive Hybridisierung zu maskieren sein [Sealey *et al.*, 1985].

## 2.2 Hybridisierung der feingerasterten cDNA-Bibliothek mit einem YAC

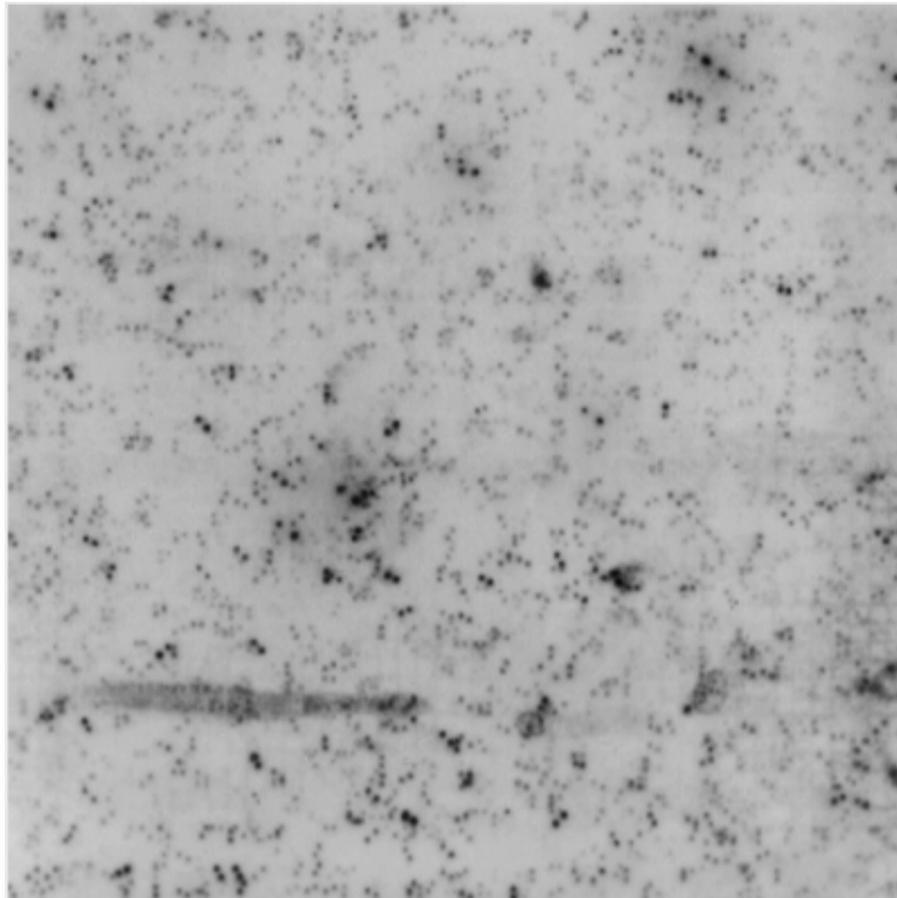
Als Sonde für die Hybridisierung der zehn feingerasterten Koloniefilter, die zusammen 184320 Klone der Keratinozyten-cDNA-Bibliothek enthielten, wurde YAC 986\_e\_10 (1440 kb) ausgewählt (Abb. 15), weil es einen relativ großen Bereich von 400 kb zwischen *FLG* und *IVL* einschloß, in dem noch kein Gen kartiert war, und weil es die Gene *FLG*, *IVL* sowie die *SPRR*-Gene enthielt, die als Differenzierungsmarker in Keratinozyten als Positivkontrolle fungierten. Da YAC 986\_e\_10 das einzige war, das diese Voraussetzungen erfüllte, wurde ein chimäres centromeres Ende akzeptiert. Für die Präparation der Sonde wurde das künstliche Hefechromosom durch ROFE abgetrennt, aus dem Gel eluiert und die YAC-DNA nach Konzentrierung und Gelfiltration radioaktiv markiert. Um unspezifische Signale durch repetitive Elemente auszublenden, wurde die denaturierte Sonde anschließend mit ei-

nem hohen Überschuß (330 µg) an ebenfalls denaturierter Gesamt-DNA des Menschen versetzt und 30 Minuten renaturiert. Darauf folgte die Hybridisierung der cDNA-Bibliothek.

Da fast die Hälfte des menschlichen Genoms aus repetitiven DNA-Sequenzen besteht, die zu einem geringeren Anteil auch in RNA transkribiert werden, mußten diese in der hochkomplexen Sonden vor der Hybridisierung maskiert werden. Dabei wurde ausgenutzt, daß die Renaturierungsgeschwindigkeit einer bestimmten DNA-Sequenz von ihrer Konzentration abhängt: Repetitive Sequenzen der Sonde sollten mit den komplementären Sequenzen der zugesetzten genomischen DNA schneller renaturieren als die in der Regel nur einmal im haploiden Genom vorkommenden genspezifischen Sequenzen und somit nicht mehr für die Hybridisierung der cDNA-Bibliothek zur Verfügung stehen.

### 2.3 Auswertung der Hybridisierung

Trotz der kompetitiven Vorgehensweise wurden auf den zehn Filtern ca. 12000 cDNA-Klone identifiziert (Abb. 18). Von diesen wurden 50 Klone zufällig ausgewählt und ihre Plasmidinserts sequenziert, ca. 600 bp von jeder Seite. Durch Vergleich der cDNA-Sequenzen mit bekannten Nucleotidsequenzen in Datenbanken wurden sie charakterisiert (s. Abschnitt III 4). Ausgewählte Inserts wurden zusätzlich für eine Hybridisierung mit den YACs des Contigs verwendet, um eine mögliche Lokalisierung der entsprechenden Gene im EDC festzustellen (s. Abschnitt III 3.1).



**Abbildung 18:** Hybridisierung der feingerasterten cDNA-Bibliothek mit einem YAC. Das Autoradiogramm zeigt einen von zehn Filtern mit 18432 doppelt aufgetragenen cDNA-Klonen der feingerasterten cDNA-Bibliothek nach der Hybridisierung mit der <sup>32</sup>P-markierten DNA von YAC 986\_e\_10. Positive Klone sind an zwei Hybridisierungssignalen, die nach einem bestimmten Muster angeordnet sind, zu erkennen.

Wie aufgrund der hohen Zahl positiver Signale zu erwarten war, offenbarte die Sequenzanalyse in mehr als 70% der cDNA-Inserts repetitive Sequenzen, die als diverse kurze und lange eingestreute DNA-Elemente (SINEs und LINEs, *short interspersed elements* und *long interspersed elements*) iden-

tifiziert wurden (Tab. 11). Zehn dieser Inserts wurden als Sonden für die Hybridisierung der YAC-Filter verwendet: Fünf enthielten die gesamte Insertsequenz und fünf nach Restriktionsspaltung nur den nicht repetitiven Anteil. Sie hybridisierten entweder mit mehreren Restriktionsfragmenten in allen YACs, zurückzuführen auf eine repetitive Sonde (Abb. 19A), oder mit keinem einzigen Fragment; die entsprechenden Gene lagen somit nicht in der Region 1q21. Auf eine weitere Analyse repetitiver cDNA-Sequenzen wurde daraufhin verzichtet, da sie offensichtlich aufgrund ihres unspezifischen Anteils nachgewiesen worden waren.

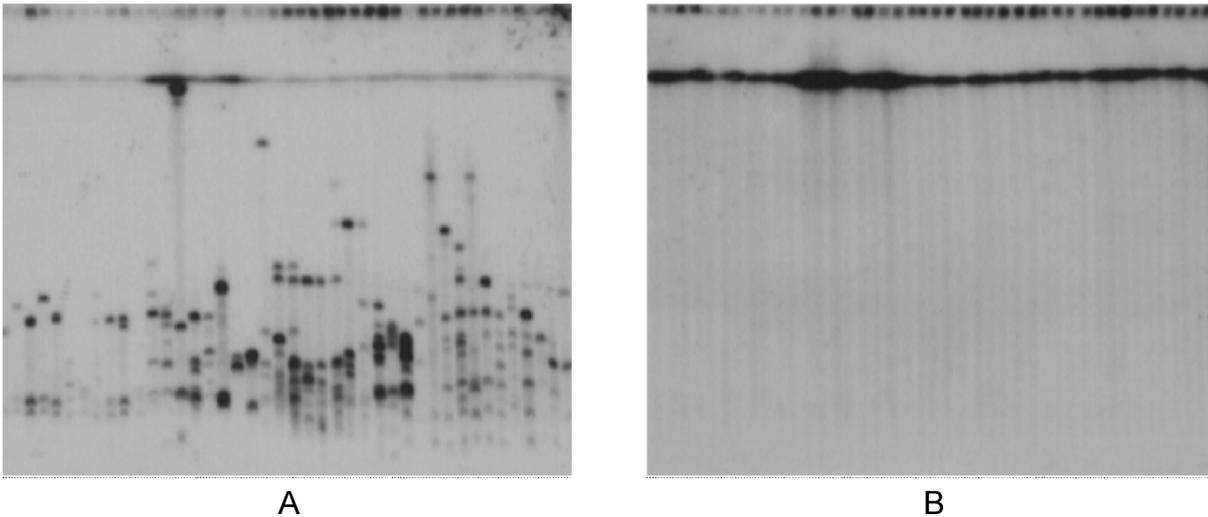
Das cDNA-Insert eines Klons stammte von ribosomaler RNA (rRNA) ab. Es konnte ebenfalls nicht auf den YACs lokalisiert werden, sondern gab starke Hybridisierungssignale mit Restriktionsfragmenten aus dem Hefegenom (Abb. 19B). Ribosomale RNA codierende Gene, die auf Chromosom 1 des Menschen lediglich in den Regionen 1q42 [Sorensen *et al.*, 1991] und 1q31 [Lomholt *et al.*, 1995] nachgewiesen wurden, sind unter Eukaryonten stark konserviert. Die Hybridisierung der Sonde mit den entsprechenden cDNA-Sequenzen wurde höchstwahrscheinlich von einer Verunreinigung mit Hefe-DNA verursacht, die bei der Isolierung des YACs nicht vollständig abgetrennt werden konnte.

Schließlich entsprachen acht cDNA-Klone den transkribierten Sequenzen bekannter Gene. Es handelte sich um die EDC-Gene

**Tabelle 11: Charakterisierung 50 ausgewählter, von YAC 986\_e\_10 identifizierter Klone**

Klon	EcoRI/XhoI-Insert-fragmente in bp <sup>a</sup>	Sequenz	Lokalisierung in 1q21
1003j24	1800	SINE-R11	n.d.
1025a 5	1200	Alu	n.d.
1025e 2	650	SPRR1A	+
1026b 9	<b>1600</b>	Alu	-
1026c 7	2200	Alu	n.d.
1026g 3	500, 1800	LINE-1	n.d.
1026j 1	700	SPRR2B	+
1105b 6	100, 550	SPRR1B	+
1126h 8	500, 800, <b>1300</b>	28S-rRNA	-
1147n18	4000	LTR5	n.d.
1154b 3	<b>1700</b>	SINE-R11	-
1160m16	2000	LTR5	n.d.
1168g 6	<b>1300</b>	GADPH	-
1175i 3	1400	LTR5	n.d.
1192o10	600	SPRR1A	+
1208j13	200, 550, 1600	LTR5	n.d.
1213k23	2300	LTR5	n.d.
1215m 3	350	Alu	n.d.
1220e19	<b>1050, 1700</b>	Plakophilin 1	-
3024i17	900	LINE-1	n.d.
3029d19	1550	LINE-1	n.d.
3029n15	2100	Alu	n.d.
3041o 4	<b>1700</b>	Alu	-
3042e15	800	SPRR2A	+
3042e22	800	NICE1	+
3042g19	600	SPRR2B	+
3046k11	750	SPRR2B	+
3052k14	750	Alu	n.d.
3065c24	<b>350, 1600</b>	Alu/AK000095	-
3065f23	350	Alu	n.d.
3067d 8	1450	Alu	n.d.
3069d13	2000	Alu	n.d.
3076n13	1800	Alu	n.d.
3077i11	350, 1150	Alu	n.d.
3089j17	<b>300, 1200</b>	Alu/AK024675	-
3093l 4	300, <b>1150</b>	Alu	-
3100b 8	<b>550</b>	VDAP2	-
3103d11	1000	Alu	n.d.
3107h16	<b>1250 (800, 450)</b>	Alu/KANGAI 1	-
3109k 8	150, 300	SPRR1B	+
3157d 6	2800	Alu	n.d.
3162e 1	1150	Alu	n.d.
3165c24	450, 750, 1900	Alu	n.d.
3183i 4	2150	Alu	n.d.
3200f 5	<b>2350</b>	Alu	-
3203g 6	<b>1950</b>	CD24 Antigen	-
3205a 4	1050	Alu	n.d.
3222a19	<b>400, 600, 1500</b>	LINE-1/ AW500154	-
3226p 8	1100	Alu	n.d.
3235j11	<b>250, 400, 500</b>	Alu/ Antileukoprotease	-

<sup>a</sup> Fett gedruckte Größen kennzeichnen die als Hybridisierungssonden verwendeten Fragmente, die Fragmente in Klammern wurden durch Restriktionsspaltung mit den Enzymen EcoRI und PvuII erhalten; n.d., nicht durchgeführt; LTR, *long terminal repeat*; GADPH, Glycerinaldehyd-3-phosphatdehydrogenase; VDAP2, *voltage-dependent anion channel 2*.



**Abbildung 19: Unspezifische Hybridisierung repetitiver und ribosomaler cDNA-Sonden mit einem YAC-Blot.** Die Autoradiogramme zeigen die Hybridisierungsergebnisse der aus den cDNA-Klonen 3065c12 und 3235j11 generierten Sonden mit einem YAC-Blot, der die mit *Sall* fragmentierte DNA der YACs des Contigs enthält. (A) Das <sup>32</sup>P-markierte, repetitive 250 bp-Insertfragment aus Klon 3235j11 hybridisiert mit allen genomischen Restriktionsfragmenten, die das gleiche repetitive Element enthalten. (B) Das ribosomale 1500 bp-Insertfragments aus Klon 3065c12 hybridisiert mit homologen ribosomalen DNA-Sequenzen des Hefegenoms, das bei den gewählten ROFE-Bedingungen nicht aufgetrennt wurde.

*SPRR1A*, *SPRR1B*, *SPRR2A* und *SPRR2B*. Darüberhinaus wurde die cDNA-Sequenz eines bisher unbekanntes Gens identifiziert (*NICE1*, von „neu identifizierte cDNA des EDC“), das auf dem als Sonde eingesetzten YAC sowie auf drei weiteren YACs nachgewiesen werden konnte (s. Abschnitt III 3.1). Damit wurde einerseits gezeigt, daß die Methode geeignet war, die in einem YAC enthaltenen Gene zu identifizieren; andererseits wurde durch die *SPRR*-Sequenzen bestätigt, daß die aus Keratinozyten gewonnene cDNA-Bibliothek Transkripte enthielt, die spät während der Differenzierung der Epidermis exprimiert werden.

Da der Anteil der EDC-spezifischen Klone unter den analysierten jedoch relativ gering war, wurde eine zweite Hybridisierung der feingerasterten Filter mit der gleichen YAC-Sonde, aber unter veränderten Kompetitionsbedingungen durchgeführt. Durch Konzentrationserhöhung der zur Maskierung zugesetzten Gesamt-DNA des Menschen auf 1 mg und Verlängerung der Renaturierungszeit bei der Vorhybridisierung auf eine Stunde sollten bei gleichbleibender Konzentration der markierten YAC-Sonde die von repetitiven Sequenzen verursachten Signale möglichst vollständig unterdrückt werden. Das Ergebnis war zwiespältig. Obwohl bei den bereits analysierten cDNA-Klonen eine leichte Verschiebung der Signalintensität hin zu den EDC-spezifischen Sequenzen zu beobachten war, konnten die repetitiven Sequenzen nicht ausreichend maskiert werden; erneut hybridisierte die Sonde mit ca. 11000 Klonen. Es zeigte sich, daß durch kompetitive Hybridisierung keine vollständige Maskierung der in der YAC-Sonde enthaltenen repetitiven Elemente zu erreichen war; die Signalintensität nahm lediglich ab. Da eine Vielzahl sowohl sequenzabhängiger als auch herstellungsbedingter Faktoren die Intensität der Hybridisierungssignale beeinflussen, wie der G/C-Gehalt, die Exongrößen und die Kopienzahl eines Gens sowie die Größe der synthetisierten cDNA und die cDNA-Konzentration der ein-

zelen Klone auf den Filtern, schien eine auf Signalintensitäten basierende Auswertung nicht erfolgversprechend zu sein.

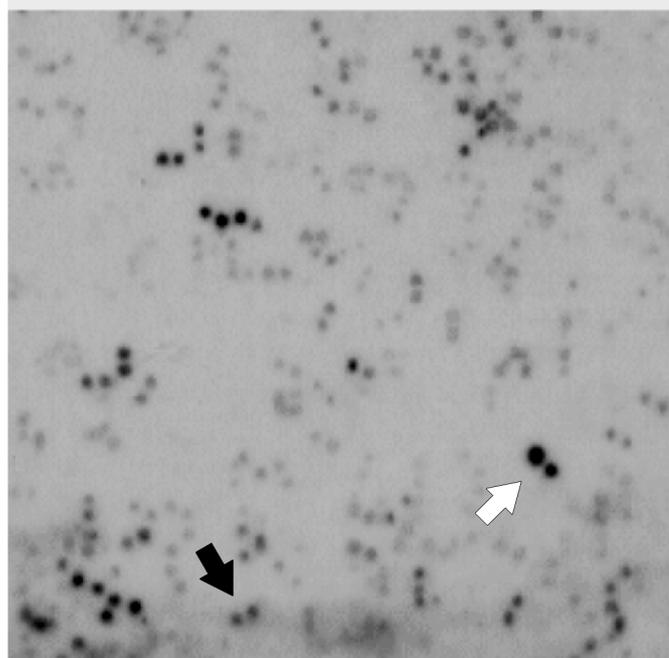
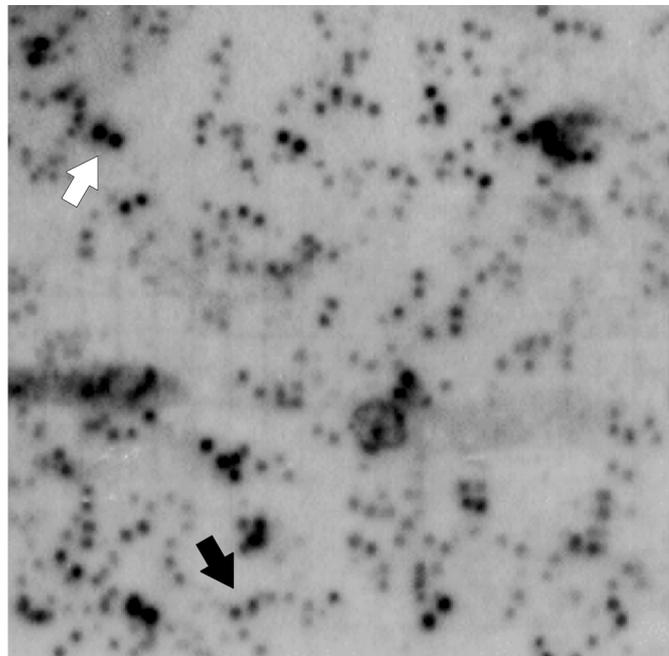
#### 2.4 Hybridisierung der feingerasterten cDNA-Bibliothek mit einem zweiten YAC

Nachdem die repetitiven Sequenzen durch kompetitive Hybridisierung nicht ausreichend maskiert werden konnten, mußten sie auf andere Weise identifiziert und die entsprechenden cDNA-Klone aussortiert werden. Dabei sollte der Umstand ausgenutzt werden, daß YACs, die im menschlichen Genom in direkter Nachbarschaft kartieren, in den enthaltenen repetitiven Elementen weitestgehend übereinstimmen. Für eine zweite Hybridisierung der zehn Filter wurde deshalb YAC 950\_e\_2#9 (690 kb) als Sonde ausgewählt, das den distalen Bereich des EDC abdeckte und mit YAC 986\_e\_10 nicht überlappte (Abb. 15). Für die Auswertung dieses Versuchs war die feingerasterte Form der cDNA-Bibliothek die entscheidende Voraussetzung. Durch den Vergleich der Hybridisierungsmuster sollten diejenigen cDNA-Klone, die, zurückzuführen auf unspezifische Hybridisierung, von beiden YACs identifiziert wurden, als repetitiv zu erkennen sein. Im Gegensatz dazu sollten Klone, die entweder mit dem einen oder mit dem anderen YAC hybridisierten, für den EDC spezifische Gensequenzen enthalten.

#### 2.5 Subtraktive Auswertung

Zwar wurden auch von YAC 950\_e\_2#9 ca. 10000 Klone detektiert; allerdings waren

986\_e\_10



959\_e\_2#9

**Abbildung 20:** Subtraktive Auswertung der Hybridisierungsergebnisse zweier YACs. Die Autoradiogramme zeigen jeweils das gleiche Sechstel eines Filters der feingerasterten cDNA-Bibliothek, der nacheinander mit  $^{32}\text{P}$ -markierter DNA der angegebenen YACs hybridisiert wurde. Die schwarzen Pfeile zeigen einen Klon, der aussortiert wurde, die weißen je einen potentiell EDC-spezifischen.

jetzt diejenigen cDNA-Klone, welche bereits mit dem ersten YAC hybridisiert hatten, durch Vergleich der Autoradiogramme leicht zu erkennen (Abb. 20); sie wurden nicht weiter berücksichtigt. Schließlich blieben 744 der von YAC 950\_e\_2#9 identifizierten Klone übrig, die idealerweise von EDC-Genen stammen sollten; 50 von ihnen wurden sequenziert.

Wie bei der ersten Hybridisierung wurden cDNA-Sequenzen der etablierten EDC-Gene, die auf dem entsprechenden YAC lagen und in Keratinozyten exprimiert wurden, wiedergefunden; in diesem Fall

**Tabelle 12: Vergleich der nicht subtraktiven mit der subtraktiven Auswertung**

	nicht subtraktiv	subtraktiv	
	YAC 986_e_10	YAC 986_e_10	YAC 950_e_2#9
detektiert	~12000	~12000	~10000
ausgewählt	~12000	2091	744
sequenziert	50	50	50
cDNA-Klone			
repetitiv	36	3	-
ribosomal	1	3	-
EDC-spezifisch	9	40	45
andere	4	4	5
-----			
etablierte EDC-Gene	4	6	3
neu identifizierte EDC-Gene	1	1	9
EDC-spezifische Klone unter den ausgewählten	15%	85%	92%

waren das *S100A4*, *S100A6* und *S100A7*. Gleichzeitig nahm die Anzahl repetitiver und ribosomaler cDNA-Inserts ab; nicht ein einziges war unter den sequenzierten zu finden (Tab. 12, rechte Spalte). Zusätzlich wurden cDNA-Sequenzen von neun Genen identifiziert, die durch Hybridisie-

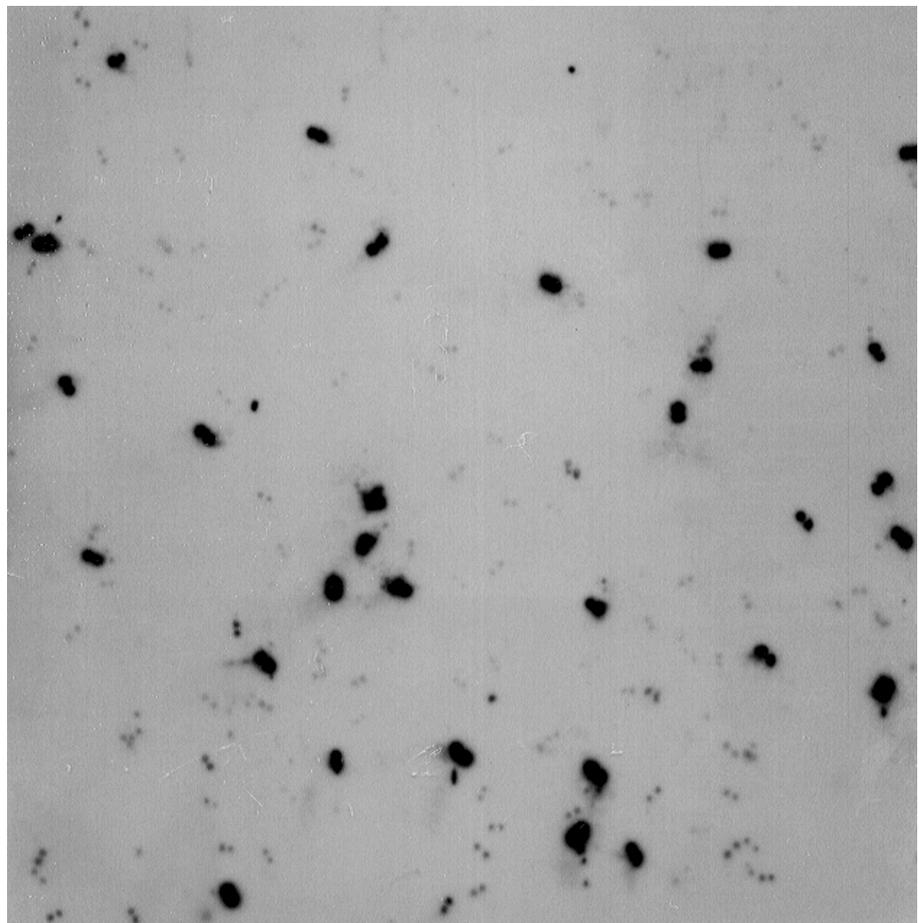
rung der entsprechenden Insertfragmente mit mehreren YACs des Contigs (s. Abschnitt III 3.1) erstmals dem EDC zugeordnet werden konnten. Sie repräsentierten *ADARI* (doppelsträngige RNA-spezifische Adenosindesaminase 1-Gen), *HAX1* (HS1-bindendes Protein X-1-Gen), *LAMRL6* (*LAMR-like 6*, Lamininrezeptor-ähnlicher Locus 6), *PSMD8L* (*PSMD8-like*, Proteasom-Untereinheit nicht ATPase 8-ähnlicher Locus), *TPM3* (Tropomyosin 3-Gen) sowie *NICE2*, *NICE3*, *NICE4* und *NICE5*.

Nach diesem überzeugenden Ergebnis wurden die Hybridisierungssignale von YAC 986\_e\_10 ein zweites Mal und diesmal subtraktiv ausgewertet. Von den 12000 cDNA-Klonen blieben nach Abzug der unspezifischen Signale 2091 als mögliche Kandidaten für den EDC übrig. Unter den 50 zusätzlich sequenzierten Klonen waren nicht nur cDNA-Sequenzen der Gene *SPRR1A*, *SPRR1B*, *SPRR2A*, *SPRR2B* und *NICE1*, sondern auch von *SPRR3* und *IVL*, die bei der ersten Auswertung nicht gefunden werden konnten (Tab. 12, mittlere und linke Spalte). Zwar stieg die Zahl der identifizierten EDC-Gene nur um zwei an, der Anteil EDC-spezifischer Transkripte unter den sequenzierten Inserts erhöhte sich jedoch von neun auf 40. Repetitive Sequenzen wurden nur noch in drei der 50 ausgewählten cDNA-Klone gefunden. Daß sie überhaupt noch auftauchten, ist wahrscheinlich auf seltener vorkommende repetitive Elemente in YAC 986\_e\_10 zurückzuführen, die in YAC 950\_e\_2#9 nicht enthalten sind.

Der Anstieg ribosomaler cDNA-Sequenzen wies darauf hin, daß lediglich die Sonde aus YAC 986\_e\_10 mit Hefe-DNA verunreinigt war.

Nur wenige der subtraktiv ausgewählten cDNA-Klone entsprachen keiner der bisher erwähnten Kategorien, repetitiv, ribosomal oder EDC-spezifisch. Bei YAC 986\_e\_10 könnte das Auftreten dieser Klone auf das chimäre centromere YAC-Ende zurückgeführt werden. Da die ausschließliche Lokalisierung von YAC 950\_e\_2#9 in der Region 1q21.3 durch FISH gesichert war (Abb. 17), mußte für die scheinbar fälschlich identifizierten Klone eine andere Erklärung gefunden werden. Folgende Ursachen konnten entweder aufgezeigt werden oder sind möglich: eine Kontamination eines Klons der gerasterten cDNA-Bibliothek mit einem stark unterrepräsentierten EDC-spezifischen Klon, der durch die hohe Sensitivität der Autoradiographie zwar nachgewiesen wurde, weitere Untersuchungen aber nicht beeinflusste; eine nicht identifizierte repetitive Sequenz im zentralen Bereich größerer Inserts, der bei der Sequenzierung nicht erreicht wurde; ein unspezifisches Hybridisieren der Sonde, verursacht durch nicht vollkommen einheitliche Bedingungen während der Hybridisierung bzw. während des Waschens der zehn Filter.

Um einen Vergleich der konventionellen mit der subtraktiven Hybridisierung zu ermöglichen, wurde der Anteil EDC-spezifischer cDNA-Sequenzen unter den detektierten bzw. subtraktiv ausgewählten Klonen bestimmt. Die von EDC-Genen stammenden cDNA-Inserts wurden radioaktiv markiert und als Sonden für die Hybridisierung der Filter der feingerasterten cDNA-Bibliothek



**Abbildung 21:** Identifizierung der die *SPRR2*-Sequenz enthaltenden cDNA-Klone. Das Autoradiogramm zeigt den Filter der feingerasterten cDNA-Bibliothek, der in Abb. 19 nach der Hybridisierung mit YAC 986\_e\_10 dargestellt ist, nach der Hybridisierung mit der  $^{32}\text{P}$ -markierten *SPRR2*-Sonde. Die Mehrzahl der schwachen Signale ist auf eine Kreuzhybridisierung der Sonde mit *SPRR1*-cDNA-Sequenzen zurückzuführen.

verwendet (Abb. 21). Die für *IVL*, *SPRR1*, *SPRR2*, *SPRR3* und *NICE1* spezifischen Sonden identifizierten 1777 cDNA-Klone. Das entspricht einem Anteil von 15% aller mit YAC 986\_e\_10 hybridisie-

renden 12000 Klone, aber 85% der 2091 subtraktiv ausgewählten Klone. YAC 950\_e\_2#9 übertraf dieses Ergebnis sogar noch: Insgesamt 686 der 744 ausgewählten cDNA-Klone wurden von den Sonden der zwölf im EDC kartierenden Gene identifiziert; das entspricht einem Anteil von 92% (Tab. 12).

**2.6 Hybridisierungsergebnisse zweier weiterer YACs aus der centromeren Region**

Durch zwei zusätzliche Hybridisierungen sollte die proximale Ausdehnung des Genkomplexes bestimmt werden. Als Sonden verwendet wurden YAC 764\_a\_1#2 (1300 kb), dessen distales Ende die bekannten EDC-Gene *S100A10* und *THH* sowie das kurz zuvor kartierte *S100A11*-Gen (s. Abschnitt III 3.1) enthielt, und YAC 934\_g\_9 (1270 kb), das die weiter centromer gelegene Region abdeckte. Den Erwartungen entsprechend waren unter den 423 Klonen, die ausschließlich von YAC 764\_a\_1#2 identifiziert wurden, cDNA-Sequenzen für *S100A10*, *S100A11* und *THH* zu finden. Darüberhinaus konnten durch Hybridisierung ausgewählter cDNA-Inserts mit den YAC-Filtern sieben Gene, *ANXA9* (Annexin A9-Gen), *PIP5K1A* (Typ I Phosphatidylinosit-4-phosphat-5-kinase  $\alpha$ -Gen), *PSMB4* (Pro-

**Tabelle 13: Subtraktive Hybridisierungsergebnisse der YACs 764\_a\_1#2 und 934\_g\_9**

	YAC 764_a_1#2	YAC 934_g_9
detektiert	8000	5000
cDNA-Klone ausgewählt	423	40
sequenziert	20	20
-----		
neue EDC-Gene	7	4

teasom-Untereinheit  $\beta$ 4-Gen), *PSMD4* (Proteasom-Untereinheit nicht-ATPase 4-Gen), *NICE6*, *NICE7* und *NICE8* erstmals dem EDC zugeordnet werden (s. Abschnitt III 3.1). Für YAC 934\_g\_9 blieben nach Abzug der unspezifisch detektierten cDNA-Klone lediglich 40 für die Sequenzierung übrig. Die Kartierung ausgewählter

cDNA-Inserts (s. Abschnitt III 3.1) identifizierte vier weitere in Keratinozyten exprimierte Gene der Region 1q21: *PIAS3* (Proteininhibitor des aktivierten Stat3-Gen), *RBM8* (RNA-bindendes Motiv Protein 8-Gen), *NICE9* und *NICE10*. Die Ergebnisse der beiden YACs sind in Tabelle 13 zusammengefaßt. Es fällt auf, daß sowohl die Anzahl der detektierten cDNA-Klone als auch die der nach Subtraktion verbleibenden Klone abnimmt. Eine verminderte Zahl an Hybridisierungssignalen könnte auf YAC-Sonden mit weniger repetitiven Elementen zurückzuführen sein; sie könnte aber auch ein Zeichen sinkender Sensitivität der Hybridisierungsfilter durch wiederholtes Verwenden sein. Interessanter scheint die zum Centromer hin stark abnehmende Zahl subtraktiv ausgewählter Klone zu sein, die auf eine im Vergleich zum restlichen EDC reduzierte Expressionsaktivität der Gene und möglicherweise auf ein Ende des Genkomplexes hinweist.

**2.7 Mögliche Fehlerquellen und ihre Vermeidung**

Trotz des positiven Ergebnisses bei der Identifizierung neuer Gene deckte die Analyse der Klone auch mögliche Probleme auf, die einerseits auf die Herstellung der feingerasterten cDNA-Bibliothek, andererseits auf die Hybridisierung mit einem YAC zurückgeführt werden konnten. Die bei der Konstruktion einer cDNA-Bibliothek unvermeidlich entstehenden chimären Sequenzen wurden in ca. 5% der

cDNA-Inserts über zwei enthaltene Poly(A)<sup>+</sup>-Schwänze bzw. über Sequenzen, die von zwei unterschiedlichen Genen stammten, nachgewiesen. Da die in einem chimären Klon enthaltenen cDNA-Sequenzen über eine Restriktionsschnittstelle miteinander verbunden sind, können sie durch Restriktionsspaltung auch wieder getrennt werden und sind dann einzeln zu charakterisieren. Ebenfalls ca. 5% der cDNA-Klone vermehrten sich im ampicillinhaltigen Medium nicht, was auf eine durch instabile Transformation verursachte Elimination des Plasmids, welches das Ampicillinresistenz-Gen enthielt, schließen läßt. Durch erneute Transformation mit aus der Originalkultur isolierter DNA sollte dieses Problem zu lösen sein. Eine andere Fehlerquelle war durch den Herstellungsprozeß der feingerasterten cDNA-Bibliothek bedingt. Beim Auflesen wurden aus zwei Klonen bestehende Kolonien vom eingesetzten Roboter nicht immer als solche erkannt; ca. 5% der cDNA-Klone waren auf diese Weise kontaminiert. Der während der Herstellung der cDNA-Bibliothek visuell ermittelte Wert wurde durch zwei Plasmide im Isolat entsprechender Klone, die bei der weiteren Analyse Extrabanden verursachten, bestätigt. Durch einen zusätzlichen Trennungsschritt sind auch diese Sequenzen problemlos zu charakterisieren. Die von chimären Sequenzen eines YACs identifizierten Klone werden spätestens dann entdeckt, wenn die Kartierung in der entsprechenden Region, die für alle cDNA-Inserts obligatorisch ist, nicht gelingt. Das gleiche gilt für falsch positive Klone aufgrund in der Sonde enthaltener repetitiver bzw. ribosomaler Sequenzen.

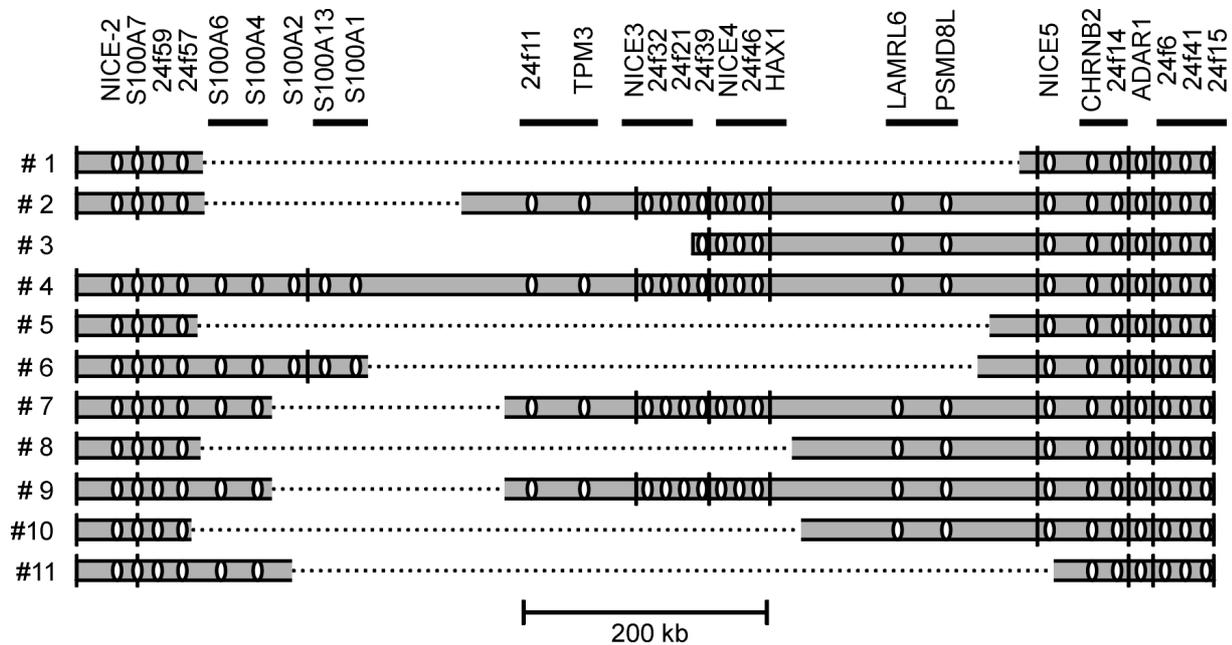
### 3 Kartierung der neuen EDC-Gene

Da die Region 1q21 des menschlichen Genoms relativ häufig an chromosomalen Umlagerungen beteiligt ist [Hoggard *et al.*, 1995; Weterman *et al.*, 1996; Forus *et al.*, 1998] und das Zusammensetzen des Contigs mehrere rearrangierte YACs erkennen ließ, wurde die Kartierung im YAC-Contig mit der genomischen Restriktionskartierung kombiniert, um die Lokalisierung der Gene in der Region 1q21 sicherstellen und ihre Anordnung festzulegen. Dabei mußte allerdings beachtet werden, daß bei der Hybridisierung von cDNA-Sonden nicht nur die funktionellen Gene nachgewiesen werden, sondern auch mögliche Pseudogene oder andere verwandte DNA-Sequenzen. Als Hybridisierungssonden dienten die cDNA-Inserts bzw. Insertfragmente der von den YACs identifizierten Klone der feingerasterten Keratinozyten-cDNA-Bibliothek, alle bei der Konstruktion des YAC-Contigs verwendeten Marker sowie weitere spezifische Marker für *S100*-Gene und drei neu zu kartierende Gene. Die cDNA-Sonden wurden mit den Restriktionsenzymen EcoRI und XhoI aus den Plasmiden herausgeschnitten, nach elektrophoretischer Trennung über ein Agarosegel isoliert und für die Hybridisierung der Southern-Blots radioaktiv markiert. Die YAC-Filter enthielten die mit dem Restriktionsenzym Sall geschnittene DNA der 24 YACs des EDC-Contigs sowie der elf aus YAC 950\_e\_2 isolierten Klone und waren dieselben, die zur Erstellung des Contigs verwendet wurden. Ebenso waren die zur Herstellung der genomischen Filter verwendete DNA der H2LCL-Zelllinie und die für die Restriktions-

spaltung eingesetzten Enzyme bereits bei der Erstellung der ersten Restriktionskarte der Region 1q21 genutzt worden [Volz *et al.*, 1993]. Die Auftrennung der hochmolekularen Restriktionsfragmente erfolgte durch ROFE.

### 3.1 YAC-Restriktionskarte

Im Vergleich zum YAC-Contig konnte die Auflösung durch Berücksichtigung der Sall-Restriktionsfragmentgrößen (Abb. 13) und der Abbauprodukte von YAC 950\_e\_2 deutlich erhöht werden. Insbesondere die aus YAC 950\_e\_2 isolierten Kolonien 950\_e\_2#1 bis 950\_e\_2#11, die aufgrund von Deletionen einen unterschiedlichen DNA-Gehalt aufwiesen, ermöglichten eine genauere Positionierung der Marker in der distalen Region des Contigs (Abb. 22). Hybridisierten die DNA-Sonden mit mindestens zwei der 24 YACs des Contigs, war die Lokalisierung in der Region bereits



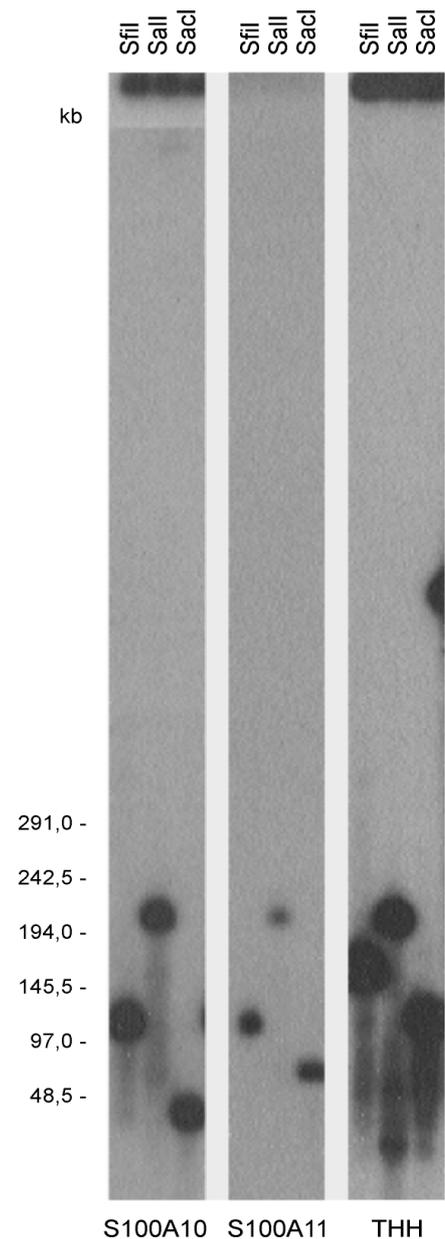
**Abbildung 22:** Restriktionskarte der aus YAC 950\_e\_2 isolierten Klone. Graue Balken repräsentieren die YAC-DNA, gestrichelte horizontale Linien deklinierte Abschnitte, schwarze vertikale Linien die mit dem Enzym Sall erhaltenen Restriktionsschnittstellen. Weiße Ovale symbolisieren die Hybridisierung der oben angegebenen Sonde mit den entsprechenden Fragmenten. Die Reihenfolge der mit einem Balken unterlegten Loci konnte nicht aufgelöst werden.

ohne Verifizierung auf der genomischen Restriktionskarte ausreichend gesichert. Wurden mehr als zwei YACs identifiziert, konnte oftmals das entsprechende intakte Sall-Restriktionsfragment von der vorherrschenden Fragmentgröße abgeleitet werden. Davon abweichende kleinere Fragmente zeigten in der Regel YAC-Enden an, konnten aber, ebenso wie zu große Fragmente, auch durch Deletionen oder chimäre YAC-Abschnitte verursacht werden (Abb. 24B).

Die Kartierung der distalen *S100*-Gene und die dabei nachgewiesenen Sall-Fragmente stimmten mit früheren Ergebnissen [Schafer *et al.*, 1995; Wicki *et al.*, 1996a; South *et al.*, 1999] überein. Neben *S100A6*, *S100A8* und *S100A9*, die bereits innerhalb des YAC-Contigs lokalisiert worden waren, wurden zur Erhöhung der Markerdichte *S100A1*, *S100A2* und *S100A13* kartiert sowie die bei der sub-

traktiven Hybridisierung identifizierten *S100A4* und *S100A7* (Abb. 24B). Auf der centromeren Seite wurde das neu identifizierte *S100A11* auf dem gleichen *Sall*-Fragment wie *S100A10* und *THH* nachgewiesen. Die genaue Position zwischen *S100A10* und *THH* konnte durch Verwendung eines weiteren Restriktionsenzym zur Spaltung der YAC-DNA – nur *S100A10* und *S100A11* hybridisierten mit einem gemeinsamen *SfiI*-Fragment (Abb. 23) – und durch Hinzuziehen der genomischen Restriktionskarte – hier lag *S100A11* auf den gleichen Fragmenten wie *THH* (Abb. 24A) – festgelegt werden und war bereits vor der Durchführung der subtraktiven Hybridisierung veröffentlicht worden [Wicki *et al.*, 1996b]. Innerhalb des ursprünglichen EDC einschließlich *S100A10* im proximalen und *S100A1* im distalen Bereich [Mischke *et al.*, 1996] konnten alle *Sall*-Fragmente und die enthaltenen Loci ohne Widerspruch zur genomischen Restriktionskarte positioniert werden. Mit *NICE1* zwischen *FLG* und *IVL* sowie *NICE2* zwischen *S100A8* und *S100A7* wurden in dieser Region zwei neue Gene kartiert (Abb. 24B). Obwohl die Sonden von *NICE2*, *S100A7* und 24f59 – wahrscheinlich zurückzuführen auf eine Amplifizierung im Genom – jeweils mit den gleichen zwei Restriktionsfragmenten hybridisierten, ermöglichten die unterschiedlichen Signalintensitäten eine Auflösung der Reihenfolge: *NICE2* zeigte die stärkere Reaktion mit dem proximalen *Sall*-Fragment, bei *S100A7* war die Intensität beider Signale gleich, und 24f59 reagierte verstärkt mit dem distalen Restriktionsfragment.

Die Feinkartierung der weiter telomer gelegenen Marker einschließlich der acht neu identifizierten EDC-Gene basierte hauptsächlich auf den aus YAC 950\_e\_2 isolierten Kolonien 950\_e\_2#1 bis 950\_e\_2#11. Diese enthielten zwar bis auf Klon 950\_e\_2#3 alle die äußeren, das YAC begrenzenden Marker, aber im Vergleich zum größten Klon 950\_e\_2#4 ließen sie Deletionen erkennen, an denen jedesmal die Region zwischen *S100A1* und 24f11 beteiligt war. Die unterschiedlichen Ausdehnungen der in den einzelnen Klonen eliminierten DNA-Sequenzen ergab die in Abbildung 22 dargestellte Reihenfolge



**Abbildung 23: Lokalisierung des *S100A11*-Gens durch YAC-Restriktionskartierung.** Die Autoradiogramme zeigen die Hybridisierungen spezifischer Sonden für *S100A10*, *S100A11* und *THH* mit einem Southern-Blot, der die aufgetrennten *SfiI*-, *Sall*- und *SacI*-Restriktionsfragmente von YAC 764\_a\_1#1 enthält. Ein 210 kb-*Sall*-Fragment wurde von allen drei <sup>32</sup>P-markierten Sonden nachgewiesen, ein 115 kb-*SfiI*-Fragment nur von den Sonden für *S100A10* und *S100A11*. Da *S100A11* und *THH* aber nicht *S100A10* auf einem gemeinsamen *BsiWI*-Fragment liegen (s. Abb. 24A), kartiert *S100A11* zwischen *S100A10* und *THH*.

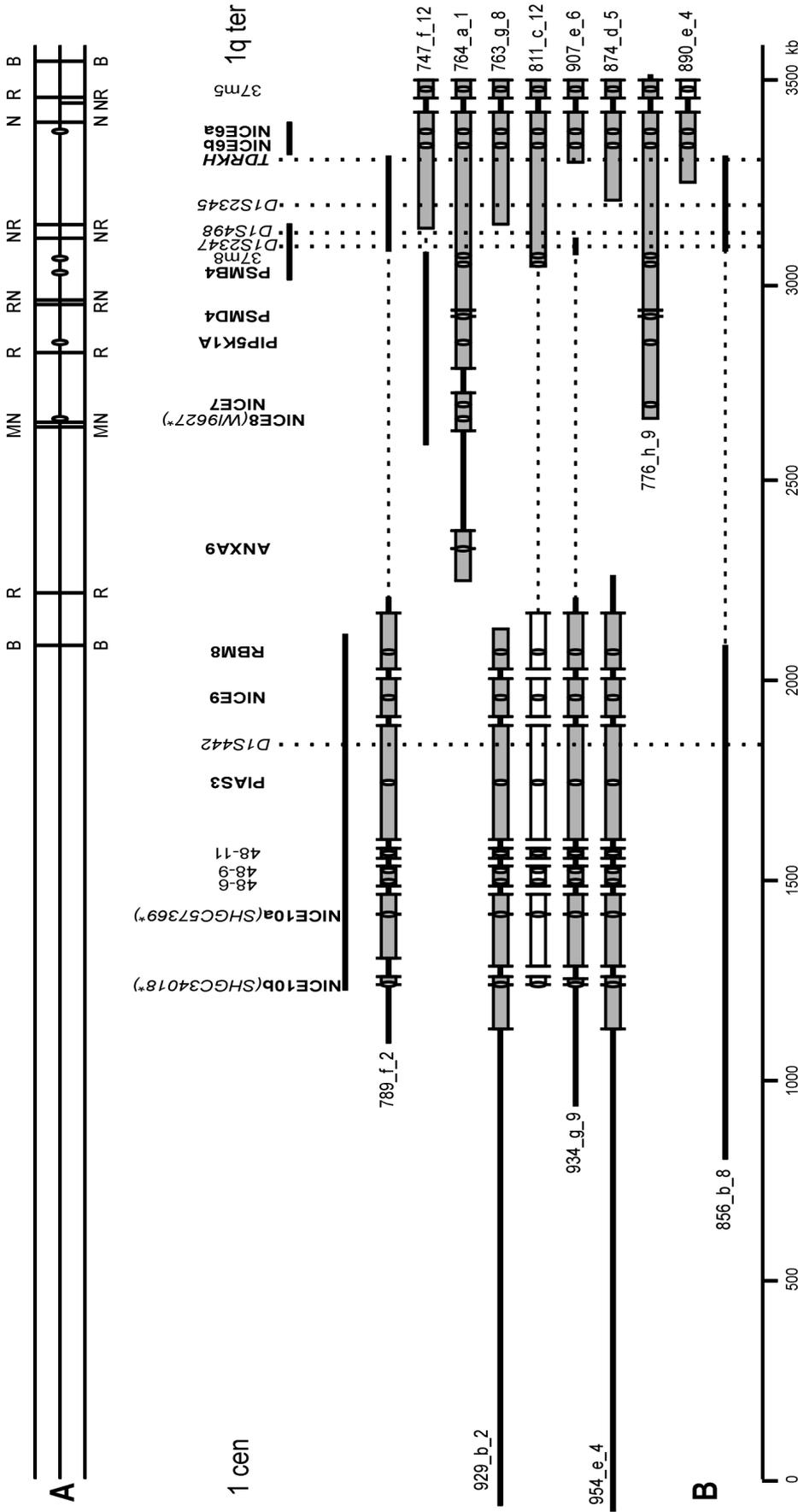
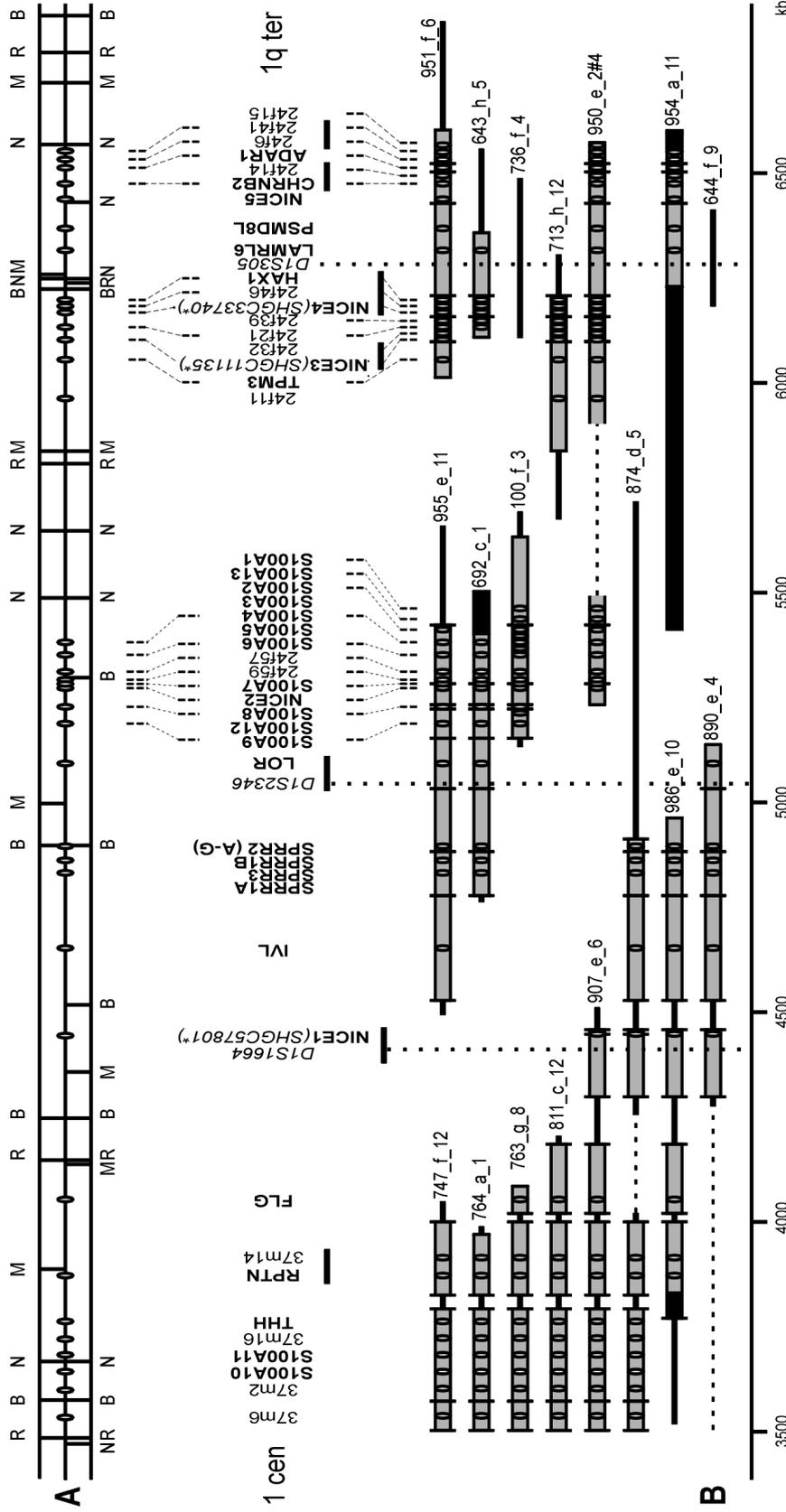


Abbildung 24: Integrierte Karte des EDC. Auf dieser Seite links ist das centromere Ende der Karte, auf Seite 83 rechts das telomere. Die Ergebnisse der genomischen und der YAC-Restriktionskartierung führten zu der angegebenen Anordnung der Gene (fett), STS-Marker (kursiv) und weiterer Loci der Region 1q21. STS-Marker, die in den *MICE*-Genen enthalten sind, aber nicht getestet wurden, sind mit einem Sternchen versehen. Die genauen Positionen der Loci 37m2/37m16, der zwei *SPRR7*-Gene und von *S100A13/S100A1* wurden bereits zuvor bestimmt [South et al., 1999]. Die Reihenfolge der mit einem Balken unterlegten Loci konnte nicht aufgelöst werden. (A) Genomische Restriktionskarte des EDC. Die Karte umfaßt 4,5 Mb der Region 1q21 mit den in der H2LCL-Zelllinie enthaltenen Restriktionsschnittstellen für NotI (N), NruI (R), MluI (M) und BsiWI (B). Die zwei Bahnen repräsentieren die zwei Haplotypen der Zelllinie. Weiße Ovale symbolisieren die Hybridisierung der unten angegebenen Sonde mit den entsprechenden Fragmenten, die aus der Restriktionsspaltung mit einem bzw. mit zwei Enzymen hervorgegangen sind. Die genaue Kartierung von *S100A10*, *THH*, *FLG*, *IVL*, *SPRR3*, *SPRR1*, *SPRR2*, *LOR*, *S100A9*, *S100A8* und *S100A6* wurde durch Restriktionsspaltung mit zusätzlichen Enzymen erreicht und von der etablierten Karte übernommen [Volz et al., 1993; Mischke et al., 1996]. (Fortsetzung auf Seite 83)



(Fortsetzung von Seite 82) **(B)** YAC-Restriktionskarte des EDC. Graue Balken repräsentieren die in den YACs nachgewiesenen genomischen Sall-Fragmente, schwarze Balken rearrangierte Restriktionsfragmente; gestrichelte horizontale Linien kennzeichnen deletierte Abschnitte. Schwarze vertikale Striche zeigen die die einzelnen Fragmente begrenzenden Restriktionsschnittstellen an. Die Länge der grauen und schwarzen Balken zusammen mit den schwarzen Linien (unidentifizierte Fragmente) entspricht den durch ROFE ermittelten YAC-Größen bzw. Sall-Fragmentgrößen. Bei abweichenden kleineren Restriktionsfragmenten, die nur in einem YAC nachgewiesen wurden, handelt es sich wahrscheinlich um Endfragmente; ihnen fehlt der Marker für die zweite Sall-Schnittstelle. Weiße Ovale symbolisieren die Hybridisierung der oben angegebenen Sonde mit den entsprechenden Fragmenten, vertikale gestrichelte Linien ein positives PCR-Signal mit den entsprechenden YACs. Die Gene *S100A3*, *S100A5* und *S100A12* (graue Ovale) wurden zuvor auf YAC 100\_f\_3 kartiert [Schaefer et al., 1995; Wicki et al., 1996b]. Schwache Hybridisierungssignale sind durch einen weißen Balken gekennzeichnet. Um die Reihenfolge der auf YAC 950\_e\_2 gelegenen Gene zu bestimmen, wurden alle aus der ursprünglichen YAC-Kultur isolierten Klone (Abb. 22) mit einbezogen. Die in zwei von vier YACs enthaltene Sall-Schnittstelle innerhalb des *NICE1*-Gens weist auf einen Polymorphismus hin.

der Marker von *NICE2* bis 24f15. In Verbindung mit weiteren YACs des Contigs konnte darüberhinaus die Anordnung *S100A2* – *S100A13/S100A1* (YAC 955\_e\_11), 24f11 – *TPM3* (YAC 951\_f\_6), *NICE3/24f32* – 24f21 und *LAMRL6* – *PSMD8L* (YAC 643\_h\_5) sowie 24f6/24f41 – 24f15 (YAC 954\_a\_11) gelöst werden, und bis auf die von der Deletion betroffenen Region konnten die intakten Sall-Fragmente bestimmt werden (Abb. 24B). Die Auflösung folgender Loci, die jeweils auf identischen Fragmenten dicht beieinander liegen, war mit Hilfe der YAC-Restriktionskarte nicht möglich: *S100A6/S100A4*, *S100A13/S100A1*, *NICE3/24f32*, *NICE4/24f46/HAX1* und 24f6/24f41. Da die *S100*-Gene von *S100A9* bis *S100A1* auf einer separaten Restriktionskarte von YAC 100\_f\_3 [Schafer *et al.*, 1995; Wicki *et al.*, 1996b] und innerhalb eines PAC/BAC/Cosmid-Contigs [South *et al.*, 1999] kartiert wurden, blieben von den Genen aus der telomeren EDC-Region einzig *NICE4* und *HAX1*, die nicht exakt lokalisiert werden konnten.

Anders sah es auf der centromeren Seite des YAC-Contigs aus. Obwohl die Region proximal von *S100A10* nahezu die Hälfte des Contigs ausmachte, waren hier keine Gene und nur sieben Hybridisierungsmarker lokalisiert. Mit der Kartierung von 13 den EDC erweiternden Loci kristallisierten sich drei Bereiche der proximalen Region heraus, die genauer charakterisiert werden konnten. Von *S100A10* bis *PSMB4/37m8* hybridisierten die DNA-Sonden mit mindestens drei YACs, woraufhin die Reihenfolge der intakten Sall-Fragmente – von *S100A10* bis *NICE6* in Übereinstimmung mit dem PAC/BAC/Cosmid-Contig [South *et al.*, 1999] – festgelegt werden konnte (Abb. 24B). Lediglich die auf den gleichen Restriktionsfragmenten liegenden Loci *PSMB4* und 37m8 konnten nicht getrennt werden. Es folgten fünf Gene, die entweder auf den YACs 764\_a\_1 und 776\_h\_9 gemeinsam nachgewiesen wurden (*PSMD4*, *PIP5K1A* und *NICE7*) oder nur auf YAC 764\_a\_1 (*NICE8* und *ANXA9*). Obwohl ihre Anordnung aufgelöst wurde, können Umlagerungen in diesem Bereich nicht ausgeschlossen werden, da die beiden YACs, vergleichbar mit der Region um *S100A1* und 24f11, keine einheitlichen Fragmentgrößen aufwiesen und YAC 764\_a\_1 darüberhinaus äußerst instabil war (s. Abschnitt III 1.3). Schließlich wurden acht Marker, darunter die vier Loci *PIAS3*, *RBM8*, *NICE9* und *NICE10*, gemeinsam auf den vier am weitesten centromer liegenden YACs kartiert. Ihnen konnten zwar intakte Sall-Fragmente zugeordnet werden, aufgrund der einheitlichen Hybridisierungsmuster war es aber nicht möglich, eine Reihenfolge zu bestimmen (Abb. 24B).

**Tabelle 14: Rearrangierte YACs**

YAC	Umlagerungen		
	integrierte Karte	Lioumi <i>et al.</i> , 1998 <sup>a</sup>	STS-Daten <sup>b</sup>
643_h_5	+		+
644_f_9	+	n.a.	
692_c_1	+	n.a.	
736_f_4	+	n.a.	+
747_f_12	+		+
789_f_2	+		
811_c_12	+		+
856_b_8	+	n.a.	+
874_d_5	+	+	
890_e_4	+	n.a.	
929_b_2	+	n.a.	+
950_e_2	+	n.a.	
954_e_4			+
954_a_11	+	n.a.	+
955_e_11	+		
986_e_10	+	+	

<sup>a</sup> Karte der Region 1q21 aus fragmentierten YACs; <sup>b</sup> Ergebnisse der STS-Kartierung des CEPH und des Whitehead Institute/MIT; +, YAC ist rearrangiert; n.a., nicht analysiert.

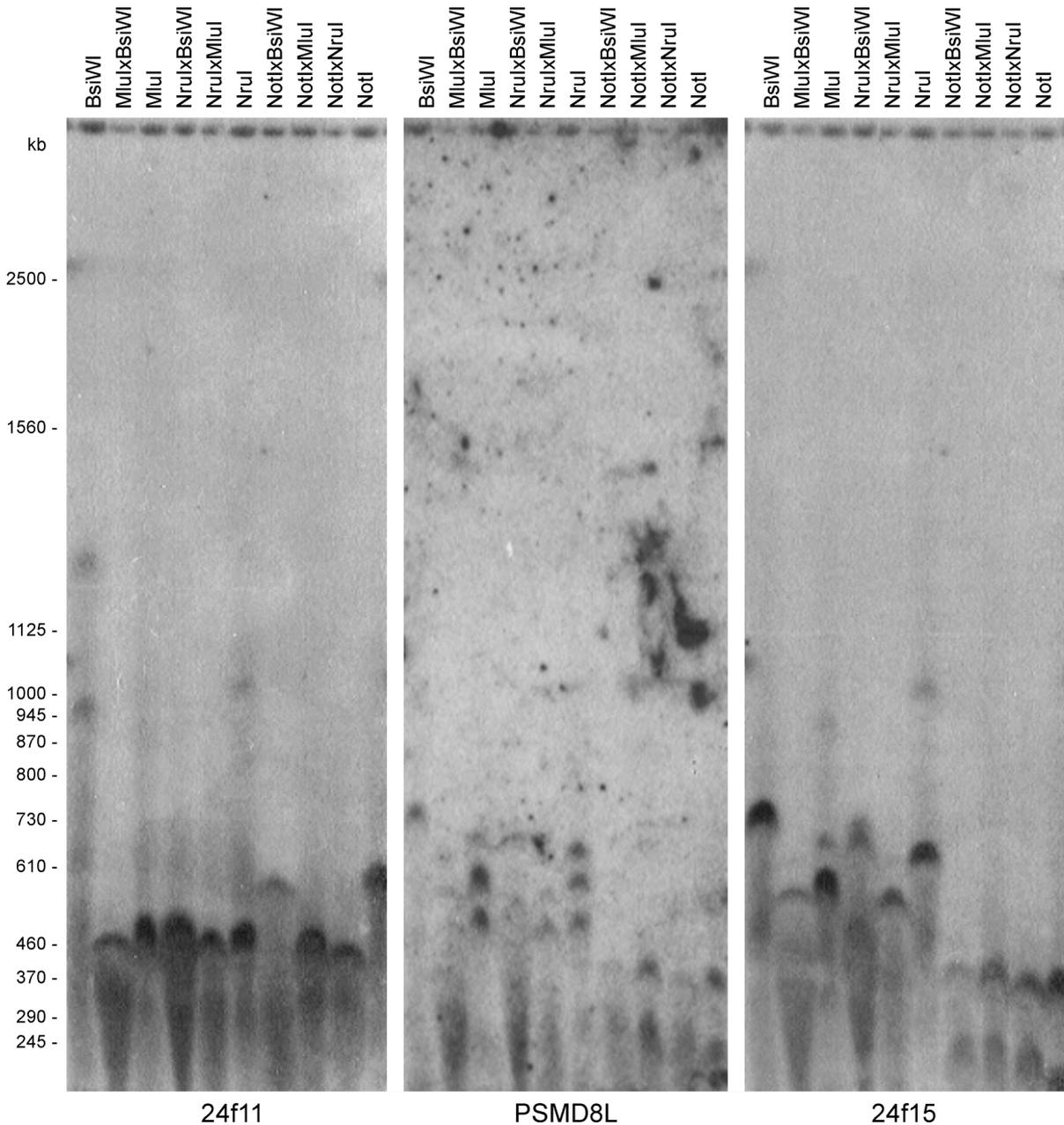
Neben der Feinkartierung der DNA-Marker ermöglichte die SalI-Restriktionskarte auch eine genauere Charakterisierung der einzelnen YACs, wie anhand der aus YAC 950\_e\_2 isolierten Kolonien demonstriert werden konnte. Umlagerungen, die bereits bei der Erstellung des YAC-Contigs aus dem Vergleich mit der genomischen Restriktionskarte abgeleitet werden konnten, wurden mit Hilfe der SalI-Fragmente verifiziert (Abb. 24B, Tab. 14). Deletionen wurden in den YACs 874\_d\_5 (*FLG*), 890\_e\_4 (von 37m6 bis *FLG*) und 950\_e\_2 (zwischen *S100A1* und 24f11) nachgewiesen. Umlagerungen am telomeren Ende offenbarten die YACs 692\_c\_1 und 954\_a\_11, am centromeren Ende die YACs 986\_e\_10 und ebenfalls 954\_a\_11. Die YACs 643\_h\_5, 747\_f\_12, 955\_e\_11 und wiederum 874\_d\_5 waren aufgrund der Diskrepanz zwischen ihrer Größe und dem Markergehalt rearrangiert; der betroffene Abschnitt konnte nicht genauer eingegrenzt werden. Im Gegensatz zu den vorher genannten war YAC 811\_c\_12 kleiner als von den enthaltenen Markern zu erwarten gewesen wäre. Da die Hybridisierungssignale im centromeren Bereich aber durchgängig schwach waren, handelte es sich wahrscheinlich um ein Gemisch zweier YACs, die aus dem gleichen Ursprungs-YAC hervorgegangen waren und von denen einer stark unterrepräsentiert war – und möglicherweise eine Deletion von *ANXA9* bis *PSMD4* aufwies. Weitere Umlagerungen der YACs 644\_f\_9, 736\_f\_4, 789\_f\_2, 856\_b\_8 und 929\_b\_2, welche die Regionen nahe der STS-Marker betrafen, deckte die integrierte Karte auf (s. Abschnitt III 3.3).

### 3.2 Genomische Restriktionskarte

Da die YAC-Restriktionskarte zwei Abschnitte der chromosomalen Region 1q21 offenlegte, die nur von jeweils einem, dazu noch instabilen YAC abgedeckt waren und somit möglicherweise nicht kontinuierlich in den YACs enthalten waren, sollten die Entfernungen und die Positionen ausgewählter Loci mit Hilfe einer Restriktionskarte der H2LCL-Zelllinie des Menschen verifiziert werden. Als Ausgangspunkt dienten die Restriktionsfragmente der bekannten Genkarte des EDC, die sich von *S100A10* bis *S100A6* erstreckte [Volz *et al.*, 1993]. Zur Überbrückung größerer Entfernungen wurden die selten schneidenden Restriktionsenzyme NotI, NruI, MluI und BsiWI für die Fragmentierung der DNA eingesetzt; durch Doppelrestriktion mit den gleichen Enzymen wurde die Auflösung erhöht.

Zwischen den bekannten EDC-Genen *S100A10* und *S100A1* wurden fünf Loci und drei weitere DNA-Marker kartiert (Abb. 24A). *S100A11* und 37m16 wurden auf den gleichen Fragmenten wie *THH* lokalisiert. Die von der *NICE1*-Sonde identifizierten Fragmente unterschieden sich lediglich in einem von zwei BsiWI-Fragmenten, die auf die zwei Haplotypen der H2LCL-Zelllinie zurückgeführt wurden, vom Hybridisierungsmuster für *IVL*. *NICE2*, *S100A7* und 24f59 wurden zusammen mit *S100A8* positioniert. Die Sonden für 24f57 und *S100A4* identifizierten schließlich wie *S100A6* die am weitesten distal gelegenen Restriktionsfragmente der etablierten Karte [Volz *et al.*, 1993].

In Richtung des Telomers konnte die genomische Restriktionskarte durch die Positionierung von acht neuen EDC-Genen und fünf zusätzlichen Loci erweitert werden. Dabei bildeten sich drei Gruppen heraus, die jeweils mit gleichen Restriktionsfragmenten hybridisierten (Abb. 25). Über zwei

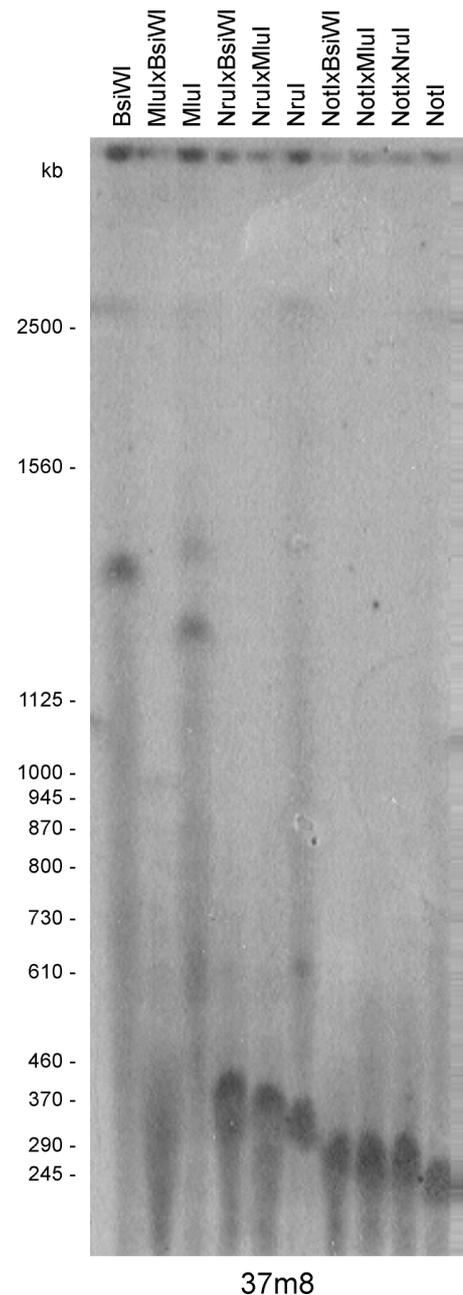


**Abbildung 25: Genomische Restriktionskartierung der telomeren EDC-Region.** Die Autoradiogramme zeigen denselben, die Restriktionsfragmente der H2LCL-Zelllinie enthaltenden Southern-Blot nach Hybridisierung mit den <sup>32</sup>P-markierten Sonden für 24f11, *PSMD8L* und 24f15. *HAX1*, *NICE3*, *NICE4*, *TPM3*, 24f21 und 24f46 identifizierten die gleichen 1q21-spezifischen Fragmente wie 24f11, *LAMRL6* die gleichen wie *PSMD8L* und *NICE5*, *CHRNB2*, *ADAR1* und 24f6 die gleichen wie 24f15. Die für die Restriktionsspaltung eingesetzten Enzyme sind angegeben. Als Größenstandard wurden die Chromosomen von *S. cerevisiae* WAY5-4 verwendet.

gemeinsame BsiWI-Fragmente mit *S100A4* verbunden waren 24f11, *TPM3*, *NICE3*, *NICE4*, 24f21, 24f46 und *HAX1*, deren Sonden vorher unbekannte NotI-, NruI- und MluI-Fragmente identifizierten. Von diesen wurde wiederum jeweils eines von zwei NruI- und MluI-Fragmenten auch von den sich anschließenden Gruppen erkannt. Ansonsten lagen die Loci *LAMRL6* und *PSMD8L* einerseits sowie *NICE5*, *ADAR1*, 24f6 und 24f15 andererseits auf neu identifizierten Fragmenten, die bis auf eines von zwei NotI-Fragmenten übereinstimmten. Insgesamt wurde der Bereich, in dem EDC-Gene lokalisiert

waren, um ca. 1 Mb in Richtung 1q22 ausgedehnt, das am weitesten telomer gelegene (BsiWI-) Fragment der Restriktionskarte deckte weitere 400 kb ab (Abb. 24A). Dabei konnte die Kartierung der fünf *NICE*-Gene sowie von *HAX1* und *ADARI* innerhalb des EDC aufgrund der Hybridisierung ihrer Sonden ausschließlich mit 1q21-spezifischen Restriktionsfragmenten zweifelsfrei nachgewiesen werden. Im Gegensatz dazu identifizierte die *TPM3*-Sonde weitere Restriktionsfragmente der H2LCL-Zelllinie, die auf verwandte Sequenzen innerhalb des menschlichen Genoms schließen lassen [MacLeod *et al.*, 1986]. Ein ähnliches Resultat gab es für *LAMRL6*, zurückzuführen auf eine Vielzahl an Pseudogenen [Jackers *et al.*, 1996b], u.a. auf den Chromosomen 3, 12, 14 und X [Bignon *et al.*, 1991], sowie ein funktionsfähiges Lamininrezeptor-Gen (*LAMRI*) auf Chromosom 3p21.3 [Jackers *et al.*, 1996a]. Schließlich zeigte auch die *PSMD8L*-Sonde eine Kreuzhybridisierung mit zusätzlichen Restriktionsfragmenten (Abb. 25), die möglicherweise vom *PSMD8*-Gen stammen, dessen Sequenz in einem auf Chromosom 19 kartierenden PAC enthalten ist (GenBank Acc AC005789).

Auf der anderen Seite des Genkomplexes wurden acht Gene und drei weitere Marker erstmals auf genomischen Restriktionsfragmenten kartiert (Abb. 24A). 37m2 hybridisierte mit den gleichen Restriktionsfragmenten wie *S100A10*. Daran anschließend hatten *NICE6a*, 37m8, das die gleichen Fragmente identifizierende *PSMB4*, *PIP5K1A* und *NICE8* lediglich die zwei *MluI*-Fragmente mit *S100A10* gemeinsam. Sie hybridisierten außerdem mit unterschiedlichen *NotI*- und *NruI*-Fragmenten, aber mit dem gleichen *BsiWI*-Fragment (Abb. 26). Die weiter centromer kartierenden Loci *PIAS3*, *RBM8*, *NICE9*, *NICE10a*, *NICE10b* und 48-11 identifizierten zwar gemeinsame Restriktionsfragmente, es konnte jedoch keine Verbindung zu dem zuvor erwähnten proximalen *BsiWI*-Fragment hergestellt werden; das gelang schließlich durch die Integration der genetischen Marker (s. Abschnitt III 3.4).



37m8  
Abbildung 26: Detektion der centromeren Fragmente der genomischen Restriktionskarte des EDC. Das Autoradiogramm zeigt die von der  $^{32}\text{P}$ -markierten 37m8-Sonde auf dem Southern-Blot detektierten Restriktionsfragmente der H2LCL-Zelllinie. Die gleichen *BsiWI*- und *MluI*-Fragmente wurden von den spezifischen Sonden für *PSMB4*, *PIP5K1A*, *NICE6a* und *NICE8* erkannt. Die für die Restriktionsspaltung eingesetzten Enzyme sind angegeben. Als Größenstandard wurden die Chromosomen von *S. cerevisiae* WAY5-4 verwendet.

### 3.3 Fusion der physikalischen Karten

Durch Kombination der genomischen und der YAC-Kartierungsergebnisse wurde eine physikalische Karte des EDC erstellt, welche die Vorteile der beiden Methoden – die Verlässlichkeit der genomischen Kartierung und die hohe Auflösung der YAC-Restriktionskarte – miteinander verband. Die sich daraus ergebende Anordnung der einzelnen Loci ist in Abbildung 24 dargestellt.

Abgesehen von der höheren Auflösung des YAC-Contigs stimmte die Reihenfolge der Loci auf beiden Karten überein. Selbst in den Regionen mit erhöhter Umlagerungsaktivität der YACs war die Anordnung ohne Widerspruch möglich. Abweichungen wurden jedoch bei den Entfernungen zwischen einzelnen Loci festgestellt. Der aus den genomischen Restriktionsfragmentgrößen resultierende Abstand zwischen *NICE2* und *ADARI* betrug mindestens 1,1 Mb, wohingegen YAC 950\_e\_2#4, das die beiden Gene enthielt und zuvor keine Umlagerung hatte erkennen lassen, nur 850 kb umfaßte. Da die Charakterisierung der aus YAC 950\_e\_2 isolierten Kolonien bereits in zehn von elf YACs interne Deletionen aufdeckte (Abb. 22), sollte das gleiche auch für YAC 950\_e\_2#4 zu erwarten sein. Aus der genomischen Entfernung und der genauen Position der Gene auf der fusionierten physikalischen Karte ergibt sich für YAC 950\_e\_2#4 eine Deletion von 400 kb. Ein ähnliches Bild zeichnete sich im proximalen Bereich ab. Hier zeigte der Vergleich der physikalischen Karten, daß *ANXA9* auf YAC 764\_a\_1 offenbar weiter von den proximal angrenzenden Loci des Contigs entfernt lag als ursprünglich angenommen.

### 3.4 Integration der genetischen Marker

Der Nachweis der STS-Marker auf den YACs der Region 1q21 durch PCR hatte bei der Erstellung des Contigs lediglich eine grobe physikalische Kartierung erlaubt, da die exakten Positionen der YACs und somit auch der STS-Marker noch nicht eindeutig festgelegt werden konnten. Die Lokalisierung aller Hybridisierungsmarker einschließlich der neu identifizierten EDC-Gene auf einzelnen Restriktionsfragmenten der physikalischen Karte ermöglichte jetzt eine genauere Kartierung der YACs – als Folge konnten auch die Positionen der hochpolymorphen STS-Marker auf der Sall-Restriktionskarte bestimmt werden (Abb. 24).

In der telomeren Region wurde D1S305 zwischen *NICE4/24f46/HAX1* und *LAMRL6*, auf dem verkürzten *LAMRL6*-Fragment (140 kb) von YAC 643\_h\_5 lokalisiert. Es folgte zentral im EDC der STS-Marker D1S2346, der in einem Bereich von 190 kb zwischen den distalen Enden der YACs 890\_e\_4 und 986\_e\_10, nahe *LOR*, kartierte, sowie D1S1664, der zusammen mit *NICE1* zwischen den *IVL*- und *FLG*-Fragmenten liegt. Als nächstes wurde D1S2345 in einem 60 kb-Intervall zwischen den proximalen Enden der YACs 763\_g\_8 und 874\_d\_5 positioniert. D1S2347 und D1S498 wurden gemeinsam zwischen D1S2345 und *PSMB4/37m8* plziert. Da ihre Position fast ausschließlich mit Hilfe rearrangierter YACs festgelegt wurde, ist sie noch zu verifizieren. D1S442 wurde schließlich als einziger STS-Marker auf den vier centromeren YACs nachgewiesen. Zusätzlich wurden sieben genspezi-

fische STS-Marker identifiziert, deren Sequenzen in fünf der neuen Gene enthalten waren (Abb. 24, Tab. 17).

### 3.5 Kartierung weiterer Gene der Region 1q21

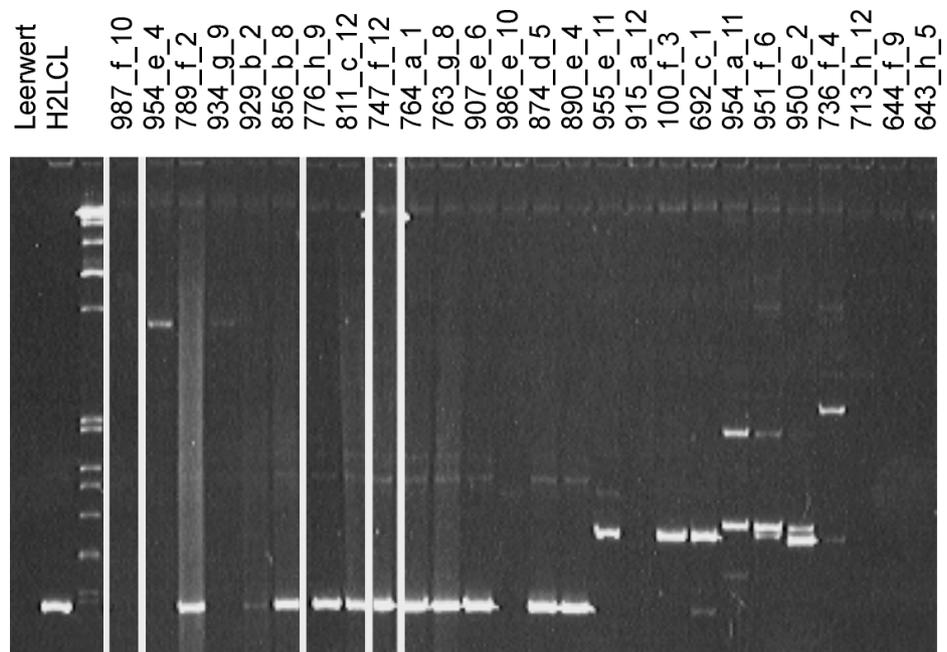
Neben den bei der subtraktiven Hybridisierung identifizierten Genen gab es noch verschiedene Kandidatengene für den EDC bzw. die Region 1q21, die für kooperierende Arbeitsgruppen kartiert wurden. Zu ihnen zählten Gene, die eine strukturelle Verwandtschaft mit den Genfamilien des EDC erkennen ließen oder denen eine Funktion in epidermalen Geweben zugeordnet war, und Gene, deren Lokalisierung in der Region 1q21 eine Zugehörigkeit zum EDC nahelegten, deren Bedeutung für die Entwicklung der Epidermis aber unklar war.

Aufgrund struktureller Gemeinsamkeiten mit Profilaggrin und Trichohyalin sowie der Lokalisierung des entsprechenden Gens der Maus in der orthologen Region auf Mausechromosom 3 (s. Abschnitt IV 5.3) wurde das Repetin-Gen (*RPTN*) [Krieg *et al.*, 1997] (s. Abschnitt IV 4) ausgewählt. Die entsprechende Sonde hybridisierte mit dem gleichen SalI-Fragment wie 37m14 (Abb. 24B); *RPTN* konnte somit zwischen *THH* und *FLG* lokalisiert werden [Huber *et al.*, in Vorbereitung].

Das Gen für die bereits in Keratinozyten beschriebene  $\beta$ 2-Untereinheit des neuronalen Nicotinrezeptors (*CHRNA2*) [Grando *et al.*, 1995] (s. Abschnitt IV 4) wurde auf den gleichen genomischen Restriktionsfragmenten wie *NICE5*, *ADARI*, 24f6 und 24f15 nachgewiesen. Die Hybridisierung der *CHRNA2*-Sonde mit dem den Marker 24f14 enthaltenden, rearrangierten SalI-Fragment von YAC

950\_e\_2#11 (Abb. 22) legte die Position zwischen *NICE5* und *ADARI* fest [Lueders *et al.*, 1999].

Das eine Tudor- und zwei K-Homologie (KH)-Domänen enthaltende *TDRKH*-Gen [Lamb *et al.*, 2000] (s. Abschnitt IV 4) wurde mittels PCR mehreren YACs der proximalen Region des Contigs zugeordnet (Abb. 27). Es wurde zwischen D1S2345 und *NICE6* kartiert (Abb. 24).

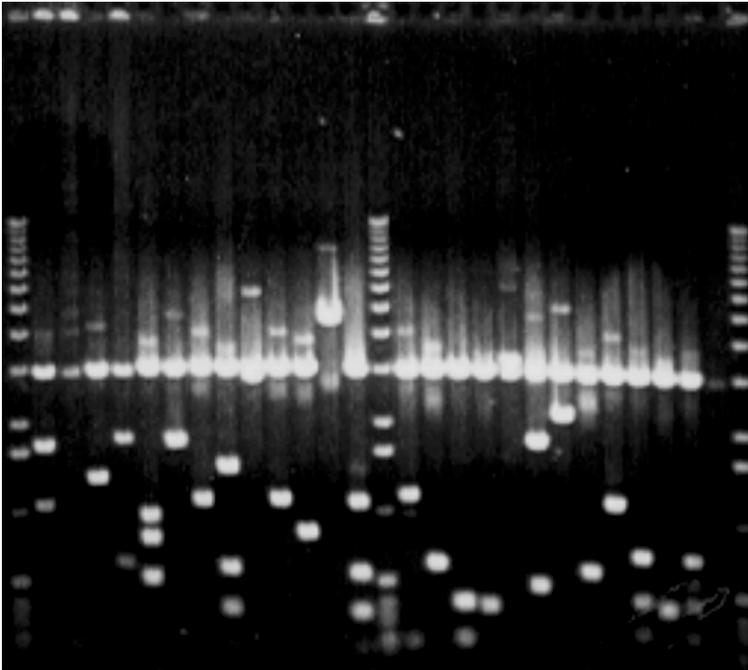


**Abbildung 27:** Nachweis der *TDRKH*-Sequenz auf den YACs des EDC-Contigs. Dargestellt ist ein mit Ethidiumbromid gefärbtes Polyacrylamidgel, auf dem das mit den spezifischen Primern für *TDRKH* durch PCR amplifizierte 120 bp-Fragment zu erkennen ist. Alle 26 Kandidaten-YACs wurden analysiert. Als Positivkontrolle wurde DNA der H2LCL-Zelllinie eingesetzt, als Größenstandard die 1kb-Leiter (dritte Bahn von links).

## 4 Charakterisierung der cDNA-Sequenzen

Für die Charakterisierung der von den YACs identifizierten cDNA-Klone mußten als erstes ihre Plasmide isoliert werden. Aus diesen wurden durch Restriktionsspaltung mit den Enzymen EcoRI und XhoI, deren Schnittstellen die klonierte cDNA einschlossen, die Inserts freigesetzt, und über ein Aga-

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29



**Abbildung 28: Restriktionsanalyse der cDNA-Klone.** Das mit Ethidiumbromid gefärbte Agarosegel zeigt die nach Restriktionsspaltung der Plasmide aufgetrennten DNA-Fragmente von 29 cDNA-Klonen. Die Vektorbande liegt bei 3 kb. Zusätzliche Banden stammen von den mit den Enzymen EcoRI und XhoI herausgetrennten cDNA-Inserts bzw. von einzelnen Insertfragmenten. Als Größenstandard (Bahnen 1, 15, 29) wurden 400 ng der 1 kb-Leiter aufgetragen. Da die 1,6 kb-Bande einem Zehntel der Menge des gesamten Größenstandards entspricht, konnte anhand der Signalintensitäten die Plasmidkonzentration der eingesetzten Lösung für die Sequenzierung abgeschätzt werden. Weitere Informationen, die aus dem Gel abgeleitet werden können, sind in Bahn 13 ein unvollständig geschnittenes Plasmid (Klon 3067n11), das möglicherweise auf eine Mutation in der Restriktionsschnittstelle hinweist, und in Bahn 28 ein Klon (3204n6), der nicht weiter analysiert werden konnte, da das Plasmid nicht ausreichend vervielfältigt wurde. Die schwächere Bande in Bahn 18 (Klon 3041m5) ist auf eine Verunreinigung mit einem zweiten Klon zurückzuführen, dessen zusätzliche schwache Signale bei der Sequenzierung als *S100A7*-Sequenz identifiziert wurden.

rosegel wurden die erhaltenen DNA-Fragmente aufgetrennt. Die anschließende Ethidiumbromidfärbung lieferte neben den Insertgrößen auch enthaltene EcoRI- und XhoI-Schnittstellen sowie die Konzentration der isolierten Plasmid-DNA (Abb. 28). Die aus dem Gel eluierten cDNA-Inserts bzw. einzelne Insertfragmente wurden nach radioaktiver Markierung für die Kartierung der entsprechenden Gene verwendet. Für die Sequenzierung der cDNA-Inserts wurde die isolierte Plasmid-DNA benötigt. Die ermittelten cDNA-Sequenzen wurden anschließend mit Hilfe diverser Computerprogramme analysiert (s. Abschnitt II 18).

Entgegen guter mikrobiologischer Praxis ging die Anzucht einzelner cDNA-Klone nicht von einer Bakterienkolonie aus; es stellte sich als vorteilhaft heraus, die Flüssigkulturen direkt aus den Mikrotiterplatten, die der Aufbewahrung der cDNA-Bibliothek dienen, anzupumpfen. Auf

diese Weise waren geringe Kontaminationen einzelner Klone mit einem zweiten, die in einigen Fällen für ein vermeintlich falsch positives Signal bei der Hybridisierung der Filter verantwortlich waren, an zusätzlichen schwachen Banden bei der Bestimmung der Insertgröße (Abb. 28) und bei der Sequenzierung zu erkennen. Da nur etwa 2% der cDNA-Klone so stark verunreinigt waren, daß eine Sequenzierung der Plasmidinserts nicht möglich war, konnten diese gezielt noch einmal isoliert werden. Im Ge-

gensatz dazu hat die Anzucht von Klonen aus einzelnen Bakterienkolonien zum Ziel, durch die Analyse mehrerer zufällig ausgewählter Kolonien einer Kultur den vorherrschenden Klon kontaminationsfrei zu isolieren. Dieses Vorgehen hätte trotz des höheren Arbeitsaufwandes eine geringe Verunreinigung eines Klons möglicherweise nicht erfaßt, was die Interpretation der Hybridisierungssignale erschwert hätte.

#### 4.1 Ermittlung der cDNA-Sequenzen

Die Insertsequenzen der ausgewählten cDNA-Klone wurden unter Verwendung zweier fluoreszenzmarkierter Primer, die an beiden Seiten der das Insert begrenzenden Vektorsequenz ansetzten, nach dem Dideoxy-Verfahren [Sanger *et al.*, 1977] bestimmt. Da für die Aussagekraft der subtraktiven Hybridisierung eine möglichst große Zahl an cDNA-Klonen analysiert werden mußte, wurden nur die terminalen Bereiche der Inserts sequenziert. Mit Hilfe eines automatischen Sequenzierers und durch anschließende manuelle Nachbearbeitung konnten bis zu 600 bp für jeden Primer sicher bestimmt werden, wie eine mehr als 99%ige Übereinstimmung mit den Datenbankeinträgen bekannter Sequenzen zeigte. Weitere 200 bp wiesen eine nachlassende, aber immer noch ausreichende Genauigkeit auf, um Sequenzen gleichen Ursprungs zu identifizieren. Dementsprechend konnten cDNA-Inserts bis zu einer Größe von etwa 1200 bp vollständig sequenziert werden. Größere Inserts wurden aus überlappenden 5'-Enden unvollständig synthetisierter cDNA-Klone zusammengesetzt, die vom gleichen Transkript stammten, aber vom Oligo(dT)-Primer ausgehend unterschiedlich weit in cDNA umgeschrieben worden waren. Verbleibende Lücken wurden, wenn vorhanden, mit ESTs desselben Gens geschlossen, die durch Sequenzvergleich in Datenbanken ermittelt wurden (BLAST [Übersicht in Altschul *et al.*, 1997]). Um Fehler bei der Sequenzierung sowie chimäre cDNA-Inserts auszuschließen, wurden die Sequenzen mehrerer Klone desselben Gens miteinander verglichen und zu einer Konsensussequenz zusammengeführt (*alignment* [Corpet, 1988]). Auf diese Weise konnten alternativ gespleißte und polymorphe Genprodukte identifiziert werden.

Von den 190 sequenzierten cDNA-Klonen (einschließlich der nicht subtraktiv ausgewählten) konnten 126 der Region 1q21 zugeordnet werden. Diese repräsentierten 40 verschiedene Transkripte, die von 33 Genen stammten: zwölf bekannten EDC-Genen, die als Positivkontrolle der subtraktiven Hybridisierung dienten, elf neuen Mitgliedern des EDC, deren Translationsprodukte bereits bekannt und teilweise charakterisiert worden waren, und zehn weiteren Genen, deren Proteine noch nicht identifiziert wurden.

#### 4.2 cDNA-Sequenzen bekannter EDC-Gene

Die cDNA-Sequenzen der zwölf bekannten EDC-Gene stammten von *IVL*, *SPRR1A/1B*, *SPRR2A/2B*, *SPRR3*, *THH*, *S100A4*, *S100A6*, *S100A7*, *S100A10* und *S100A11* (Tab. 15). Sie wiesen Übereinstimmungen von mehr als 99% mit den entsprechenden Datenbankeinträgen auf. Eine größere Abweichung (95% Übereinstimmung) wurde jedoch in vier von acht *SPRR3*-Klonen festgestellt. Hieraus

**Tabelle 15: EDC-Gene mit bekannter Funktion**

YAC-Sonde <sup>a</sup>	Gen (Transkript) <sup>b</sup>	cDNA-Größe (Klon) <sup>c</sup>	BLASTN <sup>d</sup> : GenBank Acc (Transkriptgröße), Identität	positive YACs <sup>e</sup>
<b>etablierte EDC-Gene</b>				
986_e_10	<i>IVL</i> (Involucrin)	2134 bp* (1184j17)	M13902/M13903 (2141 bp), 99%	874_d_5, 890_e_4, 955_e_11
	<i>SPRR1A/1B</i> (small proline-rich protein 1A/1B)	769 bp (1197121)/ 595 bp (1105b6)	L05187 (>596 bp), 99%/ M84757 (622 bp), 99%	692_c_1, 874_d_5, 890_e_4, 955_e_11
	<i>SPRR2A/2B</i> (small proline-rich protein 2A/2B)	665 bp (3042e15)/ 661 bp (3041j14)	M20030 (679 bp), 99%/ M21302 (682 bp), 99%	692_c_1, 874_d_5, 890_e_4, 955_e_11
	<i>SPRR3</i> (small proline-rich protein 3)	890 bp (3113I8)/ 859 bp (3162o12)	AF077374 (900 bp), 99%/95%	692_c_1, 874_d_5, 890_e_4, 955_e_11
950_e_2#9	<i>S100A4</i> (S100A4 oder CAPL)	496 bp (1007n11)	M80563 (580 bp) 99%	100_f_3, 692_c_1, 955_e_11
	<i>S100A6</i> (S100A6 oder Calcyclin)	420 bp (3103p12)	M14300 (434 bp) 100%	100_f_3, 692_c_1, 955_e_11
	<i>S100A7</i> (S100A7 oder Psoriasis)	416 bp (1008h24)	M86757 (415 bp) 99%	100_f_3, 692_c_1, 955_e_11
764_a_1#2	<i>S100A10</i> (S100A10 oder Calpactin I leichte Kette)	661 bp (3019a19)	M81457 (609 bp), 99%	747_f_12, 763_g_8, 811_c_12, 874_d_5, 907_e_6
	<i>S100A11</i> (S100A11 oder S100C)	551 bp (1007m19)	D49355 (559 bp), 100%	747_f_12, 763_g_8, 811_c_12, 874_d_5, 907_e_6
	<i>THH</i> (Trichohyalin)	1859 bp* (1056m24)	L09190 (6979 bp), 99%	747_f_12, 763_g_8, 811_c_12, 874_d_5, 907_e_6
<b>neu identifizierte EDC-Gene</b>				
950_e_2#9	<i>ADAR1</i> (RNA-spezifische Adenosin- desaminase 1, protein p136)	2625 bp (1036f7)	U18121 (6474 bp), 99%	951_f_6, 954_a_11
	<i>HAX1</i> (HS1-bindendes Protein X-1 oder Hax-1)	1084 bp (3017c12)	U68566 (1173 bp), 99%	643_h_5, 713_h_12, 951_f_6
	<i>LAMRL6</i> (34/67 kd-Lamininrezeptor oder Laminin-bindendes Protein)	1025 bp (1045i15)	J03799 (991 bp), 99%	643_h_5, 951_f_6, 954_a_11
	<i>PSMD8L</i> (26S-Proteasom-Untereinheit nicht-ATPase 8 oder p31)	973 bp (3017o23)	D38047 (928 bp), 100%	951_f_6, 954_a_11
	<i>TPM3</i> (Tropomyosin TM30nm)	2075 bp (3006f16)	X04588 (2077 bp), 99%	713_h_12, 951_f_6
764_a_1#2	<i>ANXA9</i> (Annexin A9 oder 31)	1523 bp (1185o19)	AJ009985 (1762 bp), 99%	-
	<i>PIP5K1A</i> (Typ I Phosphatidylinosit-4- phosphat-5-kinase $\alpha$ )	3028 bp* (1052i19)	U78575 (3713 bp), 99%	776_h_9
	<i>PSMB4</i> (26S-Proteasom-Untereinheit $\beta$ 4 oder HsN3)	908 bp (3004e13)	D26600 (925 bp), 99%	776_h_9, 811_c_12
	<i>PSMD4</i> (26S-Proteasom-Untereinheit nicht-ATPase 4 oder 5Sa)	1277 bp (1103I7)	U51007 (1330 bp), 99%	776_h_9
934_g_9	<i>PIAS3</i> (Proteininhibitor des aktivierten Stat3)	2387 bp* (3033I20)	AB021868 (2808 bp), 99%	789_f_2, 929_b_2, 954_e_4
	<i>RBM8</i> (RNA-bindendes Motiv Protein 8)	2235 bp* (1175I23)	AF127761 (2270 bp), 99%	789_f_2, 929_b_2, 954_e_4

<sup>a</sup> Die aufgeführten YACs wurden zur Identifizierung der den jeweiligen Genen entsprechenden cDNA-Klone eingesetzt; <sup>b</sup> die den detektierten cDNA-Klonen entsprechenden Gene bzw. Transkripte sind angegeben; <sup>c</sup> der größte identifizierte Klon ist angegeben; <sup>d</sup> BLASTN, Nucleotid-BLAST; <sup>e</sup> ausgenommen ist das jeweils als Hybridisierungssonde verwendete YAC, das in allen Fällen positiv war; die mit einem Sternchen versehenen Klone wurden nicht durchsequenziert, der Identitätswert bezieht sich bei ihnen auf die verlässliche Sequenz von ca. 600 bp.

ergab sich eine zweite, bisher unbekannte cDNA-Konsensussequenz. Im Gegensatz zu der bekannten *SPRR3*-Sequenz, die im Mittelteil 14 charakteristische, aus 24 Basenpaaren bestehende Sequenzwiederholungen aufweist [Gibbs *et al.*, 1993], fehlte dem neu entdeckten Transkript eine dieser Einheiten (Abb. 29). Weitere Unterschiede fanden sich in zehn einzelnen, substituierten Nucleotiden, von denen

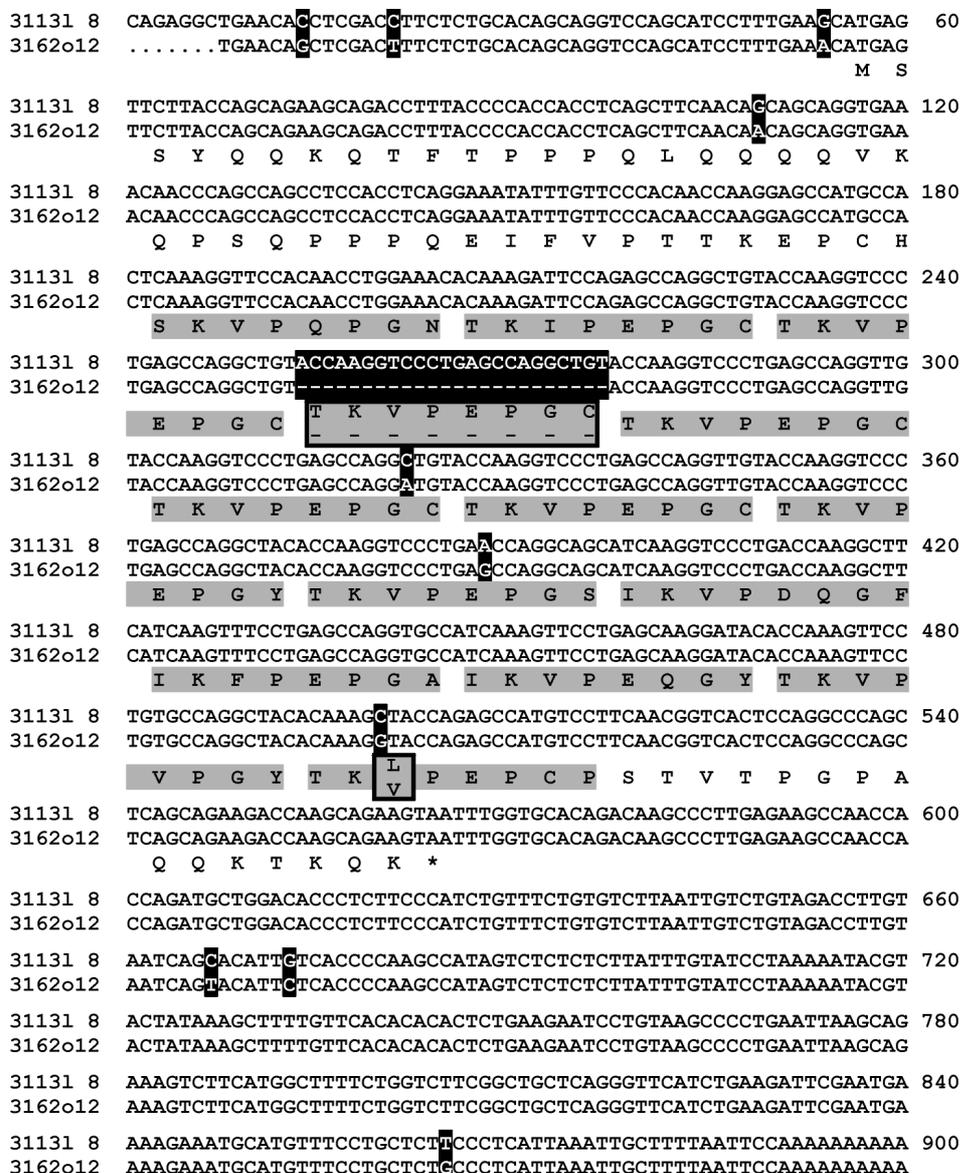


Abbildung 29: Vergleich der cDNA- und Protein-Sequenzen der zwei *SPRR3*-Transkripte. Klon 311318 enthält die bekannte *SPRR3*-Sequenz, Klon 3162o12 die neu identifizierte cDNA-Sequenz. Abweichungen in der Nucleotidsequenz sind schwarz unterlegt, Abweichungen in der abgeleiteten Aminosäuresequenz sind eingerahmt. Die Octapeptid-Wiederholung, die 14 mal in Klon 311318 und 13 mal in Klon 3162o12 enthalten ist, ist grau unterlegt.

vier innerhalb der codierenden Region lagen. Daß diese vier Ersetzungen lediglich eine ausgetauschte Aminosäure zur Folge haben, wodurch eine höhere Übereinstimmung mit dem den Wiederholungen zugrundeliegenden Konsensuspeptid TKVPEPGC erzielt wird, weist auf einen starken Selektionsdruck hin. Die ermittelte Proteinsequenz war mit dem Protein SPRC identisch, dessen Vorkommen in der Mundschleimhaut beschrieben worden war [Robinson *et al.*, 1994]. Da bisher nur ein einziges

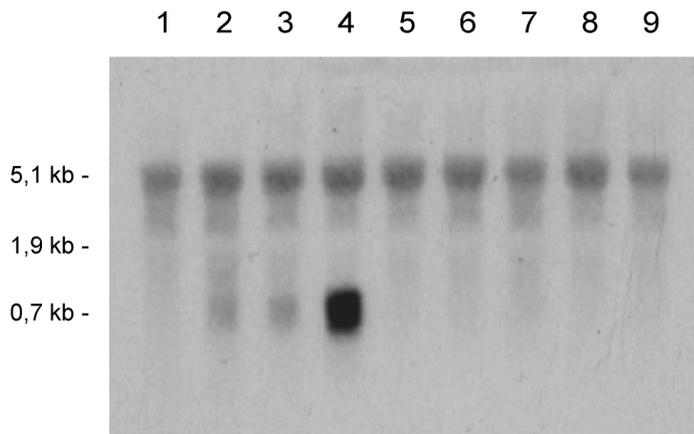
*SPRR3*-Gen im menschlichen Genom nachgewiesen wurde [Gibbs *et al.*, 1993], sind die Sequenzabweichungen höchstwahrscheinlich auf einen Polymorphismus dieses Gens zurückzuführen.

#### 4.3 cDNA-Sequenzen anderer bekannter Proteine

Von den cDNA-Sequenzen, die den 21 erstmals dem EDC zugerechneten Genen entsprachen, codierten elf die Aminosäureabfolge bekannter Proteine. Zu diesen zählten die doppelsträngige RNA-spezifische Adenosindesaminase 1 (*ADARI*), Annexin A9 (*ANXA9*), das HS1-bindende Protein X-1 (*HAX1*), das 34 kD-Vorläuferprotein des Lamininrezeptors (*LAMRL6*), der Proteininhibitor des aktivierten Stat3 (*PIAS3*), die Typ I Phosphatidylinosit-4-phosphat-5-kinase  $\alpha$  (*PIP5K1A*), die 26S-Proteasom-Untereinheiten  $\beta 4$  (*PSMB4*), nicht-ATPase 4 (*PSMD4*) und nicht-ATPase 8 (*PSMD8L*), das RNA-bindende Motiv Protein 8 (*RBM8*) und das *non muscle* Tropomyosin TM30nm (*TPM3*) (Tab. 15). Die Sequenzen der isolierten cDNA-Klone stimmten erneut zu mehr als 99% mit den entsprechenden Einträgen in den Datenbanken überein.

#### 4.4 cDNA-Sequenzen unbekannter Funktion

Unter den EDC-spezifischen Klonen wurden mehrere cDNA-Sequenzen identifiziert, denen kein Gen mit bekannter Funktion zugeordnet werden konnte. Ihr Vorkommen in der Keratinozyten-cDNA-Bibliothek und ihre Lokalisierung im EDC führten zu der Bezeichnung „neu identifizierte cDNA-Sequenzen des EDC“ (NICE). *Alignment*-Untersuchungen ergaben, daß die NICE-Sequenzen von mindestens zehn verschiedenen Genen stammten (*NICE1* bis *NICE10*), von denen fünf alternativ gespleißte Transkripte exprimierten (Tab. 16, Abb. 32). Neben der Suche nach homologen Nucleotidsequenzen (Tabs. 16 und 17) wurden bei ihnen die voraussichtlichen ORFs bestimmt; in acht Fällen war daraufhin die Charakterisierung von Proteinsequenzen möglich (Tab. 16). Zusätzlich wurde für die *NICE*-Gene der zentralen und distalen EDC-Region eine Northern-Blot-Hybridisierung durchgeführt



**Abbildung 30: Expressionsanalyse für *NICE1*.** Ein Northern-Blot, der RNA unterschiedlicher Zellpopulationen enthält, wurde mit der <sup>32</sup>P-markierten *NICE1*-Sonde hybridisiert. Pro Bahn wurden 20µg Gesamt-RNA aufgetragen, die aus den folgenden Geweben bzw. kultivierten Zellen isoliert wurde: 1) Haut des Menschen; 2) primäre Keratinozyten des Menschen, in Ca<sup>2+</sup>-freiem Medium kultiviert [Fischer *et al.*, 1999]; 3) primäre Keratinozyten des Menschen, mit 1,8 mM Strontium in proliferativem Stadium kultiviert [Praeger *et al.*, 1987]; 4) primäre Keratinozyten des Menschen, mit 1,8 mM Calcium in differenzierendem Stadium kultiviert; 5) HaCaT-Zellen (Keratinozytenzelllinie); 6) HeLa-Zellen (epitheloide Tumorzelllinie); 7) SCC4 (*squamous cell carcinoma*)-Zellen (Plattenepithelkarzinomzelllinie); 8) primäre Fibroblasten des Menschen; 9) primäre Melanozyten des Menschen. Die obere Bande ist auf eine Kreuzhybridisierung mit 28S-ribosomaler RNA (rRNA) zurückzuführen. Als Größenstandard dienen die Banden der 28S- (5,1 kb) und 18S-rRNA (1,9 kb).

führt, um die Abhängigkeit der Genexpression vom Differenzierungsstadium der Keratinozyten zu untersuchen (Tab. 18). Die Expression der *NICE*-Gene in verschiedenen Zelltypen und Geweben wurde über die Herkunft der in den Datenbanken enthaltenen EST-Sequenzen analysiert (Tab. 19).

*NICE1* war die am häufigsten vorkommende cDNA-Sequenz der neuen Gene. Die komplette cDNA besteht aus 694 bp und enthält den STS-Marker SHGC-57801. ESTs des Menschen stammten ausschließlich vom Herz. Die Expression in Keratinozyten war stark differenzierungsabhängig; die höchste Konzentration wurde nach Induktion der Differenzierung durch Calcium nachgewiesen (Abb. 30).

**A**

```

GCCCATTCAGTTGGAGAACGTAGTGAGTCTTTCAGTGGAGCCAGGGTCT 50
GGTTTGTCTGAGGAGCTCCGCGATGTCCTCTCAACAGAGCGCCGTTTCC 100
                M S S Q Q S A V S
GCCAAAGGCTTTTCCAAGGGGTCGTCCCAGGGCCCCGCTCCGTGTCCCGC 150
A K G F S K G S S Q G P A P C P A
CCCGGCGCCACCCGGCGCCCGCTCCTCCTCCTCCTGCTGTGGCTCCG 200
P A P T P A P A S S S S C C G S
GCAGGGGCTGCTGCGGCGACTCAGGCTGCTGCGGCTCCAGCTCCACCAGT 250
G R G C C G D S G C C G S S S T S
TGCTGCTGCTTCCCAAGGAGACGCCGTCGACAGCGGAGTAGTGGTTGCTG 300
C C C F P R R R R R Q R S S G C C
CTGCTGCGGGGGCGGCAGCCAGAGGTCCCAGCGCTCCAACAACCGGAGCT 350
C C G G G S Q R S Q R S N N R S
CAGGATGCTGCTCCGGCTGCTGAGAGGCCCGCAACCCCCAGCGCTGCGCT 400
S G C C S G C *
AGAGAAACCCGCCAGCCAGAGCGGGCCCCGCCCGCTGCGGCTCCCACG 450
CGGGGCTGGGCCTCGGAGTTTGCCCCGTAAAGCGAATTGCACTTTGATGT 500
TCAGAAACCCACTTTGTTCTCAGCCACGAAAACCTCCCTGACCCCGATGT 550
GATTTTCTCCCCGGGGATTTCGAGAGCCATGCGTGGGACACTGGACCCTA 600
CTGTCTACACGGGCTTGCACACAGCAGGTGCTCAGCAAATGCCTATTGAT 650
TTGATTGTCTTTGAAGATGTCATAATAAAGCTTCTACCTCCTGAAAAAA 700
    
```

**B**

```

NICE1 10 20 30
MSSQQSAVSAKGFSGSSQGPAPCPAPAPT-----PAPASS--SSC
xp5   MSCQQNQQQCQPPPKCPPKCTPKCPPKCPPKCLPQC PAPCSPAVSSC
      10 20 30 40
NICE1 40 50 60 70
CG--SGRGCCGDS--GCCGSSSTS--CF--PR--RRRRQRSSGCC
xp5   CGPIISG--GCCGPPSSGGCCNSGAGGCCLSHHRPRLFHRRRHQSPDCC-
      50 60 70 80 90
NICE1 80 90 99
CGGGSQRSQRSNNRS--SG--C
xp5   -----ESEPSSG--SGCCHSSGGCC
      100 110
    
```

Abbildung 31: *NICE1*-Sequenzanalyse. (A) cDNA-Sequenz und die aus dem größten ORF abgeleitete Aminosäuresequenz. Das vermeintliche Methionin-Startcodon und das Stopcodon sind grau unterlegt. Das Polyadenylierungssignal ist unterstrichen. (B) Vergleich der von *NICE1* und *xp5* abgeleiteten Proteinsequenzen. Identische Aminosäuren sind schwarz unterlegt, homologe grau; Lücken sind durch Striche gekennzeichnet. *Alignment* der Sequenzen wurde unter <http://vega.crbm.cnrs-mop.fr/bin/align-guess.cgi> durchgeführt.

### III ERGEBNISSE

Das vorhergesagte Protein setzte sich aus 99 aa zusammen (Abb. 31A), war wie Loricrin [Hohl *et al.*, 1991b] reich an Serin, Cystein und Glycin und enthielt mehrere Glutamin- und Lysinreste, die durch Transglutaminasen vernetzt werden könnten. Darüberhinaus zeigte es Ähnlichkeiten (36% Übereinstimmung) mit der vom xp5-Gen hergeleiteten Proteinsequenz (Abb. 31B); xp5 kartiert ebenfalls im EDC [Zhao & Elder, 1997].

**Tabelle 16: Charakterisierung der NICE-Sequenzen**

Gen	cDNA-Größe <sup>a</sup>	BLASTN <sup>b</sup> : Sequenz, GenBank Acc (Größe), Überlappung (Identität)	BLASTX <sup>c</sup> : Protein, GenBank Acc (Größe), Überlappung (Identität)	ORF <sup>d</sup>	Proteindomänen <sup>e</sup>
<i>NICE1</i>	694 bp	EST AI096376 (482 bp), 479 bp (99%)	skin-specific protein xp5, NP055172 (110 aa), 99 aa (36%)	99 aa	-
<i>NICE2</i>	2409 bp	-	-	-	-
<i>NICE3</i>	1391 bp	cDNA HSPC012, AF077036 (1636 bp), 1391 bp (100%)	Protein aus HSPC012, AAD27769 (219 aa), 197 aa (100%)	>197 aa (253 aa)	-
	1554 bp	cDNA HSPC012, AF077036 (1636 bp), 1554 bp (99%)	Protein aus HSPC012, AAD27769 (219 aa), 219 aa (100%)	219 aa	-
<i>NICE4</i>	1554 bp	cDNA HSPC012, AF077036 (1636 bp), 1540 bp (99%)	Protein aus HSPC012, AAD27769 (219 aa), 201 aa (100%)	201 aa	-
	3976 bp	cDNA KIAA0144, D63478 (3411 bp), 2944 bp (100%)	Protein aus KIAA0144, BAA09765 (983 aa), 968 aa (100%)	1167 aa	Ubiquitin assoziierte/ TS-N Domäne
<i>NICE5</i>	880 bp	EST AA769605 (701 bp), 680 bp (99%)	<i>D. melanog.</i> Genprodukt 25E8.2, CAA15712 (394 aa), 136 aa (73%)	>141 aa	-
<i>NICE6a</i>	2543 bp	cDNA KIAA0488, AB007957 (6388 bp), 1910 bp (99%)	<i>C. elegans</i> Genprodukt F25H2.2, T21345 (578 aa), 423 aa (37%)	484 aa	PhoX homologe Domäne, ezrin/radixin/moesin (ERM) Proteindomäne
<i>NICE6b</i>	1311 bp	cDNA FLJ10533, AK001395 (2839 bp), 1307 bp (99%)	-	-	-
<i>NICE7</i>	2368 bp	EST AW370938 (605 bp), 599 bp (99%)	-	(79 aa)	-
	1277 bp	EST AI936722 (503 bp), 503 bp (99%)	-	79 aa	-
<i>NICE8</i>	1267 bp	cDNA DKFZp434F1535, AL137324 (1374 bp), 1267 bp (100%)	-	-	-
<i>NICE9</i>	1103 bp	<i>NOTCH2</i> -mRNA, AF308601 (9722 bp), 672 bp (98%)	<i>NOTCH2</i> , AAA36377 (2471 aa), 224 aa (9%)	236 aa	Calcium-bindende EGF-ähnliche Domäne
<i>NICE10a</i>	2253 bp	cDNA 24976, AF131738 (1892 bp), 1873 bp (97%)	Protein aus cDNA FLJ20719, BAA91343 (766 aa), 143 aa (97%)	>244 aa	-
<i>NICE10b</i>	1671 bp	EST AW363847 (668 bp), 650 bp (99%)	-	-	-

<sup>a</sup> Die Größe des am weitesten synthetisierten Klon (Abb. 32) ohne Poly(A)<sup>+</sup>-Schwanz ist angegeben; <sup>b</sup> BLASTN, Nucleotid-BLAST; <sup>c</sup> BLASTX, übersetzender BLAST; <sup>d</sup> offene Leserahmen ab 75 aa wurden berücksichtigt, unvollständige ORFs (s. Text) sind mit > gekennzeichnet; in Klammern die sich aus den aufgeführten cDNA-Sequenzen (Abb. 32) ergebenden Proteingrößen; <sup>e</sup> Ergebnisse der mit der Proteinsequenz durchgeführten Suche nach konservierten Domänen (*Conserved Domain Search*); TS-N, N-Terminus des *elongation factor TS*; EGF, *epidermal growth factor*.

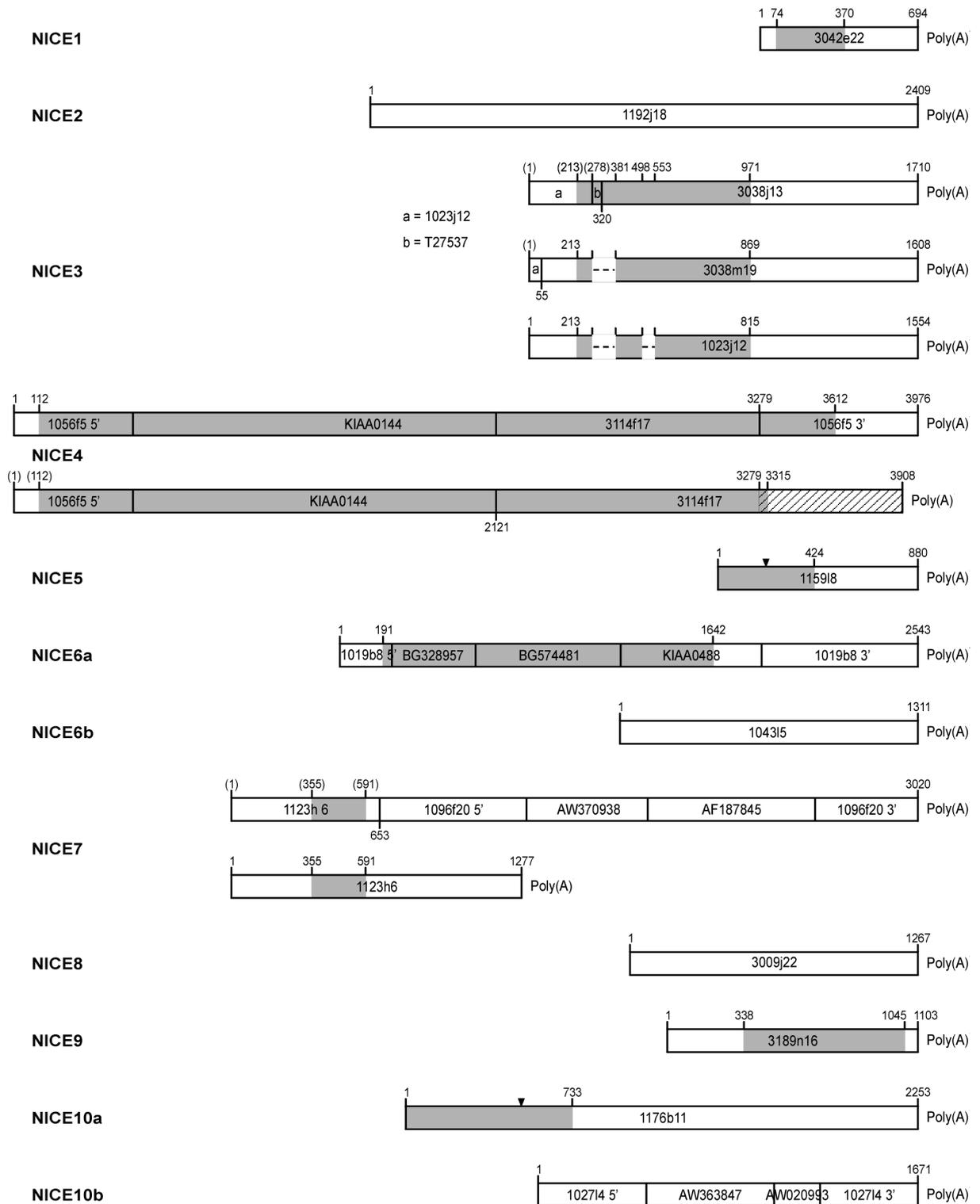


Abbildung 32: *NICE*-Genprodukte. Der jeweils größte identifizierte cDNA-Klon für jedes der Transkripte ist schematisch dargestellt. Das erste und letzte Nucleotid der jeweiligen cDNA-Sequenz, der ORFs und, bei alternativem Spleißen, der enthaltenen Exons ist angegeben. Bei vollständigen cDNA-Sequenzen stehen die Nucleotide der ergänzten Sequenzen in Klammern, der Beginn des entsprechenden cDNA-Klons ist unter dem Transkript angegeben. ORFs ab 75 aa sind grau unterlegt, bei offensichtlich unvollständigen ORFs ist der erste Methioninrest in der Sequenz des cDNA-Klons mit einem Dreieck gekennzeichnet. Nicht übereinstimmende Nucleotidsequenzen alternativ gespleißter Transkripte eines Gens sind schraffiert. Die Transkripte enthalten die Bezeichnungen der entsprechenden cDNA-Klone sowie die GenBank Acc der für die Vervollständigung verwendeten Sequenzen überlappender ESTs. NICE6a und NICE6b sowie NICE10a und NICE10b zeigen keine Übereinstimmung in der Sequenz und stammen möglicherweise von unterschiedlichen Genen.

Alle die NICE2-Sequenz enthaltenden cDNA-Klone waren unvollständig, sie enthielten keinen signifikanten ORF. Die Suche nach homologen Sequenzen ergab einzig eine 85%ige Übereinstimmung mit dem Hybridisierungsmarker 24f59, der aus einem YAC des EDC-Contigs subkloniert worden war. Die Expression in Keratinozyten war durchgängig schwach. Offensichtlich handelt es sich um ein größeres nicht translatiertes 3'-Ende.

Drei unterschiedlich gespleißte Produkte des *NICE3*-Gens wurden in Keratinozyten identifiziert. Obwohl nur eine cDNA komplett war, konnten auch die anderen zwei Sequenzen mit Hilfe überlappender Klone vervollständigt werden; die daraus resultierenden Größen lagen bei 1710 bp (Klon 3038j13 mit Sequenzergänzung), 1608 bp (Klon 3038m19 mit Sequenzergänzung) und 1554 bp (Klon 1023j12). In der 3'-terminalen Region stimmten die cDNA-Sequenzen überein, alle drei enthielten die STS Marker SHGC-11135 und TIGR-A002G29. Abweichungen fanden sich innerhalb der codierenden Region: cDNA 3038m19 fehlte Exon 279-380, cDNA 1023j12 die Exons 279-380 und 499-552.

Weitere gespleißte Exons, die außerdem in unterschiedlichen Kombinationen zu beobachten waren, konnten in ESTs des gleichen Gens identifiziert werden: Exon 202-278 in EST AA354455 (GenBank Acc) aus T-Zellen, Exons 202-278 und 279-380 in EST AA463392 aus dem Fötus, Exon 381-498 in EST AA057488 aus dem Dickdarm, Exons 279-380 und 381-498 in EST Z42265 aus dem Gehirn. Insgesamt hatten die ESTs dieses Gens ihren Ursprung in den unterschiedlichsten Geweben. In Keratinozyten war die Genexpression schwach und nicht vom Differenzierungsgrad der Zellen abhängig. Von den drei vorhergesagten Proteinen (253 aa, 219 aa, 201 aa) war das 253 aa-Protein nahezu identisch mit der ebenfalls von einer cDNA hergeleiteten Aminosäuresequenz HSPC012.

Auch vom *NICE4*-Gen kamen in

**Tabelle 17: In den *NICE*-Genen enthaltene STS-Marker**

Gen	STS-Marker, GenBank Acc (Größe), Identität	positive YACs <sup>a</sup>
<i>NICE1</i>	SHGC-57801, G41921 (110bp), 100%	890_e_4, 874_d_5, 907_e_6, <b>986_e_10</b>
<i>NICE2</i>	-	100_f_3, 692_c_1, 955_e_11, <b>950_e_2</b>
<i>NICE3</i>	SHGC-11135, G13549 (186bp), 100%; TIGR-A002G29 G26404 (144bp), 100%	713_h_12, 951_f_6, <b>950_e_2</b>
<i>NICE4</i>	SHGC-33740, G29465 (127bp), 100%	643_h_5, 713_h_12, 951_f_6, <b>950_e_2</b>
<i>NICE5</i>	-	951_f_6, 954_a_11, <b>950_e_2</b>
<i>NICE6a</i>	-	747_f_12, 763_g_8, <b>764_a_1</b> , 776_h_9, 811_c_12, 874_d_5, 890_e_4, 907_e_6
<i>NICE6b</i>	-	747_f_12, 763_g_8, <b>764_a_1</b> , 776_h_9, 811_c_12, 874_d_5, 890_e_4, 907_e_6
<i>NICE7</i>	-	<b>764_a_1</b> , 776_h_9
<i>NICE8</i>	WI-9627, G05593 (276 bp), 98%	<b>764_a_1</b>
<i>NICE9</i>	-	789_f_2, 929_b_2, <b>934_g_9</b> , 954_e_4
<i>NICE10a</i>	SHGC-57369, G37462 (305 bp), 92%	789_f_2, 929_b_2, <b>934_g_9</b> , 954_e_4
<i>NICE10b</i>	SHGC-34018, G27852 (131 bp), 100%	789_f_2, 929_b_2, <b>934_g_9</b> , 954_e_4

<sup>a</sup> Die fettgedruckten YACs wurden zur Identifizierung der den jeweiligen Genen entsprechenden cDNA-Klone eingesetzt. Von den instabilen YACs 764\_a\_1 und 950\_e\_2 wurden die isolierten Kolonien #1 bzw. #9 als Hybridisierungssonden eingesetzt.

Keratinocyten zwei verschiedene Expressionsprodukte vor. Die entsprechenden cDNA-Sequenzen unterschieden sich in ihren 3'-Enden, wobei Klon 3114f17 den STS-Marker SHGC-33740 einschloß. Die Größen der vollständigen Transkripte waren 3976 bp für Klon 1056f5, der offenbar die komplette cDNA enthielt, und 3913 bp für Klon 3114f17, der nur die 3'-Region

**Tabelle 18: Northern-Expressionsanalyse für die Gene *NICE1-NICE5***

Gen	Haut	primäre Keratinocyten			HaCaT-Zellen	HeLa-Zellen	SCC4-Zellen	Fibroblasten	Melanocyten
		-Ca <sup>2+</sup>	+Sr <sup>2+</sup>	+Ca <sup>2+</sup>					
<i>NICE1</i>	o	+	+	++	-	-	o	-	-
<i>NICE2</i>	o	o	o	o	o	o	o	o	o
<i>NICE3</i>	o	o	o	o	o	+	+	+	o
<i>NICE4</i>	-	-	-	o	o	+	+	o	o
<i>NICE5</i>	o	o	o	o	o	+	+	o	+

++ starke bis sehr starke Expression, + mittlere Expression, o schwache bis sehr schwache Expression; - Expression nicht nachweisbar; HaCaT, Keratinocytenzelllinie; HeLa, epitheloide Tumorzelllinie; SCC4 (squamous cell carcinoma), Plattenepithelkarzinomzelllinie.

abdeckte und mit der überlappenden Sequenz von cDNA 1056f5 aufgefüllt wurde. Für die Ergänzung des mittleren Teils der Sequenz wurde die überlappende cDNA KIAA0144 herangezogen, ein weiteres *NICE4*-Genprodukt, das aus einer Knochenmarkzelllinie isoliert worden war [Nagase *et al.*, 1995]. KIAA0144 (3416 bp) zeigte lediglich eine geringe Abweichung in der nicht codierenden 5'-terminalen Region und enthielt ein drittes 3'-Ende. Dementsprechend wiesen auch die vorhergesagten Proteine – sie bestanden aus 1167 aa und 1068 aa sowie aus 984 aa für KIAA0144 – unterschiedliche C-Termini auf. ESTs des *NICE4*-Gens deckten eine Vielzahl an Geweben ab, in Keratinocyten konnte nur eine geringe mRNA-Konzentration nachgewiesen werden.

Für *NICE5* wurde eine unvollständige Konsensussequenz ermittelt, eine Vielzahl an ESTs überlappten mit dem 5'-Ende der Sequenz. Die größte isolierte cDNA kam vom Poly(A)<sup>+</sup>-Schwanz ausgehend auf 880 bp. Die sich aus der Basenabfolge ergebende größtmögliche Proteinsequenz bestand aus 144 aa; ihr fehlte offensichtlich der N-Terminus mit dem initialen Methioninrest. Sie zeigte eine mehr als 70%ige Übereinstimmung mit zwei Proteinen, welche dem *Drosophila melanogaster* Gen EG:25E8.2 und dem *Caenorhabditis elegans* Gen F25H2.8 (GenBank Acc Z79754) entsprachen. Die EST-Analyse ergab eine breite Gewebeverteilung der Expressionsprodukte, in Keratinocyten war die Expression nur schwach.

Zwei cDNA-Sequenzen, die keine Gemeinsamkeiten erkennen ließen und mit unterschiedlichen Abschnitten der cDNA KIAA0488 [Seki *et al.*, 1997] überlappten, wurden für das *NICE6*-Gen ermittelt. Im Vergleich zu KIAA0488 ließ Klon 1019b8 (*NICE6a*), dessen Mittelteil aus ESTs des gleichen Gens zusammengesetzt wurde und dessen voraussichtliche Sequenz 2543 bp umfaßt, eine entgegengesetzte Leserichtung erkennen. *NICE6a* war die einzige cDNA dieses Gens, die einen größeren ORF enthielt; das daraus resultierende Protein bestand aus 484 aa und stimmte zu 37% mit dem *Caenorhabditis elegans* Genprodukt F25H22 überein. Im Gegensatz dazu deckte die wahrscheinlich unvollständige *NICE6b*-Sequenz (1311 bp) das nicht translatierte 3'-Ende eines Transkripts ab, welches als cDNA FLJ10533 identifiziert wurde. Da *NICE6a* und *NICE6b* unterschiedliche Transkripte darstellen und bezogen auf ihre jeweilige Übereinstimmung mit der KIAA0488-Sequenz unterschiedliche Lese-

**Tabelle 19: EST-Analyse für die *NICE*-Gene**

Gen	EST-Cluster (UniGene)	Lokalisierung (GenMap99)	EST-Ursprung (UniGene)
<i>NICE1</i>	Hs.110196	D1S514-D1S2635	Herz
<i>NICE2</i>	-	-	-
<i>NICE3</i>	Hs.31989	D1S514-D1S2635	u.a. Bauchspeicheldrüse, Blut, Dickdarm, Embryo, Gallenblase, Gebärmutter, Gehirn, Haut, Herz, Hoden, Leber, Lunge, Magen, Muskel, Nebenschilddrüse, Niere, Plazenta
<i>NICE4</i>	Hs.8127	D1S514-D1S2635	u.a. Bauchspeicheldrüse, Blut, Dickdarm, Embryo, Gebärmutter, Gehirn, glatte Muskulatur, Herz, Hoden, Knochen, Lunge, Magen, Niere, Plazenta, Thymus
<i>NICE5</i>	Hs.107538	D1S514-D1S2635	u.a. Bauchspeicheldrüse, Dickdarm, Embryo, Gallenblase, Gebärmutter, Gehirn, Herz, Hoden, Lunge, Magen, Niere, Plazenta, Schilddrüse, Vorhaut
<i>NICE6</i>	Hs.67619	D1S514-D1S2635	u.a. Blase, Blut, Dickdarm, Embryo, Gebärmutter, Gehirn, Herz, Hoden, Lunge, Magen, Muskel, Niere, Plazenta, Thymus
<i>NICE7</i>	Hs.22065	D1S514-D1S2635	u.a. Bauchspeicheldrüse, Blase, Blut, Dickdarm, Gebärmutter, Gehirn, Haut, Herz, Hoden, Keimzelle, Knochen, Lunge, Magen, Milz, Nebenschilddrüse, Niere, Plazenta
<i>NICE8</i>	Hs.78524	D1S514-D1S2635	u.a. Bauchspeicheldrüse, Dickdarm, Gebärmutter, Gehirn, Haut, Herz, Hoden, Keimzelle, Knochen, Leber, Lunge, Muskel, Nebenschilddrüse, Niere, Plazenta
<i>NICE9</i>	Hs.8121 ( <i>NOTCH2</i> )	D1S418-D1S514	u.a. Bauchspeicheldrüse, Blase, Blut, Dickdarm, Embryo, Gallenblase, Gebärmutter, Gehirn, Haut, Herz, Hoden, Knochen, Leber, Lunge, Magen, Muskel, Nebenschilddrüse, Niere, Plazenta
<i>NICE10</i>	Hs.218329	D1S514-D1S2635	u.a. Bauchspeicheldrüse, Blase, Blut, Dickdarm, Embryo, Gallenblase, Gebärmutter, Gehirn, Haut, Herz, Hoden, Knochen, Leber, Lunge, Magen, Muskel, Niere, Plazenta

richtungen aufweisen müßten, stammen die cDNA-Sequenzen möglicherweise nicht vom selben Gen. Die KIAA0488-Sequenz wäre demnach chimär. ESTs dieses Gens bzw. dieser Gene hatten ihren Ursprung in diversen Geweben.

Eindeutig auf ein Gen konnten die zwei *NICE7*-Sequenzen zurückgeführt werden, wobei der aus 1277 bp bestehende Klon 1123h6 ein vollständiges Transkript mit einem ORF von 79 aa repräsentierte. Die Sequenz von Klon 1096f20 wurde mit Hilfe überlappender ESTs vervollständigt; sie enthielt lediglich ein verlängertes 3'-Ende und ergab unter Berücksichtigung des 5'-Endes von Klon 1123h6 eine Transkriptgröße von 3020 bp. *NICE7*-ESTs kamen in den unterschiedlichsten cDNA-Bibliotheken vor.

Die größte vom *NICE8*-Gen identifizierte cDNA-Sequenz setzte sich aus 1267 bp zusammen. Sie entsprach dem 3'-Ende der cDNA DKFZp434F1535 und schloß den STS-Marker WI-9627 ein. Da sie keinen signifikanten ORF enthielt und diverse ESTs mit dem 5'-Ende überlappten, handelt es sich

wahrscheinlich erneut um ein nicht codierendes 3'-Ende. Exprimierte Sequenzen dieses Gens fanden sich ubiquitär im menschlichen Organismus.

Aus 1103 bp bestand die NICE9-Sequenz, die über einen Bereich von 672 bp zum *NOTCH2*-Transkript des Menschen homolog war. Sie codierte ein 236 aa-Protein, dessen Aminosäuresequenz eine Domäne des Notch2 enthielt. Die identifizierten ESTs des *NICE9*-Gens, die aus den unterschiedlichsten Geweben stammten, werden in Datenbanken aufgrund dieser partiellen Übereinstimmung möglicherweise zu Unrecht dem *NOTCH2*-Gen zugeordnet.

Wie bei den NICE6-Sequenzen wurden auch die ESTs, die mit Teilen der cDNA-Sequenzen NICE10a bzw. NICE10b übereinstimmten, einem einzigen Gen zugerechnet, obwohl sie keine Gemeinsamkeiten aufwiesen. NICE10a umfaßte 2253 bp, welche die partielle cDNA 24976 und den STS-Marker SHGC-57369 beinhalteten. Die Sequenz schloß außerdem einen ORF ein, der sich mit Hilfe überlappender ESTs über das 5'-Ende hinaus verlängern ließ. Die gesamte codierende Sequenz ergab ein 244 aa-Protein, das demzufolge wahrscheinlich unvollständig war. Für NICE10b, vermutlich ein weiteres relativ großes nicht translatiertes 3'-Ende eines Transkripts, konnten lediglich ESTs desselben Gens ermittelt werden, welche die 1671 bp-cDNA-Sequenz komplettierten und auch mit dem 5'-Ende überlappten. Mit SHGC-34018 war erneut ein STS-Marker enthalten. ESTs für NICE10a und NICE10b waren ebenfalls in cDNA-Bibliotheken weitverbreitet.