

5 ERGEBNISSE

5.1 Ergebnisse aus dem PAC-Klonbanken Screening

Um 11p13 spezifische Klone zu identifizieren, wurden „High Density Grid Filter“ der RPCI 1,3-5 PAC-Klon-Bibliothek hybridisiert. Die einzelnen PAC-Banken bestehen aus je 115.200 (RPCI 1), 75.513 (RPCI 3), 105.251 (RPCI 4) und 142.773 (RPCI 5) Klonen, die doppelt auf die Filter aufgebracht sind. Während die RPCI1-Bank noch 25 % nichtrekombinante Klone enthält, konnte bei den nachfolgenden RPCI-Banken die Anzahl der nichtrekombinanten Klone auf 1 – 8 % reduziert werden (Ioannou et al. 1994). Als Proben wurden PCR-Produkte einer Alu-PCR von 11p13 spezifischen YAC-Klonen (21 Proben), von YAC-Klonen isolierte Endproben (12 Proben), gereinigte YAC-DNA (2 Proben) sowie EST-Proben, anonyme cDNA-Proben und genomische Proben der Region 11p13 verwendet (46 Proben). Letztere wurden in drei Pools verteilt auf die Filter hybridisiert. Die Alu-PCR-Proben und YAC-DNAs wurden von Beate Gawin zur Verfügung gestellt. Alle 26 YAC-Klone, aus denen Hybridisierungsproben generiert wurden, waren zuvor von Gawin et al. (1995) in Form eines 8 Mb großen Contigs auf 11p13–14.1 kartiert worden. Die verwendeten Marker (siehe 3.3) sind ebenfalls in Gawin et al. (Gawin et al. 1995) oder in GDB (Genome Data Base) beschrieben. Insgesamt konnte hiermit eine 11p13-spezifische Bibliothek aus 509 PAC-Klonen erstellt werden. Diejenigen PAC-Klone (297), die mit YAC-Alu-PCR-Proben identifiziert wurden, konnten entsprechend der Kartierung der YAC-Klone im YAC-Contig Abschnitten der genomischen Sequenz zugeordnet werden.

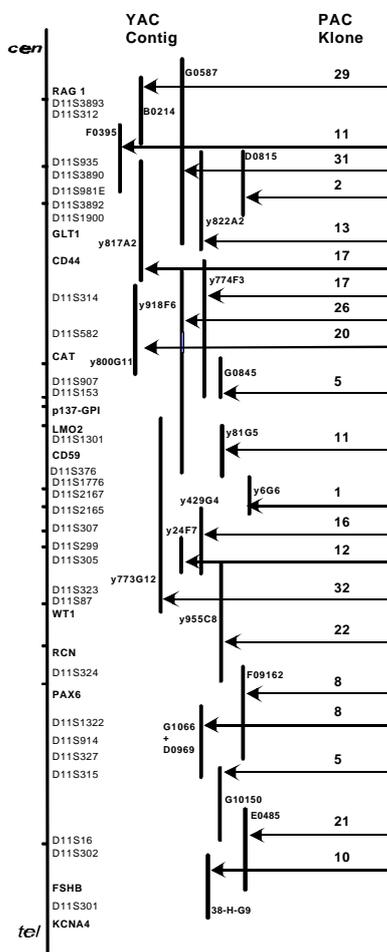


Abb. 1: Verteilung der PAC-Klone, die mit YAC-Alu-PCR-Proben identifiziert wurden.

Die genomische Karte (links) ist einschließlich der Notl-Schnittstellen (Querbalken), der kartierten Gene (fett) und Marker dargestellt. Ihr ist das YAC-Contig zugeordnet (Gawin et al. 1995). Pfeile verweisen auf die YAC-Klone, aus denen die entsprechende Anzahl an PAC-Klonen (rechts) identifiziert wurde.

Die Anzahl der PAC-Klone korreliert hierbei nicht mit der Länge des entsprechenden YAC-Klons, von dem die Hybridisierungsprobe stammt. Der Grund liegt wahrscheinlich in der Verwendung von YAC-Alu-PCR-Proben für die Hybridisierung und der damit verbundenen ungleichmäßigen Verteilung der Alu-Sequenzen auf genomischer Ebene. Aus Regionen, in denen Alu-Sequenzen häufiger vorkommen, kann eine engmaschigere Auswahl genomischer Sequenzen mit den verwendeten Alu-PCR-Primern amplifiziert werden. Mit diesen Proben kann daher auch eine im Verhältnis größere Anzahl an PAC-Klonen identifiziert werden.

Um eine bessere Abdeckung zu erreichen, wurde mit weiteren Proben auf die „High Density Grid“ Filter hybridisiert. Mit Proben aus kompletter YAC-DNA sowie YAC-Endproben konnten zusätzlich 44 PAC-Klone identifiziert werden. Weitere 168 PAC-Klone wurden mit EST-Proben, genomischen Proben, oder cDNA-Proben, die für 11p13 spezifisch waren, identifiziert. Für die Hybridisierungen wurden die Marker in drei Pools zusammengefaßt, so daß die hieraus erhaltenen PAC-Klone vorläufig keinem spezifischen Marker zugeordnet werden konnten. Sie wurden deshalb auch nicht in Abbildung 1 berücksichtigt. Bei der Hybridisierung mit cDNA ergaben einige Klone mit allen Pools positive Signale. Dies ist vermutlich auf Kreuzhybridisierungen mit Sequenzen aus dem pUC-Linker des pCYPAC2N-Vektors und Sequenzen des pBS-Vektors, worin einige der cDNA-Klone kloniert waren, zurückzuführen. Bei PAC-Klonen ohne Insert bleibt der pUC-Linker erhalten. Diese Klone führen deshalb zu falschen positiven Signalen auf den „High Density Grid“ Filtern. In einer weiteren Hybridisierung, bei der der linearisierte pBS-Vektor als Probe verwendet wurde, konnten die falsch positiven Signale aussortiert werden. Dies wird in Abbildung 2 verdeutlicht.

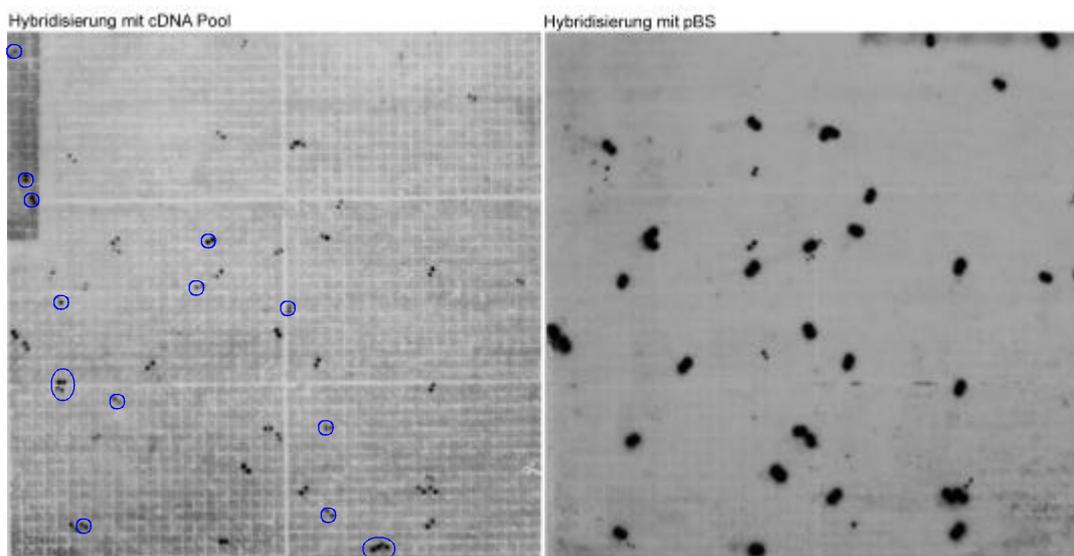


Abb. 2: Hybridisierung von cDNA-Klonen auf „High Density Grid“ Filter.

Die Klone auf den „High Density“ Grid Filtern sind jeweils doppelt aufgetragen, weshalb nach der Hybridisierung doppelte Signale zu sehen sind. Zur Unterscheidung der positiven und falsch positiven Signale bei der Auswertung der Hybridisierungen mit den cDNA-Pools (links) wurde eine Hybridisierung mit dem pBS-Vektor als Probe (rechts) durchgeführt. Anschließend wurden bei den Hybridisierungsergebnissen aus dem cDNA-Pool nur die 15 zusätzlichen Signale markiert (Kreise) und deren Koordinaten zur Bestimmung des jeweiligen PAC-Klons ausgewertet.

5.2 Erstellung des PAC-Contigs

5.2.1 Hybridisierungen auf Koloniefilter

Alle aus den zuvor beschriebenen Experimenten stammenden PAC-Klone wurden auf Koloniefilter aufgebracht, um die DNA für Hybridisierungen zugänglich zu machen. Über die Methode der DOP-Vektor-PCR (Wu et al. 1996) oder der „linker ligated“ PCR (Kere et al. 1992) wurden jeweils das rechte sowie das linke Ende der PAC-Klone isoliert. Um nach überlappenden Klonen zu suchen, wurden die isolierten PAC-Enden auf die zuvor erstellten Koloniefilter hybridisiert. Zusätzlich dienten die 11p13-spezifischen Marker als Hybridisierungsprobe, um einzelne PAC-Klone genau lokalisieren zu können. Da die Proben in einem Pool hybridisiert worden waren, mußten die Hybridisierungen auf die Koloniefilter mit diesen Proben einzeln wiederholt werden. Aufgrund der Hybridisierungsdaten wurden etwa 20 % der PAC-Klone aussortiert, da sie mit verschiedenen Proben kreuzreagierten oder ihr Endfragment bei der Hybridisierung ein zu starkes Hintergrundsignal aufwies. Ein Restriktionsverdau dieser Klone mit EcoRI zeigte, daß dieses Phänomen weitgehend auf den Verlust des Inserts zurückzuführen war. Die Endproben von 30 % der Klone konnten nur auf diese selbst zurückhybridisiert werden, überlappende Klone konnten nicht gefunden werden. Die betreffenden Klone stammten möglicherweise aufgrund falsch positiver Signale oder falscher Auswertung bei der Bestimmung der Koordinaten auf den „High Density Grid“ Filtern nicht von Chromosom 11p13. Ein Beispiel für eine Hybridisierung auf Koloniefilter ist in Abbildung 3 dargestellt.

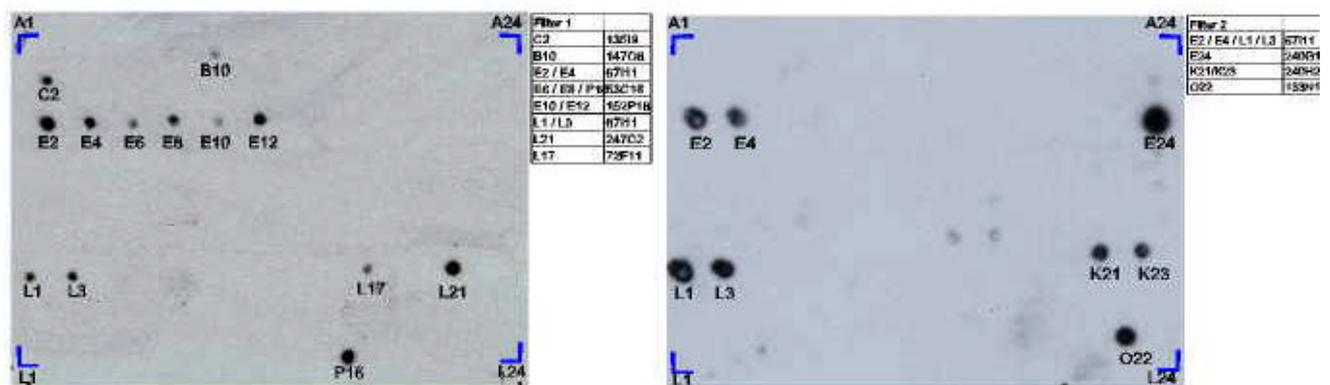


Abb. 3: Hybridisierung der beiden Endproben SP6 (links) und T7 (rechts) des PAC-Klons 67111 auf Koloniefilter. Neben dem PAC-Klon (67117), aus dem die Endproben isoliert wurden, werden zusätzliche Signale überlappender Klone gefunden. Die Umrechnung der Koordinaten in die identifizierten Klone ist in der jeweils nebenstehenden Tabelle angegeben.

Die Auswertung der Hybridisierungsergebnisse aus ca. 1000 Proben erfolgte mit Hilfe des Programms SAM V2.5. Dieses Programm schlägt Contigs vor, indem es Wahrscheinlichkeiten für die Reihenfolge der eingesetzten Marker und Endproben berechnet. Die Option, daß die Endproben jeweils am äußeren Ende des Ursprungsklons zu liegen kommen, wird nur zweitrangig und die Kenntnis der Reihenfolge der genomischen Marker gar nicht berücksichtigt, das heißt die größtmögliche Wahrscheinlichkeit des vorgeschlagenen Contigs wird zugunsten einer schnelleren Rechenoperation

Filter eindeutigere Ergebnisse zu erzielen. Die Anordnung der PAC-DNA auf den Southern Blots, sowie die Auswahl der verwendeten Hybridisierungsproben orientierte sich an den Ergebnissen, die bei den Hybridisierungen auf Koloniefilter erzielt worden waren. Bei einer Hybridisierung auf Southern Blots mit EcoRI verdauter PAC-DNA konnte einfacher zwischen positiven und falsch positiven Signalen unterschieden werden. Bei überlappenden Klonen sollte immer ein EcoRI-Fragment gleicher Länge ein positives Signal ergeben. Das EcoRI-Fragment des Klons, von dem die Endprobe stammt, ist meist kürzer oder länger als das entsprechende genomische EcoRI-Fragment, da die Inserts mit Hilfe eines Mbol-Partialverdaus in den PAC-Vektor kloniert wurden. Abbildung 5 zeigt ein Beispiel einer solchen Hybridisierung.

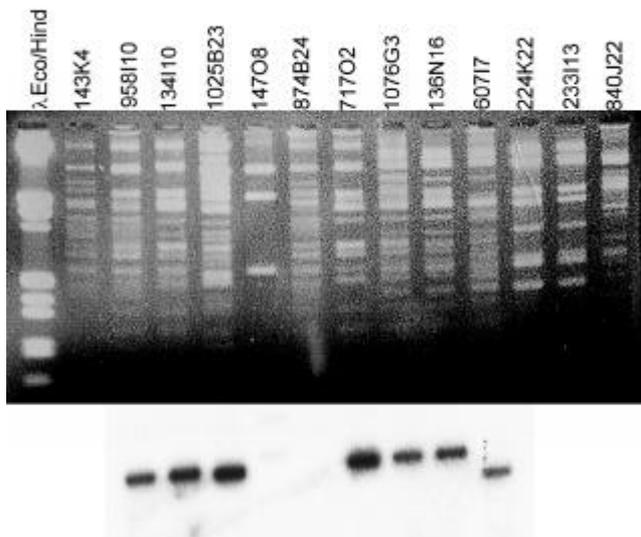


Abb. 5: Auftrennung von EcoRI verdauter PAC-DNA (oben) und anschließende Hybridisierung mit der Endprobe 60717SP6.

Es werden sechs zusätzliche überlappende Klone identifiziert (958I10, 134I10, 1025B23, 717O2, 1076G3, 136N16). Das Endfragment des Klons 607I6 ist kleiner, da es nur einen Teil des genomischen EcoRI-Fragments enthält.

An dieser Stelle wurden auch einige der Hybridisierungen mit EST-Markern, die zuvor auf Koloniefilter hybridisiert worden waren, zur Bestätigung auf Southern Blots wiederholt und zusätzlich PAC-Klone via PCR getestet, ob sie für die Marker M16, D11S2164, D11S305, D11S935 und D11S582 positiv waren. Hierbei wurden nur diejenigen PAC-Klone getestet, die aufgrund vorangegangener Ergebnisse in der Nähe dieser Marker kartiert worden waren.

Durch die Auswertung der Hybridisierungsdaten der Endproben auf Southern Blots sowie der Ergebnisse aus den PCR-Analysen mit SAM V2.5 konnten vorerst 11 Contigs erstellt werden, die zwischen dem Marker D11S3890 und dem Gen PAX6 lokalisiert waren. Ergebnisse aus den Hybridisierungen mit EST-Markern, genomischen Proben und anonymen cDNA-Proben erlaubten eine Verankerung der Contigs mit der genomischen Sequenz. Ein Beispiel für ein vorläufiges PAC-Contig, das aus Hybridisierungsdaten auf Southern Blots mit Hilfe von SAM V2.5 erstellt werden konnte, ist in Abbildung 6 dargestellt.

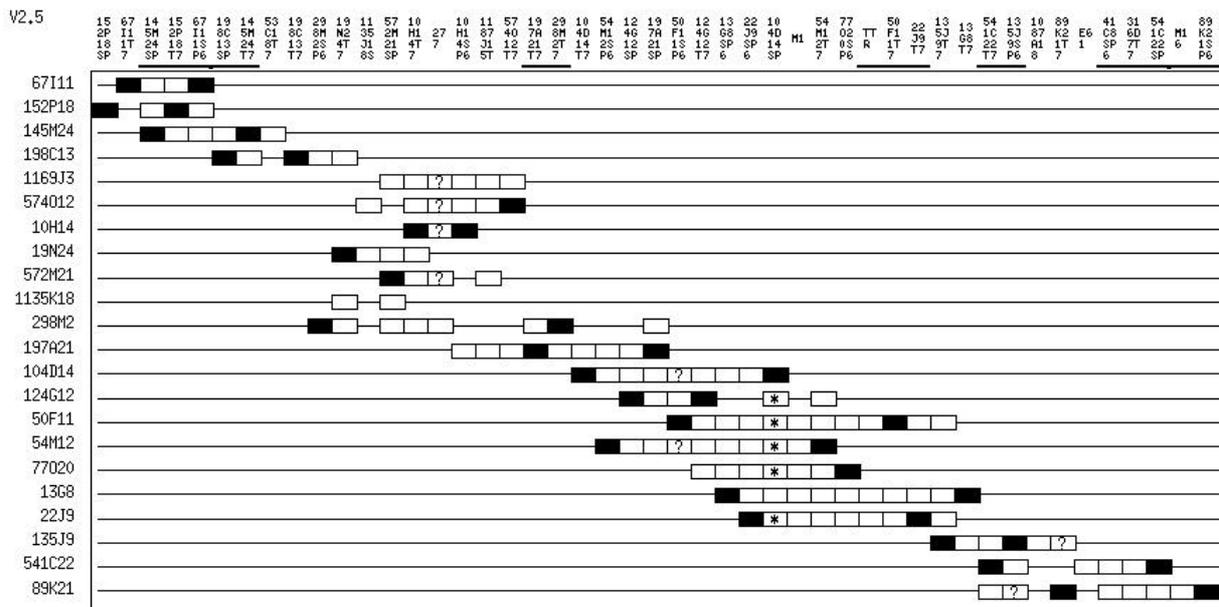


Abb. 6: Auswertung der Daten aus den Hybridisierungen auf Southern Blots mit EcoRI verdauter PAC-DNA mit SAM V2.5. Über wenige Hybridisierungssignale sind drei Contigs miteinander verknüpft: Contig 1 (67I11 – 198C13) mit Contig 2 (1169J3 – 22J9) über die Hybridisierung mit der Probe 19N24T7, Contig 2 mit Contig 3 (135I9 – 98K21) über die Hybridisierung mit den Proben 135I9T7 und 13G8T7. Nicht durchgehende Hybridisierungssignale sind hier auf noch fehlende Hybridisierungen zurückzuführen. (Erklärungen zu den Symbolen siehe Abbildung 4).

Es zeigten sich auch nach der Auswertung der Hybridisierungsdaten auf Southern Blots noch fehlende Signale innerhalb einzelner Klone. Dies war zum größten Teil auf noch fehlende Hybridisierungen zurückzuführen, da der entsprechende Klon nicht auf dem Filter vorhanden war. Konnte trotz Wiederholung dieser Hybridisierung auf neue Filter mit dem entsprechenden Klon kein positives Signal festgestellt werden, wurde der betreffende Klon als deletiert angesehen. Da auch bei den Hybridisierungen auf Southern Blots teilweise eine hohe Hintergrundhybridisierung aufgrund repetitiver Sequenzen Probleme bei der Auswertung bereitete, wurden immer noch positive Signale außerhalb der Contigs gefunden. Teilweise konnte das Problem der repetitiven Sequenzen durch Generieren einer neuen Endprobe unter Verwendung eines alternativen Enzyms (siehe End-PCR 4.4.3) umgangen werden. War dies nicht der Fall, wurde der betreffende Klon nicht in die endgültige Version des erstellten PAC-Contigs miteinbezogen.

5.2.3 Identifizierung neuer PAC-Klone mit PAC-Endproben

Um Lücken zwischen den einzelnen PAC-Contigs zu schließen, mußten zusätzliche Klone aus der PAC-Bibliothek identifiziert werden. Hierzu wurden Endproben der PAC-Klone aus den bisher erstellten Contigs gewählt, mit denen in der 11p13-spezifischen PAC-Bibliothek keine weiteren überlappenden Klone gefunden werden konnten. Insgesamt wurden 109 zusätzliche PAC-Klone gefunden und die 11p13 spezifische PAC-Klon-Bibliothek konnte auf 618 Klone erweitert werden. Von diesen Klonen wurden ebenfalls Southern Blots mit EcoRI verdauter PAC-DNA hergestellt und mit Endproben von Klonen benachbarter Contigs hybridisiert. Aus den positiv identifizierten Klonen

wurden Endproben isoliert und wiederum auf die benachbarten Contigs hybridisiert. Hiermit konnten alle Lücken geschlossen und eine Region von 4,5 Mb abgedeckt werden. Zusätzlich wurde eine höhere Abdeckung der schon vorhandenen Contigs erzielt.

5.2.4 Bestimmung der Insertgröße der PAC-Klone

Die Graphiken der Contigs, die mit Hilfe von SAM V2.5 erstellt werden konnten, zeigen keine reale Darstellung der Größenverhältnisse. Daher wurde zur Bestimmung der Insertgröße der PAC-Klone ein Restriktionsverdau der PAC-DNA mit NotI auf einem Pulsfeld-Gel aufgetrennt. Schnittstellen für NotI befinden sich jeweils flankierend der Klonierungsstelle, so daß das Insert mit voller Länge herausgeschnitten werden kann. Mit Hilfe eines Größenmarkers für DNA-Fragmente zwischen 15 und 300 kb (Mid Range Marker PFG/New England Biolabs) konnte die Größe der Insertfragmente bestimmt werden. In Abbildung 7 ist das Ergebnis eines solchen Pulsfeld-Gels gezeigt.

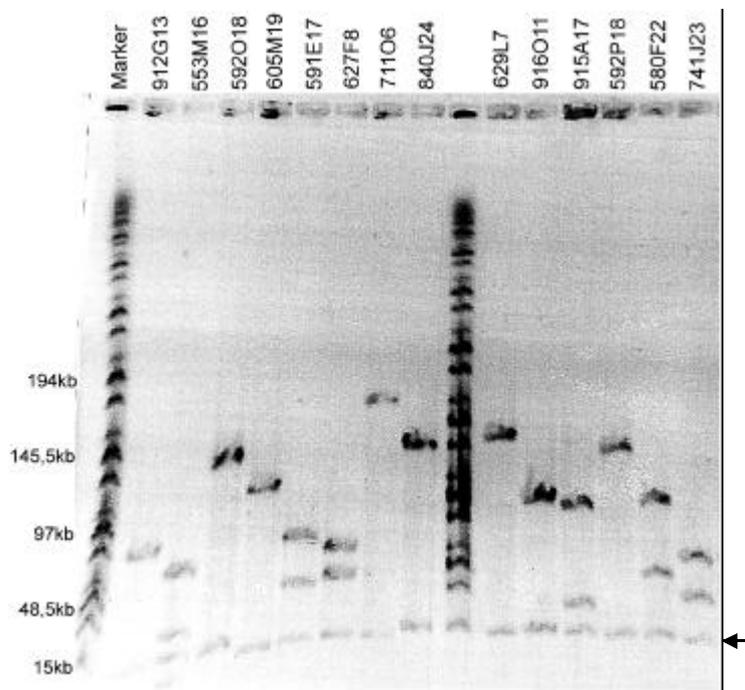


Abb. 7: Auftrennung eines Restriktionsverdaus von PAC-DNA mit NotI auf einem Pulsfeld-Gel.

Das 17 kb Fragment (Pfeil) entspricht der Vektorbande des Vektors pCYPAC2N, die übrigen dem jeweiligen Insert. Im Falle der Klone 553M16, 591E17, 627F8, 915A17 und 580F22 besitzt das Insert eine zusätzliche interne Schnittstelle für NotI, das heißt es entstehen neben dem Vektorfragment zusätzlich zwei oder mehr Fragmente.

Enthält ein PAC-Klon interne Schnittstellen für NotI, konnte diese Information bei einer späteren Kartierung genutzt werden, da von der Region eine Restriktionskarte der NotI-Schnittstellen bekannt war (Gessler und Bruns 1989).

Bei einem Verdau mit NotI oder, wie vorher beschrieben, mit EcoRI konnten Klone ohne Insert identifiziert werden. Der Anteil dieser Klone betrug hierbei etwa 2 %. Diese Zahl kann aber nicht auf die Gesamtzahl der identifizierten PAC-Klone bezogen werden, da zuvor aufgrund von Hybridisierungsartefakten schon ein Teil der Klone von weiteren Analysen ausgeschlossen worden war.

Die Orientierung der einzelnen Fragmente aus dem Verdau mit NotI des jeweiligen Klons konnte durch Hybridisierung auf die getrockneten Pulsfeld-Gele festgelegt werden. Hierbei wurden entweder

Endproben überlappender Klone oder Marker verwendet, die bei einer vorhergehenden Hybridisierung diesen Klon schon als positiv identifiziert hatten.

5.2.5 Erstellung des sequenzierfertigen 4,5 Mb PAC-Contigs

Mit Hilfe der mit SAM V2.5 erstellten Graphiken und den Ergebnissen aus den Pulsfeld-Gelanalysen wurde eine vollständige Karte der überlappenden PAC-Klone gezeichnet. Hierbei wurden nur diejenigen Klone berücksichtigt, die aufgrund der Hybridisierungsergebnisse keine Deletionen, Insertionen oder Doppelinserts aufwiesen. Das PAC-Contig erstreckt sich über 4,5 Mb von den Markern D11S935 und D11S3890 bis zu einem Cosmid-Contig distal des Gens für PAX6, dessen Sequenz zu Beginn der Arbeit bereits bekannt war (GenBank: Z83301, Z83306-83309, Z86001 und Z95332).

Mit Ausnahme von fünf Abschnitten, die jeweils unter 50 kb liegen, konnte eine mindestens zweifache Abdeckung erreicht werden. Zur Erstellung des Contigs zwischen dem Marker D11S935 und dem Gen PAX6 wurden 113 der insgesamt 618 Klone herangezogen. Die übrigen 11p13-spezifischen Klone einschließlich der Hybridisierungsergebnisse, die nicht hier aufgeführt wurden, sind in 11DB, einer für Chromosom 11 eingerichteten Datenbank, unter <http://chr11.bc.ic.ac.uk> zugänglich.

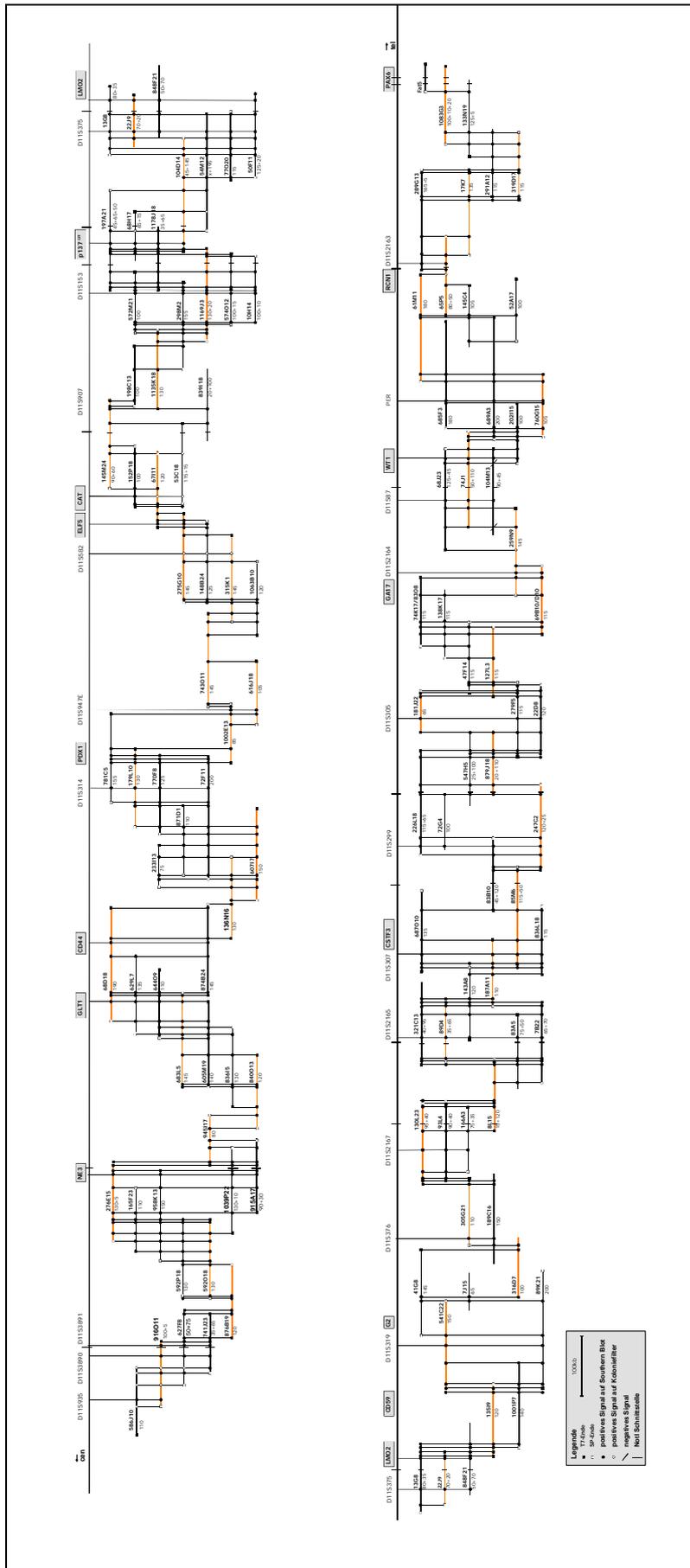


Abb. 8: 4,5 Mb PAC-Contig einschließlich der als Hybridisierungsproben verwendeten Marker und NotI Schnittstellen. Gene sind mit grauen Rechtecken unterlegt. Die Hybridisierungsdaten sind durch vertikale Linien, PAC-Klone durch horizontale Linien dargestellt, wobei der Name der PAC-Klone oberhalb und die Größe unterhalb der Linie zu finden ist. Orangefarben sind diejenigen PAC-Klone, die das Contig mit der kleinsten Überlappung, den „minimal tiling path“, bilden.

5.2.6 Genaue Lokalisation bekannter Gene mit Hilfe des PAC-Contigs

Die beiden kürzlich auf 11p13 kartierten Gene ELF5 (Zhou et al. 1998) und PDX1 (Ling et al. 1998) konnten in das erstellte PAC-Contig integriert und damit genauer lokalisiert werden.

Eine cDNA-Probe des Transkriptionsfaktors ELF5 (zur Verfügung gestellt von Adam Hart) zeigte mit den PAC-Klonen 594L9, 1050H16, 67111, 275G10, 148B24 und 14K24 ein positives Hybridisierungssignal. Entsprechend der Lokalisation dieser Klone im PAC-Contig konnte das Gen zwischen dem Katalasegen (CAT) und dem Marker E748/9 (D11S582) kartiert werden.

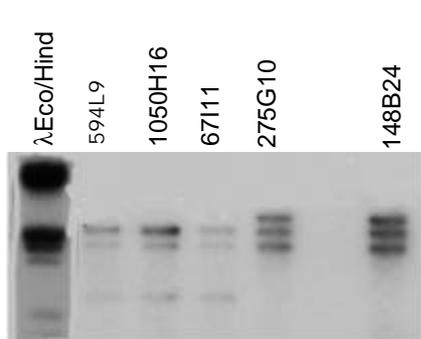


Abb. 9: Hybridisierung von ELF5 auf Southern Blots mit EcoRI verdauter PAC-DNA.

Vier verschiedene Fragmente zeigen ein Signal mit der Probe ELF5, wobei nur zwei davon bei allen genannten Klonen positiv waren. Diese beiden Fragmente aus dem EcoRI-Restriktionsverdau entsprechen einem Exon, das im Bereich der Überlappung der gezeigten PAC-Klone liegt. Weitere Exons, entsprechend den zusätzlichen Fragmenten, liegen entweder nur auf den PAC-Klonen 275G10 und 148B24 oder auf den PAC-Klonen 594L9, 1050H16 und 67111.

Für das Gen der X-Komponente des Pyruvatdehydrogenasekomplexes PDX1 wurden von Ling et al. (Ling et al. 1998) die PAC-Klone 72F11, 179L10 und 136G24 identifiziert. Aufgrund der Einbindung dieser Klone (72F11, 179L10; siehe Abbildung 8) in das erstellte PAC-Contig konnte das Gen direkt der genomischen Sequenz auf 11p13 zugeordnet werden.

5.3 Ergebnisse aus dem Exontrapping

5.3.1 Prinzip des Exontrappings

Für die Identifizierung neuer Gene wurde die Methode des Exontrappings gewählt (Burn et al. 1995). Hierbei wird genomische DNA aus PAC- BAC- oder YAC-Klonen in einen sogenannten Exontrapvektor subkloniert. Dieser enthält jeweils zwei Reporterexons mit einer Spleißdonor und einer Spleißakzeptorstelle. Der hier verwendete Vektor pSPL3-B enthält Exons aus dem β -Globingen des Kaninchens, die durch Teile des tat-Gens aus dem HI-Virus unterbrochen sind. Dieses liefert sowohl den Spleißdonor als auch den Spleißakzeptor. Zusätzlich enthält der Vektor einen SV40-Replikationsursprung (ori), der eine hohe Replikation des Plasmids in eukaryontischen Zellen ermöglicht, die das für die Replikation des SV40-ori notwendige T-Antigen exprimieren. Wird der Vektor mit kloniertem genomischen Insert in eukaryontische Zellen transfiziert, so wird über den Promotor des Reportergens RNA transkribiert. Diese kann über die Spleißstellen des Exontrapvektors gespleißte Exons aus der genomischen Insert-DNA enthalten. Die RNA wird isoliert, revers transkribiert und amplifiziert und das PCR-Produkt anschließend kloniert. Der Vektor pSPL3-B enthält zur Negativselektion von Klonen ohne Insert flankierend zur Klonierungsstelle zwei Hälften einer Restriktionsschnittstelle für BstXI, die beim Fehlen eines externen Exons eine vollständige Restriktionsschnittstelle für BstXI ergeben. Das PCR-Produkt kann so vor dem Klonieren mit BstXI

verdaut werden, wonach in einer zweiten PCR nur diejenigen Produkte amplifiziert und anschließend kloniert werden, die ein zusätzliches, zwischen die zwei Reporterexons gespleißtes Exon besitzen (Church et al. 1994). Von den über diese Klonierung erhaltenen Kolonien werden dann Einzelklone gepickt und analysiert.

5.3.2 Analyse der Klone aus dem Exontrapping

Bei dem in dieser Arbeit durchgeführten Exontrapping wurden sechs PAC-Klone (104D14, 22J9, 1001P7, 247C2, 85M6, 143A8) verwendet. Die entsprechenden Klone bildeten zwei verschiedene Contigs einer Größe von jeweils etwa 350 kb. Die Klone wurden dahingehend ausgewählt, ob bekannte Gene als Positivkontrolle oder Schnittstellen für NotI als Marker für CpG-Inseln vorhanden waren. In 50 - 60 % aller Gene befinden sich CpG-Inseln in deren 5'-Promotorregion, sie können daher als Indikator für das Vorhandensein eines Gens angesehen werden.

Wie unter 4.11 beschrieben, wurde die jeweilige PAC-DNA in den pSPL3-B-Vektor subkloniert und als Klon-Pool in COS7-Zellen transfiziert. Nach Isolierung der RNA, reverser Transkription, Amplifizierung, BstXI-Restriktionsverdau und Klonierung des hieraus amplifizierten PCR-Produkts in den Vektor pDK-101 wurden jeweils 96 Einzelklone pro PAC-Klon ($6 \times 96 = 576$) gepickt und analysiert.

Zur Auswahl der Klone mit einem Insert wurde eine PCR mit den Primern SA4 und SD2 durchgeführt. Die beiden Primer binden an Sequenzen in den beiden Reporterexons. Wurde kein externes Exon dazwischen gespleißt, so wurde ein Produkt von 155 bp amplifiziert. Klone, bei denen ein Produkt dieser Größe erhalten wurde, wurden von einer weiteren Analyse ausgeschlossen. Der Anteil dieser Klone lag bei etwa 33 %. Abbildung 10 zeigt ein Beispiel der Ergebnisse aus den PCR-Analysen.

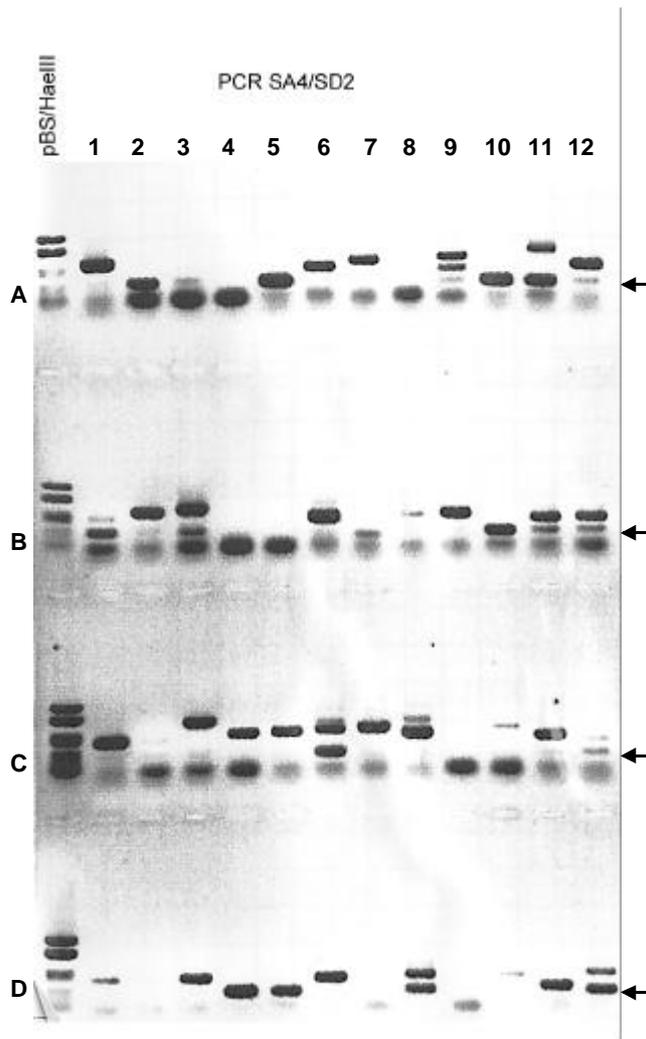


Abb. 10: Beispiel einer PCR der Primer SA4/SD2 mit den Klonen aus dem Exontrapping des PAC-Klons 22J9.

Liegt kein PCR-Produkt vor, wie bei den Klonen A8, B4, B5, C2, C8, D2, D7 und D9, so wurden beim Klonierungsschritt in den pDK-101-Vektor keine Inserts erhalten. Die Klone A2, A5, A10, B1, B10, D4, D5 und D11 enthalten kein externes Exon, da deren PCR-Produkt nicht größer als 155 bp (Pfeil) ist. Bei den Klonen A9, A11, B3, B11, B12, C6, C8 und C12 sind zwei PCR-Produkte zu sehen, es handelt sich hier vermutlich um einen Mischklon. Bei allen anderen Klonen ist zu vermuten, daß sie ein oder mehrere externe Exons enthalten.

Für die Sequenzierung der positiven Klone im Vektor pDK-101 wurde der Primer pBS7 verwendet. Es stellte sich allerdings heraus, daß ein großer Anteil der Klone ein Insert mit der pSPL3-B-Vektorsequenz enthielt.

Zur Negativselektion dieser Artefaktklone wurden Koloniefilter aus den Einzelklonen hergestellt und mit pSPL3-B-Proben hybridisiert. Die Probe stammte aus einem der Klone, der durch Sequenzierung als pSPL3-B-Insert-Klon identifiziert wurde. Zusätzliche 34 % der Klone mußten von der weiteren Sequenzierung ausgeschlossen werden.

5.3.3 Beschreibung der identifizierten Exons

Insgesamt wurden 139 Klone, die aus dem Exontrapping erhalten wurden, sequenziert. Für die weitere Analyse dieser Klone wurden die Klone danach charakterisiert, ob für die potentiellen Exons ein korrekter Spleißvorgang nachvollziehbar war. Hierfür müssen die Sequenzen der Exongrenzen aus den Reportorexons jeweils am 5'- (gaccagta) bzw. 3'-Ende (acctggaga) des klonierten Exons vorhanden sein (Church et al. 1994). Dieser Einordnung zufolge konnten 35 Klone als positiv identifiziert werden. 12 dieser Klone zeigten kein offenes Leseraster und stammen wahrscheinlich aus dem nichttranslatierten Bereich der mRNA, wobei ein Artefakt beim Exontrapping nicht vollständig ausgeschlossen werden kann. Bei 23 Klonen konnte mindestens ein offenes Leseraster identifiziert

werden. Die Ergebnisse aus einem Vergleich dieser Klone mit der Datenbank (BLAST) sind in Tabelle 1 aufgelistet.

PAC-Klon	Exon-Klon	Accession-Nummer	Exon-größe	Übereinstimmung mit:			Wahrscheinlichkeit	Übereinstimmung	Anzahl der offenen Leseraster
				EST-Klon	genomischem Klon	Gen			
85M6	85F5	AJ247008	91	/	/	/			1
	85A4	AJ247009	274	/	/	/			/
	85E5	AJ247010	55	W85798	/	/	1.5e-14	55/55	2
	85D4	AJ247011	137	T27082	/	/	3.1e-12	62/75	1
				R35500	/	/	8.1e-10	56/68	
	85C9	AJ247012	120	#					1*
	85E10	AJ247013	70	/	/	HSU15782 (CSTF3)	2.6e-20	70/70	1
	85F1	AJ247014	59	/	/	/			1
	85C4	AJ247015	130	/	/	/			2
	85G4	AJ247016	40	/	/	/			2
	85A7	AJ247017	108	/	/	/			/
	85H12	AJ247018	202	W85798	/	/	1.5e-19	87/91	1
	85F2	AJ247019	190	AA574520	/	/	1.9e-39	157/190	1
1001P7	1001A10	AJ247020	236	/	AQ120328	/	2.7e-39	130/149	/
					AQ260780	/	9.5e-25	73/74	
	1001A8	AJ247021	211	#					1
	1001B8	AJ247022	121	/		AF098863	0,026	45/53	1
	1001E10	AJ247023	101	W52510	/	/	4.1e-25	85/90	1
				AA101397	/	/	1.9e-23	79/80	
	1001E5	AJ247024	218	AA591091			1.7e-21	91/111	/
				AA501098			3.8e-13	71/91	
						AF098863 (spiC-1)	1.6e-34	152/191	
	1001D6		302			HUM5902 (CD59)	3.3e-21	77/82	3
22J9	22C3	AJ247025	252	/	/	/			/
	22A11	AJ247026	247	/	/	/			1*
	22D6	AJ247027	79	/	/	/			1*
	22F7	AJ247028	93	#					1
	22F4	AJ247029	226	/	/	/			/
	22B11	AJ247030	83	/	/	/			1
	22F9		133			HSTTG2 (LMO2)	1.6e-45	132/133	1
143A8	143G8	AJ247031	161	AI014586	/	/			/
247C2	247B10	AJ247032	101	/	/	/			2
	247D6	AJ247036	130	#					1*
	247B4	AJ247033	218	AA810522	/	/	1.8e-50	145/145	1
				AA905517	/	/	3.1e-49	139/139	
				AI49869	/	/	5.0e-48	142/142	
104D14	104C4		126			HSGPIP-137 (p137-GPI)	7.4e-43	126/126	
	104A10	AJ247034	160	/	/	/			2
	104A4	AJ247035	63	/	/	/			1
	104H10	AJ247037	93	/	/	/			1
	104D7	AJ247038	197	/	/	/			2
	104F11	AJ247039	119	#					1

Ähnlichkeit mit EST-Klonen, die repetitive Sequenzen enthalten; * Stopcodon

Exons, die keine signifikante Ähnlichkeit zu einem EST-Klon aufwiesen, wurden zu keiner weiteren Analyse herangezogen. Eine Verwendung der Exons als Hybridisierungsprobe auf Northern Blots ist aufgrund der kurzen Sequenz der Probe schwierig. Eine Verifizierung der Lage dieser Exons auf dem entsprechenden PAC-Klon ist nach der Sequenzierung des gesamten PAC-Contigs durch das Sanger Centre/UK anhand eines Datenbankvergleichs möglich. Da allerdings mit den identifizierten Exons (85E5, 85D4, 85H12, 1001E10 und 247B4), die signifikante Ähnlichkeiten zu einem anonymen cDNA-Klon aufwiesen, Northern Blot Analysen sowie in situ Hybridisierungen mit einem für das Exon repräsentativen cDNA-Klon durchgeführt wurden, wurde die Lokalisation dieser Exons auf den Ursprungsklon (85M6, 1001P7 und 247C2) mittels Hybridisierung überprüft. Hierfür wurden PCR-Produkte der klonierten Exons mit den Oligonukleotiden SA4/SD2 als Hybridisierungsprobe für PAC-Filter verwendet. Standen IMAGE-Klone mit menschlicher cDNA zur Verfügung, so wurden diese als Hybridisierungsprobe herangezogen. Die Lokalisation auf den jeweiligen PAC-Klon, mit dem das Exontrapping durchgeführt wurde, konnte für jedes der getesteten Exons bestätigt werden.

EST-Klone, die durch eine signifikante Ähnlichkeit mit den Sequenzen der isolierten Exons gefunden wurden, wurden in einem Vergleich mit der Datenbank (BLAST) herangezogen, um weitere überlappende EST-Klone zu identifizieren. Mit diesen zusätzlichen EST-Klonen konnten cDNA-Contigs konstruiert werden. Ausnahmen bildeten das Exon 85E10 sowie 1001E5 und 1001B8. 85E10 stimmt mit dem Gen der 70 kD Untereinheit von CSTF3 (cleavage stimulation factor 3) überein (GenBank: U15782). EST-Klone hierfür wurden von Beate Gawin auf 11p13 kartiert und analysiert. Die beiden Exons 1001E5 und 1001B8 sind ähnlich zu Spi-C, das aus der Maus isoliert wurde (Bemark et al. 1999). Allerdings unterscheidet sich die Leserichtung der beiden identifizierten Exons von der des Spi-C-Gens. Hinzu kommen Stopcodons in der menschlichen Sequenz und eine kryptische Spleißstelle in Exon 1001E5, das 70 Nukleotide mit Exon 1001B8 überlappt. Aus diesem Grund wurde angenommen, daß es sich hierbei um isolierte Exons eines Pseudogens homolog zum murinem Spi-C handelt. Die erstellten cDNA-Contigs der Exons 247B4, 85F2, 1001E10 sowie 85D4, 85H12 und 85E5 sind in Abbildung 11 dargestellt.

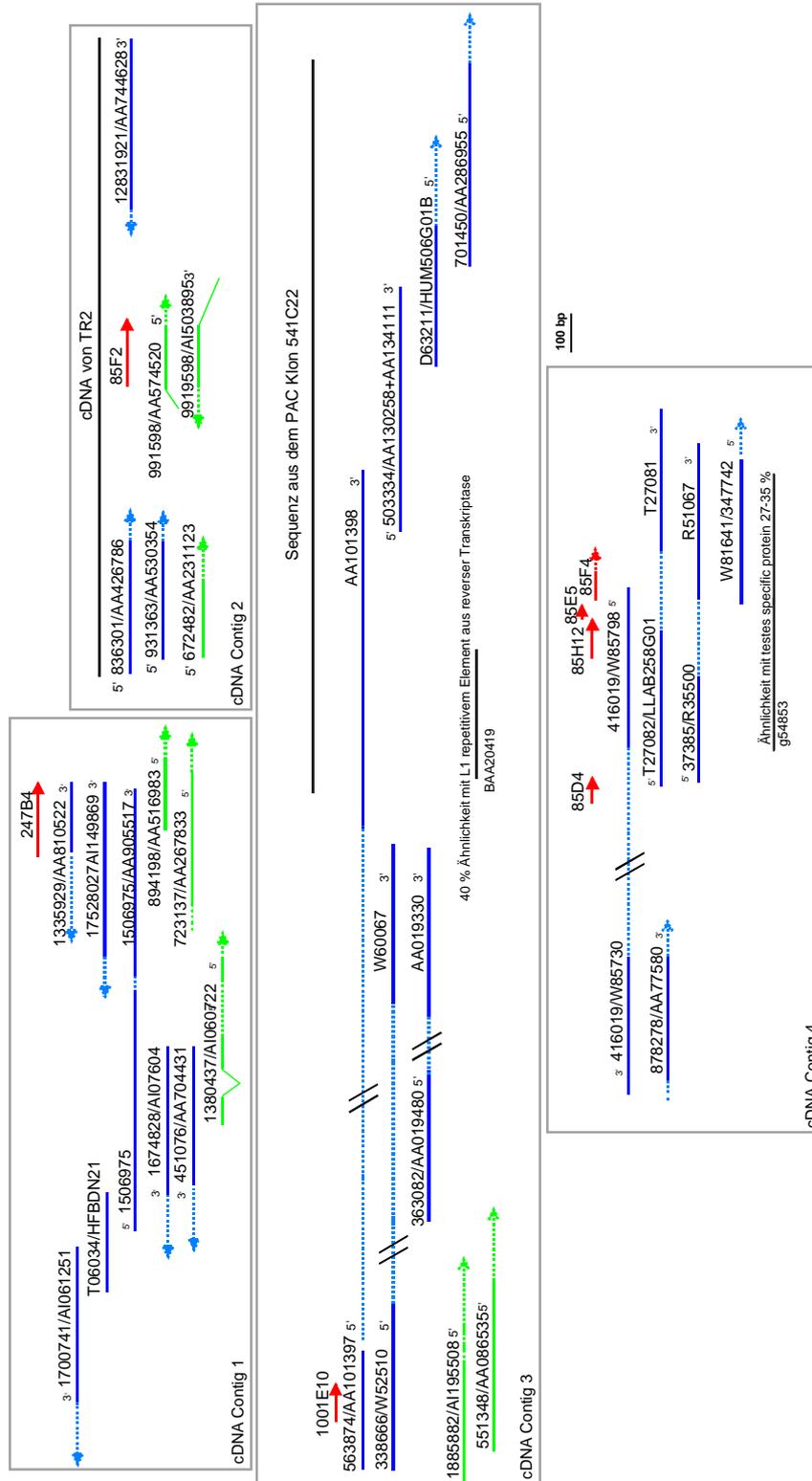


Abb. 11: cDNA Contigs der EST-Klone aus dem Exontrapping. cDNA-Klone aus der Maus sind in grüner Farbe, cDNA Klone aus dem Menschen in blauer Farbe und Exonsequenzen in roter Farbe in Leserichtung (Pfeilrichtung) dargestellt. Die 3'- bzw. 5'-Enden der cDNA-Klone sind durch unterbrochene Linien verbunden. Fehlt die Sequenz des jeweiligen 3'- oder 5'-Endes in der Datenbank, so ist dies durch einen Pfeil gekennzeichnet. Wurden Ähnlichkeiten zu bekannten Genen gefunden, so sind diese oberhalb oder unterhalb der entsprechenden Sequenz (schwarz) angegeben.

Für das Exon 247B4 konnten über die 5'-Sequenz des Klons 1506975 mehrere zusätzliche cDNAs aus der Maus sowie aus dem Menschen identifiziert werden (cDNA-Contig 1), die ein durchgehendes offenes Leseraster aufweisen. Ein Stopcodon befindet sich am Ende der 5'-Sequenz der beiden Maus-cDNA-Klone 894198 (AA516983) und 723137 (AA267833). Das 5'-Ende des murinen cDNA-Klons 1380437 (AI60772) ist vermutlich alternativ gespleißt, da die Übereinstimmung zur Sequenz aus dem humanen cDNA-Klon nicht durchgängig ist.

Das Exon 85F2 stammt aus dem Gen TR2 (persönliche Mitteilung, Manfred Gessler). Der murine cDNA-Klon 991598, der 85 % homolog zum Exon 85F2 ist, besitzt interne Spleißstellen jeweils am 5'- bzw. 3'-Ende der Homologien zu Exon 85F2, das heißt, er enthält zusätzlich zu einem Exon Anteile genomischer DNA. TR2 wird in der adulten Leber und schwach in der adulten Niere exprimiert (persönliche Mitteilung, Beate Gawin).

Für das Exon aus dem Klon 1001E10 (cDNA Contig 3) konnten ebenfalls zusätzliche humane und murine cDNAs ermittelt werden. Die Sequenz des 3'-Endes des Klons 563874 (AA101398), dessen 5'-Ende (AA101397) mit dem Exon 1001E10 gefunden wurde, stimmt mit Sequenzen aus dem PAC-Klon 541C22 überein. Da keine Spleißstellen für die 3'-Sequenz des Klons 563874 (AA101398) gefunden wurden, handelt es sich vermutlich um ein einziges Exon. Das 3'-Ende des Klons 563874 (AA101398) besitzt außerdem ein Polyadenylierungssignal, das möglicherweise alternativ genutzt werden kann, da eine Überlappung mit weiteren cDNA-Klonen gefunden wurde. Allerdings besteht auch die Möglichkeit, daß es sich bei den letztgenannten cDNA-Klonen (503334, D63211, 701450) um keine echten cDNAs sondern Artefaktklone, die genomische DNA enthalten, handelt und diese daher nicht mehr zu dieser Transkriptionseinheit gehören.

Im cDNA-Contig 4 konnten vier Exons in eine Transkriptionseinheit zusammengefaßt werden. Exon 85F4 nimmt eine Sonderstellung ein, da es nur auf einer Seite von einer korrekten Spleißstelle der Reporterexons aus dem pSPL3-B-Vektor begrenzt ist. Es wird daher nicht in der zuvor gezeigten Tabelle aufgeführt. Das 5'-Ende des Klons 416019 (W85798) in dem gezeigten cDNA-Contig ist in die gegenläufige Richtung der Leserichtung der isolierten Exons orientiert. Die 5'-Sequenz des Klons 416019 (W85798) enthält allerdings ein offenes Leseraster nur in der Orientierung der Leserichtung der Exons. Wahrscheinlich wurde der Klon 416019 in umgekehrter Richtung in den cDNA-Vektor kloniert. Gleiches könnte auch für den Klon 878278 gelten.

Von den hier gezeigten EST-Klonen wurden für jede Transkriptionseinheit jeweils ein humaner cDNA-Klon und, soweit vorhanden, ein muriner cDNA-Klon für eine Analyse auf Northern Blots herangezogen, sofern sie über das IMAGE-Konsortium zugänglich waren. Eine Ausnahme bildeten die EST-Klone aus cDNA-Contig 2, die die Sequenz des Gens TR2 repräsentieren. Eine Northern Blot Analyse hierfür wurde von Beate Gawin durchgeführt.

5.3.4 Northern Blot Analysen

Um für die beim Exontrapping ermittelten Transkriptionseinheiten ein RNA-Transkript nachzuweisen und ein Expressionsmuster zu erhalten, wurden erste Analysen auf Northern Blots durchgeführt. Als cDNA-Probe für diese Analysen wurden nur diejenigen EST-Klone gewählt, deren Übereinstimmung mit dem jeweils identifizierten Exon signifikant war (94-100% für humane EST-Klone, 80 % für murine

EST-Klone) und deren Sequenz daher als repräsentativ für das jeweils potentielle Transkript angesehen werden konnte.

Die Inserts der IMAGE-Klone wurden herausgeschnitten und als Hybridisierungsprobe verwendet. Für die Herstellung der Maus-Northern Blots wurde Gesamt-RNA aus verschiedenen Mausgeweben und Embryonalstadien isoliert und geblottet. Northern Blots mit Poly-A⁺-RNA aus menschlichen Geweben wurden von Clontech bezogen. Eine Übersicht der EST-Klone, die für eine Analyse auf Northern Blots verwendet wurden, gibt Tabelle 2.

PAC-Klon	Exon	EST Klon	Spezies	cDNA Contig (siehe Abbildung 11)
247C2	247B4	1506975/AA905517	Mensch	1
247C2	247B4	894158/AA516983	Maus	1
1001P7	1001E10	563874/AA101397	Mensch	3
1001P7	1001E10	1885882/AI195508	Maus	3
85M6	85E5	416019/W85798	Mensch	4

Für die Klone 1506975 und 894158 konnte kein Transkript im Northern Blot nachgewiesen werden.

Die Ergebnisse der Hybridisierungen der übrigen EST-Klone auf Northern Blots mit menschlicher Poly-A⁺-RNA und Gesamt-RNA aus verschiedenen Mausgeweben sind in Abbildung 12 zusammengefaßt.

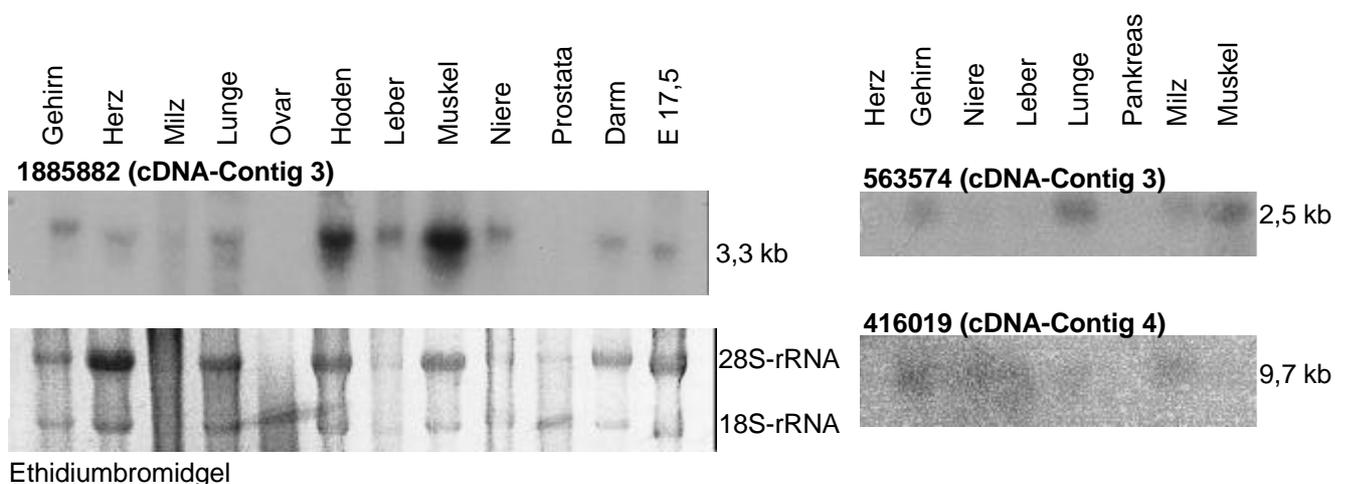


Abb. 12: Hybridisierung von cDNA-Proben auf murine (links) und humane (rechts) Northern Blots. Die jeweils verwendete Probe mit einem Hinweis auf die in Abbildung 11 gezeigten cDNA-Contigs ist oberhalb der Hybridisierungsergebnisse angegeben. Das entsprechende Ethidiumbromidgel mit der 18S- und 28S-rRNA ist unterhalb davon gezeigt. Die humanen Northern Blots sind auf β -Aktin normiert (Clontech). Die Größe des detektierten Transkripts ist jeweils rechts der Hybridisierungsergebnisse angegeben.

Mit der Probe 1885882 konnte in verschiedenen Mausgeweben ein Transkript der Größe 3,3 kb nachgewiesen werden. Eine schwache Expression ist im Gehirn, im Herzen, in der Milz, in der Lunge, im Darm und im Embryonalstadium E17,5 zu erkennen. Eine starke Expression zeigte sich im Muskel, im Hoden und in der Leber. Hier erscheint die Bande im Northern Blot zwar vergleichsweise schwach,

was aber auf eine niedrigere RNA-Konzentration im Ethidiumbromidgel zurückzuführen ist. Ebenso kann über eine Expression im Ovar keine Aussage getroffen werden, da hier die RNA-Probe degradiert war. Bei einem Vergleich der Expression des entsprechenden Transkripts beim Menschen (Probe 563574) wurde ein kleineres Transkript der Größe von 2,5 kb erhalten. Im Muskel konnte eine stärkere Expression nachgewiesen werden, außerdem im Gehirn und in der Lunge. Eine sehr schwache Expression war in der Niere, der Leber und der Milz erkennbar. Da der humane Northern Blot auf Poly-A⁺-RNA normiert war, für den murinen Northern Blot hingegen die gleiche Menge an Gesamt-RNA eingesetzt wurde, kann nur ein qualitativer Vergleich zwischen der Expression beim Menschen und bei der Maus gezogen werden.

Mit der menschlichen cDNA-Probe 416019 konnte ein Transkript der Größe 9,7 kb im Gehirn, in der Niere, in der Leber, in der Lunge und in der Milz detektiert werden. Eine entsprechende Probe für die Hybridisierung auf einen Northern Blot mit murinem Gewebe stand nicht zur Verfügung.

5.3.5 Charakterisierung durch in situ Hybridisierungen

Um weitere Informationen über das Expressionsmuster der aus dem Exontrapping ermittelten Transkripte zu erhalten, wurden in situ Hybridisierungen an Paraffinschnitten von Mausembryonen oder an ganzen Embryonen durchgeführt. Da hierfür eine cDNA-Probe aus der Maus erforderlich war, konnten diese Analysen nicht für das Transkript erstellt werden, das in cDNA-Contig 4 (Abbildung 11) gezeigt wird. In den folgenden Abbildungen sind die Ergebnisse der Hybridisierungen an Paraffinschnitten und ganzen Embryonen dargestellt.

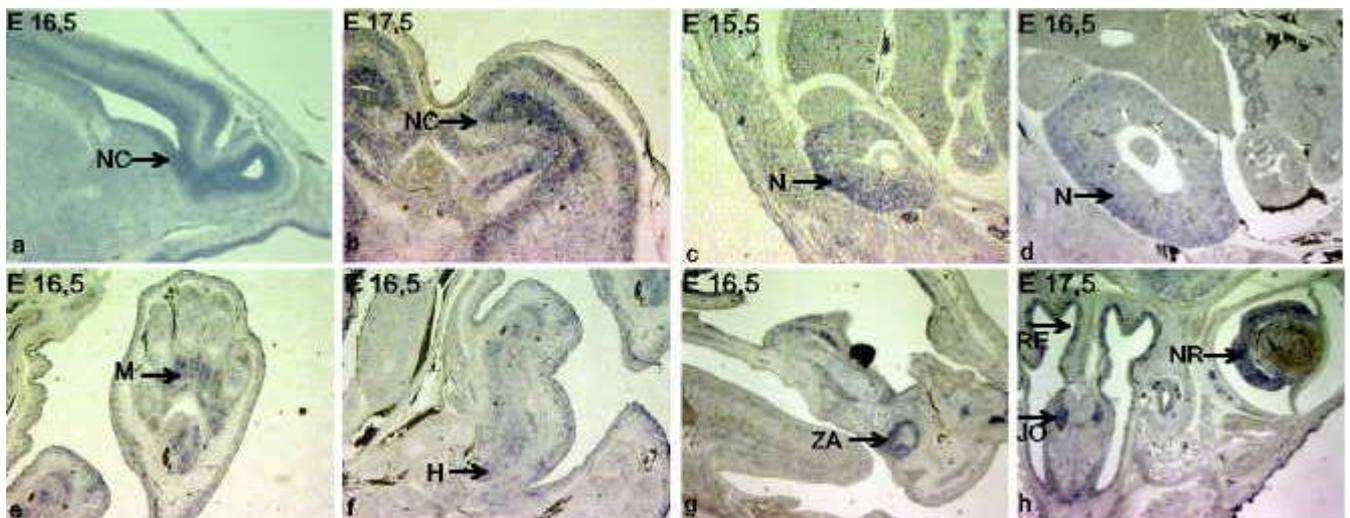


Abb. 13: In situ Hybridisierungen der Probe 894158 auf Paraffinschnitte der Embryonalstadien E15,5, 16,5 (Längsschnitte) und E17,5 (Querschnitt). Eine Expression ist im Neocortex (NC) in den Stadien E16,5 und E17,5 (a und b) zu erkennen. In c und d ist eine Expression im cortikalen Bereich der Niere (N) gezeigt. Eine schwache Expression ist im Stadium E16,5 in den Muskeln (M) der Extremitäten und im Hoden (H) nachweisbar (e und f). Eine stärkere Expression ist in den Zahnanlagen der oberen vorderen Schneidezähne (ZA), dem Jacobson'schen Organ (JO), den Riechepithelien (RE) und der neuronalen Retina (NR) zu erkennen (g und h).

Eine Expression von 894158 (cDNA-Contig 1, Abbildung 11) im Gehirn war schon ab dem Stadium E12,5 nachweisbar (E12,5 – 15,5 nicht gezeigt), wobei eine starke Expression im Vorderhirn, speziell im Neocortex, in späteren Stadien (ab E15,5) zu finden war. Außer der Expression im Gehirn, konnte eine Expression für neuronales Gewebe im Auge, in den Riechepithelien (E16,5, E17,5) und dem Jacobson'schen Organ (E17,5) detektiert werden. Ab dem Stadium E15,5 war zusätzlich eine Expression im corticalen Bereich der Niere im sich kondensierendem Mesenchym (nicht gezeigt) zu sehen, ab E16,5 in den Muskeln der Extremitäten, im Hoden und den Zahnanlagen.

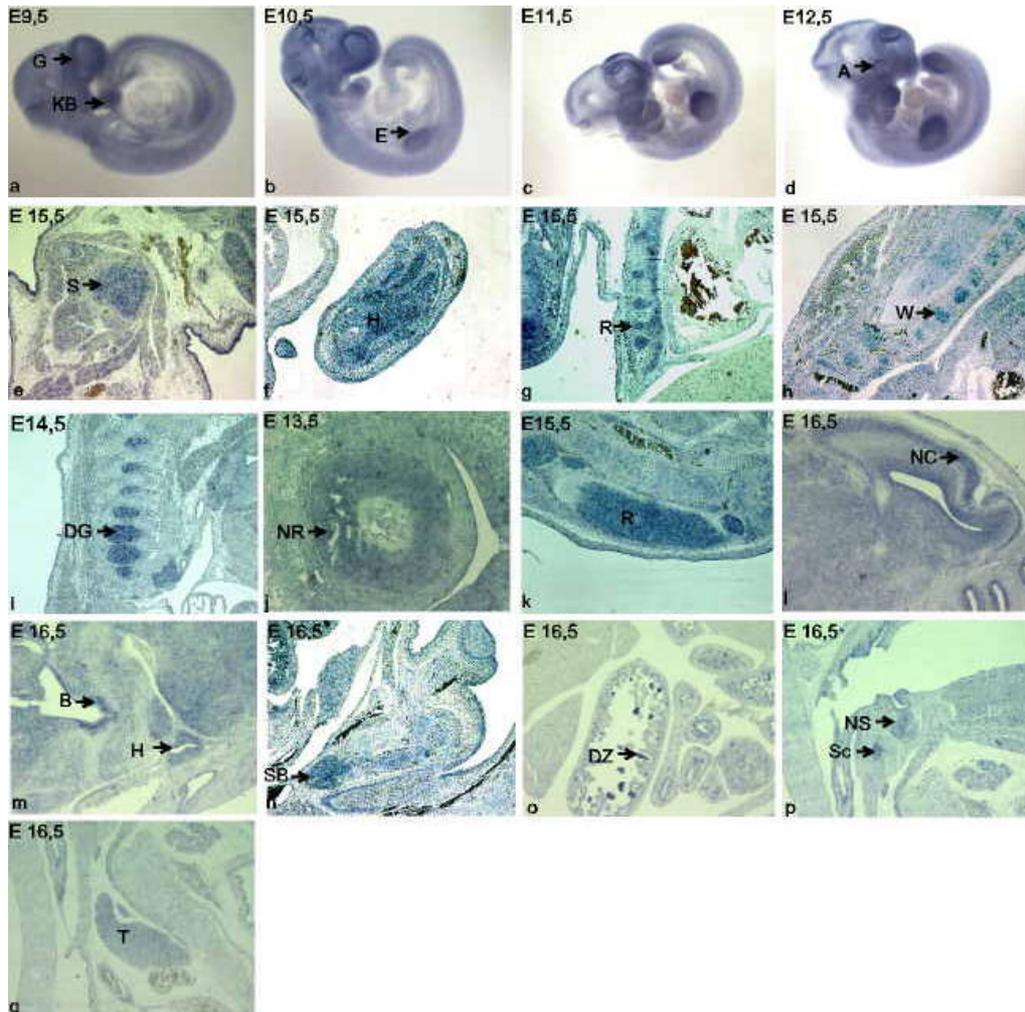


Abb. 14: In situ Hybridisierungen der Probe 551348 auf ganze Embryonen (a-d) und Paraffinschnitte (e-q). Im Embryonalstadium E9,5 (a) ist eine Expression im Gehirn (G) und den Kiemenbögen (KB) erkennbar. Zusätzlich zur Expression im Gehirn ist bei der in situ Hybridisierung auf ganze Embryonen der Stadien E10,5 bis E12,5 (b-d) eine Expression in den sich entwickelnden Extremitäten (E), im Stadium E12,5 (d) auch im Auge (A), nachweisbar. In e-h ist die Expression im Knorpelgewebe gezeigt (S Schulterblatt, H Hinterextremität, R Rippen, W Wirbel). Eine Expression im Neuronalgewebe (i – m) ist ebenso in den Hybridisierungen auf Schnitten zu finden (DG dorsale Ganglien, NR neuronale Retina, R Rückenmark, NC Neocortex, B Brückenbeuge und H Hypophyse). Im Stadium E16,5 ist eine Expression in verschiedenen Organen nachweisbar (SB Samenblase, DZ Darmzotten, NS Nebenschilddrüse, Sc Schilddrüse und T Thymus).

Die Probe 551348 zeigt eine Expression in vielen unterschiedlichen Geweben, vorrangig aber in neuronalem Gewebe (Gehirn, Rückenmark, Ganglien, neuronale Retina) und Knorpelgewebe. Die Expression im Knorpelgewebe ist ab dem Stadium E16,5 nicht mehr zu finden, da ab diesem Entwicklungszeitpunkt das Knorpelgewebe durch Knochengewebe ersetzt wird. Eine schwache Expression im Gehirn konnte ab dem Stadium E9,5 nachgewiesen werden. Eine regional stärkere Expression im Gehirn wurde im Stadium E16,5 im Neocortex, in der Brückenbeuge – der Verbindung zwischen Myelenzephalon und Metenzephalon - und in der Hypophyse nachgewiesen. Eine schwächere Expression konnte allerdings auch in der Samenblase, verschiedenen Drüsen (Schilddrüse, Nebenschilddrüse, Thymusdrüse) und in den Darmzotten festgestellt werden.

5.4 Ergebnisse aus den Sequenzanalysen

5.4.1 Datenbankvergleiche

Um weitere Transkriptionseinheiten zu erschließen, wurden Sequenzanalysen einzelner PAC-Klone aus dem erstellten Contig mit Hilfe von Computerprogrammen und Datenbankvergleichen durchgeführt. Die Sequenzierung der PAC-Klone wurde im Sanger Centre/UK durchgeführt. Soweit die Sequenzen zur Verfügung standen, wurden diese mit Hilfe des NIX Programmpakets/HGMP analysiert. NIX beinhaltet mehrere Vorhersagen für potentielle Exons, Polyadenylierungsstellen, CpG-Inseln und Promotoren, außerdem wird mit Hilfe von BLAST ein Datenbankvergleich mit Protein- und DNA-Sequenzdatenbanken durchgeführt. Vor der Analyse werden repetitive Sequenzen der genomischen DNA maskiert. Zum Zeitpunkt der Auswertung der PAC-Sequenzen waren die noch ungeordneten Contigs der Sequenzen in der Datenbank (HTGS) verfügbar. Die Größen dieser einzelnen Sequenzcontigs lagen im Durchschnitt zwischen 10 und 40 kb. Ein Beispiel für die Auswertung eines dieser Sequenzcontigs mit dem Programmpaket NIX ist in Abbildung 15 zu sehen.

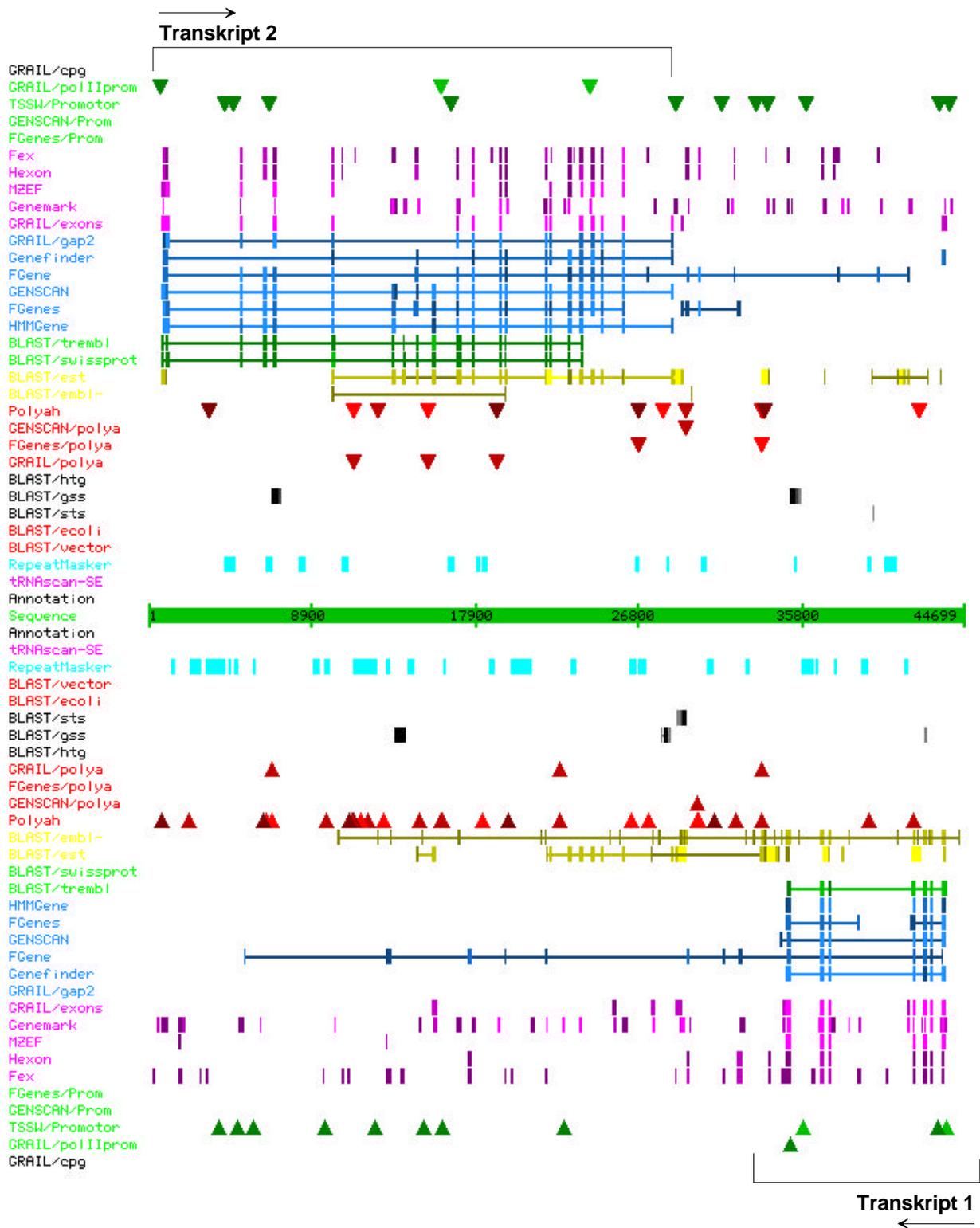


Abb. 15: Analyse des NIX Programmpakets einer 45 kb großen Sequenz aus dem PAC-Klon 1169J3. Neben den Vorhersagen für Exons (blau, violett), CpG-Inseln (schwarz), Promotoren (grün) und Polyadenylierungsstellen (rot) werden die Sequenzen mit Protein- (grün) sowie DNA-Sequenz- (gelb, schwarz) Datenbanken verglichen. Je wahrscheinlicher die Aussage, desto heller ist die Farbabstufung. Es konnte hier für beide Stränge jeweils ein Transkript vorhergesagt werden, das jeweils in unterschiedliche Richtungen abgelesen wird.

Insgesamt wurden fünf PAC-Klone vollständig entsprechend einer Sequenz von insgesamt etwa 640 kb (53C18/115+16 kb, 1169J3/130+20 kb, 541C22/150 kb, 305G21/110 kb, 316D7/100 kb) und ein PAC Klon (247C4/20 kb) teilweise analysiert. In Tabelle 3 sind die Ergebnisse aus den Sequenzanalysen zusammengefaßt.

PAC-Klon	kartierte Gene	Homologien zu bekannten Genen	EST-Cluster		Übereinstimmungen mit einzelnen EST-Klonen
			Anzahl	repräsentative EST-Klone	
53C18	CAT	Rezepin (HSU03644); DnaJ (HSHIRIP4); cca3 (Exon1, AB00216)			AI147890
1169J3	p137-GPI	cca3 (AB00216); Transkript aus Achlya ambisexualis (M23451)			
541C22	CD59, G2		2	H92965; H86320 AA130258	
316D7	G2				
305G21		Calgizzarin (D38583)			AA339509; Gen-421F09
247C2			1	AA905517	

Neben den Genen CAT (Niikawa et al. 1982), p137-GPI (Gessler et al. 1996), CD59 (Bickmore et al. 1993) und G2 (U10991), die zuvor schon auf die entsprechenden PAC-Klone lokalisiert worden waren, wurden Homologien der analysierten Sequenzen zu bisher noch nicht kartierten Genen gefunden. Es handelte sich hierbei unter anderem um Sequenzhomologien zu humanen Genen, dem DnaJ Protein (HSHIRIP 4, 78-91 % Ähnlichkeit), dem Protein für Rezepin (HSU03644, 67-94 % Ähnlichkeit) und dem Gen für Calgizzarin (D38583, 73-76 % Ähnlichkeit). Es liegt allerdings in allen drei Sequenzen mit Homologien zu diesen Genen kein durchgehendes offenes Leseraster vor, so daß es sich hier vermutlich um Pseudogene handelt. Bei einem Vergleich der genomischen Sequenz mit der des DnaJ Proteins sind keine Introns zu erkennen, dafür aber interne Polyadenylierungsstellen. Beides gilt als ein Charakteristikum für ein Pseudogen. Die Übereinstimmung der partiellen Sequenz der Rezepin-mRNA mit der genomischen PAC-DNA ist zwar nicht durchgehend wie im Falle des DnaJ Proteins, Unterbrechungen der Übereinstimmungen sind aber nicht von korrekten Spleißstellen begrenzt, sondern enthalten repetitive Sequenzen (Alu), was wiederum für ein Pseudogen spricht. Ebenfalls ohne Introns ist die zum Calgizzarin (Tanaka et al. 1995) ähnliche Sequenz, die im PAC-Klon 305G21 gefunden wurde, das heißt, es handelt sich hier wiederum um Sequenzen eines Pseudogens.

Ein EST-Cluster beschreibt mehrere EST-Klone, die eine Übereinstimmung mit den genomischen Sequenzen der PAC-Klone zeigen und untereinander homolog sind oder überlappende homologe Bereiche besitzen. Eines der EST-Cluster aus der Sequenz des PAC-Klons 541C22, das die beiden überlappenden EST-Klone H92965 und H86320 enthält, beinhaltet zum großen Teil repetitive Alu-Sequenzen. Es ist daher nicht auszuschließen, daß es sich bei diesen EST-Klonen um genomische Klone und keine echten cDNA-Klone handelt. Das zweite der EST-Cluster aus dem PAC-Klon 541C22 ist über einen cDNA Klon mit einem der im Exontrapping isolierten Exons (1001E10) verknüpft und

daher unter 5.4.3 (siehe Abbildung 11, cDNA-Contig 3) näher beschrieben. Die Consensussequenz des EST-Clusters aus dem PAC-Klon 247C2, die durch den EST-Klon AA905517 repräsentiert wird, beinhaltet vier Exons eines unbekanntes Transkripts. Die Exons, von dem eines dem Exon 247B4 (siehe Abbildung 11, cDNA-Contig 1) entspricht, besitzen ein durchgehendes offenes Leseraster.

Der EST-Klon AI147890, der mit der Sequenz aus dem PAC-Klon 53C18 übereinstimmt, ist weder gespleißt noch besitzt er ein offenes Leseraster. Die beiden einzelnen EST-Klone AA339509 und Gen-421F09 stimmen ebenfalls durchgehend mit Sequenzen aus dem PAC-Klon 305G21 überein und weisen kein offenes Leseraster auf.

In einem der sequenzierten PAC-Klone (1169J3) konnten zwei nahezu vollständige Transkripte identifiziert werden, die in gegenläufigen Richtungen abgelesen werden (siehe Abbildung 15). Die 3'-Enden beider Transkripte liegen 4074 bp auseinander. Beide Transkripte werden im folgenden näher beschrieben.

Transkript 1

Das erste der beiden Transkripte weist eine Homologie zu einem Gen (*cca3*) auf, das von Hayashi et al. (1997b) aus der Rattenleberzelllinie 3Y1 isoliert wurde. Aufgrund der Übereinstimmungen der mRNA-Sequenz sowie homologen EST-Klonen mit der genomischen Sequenz in den PAC-Klonen 53C18 (Exon1) und 1169J3 (Exon 2-17) konnte eine potentielle Exon-Intron-Struktur des Transkripts im Menschen erstellt werden.

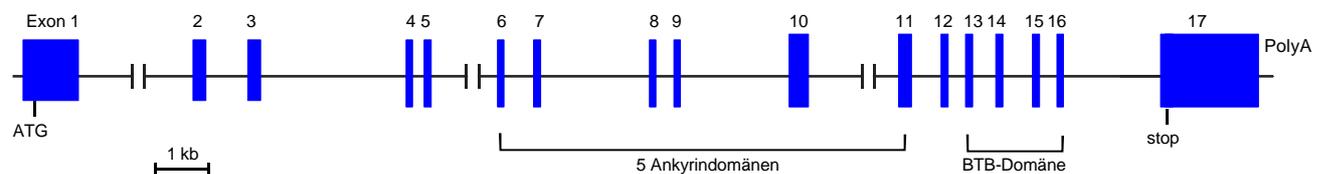


Abb. 16: Genomischer Aufbau des Transkripts 1. Die Exons sind als Kästchen dargestellt. Das Startcodon (ATG, Exon 1), das Stopcodon (Exon 17) und die Polyadenylierungsstelle sowie Homologien der cDNA zu bekannten Proteindomänen (Ankyrin: Exon 6 - 11, BTB: Exon 13 -16) sind am entsprechendem Exon eingezeichnet. Abstände zwischen den Exons, die nicht bekannt sind, sind durch zwei Querbalken gekennzeichnet.

Bei allen hier gezeigten Exons wurde eine korrekte Spleißstelle am 5'- und 3'-Ende der Exons in der genomischen PAC-Sequenz gefunden. Die Ähnlichkeiten zwischen der genomischen Sequenz des PAC-Klons und der mRNA-Sequenz des *cca3* Gens betragen zwischen 86 und 90 %. Die ersten 351 Basen des Exons 1 der *cca3*-mRNA der Ratte weisen nur lückenhafte Ähnlichkeiten mit der menschlichen Sequenz auf, das heißt, es konnte nur zwischen den Basen 2 und 40 bzw. 76 und 100 eine Übereinstimmung gefunden werden. Da diese Region in der *cca3*-mRNA nicht zum offenen Leseraster gehört, unterliegt sie keinem Selektionsdruck, das bedeutet, es können sich Mutationen anhäufen, so daß die Sequenzen bei der Ratte und dem Menschen an dieser Stelle nicht mehr konserviert sind.

Die Sequenz des Exons 1 liegt im PAC-Klon 53C18 etwa 20 kb vom SP6-Ende entfernt. Die restlichen 16 Exons liegen in der Sequenz des PAC-Klons 1169J3. Die beiden PAC-Klone überlappen nicht miteinander. Der im PAC-Contig ermittelte Abstand zwischen beiden Klonen beträgt ungefähr 80 bis 100 kb, das heißt Exon 1 und Exon 2 liegen mindestens 80 kb voneinander entfernt.

Die in silico ermittelte Länge der humanen cDNA beträgt 4,3 kb (4306 bp). Die cDNA des menschlichen Homologen ist gegenüber der 4,5 kb langen cca3-mRNA um 0,2 kb kürzer, da die Sequenz des Exons 1 nicht vollständig ist. Das offene Leseraster einer Länge von 3176 bp beginnt mit einem Startcodon bei Nukleotid 20 der ermittelten cDNA und endet mit einem Stopcodon bei Nukleotid 3196. In der cDNA des homologen Gens der Ratte hingegen existiert ein offenes Leseraster von nur 3027 Basen, 149 Nukleotide kürzer als beim Menschen. Allerdings befindet sich 158 Nukleotide vor dem Stopcodon der humanen Sequenz eine Verschiebung des Leserasters um +1. Die Homologie auf Proteinebene geht an dieser Stelle verloren. Unter Annahme eines Sequenzierfehlers in der PAC-Sequenz bleibt diese jedoch erhalten. Ein endgültiger Vergleich der Sequenzen beider Gene ist nach Fertigstellung der Endversion für die genomische Sequenz des PAC-Klons 1169J3 möglich.

Die codierende Region der humanen cDNA wurde in die Proteinsequenz übersetzt und mit Hilfe des Programms SMART V3.0 (Schultz et al. 1998) und der Datenbank PROSITE (Bairoch 1991) nach Homologien zu bekannten Domänen hin untersucht. In der Proteinsequenz der translatierten cDNA-Sequenz wurden fünf (SMART V3.0) bzw. drei Ankyrin-Domänen (PROSITE) und eine BTB Domäne (SMART V3.0, PROSITE) gefunden. Beide Domänen sind ebenfalls im cca3-Gen in gleicher Anzahl vorhanden. Ankyrin- sowie BTB-Domänen vermitteln Protein-Protein Wechselwirkungen. BTB-Domänen werden außerdem häufig am N-Terminus von Zink-Finger-Proteinen gefunden. Proteine mit BTB Domänen sind daher häufig an der Regulation der Transkription beteiligt, indem sie die Konformation von Chromatinstrukturen beeinflussen.

Transkript 2

Das zweite Transkript wurde mittels Übereinstimmungen der genomischen Sequenz mit EST-Klonen im PAC-Klon 1169J3 identifiziert. Die EST-Klone, die mit der genomischen Sequenz gefunden wurden, waren ihrerseits in einem Contig miteinander verbunden. Daraus wurde geschlossen, daß sie einem Transkript angehörten, für das sich eine potentielle Exon-Intron-Struktur ableiten läßt.

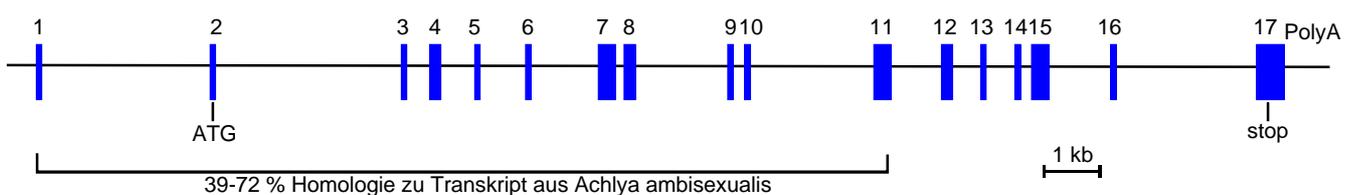


Abb. 17: Genomischer Aufbau des Transkripts 2. Auf Proteinebene homologe Bereiche zu einem Transkript aus *Achlya ambisexualis* befinden sich im Exon 1 – 11. Ein mögliches Startcodon (ATG) befindet sich im Exon 2, das Stopcodon im Exon 17. Vor dem ATG in Exon 2 befindet sich kein Stopcodon. Aufgrund der homologen Bereiche in Exon 1 zum Transkript aus *Achlya ambisexualis* ist anzunehmen, daß vor dem Exon 1 noch ein oder mehrere Exons vorhanden sind, die das eigentliche Startcodon enthalten.

Bis auf eine einzige Ausnahme waren alle EST-Klone miteinander vernetzt. Bei der Verbindung des Exons 7 mit 8 endete die bekannte Sequenz eines EST-Klons im Exon 7. Im nächsten Exon begann die Übereinstimmung der genomischen PAC-Sequenz mit einem anderen EST-Klon. Aufgrund der kurzen Distanz von 92 bp zwischen dem Exon 7 und dem potentiell neuen Exon 8 konnte angenommen werden, daß es sich möglicherweise um ein einziges Exon handelt. Da auf der 3'-Seite

von Exon 7 und auf der 5'-Seite von Exon 8 jeweils eine Spleißstelle gefunden werden konnte, wurde jedoch angenommen, daß an dieser Stelle zwei getrennte Exons vorliegen, zumindest aber die Möglichkeit des alternativen Spleißens besteht. Eine weitere alternative Spleißstelle kann in Exon 11 benutzt werden. An dieser Stelle stimmen EST-Klone entweder durchgängig mit der genomischen Sequenz überein oder aber die Übereinstimmung wird von 116 bp unterbrochen, die von Spleißstellen flankiert sind.

Ein Spleißen in silico der durch die Übereinstimmung mit den cDNA-Klonen determinierten 17 Exons zu einer potentiellen cDNA ergab eine Länge von 2896 bp bzw. 2780 bp, da im Exon 11 eine alternative Spleißstelle benutzt werden kann. In einem Vergleich dieser cDNA mit der Datenbank konnte eine Ähnlichkeit des Transkripts zu einem Cosmid Klon (AF003130; 34-72% Ähnlichkeit) gefunden werden, der die Sequenz eines partiellen Transkripts (M23451) enthält, das aus dem Pilz *Achlya ambisexualis* isoliert wurde (Mallet, Bussereau und Jacquet 1995; Schowalter, Toft und Sommer 1990; Wilson et al. 1994). Die Sequenz dieses partiellen Transkripts weist auf Proteinebene eine Ähnlichkeit zwischen 39 und 72 % zur Sequenz aus dem menschlichen Transkript auf. Zwei dazu ähnliche Proteine sind auch in den beiden Hefestämmen *Saccharomyces cerevisiae* (p53914; 41-50% Ähnlichkeit) und *Schizosaccharomyces pombe* (p87115; 32-50% Ähnlichkeit) beschrieben.

Ein offenes Leseraster konnte zwischen den Nukleotiden 222 und 2070 identifiziert werden (Länge des ORF 1848). Dieses offene Leseraster ist allerdings nur dann durchgängig, wenn in Exon 11 die alternative Spleißstelle benutzt wird. Die Protein-Sequenz des offenen Leserasters ist 616 AS lang. In einer Analyse dieser Sequenz mit SMART und PROSITE konnten keine Strukturen bekannter Domänen gefunden werden, die auf eine mögliche Funktion des Proteins schließen lassen. Das Protein enthält allerdings ein Kernlokalisationssignal am N-Terminus (PROSITE).

Für das Transkript aus *Achlya ambisexualis* wird eine Funktion als Steroidrezeptor postuliert. Es wurde über Hybridisierung mit der DNA-Bindungsdomäne des Östrogenrezeptors aus *Xenopus* und dem Hühnchen isoliert. Das identifizierte Transkript weist eine Zinkfinger-Domäne auf, die die Mitglieder der Steroidrezeptorproteinfamilie kennzeichnen. Allerdings ist die Ähnlichkeit zu den Östrogenrezeptoren, die als Probe benutzt wurden, auf einen sehr kleinen Bereich von 14 AS beschränkt (Schowalter, Toft und Sommer 1990).

Die Zinkfinger-Domäne des Proteins aus *Achlya ambisexualis* konnte in der menschlichen Proteinsequenz nicht wieder gefunden werden. Es besteht aber die Möglichkeit, daß das offene Leseraster größer ist als hier angenommen, da vor dem ersten ATG kein Stopcodon vorhanden ist. Von der ausgewerteten Sequenz des PAC-Klons lag noch keine Endversion vor. Vor dem hier als Exon 1 festgelegten Exon können also durchaus noch weitere bisher nicht identifizierte Exons liegen, die Teile der codierenden Sequenz enthalten. Dafür spricht, daß die Übereinstimmung der Sequenz des Transkripts aus *Achlya ambisexualis* mit der AS 92 beginnt, die Zinkfinger-Domäne aber zwischen den AS 69 und 93 liegt (Schowalter, Toft und Sommer 1990). Allerdings handelt es sich hier um ein partielles Transkript, von dem nur ein Teil der Proteinsequenz bekannt ist.

Von beiden hier beschriebenen Transkripten wurden EST-Klone vom IMAGE-Konsortium für weitere Analysen bezogen. Eine Übersicht aller in der Sequenzanalyse identifizierten Gene und EST-Cluster sowie der Ergebnisse aus dem Exontrapping ist in Abbildung 18 zusammengefaßt.

5.4.2 Northern Blot Analysen

Die Proben der IMAGE-Klone wurden auf RNA-Blots mit humaner bzw. muriner RNA hybridisiert, um ein Genexpressionsmuster des entsprechenden Transkripts zu erhalten.

Tabelle 4 gibt eine Zusammenfassung der cDNA Klone wieder, die für die Northern Blot Analysen, sowie später auch für die in situ Hybridisierungen verwendet wurden. Eine humane und zugleich murine cDNA-Probe stand nur für das zu M23451 homologe Transkript zur Verfügung.

PAC Klon	EST Klon	Spezies	Homologien zu
541C22	503334/AA130258	Mensch	EST-Cluster aus PAC-Klon 541C22
1169J3	2048118/AI374804	Mensch	cca3 (AB00216)
1169J3	588263/AA117635	Mensch	Transkript aus <i>Achlya ambisexualis</i> (M23451)
	558312/AA117635	Maus	

Für die Proben 503334 und 2048118 konnten keine Transkripte nachgewiesen werden. Die Hybridisierungsergebnisse der übrigen Proben auf Northern Blots sind in Abbildung 19 gezeigt.

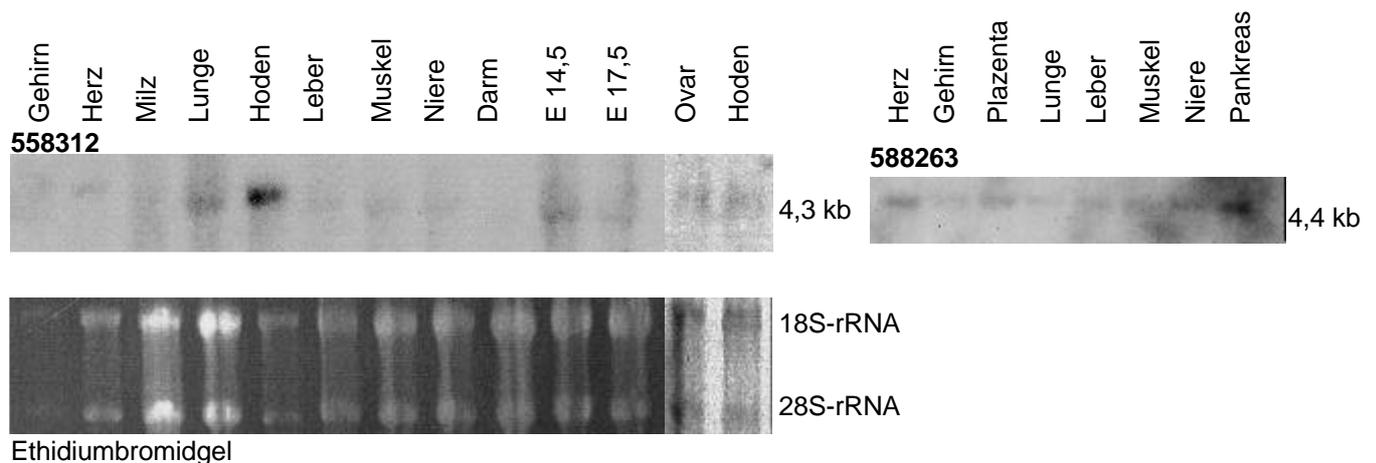


Abb. 19: Hybridisierungen auf Northern Blots mit murinem (links) und menschlichem Gewebe (rechts). Die verwendete Probe ist jeweils oberhalb der Hybridisierungsergebnisse angegeben, die Größe des detektierten Transkripts rechts daneben. Das entsprechende Ethidiumbromidgel mit 18S und 28S-rRNA ist für den Northern Blot mit Mausgeweben gezeigt (unten links). Die Bande in der Spur des Ovars stammt aus einem zweiten Northern Blot. Zum Vergleich wird ein zweites Mal die Spur mit einer RNA-Probe aus dem Hoden gezeigt, die ebenfalls auf diesen Northern Blot aufgetragen war. Der Northern Blot mit humanem Gewebe ist auf β -Aktin normiert (Clontech).

Für die Probe 558312 konnte die Expression eines Transkripts von 4,3 kb in allen präparierten Mausgeweben beobachtet werden. Die Expression war allerdings im Hoden und im Ovar stärker als in den übrigen Geweben.

Das homologe Transkript beim Menschen einer Größe von 4,4 kb konnte mit der Probe 588263 ebenfalls ubiquitär nachgewiesen werden.

5.4.3 Charakterisierung durch in situ Hybridisierungen

Wie schon zuvor bei der Analyse der cDNA-Klone aus dem Exontrapping, wurden mit Proben der IMAGE-Klone in situ Hybridisierungen an Mäusembryonen durchgeführt, um ein entwicklungsabhängiges Expressionsmuster zu erhalten. Ein zum menschlichen EST-Klon homologer Klon in der Maus konnte allerdings nur für das zum Transkript aus *Achlya ambisexualis* (M23451) ähnliche Transkript gefunden werden. Die folgenden Abbildungen zeigen Ergebnisse der Hybridisierung der RNA-Probe aus dem IMAGE-Klon 558312 auf verschiedene Embryonalstadien der Maus.

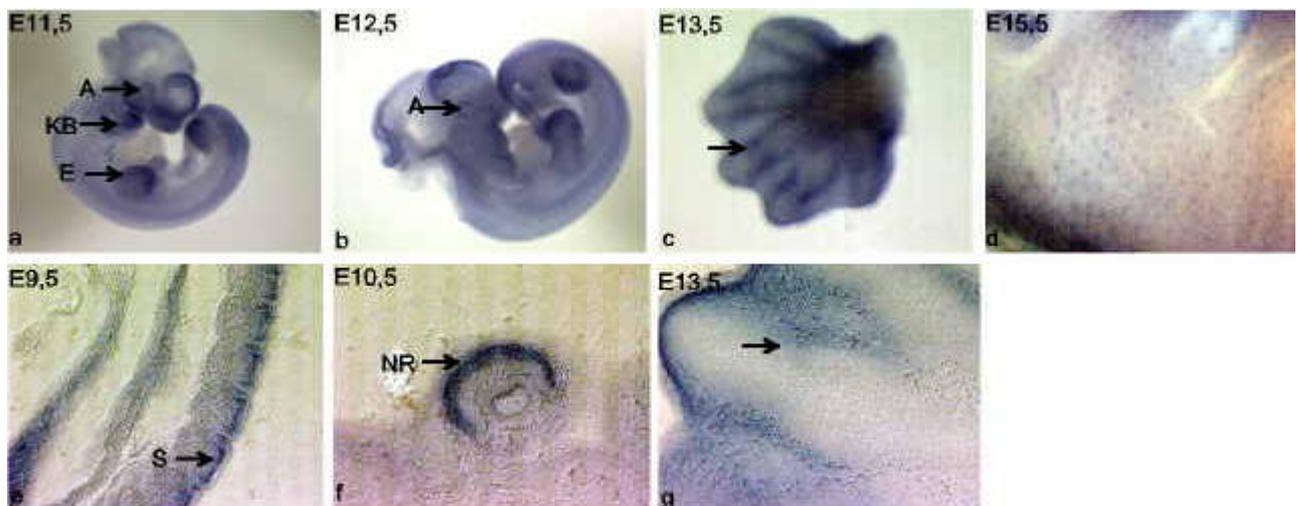


Abb. 20: In situ Hybridisierung der Probe 558312 auf ganze Embryonen (a-d) und Vibratomschnitte der zuvor hybridisierten Embryonen (e-g). Eine Expression ist im Stadium E11,5 (a) und 12,5 (b) im Gehirn, in den Kiemenbögen (KB), im Auge (A) und den Extremitäten (E) zu sehen. Im Embryonalstadium E13,5 (c) konzentriert sich die Expression in den Extremitäten um die sich entwickelnden Finger und Zehen. Positive Signale sind im Embryonalstadium E15,5 in den Haarfollikeln nachweisbar. Auf Vibratomschnitten der hybridisierten Embryonen ist eine deutliche Expression in den Somiten (S) im Stadium E9,5 (e) und im Stadium E10,5 in der neuronalen Retina zu erkennen. In g ist die Expression um die sich entwickelnden Gelenke der Vorderextremität im Längsschnitt gezeigt.

Bei den in situ Hybridisierungen der Probe 558312 auf ganze Embryonen konnte, ähnlich wie im Northern Blot, eine schwache Expression nachgewiesen werden. Allerdings ist die Expression in den getesteten frühen Embryonalstadien nicht ubiquitär, sondern auf Gehirn, neuronale Retina, den Somiten und den Extremitäten beschränkt. Die Expression in den Extremitäten konzentriert sich im Stadium E13,5 auf kleine Bereiche um die sich entwickelnden Finger und Zehen.

