

4 Ergebnisse

Im Rahmen dieser Arbeit wurden cDNA-Bibliotheken aus mRNA zweier Zell-Linien -humaner Lymphocyten leukämischer Patienten- erstellt. Unter Zuhilfenahme der Roboter-Technologie wurden die erhaltenen Klone mit Hilfe eines automatischen Picking-Roboters in Microtiterplatten gepickt, mittels PCR amplifiziert und auf Nylon-Membranen transferiert (gespottet). Daran anschließend folgte das sogenannte Oligonucleotid Fingerprinting (ONF), Hybridisierungen mit 225 radioaktiv markierten Oligonucleotiden (Dekamere mit einem oktameren Core), sowie die Clustering-Analyse. Dies ist eine Methode zur Identifizierung von Klonen mit gleichen oder ähnlichen Sequenzen (detektiert anhand gleicher bzw. ähnlicher Hybridisierungsmuster, den sogenannten Fingerprints), die dann in Clustern zusammengefaßt werden können. Nachfolgend wurden beide Bibliotheken zusammengeclustert, um so die Genexpressionsmuster zu vergleichen und Unterschiede in der Expression beider Leukämie-Zell-Linien zu finden (Co-Clustering). Zur Überprüfung der durch das Fingerprinting erhaltenen Resultate wurden ausgewählte Klone des Co-Clustering mittels Northern-Hybridisierung auf immobilisierte Gesamt-RNA hybridisiert und die so erhaltenen Ergebnisse verglichen. Einen Überblick über den Ablauf des ONF gibt die Abb. 4-1.

Die Methode des Oligonucleotid Fingerprintings eröffnet auch die Möglichkeit der Erstellung eines Unigene-Sets, eines Arrays nicht redundanter cDNA-Klone, beispielsweise der einzelnen Bibliotheken oder der in beiden Bibliotheken differentiell exprimierten Gene. Dies erlaubt die weitere Charakterisierung bestimmter differentiell exprimierter Gene, wie die nähere Identifizierung und quantitative Analyse der Expression oder aber auch die Komplex-Hybridisierung mit mRNA aus Leukämiepatienten. Diese Hybridisierungen mit dem Unigene-Set können zur Differenzierung der Diagnostik und Klassifizierung des Leukämietyps beitragen, aber auch für Screenings von Leukämiepatienten in „Drug Targeting“-Studien und der molekularen Charakterisierung der „Drug Resistance“ genutzt werden, wofür bis heute nur einige wenige Methoden zur Verfügung stehen.

Dieses Kapitel gliedert sich in 3 Teile:

- 1.) - Erstellen der Grundlagen für das Oligonucleotid Fingerprinting - Konstruktion der Bibliotheken
- 2.) - Oligonucleotid Fingerprinting-Hybridisierungen
- 3.) - Analyse und Ergebnisse.

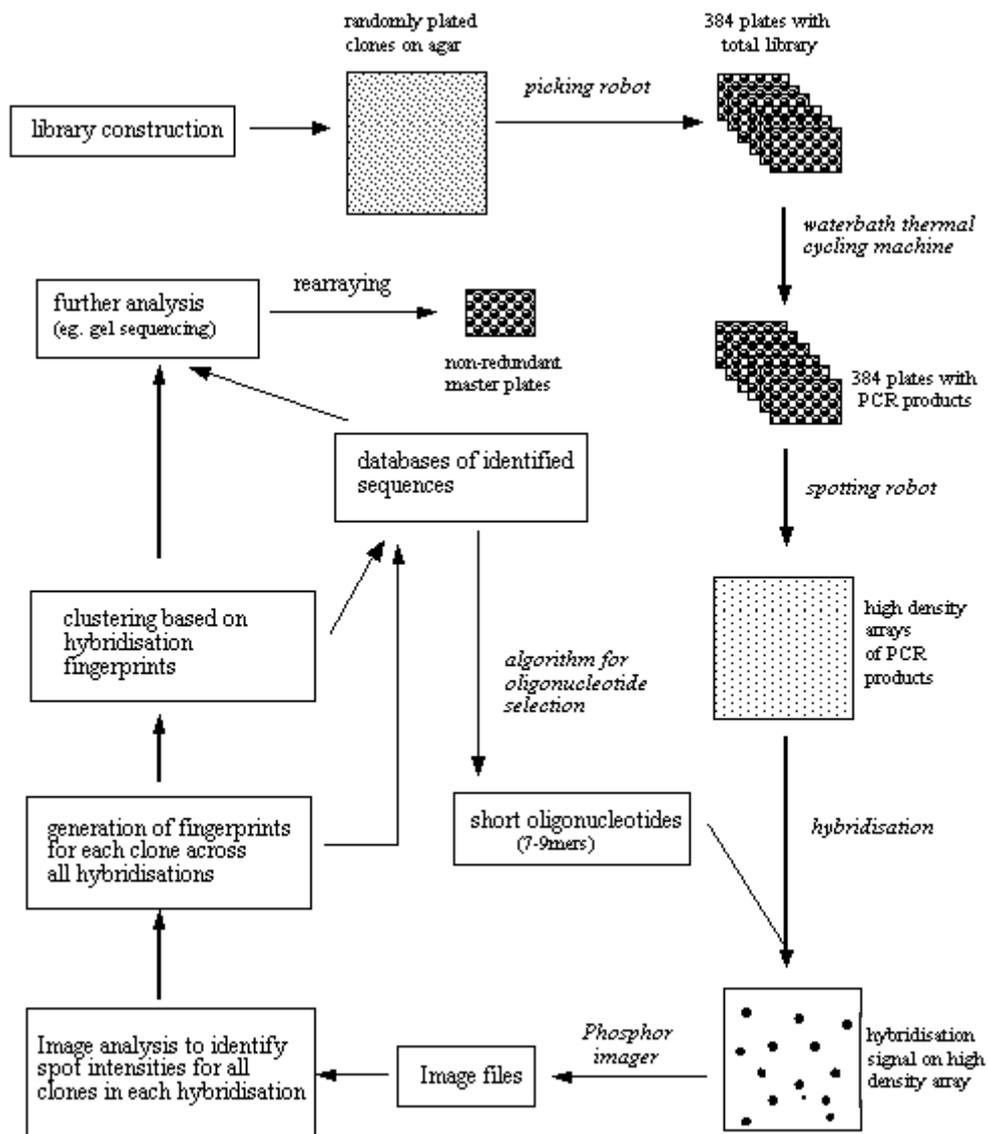


Abb. 4-1 Überblick über das ONF

Überblick über die einzelnen Schritte im Ablauf des Oligonucleotid Fingerprintings:

Ausgehend von der Konstruktion der Bibliothek werden die Klone auf Agar plattiert und mittels eines Picking-Roboters in Microtiterplatten überführt (*picking*). Die Klone werden in einer PCR (im Wasserbad oder in Thermocyclern) amplifiziert und die dabei entstandenen Fragmente mittels Spotting-Roboter auf Nylonfilter übertragen (*spotting*). Dabei entstehen hochdichte Arrays an PCR-Produkten, die dann in den Oligonucleotid Fingerprinting-Hybridisierungen untersucht werden. Die Analyse und Auswertung der Signale erfolgt mit Hilfe verschiedener bioinformatischer Programme, so daß später jeder Klon identifiziert werden kann und gleiche Klone zu sogenannten Clustern zusammengefaßt werden können.

4.1 Erstellen der Grundlagen für das Oligonucleotid Fingerprinting - Konstruktion der Bibliotheken

4.1.1 mRNA-Isolierung aus Jurkat und NKL Zell-Linien

Die Gesamt-RNA der beiden Zell-Linien Jurkat und NKL für das Erstellen der Bibliotheken wurde für diese Arbeit zur Verfügung gestellt von C. Brostjan (Department of Vascular Biology and Thrombosis Research, University of Vienna). Isoliert wurde die Gesamt-RNA nach der Methode von Chomczynski et al. (Chomczynski, P. *et al.*, 1987; Chomczynski, P., 1993), einer Phenol/Chloroform-Isolierung.

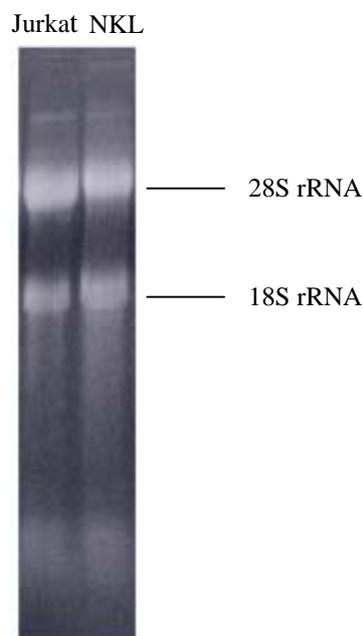


Abb. 4-2 Qualitätskontrolle präparierter RNA

Gelelektrophoretische Auftrennung der präparierten Gesamt-RNA aus Jurkat und NKL zur Qualitätskontrolle.

Deutlich zu sehen sind in der Abb. 4-2 die beiden Banden für die ribosomalen RNAs, 28S und 18S rRNA, die bei einer Fragmentgröße von etwa 5.000 bp bzw. 1.900 bp laufen. Das Verhältnis der Intensitäten/Mengen der beiden ribosomalen Banden auf dem Gel entspricht in etwa dem Faktor 2.

Die RNA kommt etwa 5 - 10x häufiger in der Zelle vor als DNA, sie macht somit den Großteil der Nucleinsäuren in den Zellen aus. Ungefähr 80% der zellulären RNA sind verschiedene rRNA-Typen, weitere 15% werden von den tRNA ausgemacht. Weniger als nur etwa 5% der zellulären RNA ist die hier interessierende mRNA-Population (Tausende verschiedener RNA-Typen). Somit ist bei einer mRNA-Präparation nur etwa eine Ausbeute an mRNA von 1- max. 5% zu erwarten.

Im Rahmen dieser Arbeit wurde zur mRNA-Isolierung und der Abreicherung aller anderen RNA-Arten der Oligotex™ mRNA Purifikation Kit (Qiagen) verwendet, mit Hilfe dessen gegenüber den beiden anderen, magnetischen Systemen, MACS und Dynal, in vorher durchgeführten Tests zwar eine etwas geringere Ausbeute an mRNA erzielt werden konnte, aber die Aufreinigung der mRNA, d.h., die Abreicherung vor allem der in der Gesamt-RNA dominierenden ribosomalen RNAs, am besten realisiert werden konnte. Hier macht man sich das Vorhandensein eines Poly-A-Tails am 3'-Ende eukaryotischer mRNA zunutze, über den die mRNA durch Bindung an Poly-T der Oligotex™-Suspension schnell isoliert werden kann.

Für die Präparation der mRNA aus den Gesamt-RNA-Populationen der beiden Zell-Linien wurden insgesamt je 120 µg Gesamt-RNA eingesetzt und in beiden Fällen ca. 1.15 µg mRNA erhalten, was einer Ausbeute von etwa 0.95% entspricht. Die gesamte Menge an mRNA wurde mit NaOAc gefällt, in 9 µl DEPC-Wasser aufgenommen und für die nun folgende cDNA-Synthese eingesetzt. Um die Effizienz der Fällung zu erhöhen und das Pellet besser sichtbar zu machen, wurde zuvor der Fällungsreaktion 1 µl Glycogen (20 mg/ml) als Carrier zugesetzt.

4.1.2 cDNA-Synthese aus der mRNA

Die reverse Transkription der mRNA in die komplementäre cDNA wurde mit dem Superscript™ Plasmid System for cDNA Synthesis and Plasmid Cloning Kit (Invitrogen) durchgeführt. Dazu wurde jeweils die gesamte isolierte mRNA (1.15 µg) eingesetzt. Um später die Fragmente directional in den Vektor klonieren zu können, ist es vorerst notwendig, eine Asymmetrie der Enden der cDNA einzuführen. Dies geschieht während der Erststrang-Synthese mit Hilfe eines *NotI*-Primer-Adapters und *SalI*-Adaptoren. Der *NotI*-Primer-Adapter enthält am 5'-Ende eine *NotI*-spezifische DNA-Sequenz, die gekoppelt ist an einen am 3'-Ende befindlichen oligo(dT)₁₅-Bereich. Durch die Addition von *SalI*-Adaptoren an die glatten Enden (*blunt ends*) der Fragmente, die während der cDNA-Synthese entstehen, kann die Effizienz der späteren Ligation erhöht werden.

Um die von der mRNA synthetisierten cDNA-Fragmente nachweisen zu können und durch den Auftrag auf das Gel einen nur geringen Verlust an cDNA zu haben, wurde der Transkriptionsreaktion 1 µl [$\alpha^{32}\text{P}$]dCTP zugegeben, wodurch die entstehenden cDNA-Fragmente radioaktiv markiert werden. Ein verdünntes Aliquot der Erststrang-cDNA wird auf ein Gel aufgetragen und gelelektrophoretisch aufgetrennt (Abb. 4-3).

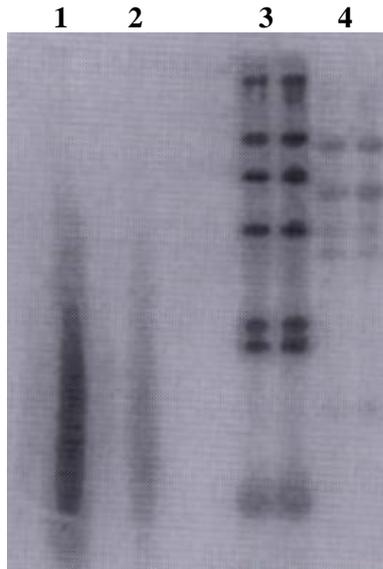


Abb. 4-3 Radioaktiv markierte cDNA nach Gelelektrophorese

Alkalische Gelelektrophorese der radioaktiv markierten cDNA am Beispiel der NKL-Präparation. Die Fragmente der nicht degradierten, heterogenen cDNA-Population decken einen Bereich der Größe von etwa 0.5 bis 4 Kb ab.

- | | |
|---|---|
| 1 | NKL cDNA |
| 2 | NK92 cDNA |
| 3 | λ HinDIII (radioaktiv gelabelter Marker, 5 bzw. 10 μ l) |
| 4 | λ BstEII (radioaktiv gelabelter Marker, 5 bzw. 10 μ l) |

4.1.3 Größenfraktionierung der cDNA-Fragmente

Anschließend an die *NotI/SalI*-Restriktion der cDNA-Fragmente wird eine Größenfraktionierung mittels Säulenchromatographie (cDNA Size Fractionation Columns, Invitrogen) durchgeführt. So können nicht eingebaute Adapter entfernt werden, die ansonsten die Ligations-effizienz vermindern könnten. Auch die kleinen Fragmente, die nach der Restriktion aus den *NotI*-Primer-Adaptoren hervorgehen und *SalI*- und *NotI*-Enden aufweisen, so daß sie die Bibliothek mit scheinbar „leeren Klonen“ kontaminieren können, werden somit entfernt. Der Einbau kleinerer Fragmente (<0.5 kb), die aus z.T. degradierter mRNA hervorgehen, wird vermindert und somit die Wahrscheinlichkeit einer Bibliothek von guter Qualität und einem höheren Anteil an Klonen voller Länge (*full length clones*) erhöht. Die Größenfraktionierung geschieht anhand der Trennung der DNA-Moleküle in Sepharose-Säulen und der sich daran anschließenden Messung der radioaktiven Zerfallsrate (Cerenkov-Counts) mittels eines Scintillationsmeßgerätes. Anhand dieser Raten läßt sich die Menge der in der entsprechenden Fraktion enthaltenen cDNA kalkulieren. Eine solche Quantifizierung der DNA-Menge ist in der Tab. 4-1 dargestellt.

Fraktion	Vol _{Fraktion} (μ l)	Vol _{total} (μ l)	Cerenkov Counts (cpm)	cDNA (ng)	C _{cDNA} (ng/ μ l)
1	77	77	16,95		
2	99	176	17,94		
3	27	203	14,94		
4	24	227	11,96		
5	23	250	9,97		
6	27	277	372,95	6,15	0,227
7	29	306	17,94		
8	30	336	78,77	1,3	0,043
9	27	363	396,88	6,55	0,24
10	29	392	1129,01	18,63	0,64
11	23	415	1295,63	21,38	0,93
12	17	432	1360,48	22,45	1,32
13	20	452	1767,6	29,17	1,46
14	21	473	2158,81	35,63	1,69
15	23	496	2573,08	42,46	1,84
16	28	524	3249,03	53,62	1,915
17	29	553			
18	31	584			
19	26	610			
20					

Tab. 4-1 Größenfraktionierung der cDNA

Die tropfenweise Größenfraktionierung der cDNA mittels Säulenchromatographie ist am Beispiel der Jurkat-Fraktionierung dargestellt.

Mit steigender Fraktionszahl werden die Fragmente, die in den jeweiligen Fraktionen enthalten sind, immer kleiner. Fraktionen (in diesem Beispiel ab Fraktion 17), mit denen ein Gesamtvolumen von mehr als 550 μ l überschritten wird, enthalten sehr kleine Fragmente, nicht ligierte Adapter und Primer-Adapter, die durch die *NotI*-Restriktion freigesetzt wurden. Diese Fraktionen werden nicht mehr zur Ligation verwendet, da diese Fragmente die Qualität der Bibliothek stark beeinträchtigen würden. Zum Erstellen einer Bibliothek ist es wichtig, Fraktionen zu verwenden, die möglichst große Fragmente enthalten, aber auch eine ausreichende Menge an DNA zur Verfügung stellen.

4.1.4 Ligation und Transformation

Die Ligation der cDNA-Fragmente erfolgte in den Expressionsvektor pQE30NST (Amp^r), einem Derivat des Vektors pQE30 (Qiagen). Dieser Vektor enthält die für die β -Galactosidase codierende Sequenz sowie 2 *lac*-Operatoren für eine durch IPTG induzierbare (His-tag) Protein-Expression. Durch das Vorhandensein der 2 *lac*-Operatoren erhöht sich die Effektivität der Bindung der Repressoren, somit wird die Repression des starken T5-Promotors sichergestellt. Die multiple Klonierungsstelle (*multiple cloning site*, MCS) des Vektors pQE30 wurde so modifiziert,

daß eine gerichtete Klonierung über die Restriktionsschnittstellen *NotI/SalI* möglich ist (Büssow, K. *et al.*, 1998).

Der Stamm *E. coli* SCS1 weist natürlicherweise keine Defizienz der β -Galactosidase-Aktivität auf, so daß hier die Blau-Weiß-Selektion zur Beurteilung der Ligations- bzw. Transformations-effizienz nicht genutzt werden kann.

Zur Kontrolle der Ligation wurde eine Vektor-Religation durchgeführt. Aufgrund der Restriktion des Vektors mit 2 Enzymen, die jeweils einen 3'- bzw. einen 5'-Überhang zur Folge haben, sollte keine Religation zu erwarten sein. Nach Transformation des Religations-Ansatzes konnten auch keine Transformanten nachgewiesen werden und somit eine Religation des Vektors ausgeschlossen werden. Der präzipitierte Ligationsansatz wurde nun per Elektroporation in den *E.coli*-Stamm SCS1, der die Plasmide pREP4 und pSE111 enthält, transformiert und auf Agarplatten über Nacht inkubiert.

Das *low copy* Plasmid pREP codiert für die Kanamycin-Resistenz (Km^r) und konstitutiv für den Lactose-Repressor (*lacI^q*), wodurch die Expression eines hohen Levels an Repressor sichergestellt werden kann, der den Operator bindet und somit streng die Protein-Expression reguliert. Das Plasmid pSE111 (Km^r) dient als Helfer-Plasmid für die Überexpression des *lacI^q*-Repressors. Weiterhin trägt das Plasmid das ArgU-Gen, das für eine seltene Arginin-tRNA codiert, so daß die Expression der Gene, die multiple AGG oder AGA enthalten, gesteigert werden kann (Schenk, P. M. *et al.*, 1995).

So wurden im Falle der Jurkat-Linie die Fraktionen 6, 8 und 9 (vgl. Tab. 4-1) gepoolt, um die notwendigen 10 ng zu erhalten, und in den Vektor ligiert. Jedoch konnten nach der Elektroporation keine Transformanten erhalten werden. So wurden die nächst kleineren Fraktionen 10 und 11 gepoolt bzw. 10 ng der Fraktion 11 in den Vektor ligiert und transformiert. Für die NKL-Bibliothek wurden die Fraktionen 8, 9 sowie 10 zur Ligation verwendet und transformiert. Die Transformationsgemische wurden auf 2YT-Agar plattiert, die die beiden Antibiotika Ampicillin und Kanamycin zur Doppelselektion enthielten.

4.1.5 Transfer der Kolonien

Da der Transfer Tausender Kolonien manuell nicht ohne großen Aufwand durchführbar ist, wurden die Transformanten mittels eines Roboters von den Agarplatten in mit den entsprechenden Medien und Zusätzen befüllte (Q-Fill II, Genetix) 384-well Microtiterplatten (MTP) inokuliert (*picking*), so daß jeder Klon nun durch Koordinaten definiert und somit adressierbar ist. Der Einsatz eines solchen Roboters erhöht die Geschwindigkeit des Kolonie-Transfers um das Zehnfache im Vergleich zum Transfer per Hand (Maier, E. *et al.*, 1994). Die Kolonien werden dabei von einer CCD-Kamera detektiert und von einer Software analysiert. Die Auswahl der zu

pickenden Kolonien erfolgte aufgrund verschiedener Parameter, wie der Größe der Kolonien (Durchmesser 0.5 - 1 mm), der Form (Grad der Rundheit), dem Kontrast (Grauwerte) und dem Abstand der Kolonien zueinander. Die Bewegungen des Roboters beruhen dabei auf einem 3-Achsen-System, so daß in x/y/z Richtung präzise jeder Punkt im dreidimensionalen Arbeitsbereich des Roboters angesteuert werden kann. In der Abb. 4-4 ist der Picking-Kopf mit dem 96-Pintool und der Kamera eines solchen Roboters dargestellt.

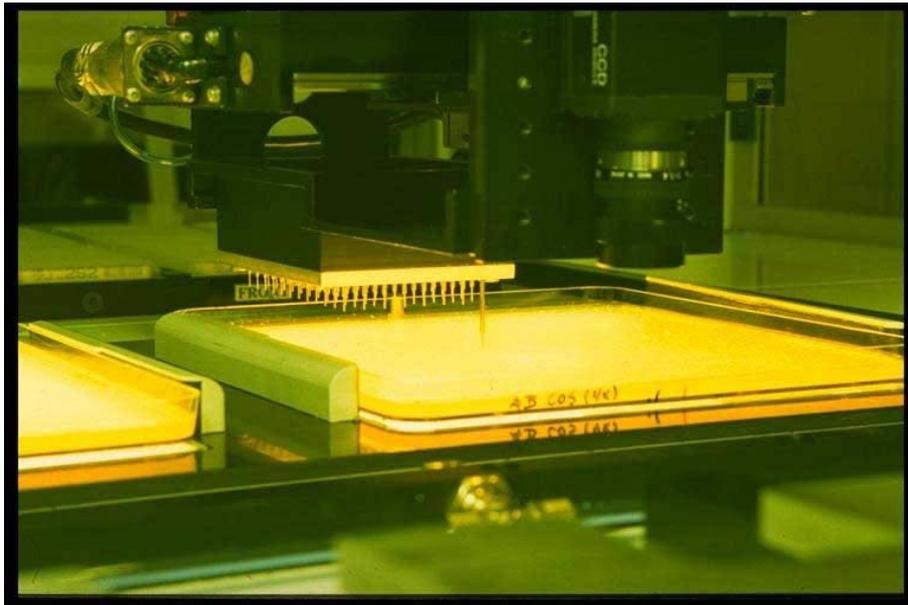


Abb. 4-4 Picking-Roboter

Picking-Kopf mit CCD-Kamera und 96-Pintool beim Picken einer Kolonie von einer Agarplatte.

Die Kolonien werden mit einer Geschwindigkeit von durchschnittlich 3.000 - 4.000 Kolonien in der Stunde gepickt. Gepickt wird mit einem 96-Pintool (*gadget*), bei dem jeder einzelne Pin federnd gelagert und individuell pneumatisch ansteuerbar ist. Die Abb. 4-5 zeigt die Momentaufnahme eines ausgefahrenen Pins beim Picken einer Kolonie.

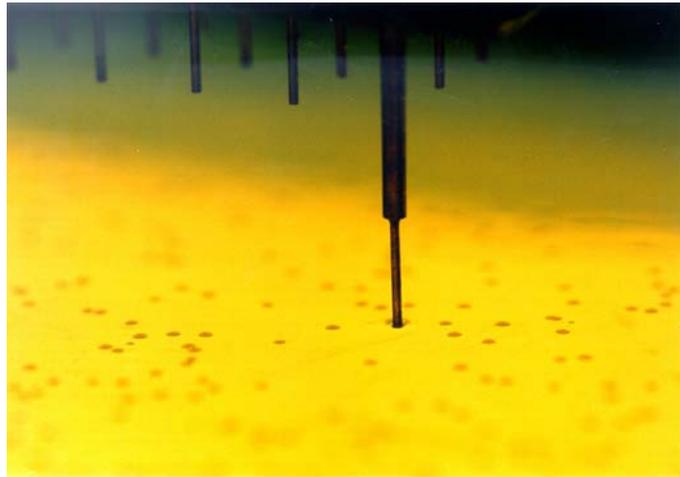


Abb. 4-5 Picking-Roboter - Detailansicht

Moment-Aufnahme eines ausgefahrenen Pins beim Picken einer Kolonie.

Insgesamt konnten für die Bibliothek Jurkat etwa 24.100 Kolonien (63 Microtiterplatten) und für die NKL-Bibliothek etwa 35.300 Kolonien (92 Microtiterplatten) isoliert werden. Die Microtiterplatten wurden anschließend ÜN bei 37°C inkubiert. Von diesen Originalen wurden dann jeweils 3 Arbeitskopien erstellt und das Original sowie alle Kopien zur Lagerung bei -80°C eingefroren.

4.1.6 Analyse der Bibliotheken

Um die optimale Qualität der Bibliotheken sicherstellen zu können, wurden vor den weiteren, aufwendigeren Arbeitsschritten einige Kontrollen durchgeführt. Um Klonierungsartefakte, wie z.B. ein Überhäufen der Bibliothek mit rRNA oder zu vielen „leeren“ Klonen auszuschließen zu können, wurden schon vor dem eigentlichen Picken der Bibliotheken jeweils etwa 100 Kolonien aus Verdünnungen des Transformations-Ansatzes ansequenziert. Dabei konnte ein breites Spektrum verschiedener cDNAs nachgewiesen werden, darunter auch Glycerinaldehyd-3-phosphat-dehydrogenase, Calmodulin, Cytochromoxidase C, Tubulin und natürlich auch einige ribosomale RNAs. Jedoch konnten keine Klonierungsartefakte festgestellt werden, und die Bibliothek wurde vollständig auf 22.4 x 22.4 cm Agarplatten ausplattiert und gepickt.

Der Umfang der Jurkat-Bibliothek (cfu total) insgesamt betrug etwa 55.650 Klone. Von dieser Gesamtzahl konnten ca. 43% der Klone gepickt werden. Beim Wachstum der Kolonien kam es wiederum zu einem etwa 4%-igen Ausfall, so daß die Bibliothek insgesamt eine Anzahl von etwa 23.100 Klonen (in 63 MTP) umfaßt. 103 Klone dieser Bibliothek wurden zur Größenbestimmung in der PCR ausgewertet (s. Abb. 4-7) Die durchschnittliche Insertgröße liegt bei 1.5 kb, wobei die Fragmente einen Bereich von 0.5 - 5 kb abdecken.

Der Umfang der NKL-Bibliothek betrug insgesamt 57.600 Klone, von denen etwa 61% gepickt werden konnten. Auch hier konnte ein Ausfall beim Wachstum nach der Inokulation von 4% beobachtet werden. Die NKL-Bibliothek umfaßt somit 34.000 Klone (in 92 MTP). 47 der Klone wurden zur Größenbestimmung in der PCR amplifiziert und untersucht. Die Fragmente dieser Bibliothek umfassen Größen von 0.5 - 4 kb.

Der Ausfall beim Picken läßt sich dadurch erklären, daß die Kolonien beispielsweise zu dicht zueinander gewachsen sind und somit vom Picken ausgeschlossen wurden. Dies gilt auch, wenn Kolonien nicht anderen, vorgeschriebenen Parametern, wie dem Grauwert oder dem Durchmesser, entsprechen. Wachsen Kolonien zu dicht am Rand der Agarplatte, so können sie von Roboter nicht erreicht werden. Ausfälle, die hingegen beim Wachstum beobachtet werden, können darauf zurückzuführen sein, daß einige Klone nach dem Inokulieren in die Microtiterplatten nicht anwachsen konnten. Auch kann es vorkommen, daß ein Pin die zu pickende Kolonie nicht getroffen hat. Der prozentuale Anteil der Klone ohne Insert beider Bibliotheken beträgt in etwa 8%.

Beide Bibliotheken sind mit 23.100 bzw. 34.000 Klonen relativ klein, was darauf zurückzuführen ist, daß sich aus Lymphocyten im Vergleich zu anderen Zell- oder Gewebetypen weniger mRNA präparieren läßt. Die Wahrscheinlichkeit, seltenere Transkripte zu finden, liegt dementsprechend etwas geringer als in Bibliotheken mit einer größeren Anzahl an Klonen.

4.1.7 PCR in großem Maßstab

Bei der Oligonucleotid Fingerprinting-Analyse werden nur sehr kurze Oligos für die Hybridisierungen benutzt. Um den Fingerprint zu generieren, werden die Hybridisierungsmuster der einzelnen Hybridisierungen ausgewertet und miteinander verglichen. Dafür ist es notwendig, hochreine DNA in hoher Konzentration einzusetzen, da in größeren Mengen noch vorhandene Bakterien-DNA oder Vektor-Sequenzen sonst zu unspezifischen Bindungen der Oligos an diese DNA und somit zu falsch positiven Signalen oder zu starken Background-Signalen führen würden. Dies geschieht mittels PCR in großem Maßstab (*large scale PCR*), wobei mit Hilfe von 384-Pin Replikatoren (Genetix) direkt aus der Bakterienkultur in den PCR-Mix inokuliert wird. Dabei werden etwa 0.2 µl der Bakterien-Suspension übertragen. Die Amplifikation erfolgt in mit 30 µl PCR-Mix befüllten 384-well PCR-Microtiterplatten (ABgene) in PTC225 Thermocyclern (MJ Research). Im Rahmen dieser Arbeit wurde die PCR in großem Maßstab von der Wasserbad-PCR auf die PCR in den Thermocyclern PTC225 umgestellt und etabliert und ein PCR-Programm an die PCR-Microtiterplatten (ABgene) optimiert. Das Befüllen der PCR-Platten erfolgte mit der Q-Fill II-Maschine (Genetix). In der Abb. 4-6 sind mehrere Blöcke solcher Thermocycler dargestellt.



Abb. 4-6 Thermocycler PTC225

Thermocycler PTC225 (MJ Resarch-Tetraden), die zur Durchführung der *large scale* PCR-Amplifikation der Bibliotheken genutzt wurden. In jeder der Tetraden sind gleichzeitig vier Microtiterplatten unabhängig voneinander amplifizierbar.

Pro Lauf kann die PCR gleichzeitig von 18 Platten durchgeführt werden. Bei drei bis vier Läufen am Tag konnte so eine gesamte Bibliothek innerhalb von 2 Tagen vollständig amplifiziert werden. In der in diesem Institut entwickelten und bisher genutzten Wasserbad-PCR konnte eine Bibliothek dieser Größe an einem einzigen Tag amplifiziert werden (bis zu 50.000 Klone). Da allerdings die Ausbeute in den Thermocyclern (Tischgeräte) etwa einer doppelten Ausbeute im Vergleich zu der Wasserbad-PCR entspricht, wurde sich für die Amplifikation der Bibliotheken in den Tischgeräten entschieden. Ein weiterer Vorteil besteht darin, daß die Microtiterplatten sich hier nicht wie in der Wasserbad-PCR verformen und so ein zusätzliches Umpipettieren in neue Microtiterplatten entfällt. Die PCR-Platten können sofort nach der PCR gespottet werden. Lediglich eine Metallplatte, die zur Stabilisierung unter den flexiblen PCR-Platten angebracht wird, ist notwendig, so daß ein fehlerfreies und stabiles Arbeiten des Roboters (*handling*) gewährleistet werden kann.

Die PCR erfolgte in einem 25 Zyklen umfassenden 2-Schritt-Programm, das die Denaturierung für 45 sec bei 94°C sowie das Annealing und die Elongation für 4 min bei 65°C beinhaltet. Ein Aliquot von 1 µl der PCR-Reaktion zufällig ausgewählter Klone wurde auf ein 1%-iges Agarosegel aufgetragen, um die Qualität und Quantität der PCR-Produkte zu kontrollieren. Die Abb. 4-7 gibt einen Überblick über die Fragmentgrößen und Konzentrationen von 82 aufgetragenen Klonen.

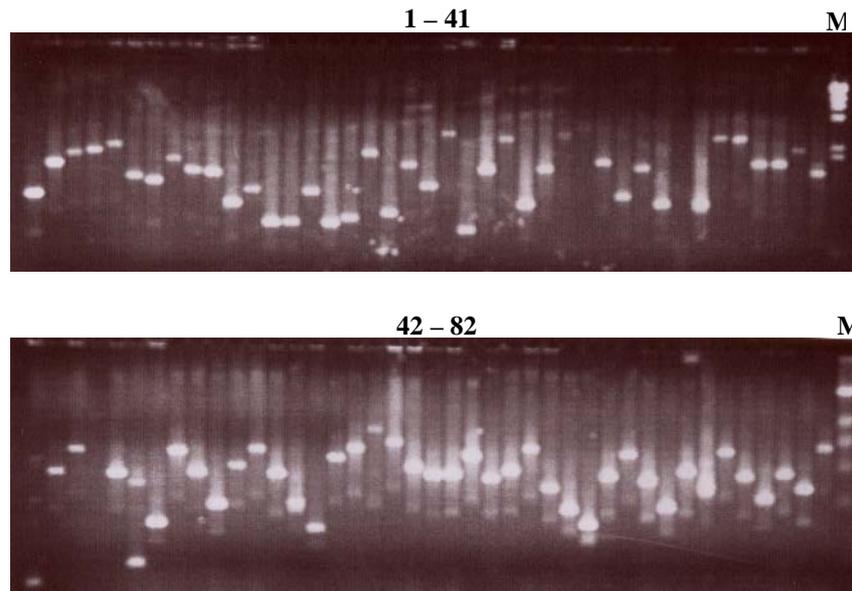


Abb. 4-7 Gelelektrophoretische Auftrennung der Fragmente nach *large scale* PCR

Gelelektrophoretische Auftrennung nach *large scale* PCR zur Kontrolle der Quantität sowie Qualität der amplifizierten Produkte, aufgetragen wurde 1 μ l der PCR-Produkte.

1-82 Auswahl einiger Klone der Jurkat-Bibliothek nach *large scale* PCR-Amplifikation

M λ HinDIII 500 ng

Einige der aufgetragenen Klone zeigen keine Amplifikation der Inserts oder tragen keine Inserts (s. Kapitel 4.1.6 Analyse der Bibliotheken), was in einer Plasmidpräparation mit anschließendem Retriktionsverdau nachgewiesen werden konnte. Der Ausfall der Amplifikation liegt ungefähr bei 8%. Die durchschnittliche Konzentration der PCR-Produkte liegt bei etwa 100 - 150 ng/ μ l.

4.1.8 Transfer der PCR-Produkte auf Membranen

Für das Screening der Bibliotheken anhand der folgenden Hybridisierungen wurden cDNA- bzw. Kolonie-Filter hoher Dichte erstellt. Dazu wurden die PCR-Produkte der Klone aus den Microtiterplatten auf Nylon-Membranen transferiert (*spotting*) und immobilisiert. Diese Filter repräsentieren alle Klone der Bibliothek. Da die DNA eine negative Nettoladung aufweist, wurden Nylon-Transfer-Membranen verwendet, deren Oberfläche positiv geladen ist (Hybond-N⁺, Nylon Transfer Membrane, Amersham) und somit die Bindung der DNA an die Filteroberfläche erleichtert. Anschließend wurden die Membranen bei 80°C gebacken und mit UV-Licht behandelt, um die DNA-Fragmente nun auch kovalent an die Membran zu binden und zu immobilisieren und damit die Sensitivität und Haltbarkeit auch nach mehrmaligem Strippen und Rehybridisieren zu gewährleisten (Nierzwicki-Bauer, S. A. *et al.*, 1990; Semikozova, O. P. *et al.*, 1992). Der Transfer der PCR-Produkte wurde mit Hilfe eines Spotting-Roboters, wie er in Abb. 4-8 gezeigt ist,

realisiert. Zum Transfer wurde ein 384-Pin Spotting-Gadget eingesetzt, bei dem -anders als beim Picking-Gadget- nicht jeder Pin einzeln ansteuerbar ist, sondern alle Pins, federnd gelagert, als eine Einheit gleichzeitig den Transfer bewerkstelligen (Meier-Ewert, S. *et al.*, 1993; Lehrach, H. *et al.*, 1997).

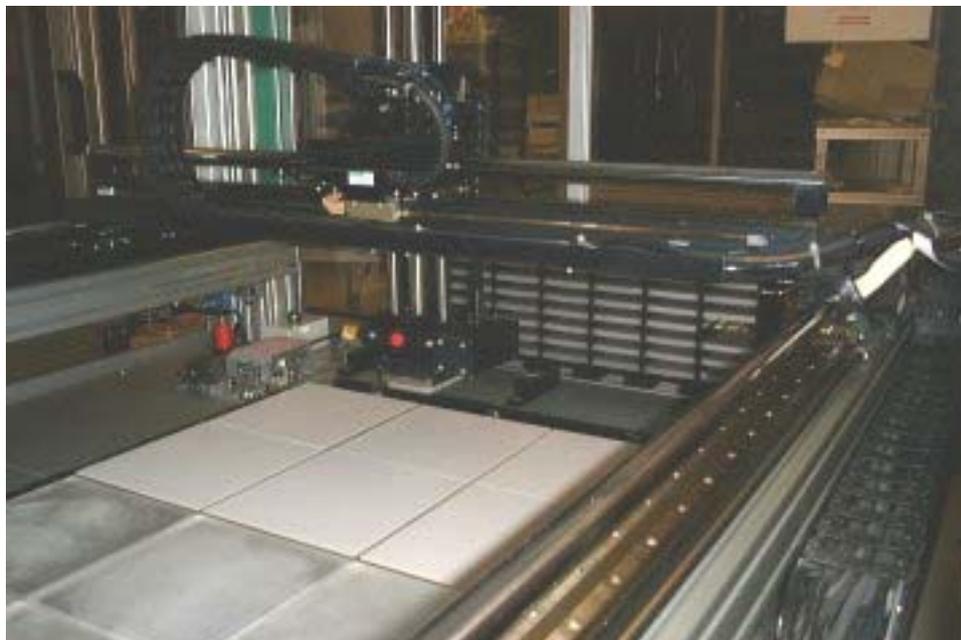


Abb. 4-8 Spotting-Roboter

Spotting-Roboter beim Transfer amplifizierter DNA aus den Microtiterplatten auf Nylon-Membranen. Aus einem Plattenstapel, in dem MTP einer Bibliothek angeordnet sind, wird nacheinander eine MTP in einen dafür vorgesehenen Plattenhalter transportiert. Nach dem automatischen Abheben des MTP-Deckels verfährt der Roboter das Pintool in die Wells der MTP (Benetzung der Pins mit DNA-Lösung) und anschließend auf eine vorher exakt berechnete Position auf der Nylon-Membran (Abgabe von Flüssigkeitstropfen durch Berührung der Membran, *spotting*). Dieser Vorgang wiederholt sich entsprechend den gewählten Parametern der Robotersoftware für verschiedene Positionen des Arrays, der Anzahl der Filter und der MTP.

Die Microtiterplatten wurden dabei nach bestimmten Parametern und in einer bestimmten Anordnung (Pattern) auf die Membranen übertragen. Hier wurde ein Pattern von 5x5 Spots in Duplikaten gespottet. Duplikate der einzelnen Klone werden gespottet, um falsch positive Signale erkennen und von der Analyse ausschließen zu können. In der Mitte eines jeden 5x5-Blockes befindet sich der sogenannte *guide dot*. An diese Position wird eine Lösung aus Lachsspermien-DNA gespottet, die in jeder Oligonucleotid-Hybridisierung Signale gibt und somit der Orientierung bei der späteren Bildanalyse dient. Anhand dieser Signale wird ein entsprechendes Gitternetz zum Finden der Signalpositionen erstellt (*grid finding*). Ein solcher 5x5-Block ist in Abb. 4-9 dargestellt.

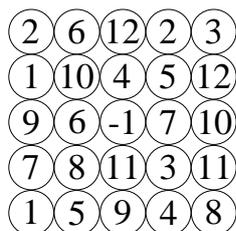


Abb. 4-9 Darstellung eines Blockes des 5x5 Patterns

Anordnung der Klone im Pattern 5x5 in Duplikaten. -1 in der Mitte des Patterns entspricht dem Guide Dot.

16x24 solcher Blöcke (entsprechend der Anzahl an Wells in einer Microtiterplatte) ergeben ein Feld. In jedes Feld können bis zu 12 Microtiterplatten gespottet werden. Ein Filter enthält 6 dieser Felder. Um die dichte Anordnung der Klone auf der Membran einmal zu veranschaulichen: der Abstand der Pins zueinander beträgt 4500 µm, was dem Abstand der Mittelpunkte der Wells der Microtiterplatten entspricht. So beträgt bei einem Spotting-Pattern von 5x5, welches hier verwendet wurde, der Abstand der Spots auf der Membran (*dot pitch*) 900 µm. Auf einen Filter der Größe 22.2 x 22.2 cm können so bis zu 27.648 Klone in Duplikaten gespottet werden. Das entspricht einer Gesamtzahl von insgesamt 55.296 Spots auf der Membran. Ein 250 µm Pin, wie er hier verwendet wurde, transferiert im Durchschnitt 2 nl der DNA-Lösung (Eickhoff, H. *et al.*, 2000), so daß per Spot bei 10-maligem Transfer und einer durchschnittlichen DNA-Konzentration der PCR-Produkte von 100 - 150 ng/µl eine DNA-Menge von etwa 2 - 3 ng auf die Membranen aufgetragen wird.

Generell ist es möglich, die einzelnen Filter mehrmals zur Hybridisierung zu verwenden. Jedoch bei über 200 notwendigen Hybridisierungen ist eine gleichbleibend gute Qualität der Signale bei der Benutzung nur einiger weniger Filter nicht mehr zu gewährleisten. Ein Filter kann max. für 10 - 20 Hybridisierungen verwendet werden, dann ist die DNA durch das Strippen und mechanische Beanspruchung weitgehend abgelöst. So wurden insgesamt von der Bibliothek Jurkat 40 Filter, von der NKL-Bibliothek 2 Sets (von jeweils 26 MTP) mit je 30 Filtern gespottet.

Die Filter einer Bibliothek sind jedoch nicht absolut identisch, was bei der späteren Bildanalyse beachtet werden muß. Verschiedene Faktoren können hier die Qualität der Filter beeinflussen. Einige Fehlerquellen können beispielsweise eine veränderte Luftfeuchte oder Temperatur sein, die die Verdunstungsrate der DNA-Lösung beeinflussen, woraus eine unterschiedliche Konzentration der PCR-Produkte in den Microtiterplatten resultieren kann und somit auch unterschiedliche Mengen der transferierten DNA. Auch die Qualität der Nylon-Membranen kann die Gleichmäßigkeit des Musters beeinflussen.

Diese Probleme von vornherein auszuschließen ist nahezu unmöglich, jedoch können sie durch genaues Einhalten der Arbeitsprotokolle und Methoden (Beachten der Qualität der Pins) weitgehend minimiert werden. Durch anschließende Kontrollen (Background-Hybridisierung) können fehlerhafte Filter erkannt und entsprechend von der Analyse ausgeschlossen werden bzw. bestimmte Phänomene *in silico* korrigiert werden, d.h. mittels Software und verschiedener Algorithmen wie der Normalisierung der Daten, auf die später noch etwas genauer eingegangen werden soll, beseitigt werden.

Um die Filter zu erstellen, wurde die Methode des Naß-Spottens angewandt, bei der die Filter vor dem Aufbringen der PCR-Produkte in NaOH getränkt werden. Das dient der Denaturierung der PCR-Produkte während des Aufbringens auf die Membran. Die Spots sind etwas größer als beim Trocken-Spotting, bei dem die Filter vorher nicht befeuchtet werden, es hat jedoch dem Trocken-Spotten gegenüber den Vorteil, daß die Filter plan aufliegen und so ein zusätzliches Fixieren der Filter im Roboter entfällt. Auch ist beim Naß-Spotten die DNA-Transferrate etwas größer, da sich die DNA-Lösung besser vom Pin auf die feuchte Membran übertragen läßt. Der Vorteil des Trocken-Spottings liegt darin, daß die Membranen nicht so stark äußeren Einflüssen, wie Schwankungen der Temperatur und der Luftfeuchte, unterliegen, so daß längere Läufe mit mehr Filtern durchgeführt werden können.

Die PCR-Produkte werden in Duplikaten gespottet als eine direkte interne Kontrolle der Filter- sowie Hybridisierungs-Qualität. Normalerweise sollten beide Duplikate bei einer Hybridisierung mit einem komplementären Oligonucleotid ein Signal geben, das idealerweise mit gleicher Signalintensität auftritt. Da dies aber in der Praxis kaum erreicht werden kann, wird zur Vermeidung falsch positiver Signale der Quotient der Intensitäten aus dem maximalen (max) und dem minimalen (min) Duplikatsignal berechnet und mit einem spezifischen Grenzwert (hier 1.5) verglichen. Bei einem Verhältnis der maximalen zur minimalen Intensität eines Duplikatpaares von kleiner als 1.5, die Abweichung der Signalintensitäten also nur gering ist, wird der Mittelwert beider errechnet und zur Analyse verwendet. Ist das Verhältnis beider Duplikatintensitäten jedoch größer als 1.5, die Duplikate nur schlecht reproduzierbar, so wird zur weiteren Analyse der minimale Wert in Betracht gezogen.

$$\frac{\max}{\min} < 1.5 \Rightarrow \text{Mittelwert}$$

$$\frac{\max}{\min} > 1.5 \Rightarrow \text{Minimum}$$

4.2 Oligonucleotid Fingerprinting-Hybridisierungen

4.2.1 Background-Hybridisierung

Vor den eigentlichen Oligonucleotid-Hybridisierungen wurden sogenannte Background-Hybridisierungen durchgeführt. Mit dieser Art Hybridisierung können alle gespotteten Klone auf dem Filter sichtbar gemacht werden, um so die Qualität des Spottens in Bezug auf die Gleichmäßigkeit der Spots hinsichtlich der Anordnung und DNA-Menge abschätzen und überprüfen zu können. Als Sonde wurde einer der für die PCR-Amplifikation verwendeten Primer (30-mer) genutzt. Dieser sollte -radioaktiv markiert- an alle Produkte binden und so auf allen Positionen, die gespottet wurden, ein Signal geben. Gleichzeitig stellt die Background-Hybridisierung eine Art Positivkontrolle für die folgenden Hybridisierungen dar. So können Klone von der Analyse ausgeschlossen werden, die nachweislich in der PCR nicht amplifiziert wurden. Weiterhin dient diese Hybridisierung zur Normalisierung der variierenden Menge an PCR-Produkt.

Filter, die größere Mängel aufweisen, werden von der weiteren Verwendung ausgeschlossen. Dazu zählen beispielsweise Spotting-Phänomene, bei denen es dazu kommen kann, daß die Randpositionen der einzelnen Felder stärker ausgeprägt sind als die gesamten Felder, größere mechanische Läsionen oder Unregelmäßigkeiten im Spotting-Pattern (Fehlen einiger Blöcke).

4.2.2 Oligonucleotid-Hybridisierung

Die Oligonucleotid-Hybridisierungen wurden durchgeführt mit 225 radioaktiv markierten Oligos, jeweils einem Pool von Dekameren mit einem oktameren Core (XNNNNNNNNX). Ausgehend vom oktameren Kern der Oligos und den vier Basen (A,C,G,T) ergibt sich eine Anzahl von $4^8 = 65.536$ möglichen Oligos, die theoretisch in den Hybridisierungen eingesetzt werden könnten. Die Auswahl der für die Hybridisierung einzusetzenden Oligos erfolgte nicht rein zufällig, was prinzipiell möglich wäre (Schmitt, A. O. *et al.*, 1997), sondern die Oligos wurden nach bestimmten Parametern ausgesucht oder kreiert (Fu, Y. X. *et al.*, 1992). Es wurden solche Oligos ausgewählt, die aufgrund ihrer Hybridisierung theoretisch die Klon-Population am besten diskriminieren können. Das bedeutet, ein Oligo sollte optimalerweise den Klon-Pool so teilen, daß etwa 50% der Klone Homologie zum Oligo aufweisen, es also binden können, und 50% der Klone nicht hybridisieren. Dies ist abhängig vom zu untersuchenden Organismus und kann in verschiedenen Organismen recht unterschiedlich sein (Herwig, R., 2000; Herwig, R. *et al.*, 2000). Man kann so für verschiedene Gruppen von Individuen (Pflanzen, Nagetiere, Mensch) Oligo-Sets erstellen, die für diese Organismen mehr spezifisch sind. Nicht verwendet werden dabei Oligos, die einen zu geringen GC-Gehalt aufweisen (AAAAAAA), da hier die Bindung des Oligos an die

Zielsequenz nicht spezifisch genug sein würde. Oligos mit einem höheren GC-Gehalt bilden stabilere Hybride mit der Zielsequenz und führen somit zu spezifischeren Signalen. Weiterhin schließt man Oligos aus, die Redundanzen zueinander aufweisen und einander zu ähnlich wären (ACGTAGTA : CGTAGTAC). Diese würden relativ ähnliche Signale geben, sind daher nicht oder nur von geringerem bioinformatischen Wert. Auch revers komplementäre Oligos (ACGTAGTA : TACTACGT) kommen nicht zum Einsatz, da die gespotteten PCR-Produkte als denaturierte Doppelstränge vorliegen und so die revers komplementären Oligos gleiche Signale geben würden.

Es gibt einige interne Faktoren und Kontrollen, anhand derer man die Spezifität und Qualität der Hybridisierungs-Signale überprüfen kann. Zum einen können Kontrollklone mit gespottet werden, deren Sequenz bekannt ist. Diese hybridisieren in den Oligonucleotid-Hybridisierungen und ergeben ein spezifisches Muster und dienen somit als interne Kontrolle.

Das Spotten von Duplikaten dient als weitere Möglichkeit zur Kontrolle, beispielsweise um falsch positive Signale ausschließen zu können. Normalerweise sollten beide Duplikate ein Signal geben bei der Hybridisierung mit einem zu ihnen komplementären Oligo, idealerweise mit gleicher Signalintensität.

Die Duplikatintensität setzt sich, wie schon erwähnt, zusammen aus dem maximalen und dem minimalen Wert des x- und des y-Duplikates eines Duplikatpaares einer Hybridisierung verglichen mit einem bestimmten Schwellenwert. Sie spiegelt den Grad der Übereinstimmung der Duplikate eines Duplikatpaares wider. Für die Analyse der Signalintensität wird zunächst das Verhältnis der Duplikatintensitäten berechnet (max : min). Ist dieser Quotient zu groß, im Allgemeinen >1.5 , so wird das Minimum der Duplikatsignale verwendet, ansonsten das arithmetische Mittel aus beiden Signalen. Die Duplikatintensität eines jeden Klons einer Hybridisierung wird in die sogenannte uhd-Datei (*unformatted hybridization data*) geschrieben, die später zur Normalisierung verwendet werden.

Duplikate können ferner dazu verwendet werden, das Hybridisierungsexperiment numerisch zu bewerten. Betrachtet man alle 27.648 Duplikatsignale auf dem Filter, so läßt sich der Korrelationskoeffizient berechnen (Herwig, R. *et al.*, 1999). Der Korrelationskoeffizient ist ein Wert der Übereinstimmung der Signalintensitäten aller Duplikate x und y einer Hybridisierung und dient dem Ausschluß falsch positiver Signale. Bei optimaler Übereinstimmung beider Duplikatsignale liegt der Wert bei 1.00, die Duplikatsignale sind perfekt korreliert, was praktisch jedoch kaum erreichbar ist. Nimmt der Korrelationskoeffizient kleine Werte an, so besteht eine schlechte Korrelation der Duplikatsignale. Liegt der Korrelationskoeffizient unter 0.6 (empirisch ermittelter Wert), so wird das gesamte Experiment verworfen. Eine ausreichende Übereinstimmung beider Duplikatintensitäten ist noch bei einem Wert von etwa 0.8 gewährleistet, Werte über 0.9 stellen eine für die Praxis sehr gute Übereinstimmung dar. Ein Beispiel für einen Korrelationsplot zeigt Abb. 4-11.

Die hybridisierten Filter wurden auf PhosphorImager-Screens exponiert und anschließend am (PhosphorImager 445SI Version 4.0, Molecular Dynamics) gescannt (Johnston, R. F. *et al.*, 1990). Die so erhaltenen Bilddaten wurden als TIFF-Dateien (*tagged image file format*) gespeichert und können mit verschiedenen Visualisierungsprogrammen betrachtet und mit Bildanalyseprogrammen (Image Quant Tools Version 2.00 build 030, Molecular Dynamics; Visual Grid[®] 2.01 oder XDigitise) ausgewertet werden. Hierbei wird jedem Spot ein numerischer Wert zugeordnet, der proportional zur Menge an gebundenem radioaktiv markiertem Material ist. Dieser Intensitätswert gibt also wieder, ob das Oligo in der cDNA-Sequenz vorhanden ist oder nicht. In der Abb. 4-10 ist ein Ausschnitt eines Filters nach einer Oligonucleotid-Hybridisierung gezeigt.

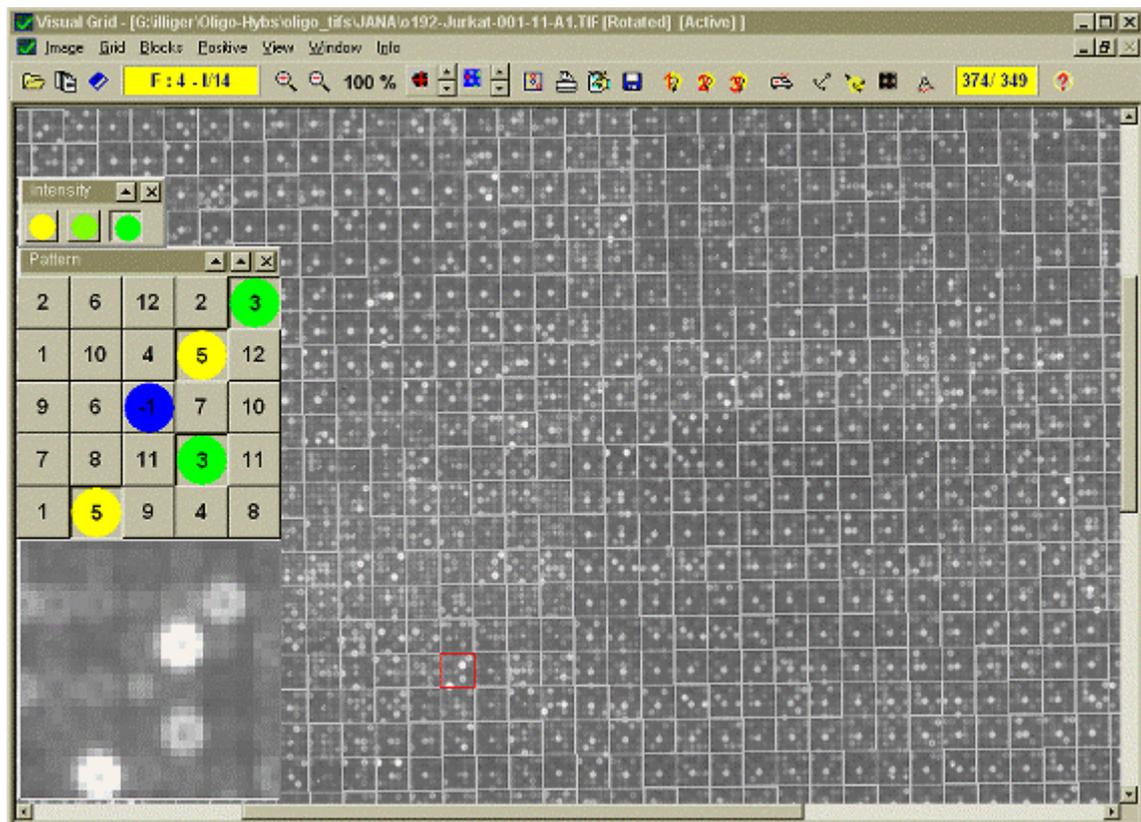


Abb. 4-10 Oligonucleotid-Hybridisierung

Die Abbildung zeigt einen Ausschnitt eines Filters nach einer Oligonucleotid-Hybridisierung. Ein ausgewerteter Block ist vergrößert dargestellt und im Pattern visualisiert.

Die Übereinstimmung der Signalintensitäten läßt sich graphisch in einem Diagramm darstellen, in dem die jeweils zusammengehörenden Signalintensitäten der Duplikate gegeneinander abgetragen sind. Bei optimaler Korrelation liegt der Wert, wie schon erwähnt, bei 1. Berechnet man nun den Wert für alle Duplikate eines Filters in einer Hybridisierung, so ergibt sich in der Praxis im günstigen Fall eine geringe Streuung um die Winkelhalbierende. Dieser Graph kann als Maß für die Güte der Hybridisierung angesehen werden. Ein solcher Fall ist in der Abb. 4-11 dargestellt.

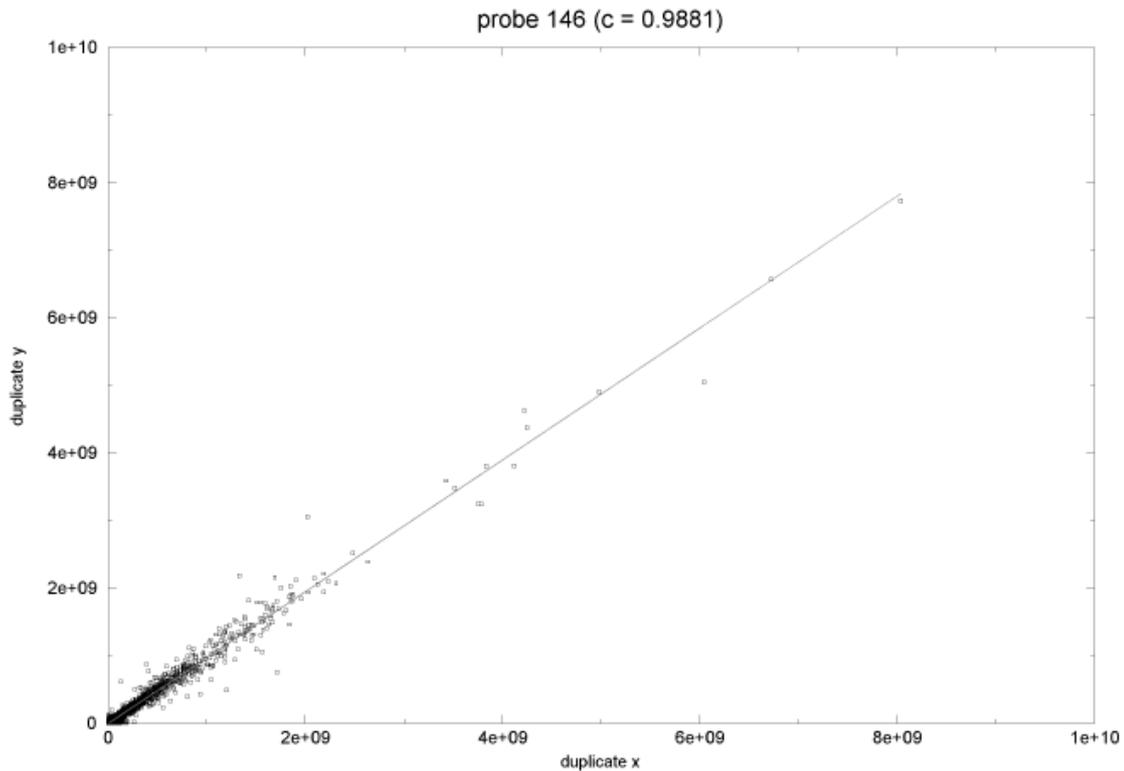


Abb. 4-11 Korrelation der x- und y-Duplikate

Darstellung der Übereinstimmung der Korrelation der Signalintensitäten der x- und y-Duplikate in der Hybridisierung des Oligos 146 mit der Sequenz NGAGGAGGAN. Ein Quadrat stellt hier das x/y-Verhältnis dar. In dieser Hybridisierung wurde eine Korrelation von 0.9881 erreicht.

4.2.3 Longprobe-Hybridisierung

Die Longprobe-Hybridisierung ermöglicht es, alle Klone einer cDNA-Bibliothek zu identifizieren, die ein bestimmtes Gen repräsentieren oder eine einem Gen sehr ähnliche Sequenz aufweisen. Alle positiven (Signal gebenden) Klone, die sich so mit einer Sonde identifizieren lassen, gehören einem Gen oder einer Gengruppe an und bilden ein Cluster. Diese Hybridisierungen dienen der Überprüfung und Validierung der Clustering-Ergebnisse, sowie der Anpassung der Parameter, die für die Clustering-Prozedur verwendet werden. Die Ergebnisse der Hybridisierungen werden verglichen mit den Resultaten der Clustering-Analyse, d.h., die Klone, die durch die Hybridisierung mit einer bestimmten Sonde identifiziert wurden, sollten sich nach dem Clustern auch in dem selben Cluster wiederfinden lassen. Zwei Möglichkeiten dieser Art Überprüfung wurden angewandt. Die eigentliche Longprobe-Hybridisierung wird durchgeführt mit Sonden bekannter DNA-Sequenz. Die zweite Möglichkeit, die Rückhybridisierung (oder auch Backhybridisierung), beruht auf der Hybridisierung nicht näher definierter, zufällig aus der

Bibliothek ausgewählter Klone. Die Hybridisierungen können entweder nicht-radioaktiv auf Kolonie-Filtern oder mit radioaktiv markierten Sonden auf PCR-Filtern (cDNA-Filter) anschließend an die Oligonucleotid-Hybridisierungen durchgeführt werden.

Es gibt einige Parameter, die das Ergebnis des Clusterings -die Clustergröße- beeinflussen (Herwig, R. *et al.*, 1999). Einerseits muß man bei der Wahl der Parameter darauf achten, daß die Cluster nicht zu groß werden dürfen, d.h. zu viele nur wenig ähnliche Klone enthalten. Andererseits dürfen die Parameter nicht zu stringent gewählt werden, so daß sehr viele kleinere Cluster entstehen. Die Klone wären dann zwar hoch ähnlich, aber eigentlich gleiche Klone könnten aufgrund nur geringfügiger Abweichungen in ein anderes Cluster geordnet werden. Um die Parameter richtig wählen zu können und eine Möglichkeit zur Überprüfung der Clustering-Ergebnisse zu erhalten, wurden in dieser Arbeit 25 Sonden bekannter Sequenz (von 500 bp bis etwa 2 kb Länge) hybridisiert. Einerseits wurden Proben hybridisiert, die der Gruppe der sogenannten *house keeping genes* (Haushaltsgene) angehören, dazu zählen beispielsweise die Elongationsfaktoren, ribosomale Proteine, GAPDH, Calmodulin oder Tubulin. Die *house keeping* Gene weisen eine relativ stabile Expression auf und dienen so als eine Art Marker für alle Zellen eines Organismus, unabhängig vom Zelltyp oder der Art des Gewebes. Andererseits wurden auch T- oder NK-Zell spezifische Gene zur Hybridisierung verwendet, solche Gene wie NKG5, L-Plastin oder T200 *leucocyte common antigen* CD45. Die Auswertung der positiven Signale erfolgte mit dem Programm Visual Grid[®], Version 2.01. Sind Sequenzen homolog zu dem verwendeten Kontrollgen, müssen sich die Ergebnisse dieser Hybridisierungen (Signale) nach dem Clustering wiederfinden lassen, d.h., sie werden Signale geben und später auch in ein selbes Cluster geordnet werden. Ein Beispiel für eine radioaktive Longprobe-Hybridisierung ist in Abb. 4-12 gezeigt.

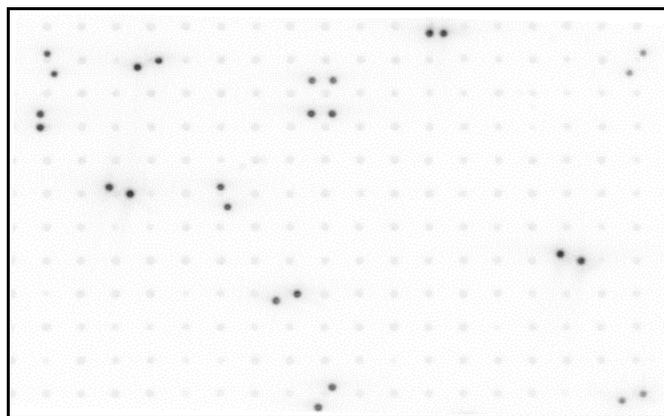


Abb. 4-12 Longprobe-Hybridisierung – Ausschnitt eines Filters

Ausschnitt eines cDNA-Filters hybridisiert mit einer radioaktiv markierten Longprobe. Die grauen Spots sind die Guide Dots, deutlich zu sehen sind einige positive Signale in Duplikaten.

Anhand der erhaltenen Ergebnisse wurde ein erstes Clustering durchgeführt und mit den Ergebnissen der Longprobe-Hybridisierungen verglichen. Die Klone, die bei der Hybridisierung mit einer Longprobe ein Signal geben, repräsentieren dieses Gen und müssen dann auch gemeinsam in einem Cluster wiedergefunden werden. Jedoch konnte in diesem ersten Versuch nur eine sehr schlechte Übereinstimmung gefunden werden, d.h., die Klone, die aufgrund der positiven Signale bei den Hybridisierungen identifiziert werden konnte, konnten nur ungenügend in einem Cluster wiedergefunden werden. Selbst mit veränderten Stringenz-Parametern konnte keine zufriedenstellende Cluster-Zuordnung erreicht werden. Zu wenige Klone, nur etwa die Hälfte, die zu jeweils einer Probe homolog waren, d.h., in der Longprobe-Hybridisierung Signale gaben, konnten auch in dem jeweiligen Cluster wiedergefunden werden. Es wurde vermutet, daß in der DIG-Hybridisierung zu wenige Klone detektiert wurden. Die Experimente wurden daraufhin wiederholt, diesmal wurden jedoch die Sonden radioaktiv markiert und auf die PCR-Filter hybridisiert. Anhand dieser Daten wurde ein erneuter Clustering-Versuch unternommen. Nun konnten die meisten der in der Hybridisierung detektierten Klone in einem Cluster geordnet wiedergefunden werden. Nur wenige Klone splitteten in andere Cluster oder verblieben als Singletons. Als ein Beispiel zur Verdeutlichung der Auswertung der Longprobe-Hybridisierungen ist die Hybridisierung der Jurkat-Bibliothek mit der DNA des ribosomalen Proteins 60S L4 dargestellt.

60S ribosomal protein L4					
1	clones in cluster	14	size	81	1
76	clones in cluster	16	size	78	76
28	clones in cluster	75	size	28	28
1	clones in cluster	181	size	13	1
2	clones in cluster	697	size	4	2
1	clones in cluster	1139	size	2	1
1	clones in cluster	1219	size	2	1
1	clones in cluster	1355	size	2	1
4	clones in cluster	2231	size	8276	4
Total					115
115/116 clones are split into 9 clusters					

Daraus ist folgendes ersichtlich: 116 Klone wurden mit der DNA-Sonde für das ribosomale Protein 60S L4 in der Longprobe-Hybridisierung detektiert. 115 Klone davon konnten im Clustering wiedergefunden werden. 76 der 115 Klone konnten im Cluster 16 zusammengefaßt werden, dessen Gesamtgröße 78 Klone beträgt. Die zwei zur Gesamtzahl 76 zusätzlichen Klone sind entweder falsch positive Klone oder solche, die in der Longprobe-Hybridisierung nicht detektiert wurden. Weitere 28 Klone fallen in das Cluster 75, das mit einer Größe von 28 Klone ein reines Cluster darstellt. Einzelne Klone lassen sich in anderen Clustern wiederfinden, 4 Klone verbleiben als Singletons, von denen insgesamt 8.276 in der Jurkat-Bibliothek vorkommen.

Die DIG-Hybridisierungen auf den Kolonie-Filtern stellten sich als nicht sensitiv genug heraus, alle Klone zu detektieren, die ein Gen repräsentieren, und gaben somit keine verlässlichen Ergebnisse. Die radioaktiven Longprobe-Hybridisierungen auf den PCR-Filtern, die anschließend an alle Oligonucleotid-Hybridisierungen durchgeführt wurden, stellten sich als signifikant sensitiver und damit verlässlicher heraus. Obgleich die Longprobe-Hybridisierung auf Kolonie-Filtern nicht nur als Kontrolle des Clusterings dient, sondern gleichzeitig auch zur Validierung der Übereinstimmung zwischen den Kulturplatten und den entsprechenden PCR-Platten genutzt wird, konnten die Daten der nicht radioaktiven DIG-Hybridisierungen auf Kolonie-Filtern in dieser Arbeit keine Anwendung finden. In der Tab. 4-2 sind die Häufigkeiten der Signale der radioaktiven und der DIG-Hybridisierungen der Longprobe-Hybridisierungen im Vergleich dargestellt. Es fällt auf, daß z.T. um ein Vielfaches mehr Klone in den radioaktiven Hybridisierungen gefunden wurden.

Longprobe	Σ radioaktiv : Σ DIG	
	Ju	NKL
Hemoglobin G.-A.+G.+G.	16,0	1,6
Hemoglobin beta chain	29,0	1,0
Elongation factor 1 alpha	17,1	5,3
Elongation factor 1 beta	4,1	1,7
Elongation factor 1 gamma	7,1	2,9
40S ribosomal protein S4 isoform	1,8	0,3
40S ribosomal protein S6	2,0	2,0
40S ribosomal protein S3	1,4	4,4
60S ribosomal protein L4	23,2	9,0
60S ribosomal protein L6	3,3	9,0
Glycerinaldehyd-3-phosphate DH	2,4	2,0
Cytochrome C oxi. polypeptide III	10,8	5,8
Heat shock cognate 71 KD protein	14,9	6,3
Thymosin beta-10	n.d.	14,5
Tubulin beta-1 chain	3,1	2,2
Calmodulin	1,5	1,6
Anion channel	15,0	n.d.
TNF-Receptor	25,0	9,4
beta-Actin	14,0	3,3
HLA-E	58,4	21,1
NFkB DNA binding subunit	15,0	1,8
NFkB transcription factor p65 subunit	1,5	n.d.
human NKG5	1,0	1,3
MAP kinase activared protein kinase	111,0	n.d.
mRNA for T cell leukemia	104,5	17,6
lymphocyte cytosolic protein 1 (L-plastin)	5,5	n.d.
T200 leucocyte common antigen CD45	n.d.	0,7

Tab. 4-2 Signalthäufigkeit der Longprobe-Hybridisierungen

Angegeben ist das Verhältnis der Signalthäufigkeit der Longprobe-Hybridisierungen radioaktiv markierter Sonden im Vergleich zu Digoxigenin markierten Sonden. Die wesentlich höhere Sensitivität der radioaktiven Detektion wird deutlich sichtbar. (n.d.: nicht bestimmt)

Die Häufigkeit der Signale, die sich nach der Hybridisierung verschiedener als Longprobe verwendeter Gene ergeben, kann Aufschluß über die relative Häufigkeit der Expression einer mRNA in einer Zellpopulation oder eine veränderte Genexpression gegenüber einem vergleichbaren Zelltyp geben. In der Tab. 4-3 sind einige der hybridisierten Proben im Vergleich zu humanen dendritischen Zellen des peripheren Blutes (Herwig, R. *et al.*, 1999) dargestellt. Dabei sind einige Häufigkeiten der T-Zell-Linie Jurkat und der NK-Zell-Linie NKL durchaus vergleichbar mit den Werten, die für die dendritischen Zellen ermittelt wurden. Einige jedoch weisen signifikante Unterschiede hinsichtlich ihrer Häufigkeit in der jeweiligen Zellpopulation auf. So beispielsweise ist dies beim Elongationsfaktor 1α der Fall, der in der Zell-Linie Jurkat etwa 5.4 mal häufiger und in NKL 2.75 mal häufiger vorkommt, was auf eine Erhöhung der Translationsaktivität hinweisen kann. Auch das Hitzeschock-Protein *heat shock cognate 71KD protein* ist in den T-Zell- bzw. NK-Zell-Bibliotheken mit einer relativen Häufigkeit von 1.104 bzw. 0.671 um ein Vielfaches stärker exprimiert als in den dendritischen Zellen (0.037), ebenso ist dies beim TNF-Rezeptor der Fall.

mRNA	relative Häufigkeit in der mRNA-Population (%)		rel. Häufigkeit in % (Herwig, R. <i>et al.</i> , 1999)
	Jurkat	NKL	humane dendritische Zellen
Elongationsfaktor 1α	4.757	2.406	0.876
Elongationsfaktor 1β	0.334	0.077	0.088
40S rRNA S6	0.444	0.185	0.241
60S rRNA L4	0.504	0.371	0.113
GAPDH	0.739	0.506	0.108
Calmodulin	0.052	0.062	0.042
Heat shock cognate 71KD Protein	1.104	0.671	0.037
Tubulin β	0.404	0.215	0.272
TNF-Rezeptor	0.109	0.221	0.016

Tab. 4-3 Relative Häufigkeit der Expression einiger als Longprobe hybridisierter Gene

Dargestellt sind einzelne Gene in ihrer relativen Häufigkeit der mRNA in den Zell-Linien Jurkat und NKL im Vergleich zu humanen dendritischen Zellen (Herwig, R. *et al.*, 1999).

4.2.4 Normalisieren der Daten

Viele verschiedene Faktoren und Bedingungen im experimentellen Ablauf des Oligonucleotid Fingerprintings haben Einfluß auf die Hybridisierung der Oligos an die immobilisierte DNA und somit auch auf die Signalintensitäten (Rohdaten).

Zum einen wird die Bibliothek auf verschiedene Membranen gespottet, um mehrere Kopien für die über 200 Oligonucleotid-Hybridisierungen zur Verfügung zu stellen.

Eine Bibliothek beinhaltet häufig eine größere Anzahl von Klonen, als auf einen Filter gespottet werden kann (27.648 Klone in Duplikaten je Filter), so daß diese Bibliothek in zwei Sets aufgeteilt gespottet werden muß. Die Qualität der einzelnen Filter aber kann variieren, so daß selbst die Wiederholung der Hybridisierung ein und desselben Oligos abweichende Signalintensitäten geben kann.

Weiterhin können diese Veränderungen der Intensitäten aus den Mengenunterschieden an immobilisierter DNA auf den Membranen resultieren, die entweder auf unterschiedliche Mengen an amplifiziertem PCR-Produkt (oder auf einen Ausfall bestimmter Klone) zurückzuführen sind oder aber auf eine unterschiedliche Transfereffizienz der einzelnen Pins. Sehr starke Signale können zu einem Überstrahlen benachbarter Positionen führen.

Auch Unterschiede in der Aktivität der gelabelten Proben, lokale Veränderungen (Verschmutzungen, Klone an Randpositionen, lokaler Hintergrund) auf der Membran, die während der Hybridisierung auftreten, oder Variationen der Bedingungen der Hybridisierungen und der Exposition der Filter auf den Screens können wesentliche Faktoren mit Einfluß auf die Hybridisierungsergebnisse darstellen (Schuchhardt, J. *et al.*, 2000).

Eine weitere Fehlerquelle, die Störungen in der Datenanalyse bewirken kann, liegt darin begründet, daß gleiche Versuchsbedingungen bei der Hybridisierung verschiedener Oligos eingesetzt werden, obwohl jedes ein unterschiedliches Temperaturoptimum aufweist.

Das Spotten der Klone in Duplikaten, was einer Wiederholung des Hybridisierungsexperiments entspricht, dient als erste Möglichkeit der Überprüfung der Signale. Um aber all die möglichen Fehler auszugleichen, die die Signalintensität beeinflussen können, müssen die erhaltenen Rohdaten der Signalintensitäten normalisiert werden. Die Normalisierung erfolgt in zwei wesentlichen Schritten:

- a) der Normalisierung innerhalb eines Filters für alle Klone und
- b) der Normalisierung über alle Oligos für jeden einzelnen Klon.

Eine relativ einfache und robuste Methode der Normalisierung ist das sogenannte *double ranking*, bei dem jeder Signalintensität eines Klones ihr Rang unter allen Signalintensitäten zugeordnet wird, erstens unter allen Klonen eines Filters in einer Hybridisierung und zweitens unter allen Oligonucleotid-Hybridisierungen für jeden Klon (Milosavljevic, A. *et al.*, 1995). Das stärkste Signal erhält den Wert=1, das schwächste den Wert=0. So wird dem zweitstärksten Signal der Wert $1-1/N$, dem dritten $1-2/N$ usw., bzw. für alle Hybridisierungsexperimente wiederum dem stärksten Signal der Wert=1, dem zweiten $1-1/p$, dem dritten $1-2/p$ usw. zugeordnet. Dabei stellt N die Anzahl der Klone auf der Membran und p die Anzahl der Oligonucleotid-Hybridisierungen dar. Ein Nachteil dieser Methode ist der recht hohe Verlust an Information, da jeder Klon jetzt keinen exakten Wert mit den relativen Differenzen der einzelnen Intensitäten mehr aufweist, sondern ihm lediglich ein Rang innerhalb aller Intensitäten zugeordnet wurde (Herwig, R. *et al.*, 1999).

Aus vorangegangenen Experimenten ist bekannt, daß aufgrund experimenteller Ursachen, beispielsweise dem Ausfall der PCR-Amplifikation oder anderen -oben erwähnten- Faktoren, etwa 15 - 25% der originalen cDNA-Klone mit den schwächsten Signalintensitäten von der weiteren Analyse ausgeschlossen werden sollten, da diese sonst infolge des Überhäufens mit zu geringen oder fehlenden Signalintensitäten das Ergebnis des Clusterings verfälschen würden. Dies kann anhand einer Background-Hybridisierung evaluiert werden, die als Sonde einen der in der PCR verwendeten Primer nutzt, um alle auf einem Filter gespotteten Klone sichtbar zu machen und so auf die jeweilige DNA-Menge schließen zu können. Im Wesentlichen besteht die Normalisierungs-Prozedur aus den Punkten:

- Berechnen des Verhältnisses der Duplikatintensitäten (max : min) aus den Rohdaten,
- Ersetzen der Klon-Signale für jeden Filter durch Ränge,
- Ermitteln des durchschnittlichen Ranges eines jeden Klons für alle Filter, Sortieren, Verwerfen der 15 - 25% schwächsten Werte der Intensitäten,
- erneutes Ranking der Ränge unter allen Oligonucleotid-Hybridisierungen.

4.3 Analyse und Ergebnisse

4.3.1 Clustering der Fingerprints

Das Muster der Hybridisierung der Oligos an die DNA ist spezifisch für die jeweilige cDNA-Sequenz und ergibt einen einzigartigen Fingerprint. Hybridisiert man eine cDNA-Sequenz mit einer Reihe von Proben (Oligos), so entsteht ein spezifisches Muster aus numerischen Werten, der Fingerprint. Ähnliche, das selbe Gen repräsentierende cDNA-Sequenzen werden somit ähnliche Fingerprints aufweisen.

Aufgabe der mathematischen Clustering-Analyse ist es, gleiche oder ähnliche Fingerprints zu Gruppen, ein Gen repräsentierender Sequenzen, zusammenzufassen (Cluster) und unähnliche voneinander zu trennen. Dazu werden die Klon-Ähnlichkeiten anhand der Fingerprints nach einem modifizierten sequentiellen Algorithmus (*sequential k-means*) paarweise miteinander verglichen. Die Berechnung der paarweisen Klon-Ähnlichkeit erfolgt dabei durch ein informationstheoretisches Maß, der Transinformation (Herwig, R. *et al.*, 1999). Diese bewertet, wieviel Information beide Fingerprints voneinander beinhalten. Ein Fingerprint wird dabei einer Gruppe von Fingerprints zugeordnet, wenn er genügend Information über den die Gruppe repräsentierenden Zentroid (Mittelwertvektor) beinhaltet, was anhand von Schwellenwerten entschieden wird. Bekommt eine Gruppe von Fingerprints einen weiteren Fingerprint zugeordnet, so wird der Zentroid als Clusterrepräsentant neu berechnet. Am Ende des Algorithmus wird jeder Fingerprint dem ihm am ähnlichsten Zentroid zugeordnet. Idealerweise liefert dann das Clusterergebnis die Anzahl der verschiedenen Gene in der Bibliothek (Anzahl der verschiedenen Cluster) und die entsprechende Expressionshäufigkeit der Gene (Größe der Cluster).

Innerhalb eines Clusters werden die Klone in Abhängigkeit ihrer Ähnlichkeit zum Consensus-Fingerprints aller Mitglieder geordnet (Meier-Ewert, S. *et al.*, 1998). Dies erleichtert das Identifizieren repräsentativer Klone der entsprechenden Cluster z.B. für ein späteres Neuordnen (*rearranging*) zum Erstellen eines Unigene-Sets. So ist der sogenannte 1.Wahl-Klon derjenige Klon, der die größte Übereinstimmung mit dem Consensus-Fingerprint aufweist.

Die Klone eines Clusters müssen nicht zwingendermaßen das gleiche Gen repräsentieren. Da nicht nur identische, sondern auch ähnliche Sequenzen in ein Cluster geordnet werden, ist es prinzipiell möglich, daß es sich dabei um Mitglieder einer Genfamilie oder auch verschiedene Spleißvarianten eines Transkripts handeln kann. Stringentere Parameter oder der Einsatz weiterer, spezifischerer Oligos könnten beim Clustering diese Varianten aber wieder voneinander trennen und den verschiedenen Clustern zuordnen. Generell ist es jedoch nicht möglich, im Umfang einer solch großen Datenmenge das eine optimale Clustering-Ergebnis zu finden, da die Berechnungskomplexität zu groß wäre. Vielmehr müssen die Parameter so gewählt werden, daß eine gute Lösung in annehmbarer Zeit und Aufwand erzielt werden kann.

Die Clustering-Ergebnisse werden anhand der Longprobe-Hybridisierung evaluiert. Das heißt, alle mit einer Longprobe-Hybridisierung gefundenen positiven Klone repräsentieren ein gleiches Gen und sollten somit einen gleichen oder zumindest ähnlichen Fingerprint aufweisen und auch in einem Cluster wiedergefunden werden. So werden Clustering-Tests mit verschiedenen Grenzwerten und Parametern (beispielsweise der Mindestähnlichkeit der Sequenzen in einem Cluster) durchgeführt, um die optimalen Parameter herauszufinden, damit ähnliche Sequenzen in ein Cluster geordnet werden oder falsch dazugeordnete Sequenzen wieder aus dem Cluster entfernt werden können. Werden zu strenge Parameter gewählt, werden zu wenige Klone in ein und dasselbe Cluster geordnet und es kommt zu einer Häufung des Cluster-Splittings, auf das später noch genauer eingegangen wird (s.Kapitel 4.3.2.1 Cluster-Splitting). Wählt man andererseits weniger stringente Parameter, so werden zu viele, weniger ähnliche Klone einem Cluster zugeordnet.

Aufgrund des Zusammenfassens gleicher Klone, die für ein und dasselbe Gen kodierende Sequenzen enthalten, in ein Cluster wird eine Bibliothek um ein Vielfaches normalisiert. Die Repräsentanten dieser Cluster bilden dann ein gering oder nicht redundantes Set an Genen, die wiederum auf Filter gespottet und analysiert werden können, um neue Gene zu finden. Durch das Identifizieren der Klone anhand eines Datenbankabgleichs der Fingerprints und das starke Verringern der Redundanz wird es nun möglich, nur selten exprimierte oder noch unbekannte Gene ausfindig zu machen und mit geringerem Aufwand zu sequenzieren.

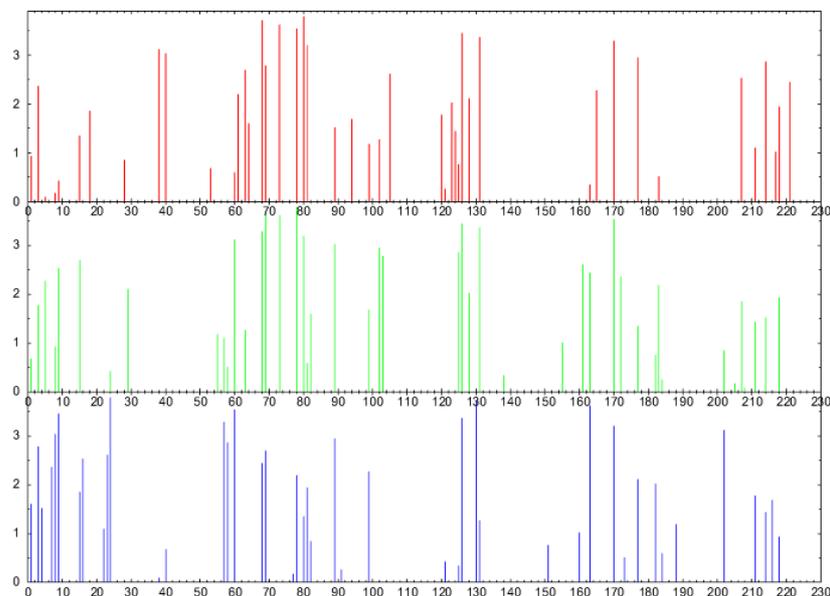


Abb. 4-13 Fingerprints der Klone eines Co-Clusters

Darstellung des Fingerprints der Klone des Co-Clusters 481, das für das CD3E-Gen (epsilon Polypeptid (TiT3-Komplex)) codiert. Auf der Abszisse sind die hybridisierten Oligos abgetragen, die Peaks auf der Ordinate entsprechen den Signalintensitäten.

Basierend auf der bekannten Sequenz der hybridisierten Oligonucleotide ergibt sich ein experimenteller Fingerprint der Klone. Ein solcher Fingerprint ist in der Abb. 4-13 dargestellt. Das Clustering erfolgte automatisch mit Hilfe eines Verfahrens, das anhand der Ähnlichkeiten der experimentellen Fingerprints zueinander mittels eines sequentiellen *k-means* Algorithmus die Daten klassifiziert (Herwig, R., 2000). Die Clustergröße gibt dabei die Anzahl der einem Cluster angehörenden Klone an, was in etwa der Menge an in der Zellpopulation exprimierten Transkripten entspricht. Als erste Überprüfung der Clustering-Ergebnisse dienten, wie bereits beschrieben, die Longprobe-Hybridisierungen. Die Ergebnisse müssen jedoch auch anhand von anderen, direkten Methoden nachvollziehbar sein und kontrolliert werden. In dieser Arbeit wurden zur Validierung dieser Ergebnisse -speziell der Co-Clustering-Analysen- Northern-Hybridisierungen durchgeführt, worauf später noch genauer eingegangen werden soll.

Anhand der Clustering-Analysen der einzelnen Bibliotheken Jurkat und NKL ergaben sich die folgenden Ergebnisse, die auch in der Tab. 4-4 detailliert dargestellt sind. Die Klone der NKL-Bibliothek konnten in 2.168 Cluster mit einer Mindestgröße von 2 Klonen und 10.926 Singletons, die Klone der Jurkat-Bibliothek in 2.230 Cluster und 8.276 Singletons geordnet werden. Die größten Cluster umfassen jeweils 305 Klone in der Bibliothek Jurkat bzw. 1.181 Klone in der Bibliothek NKL.

	Singletons	Clustergröße					total	>/=5
		2-4	5-10	11-100	101-1000	>1000		
NKL-Cluster	10926	1671	279	201	16	1	2168	497
Jurkat-Cluster	8276	1690	308	223	9	0	2230	540

Tab. 4-4 Verteilung der Clustergrößen der Bibliotheken Jurkat und NKL

4.3.2 Co-Clustering und Datenbank-Matching

Als erstes wurde jede einzelne der beiden Bibliotheken für sich geclustert, d.h., Klone, die einen ähnlichen Fingerprint aufweisen wurden in eine Gruppe, die ein bestimmtes Gen repräsentiert, zusammengefaßt. Aufgrund der Anzahl der in einem Cluster enthaltenen Klone können relative Angaben über die Häufigkeit eines Transkripts in der jeweiligen Zellpopulation gemacht werden. Um nun die Expressionsmuster beider Bibliotheken miteinander vergleichen zu können und Gene zu identifizieren, die differentiell exprimiert und somit spezifisch für ein Gewebe oder eine

Erkrankung sind, wurde im Anschluß an das Clustering ein sogenanntes Co-Clustering durchgeführt.

Das Prinzip der Co-Clustering-Prozedur ist im Wesentlichen das gleiche wie das des einfachen Clusterings. Der Unterschied besteht darin, daß als Cluster-Mittelpunkt diesmal der experimentelle Consensus-Fingerprint des kalkulierten Clusters der Bibliothek 1 initialisiert wird und der sequentielle Algorithmus mit den Consensus-Fingerprints der Bibliothek 2 begonnen wird (Herwig, R., 2000). Das Co-Clustering erfolgt also anhand der Ähnlichkeiten der experimentellen Consensus-Fingerprints der jeweiligen Cluster zueinander. Dabei werden Cluster mit einem stark ähnlichem oder identischem Consensus-Fingerprint, die also jeweils in der anderen Bibliothek ein Homolog fanden, in ein Co-Cluster zusammengefaßt. Aus den Ergebnissen geht die Gesamtzahl an Klonen, die dem Cluster angehören, hervor sowie auch die Anzahl der Klone, die jeweils aus den unterschiedlichen Bibliotheken stammen. So kann man die Expressionsmuster beider Bibliotheken direkt miteinander vergleichen und Hinweise über das Expressionsverhältnis interessierender Gene erhalten und darüber, welche Gene in der einen oder in der anderen Zellpopulation differentiell transkribiert werden. Diese Strategie erweist sich als besser als das einfache Clustering aller Sequenzen beider Bibliotheken auf einmal, besonders dann, wenn beide Bibliotheken sehr unterschiedliche Fragmentlängen der cDNA-Fragmente aufweisen. Aufgrund dieser differierenden Fragmentlängen ergeben sich zum Teil stark unterschiedliche Fingerprints der gleichen Fragmente in beiden Bibliotheken, so daß dabei häufiger das Problem auftreten würde, daß nur wenige Cluster errechnet werden, die Klone beider Bibliotheken beinhalten und eine größere Anzahl an Clustern, die nur Klone einer Bibliothek aufweisen.

Um die Repräsentation der beiden Bibliotheken in jedem kalkulierten Cluster zu testen, wird ein zweiseitiger Binomialtest durchgeführt. Um nun auszuwerten, ob ein Cluster differentiell exprimiert ist, wird der exakte p-Value-Wert des Binomialtests für jedes Cluster mit einer Mindestgröße von 5 Mitgliedern errechnet. Zur Identifizierung der Co-Cluster wurden die experimentellen Fingerprints mit theoretischen Fingerprints, die anhand von *in silico*-Hybridisierungen der Oligos von Genen / existierenden Sequenzen aus den Datenbanken erstellt wurden, verglichen. Dabei erhält man ein Ergebnis, wie es am Beispiel des Co-Clusters 5 und 10 einmal dargestellt ist.

cluster 5 size=1114 NKL=541 Jurkat=573 p-value=1.00e-02

- 1.) p-value=1.822e-08 probe hits=30 /ug=Hs.110444 /gb=AA844372
/gene= Translation elongation factor 1-alpha-1,
- 2.) p-value=2.221e-08 probe hits=29 /ug=Hs.110444 /gb=AA062651
/gene= Translation elongation factor 1-alpha-1,
- 3.) p-value=2.894e-08 probe hits=29 /ug=Hs.110444 /gb=W04347
/gene= Translation elongation factor 1-alpha-1,
- 4.) p-value=3.558e-08 probe hits=31 /ug=Hs.110444 /gb=AA338473
/gene= Translation elongation factor 1-alpha-1,
- 5.) p-value=3.558e-08 probe hits=31 /ug=Hs.110444 /gb=AA324189
/gene= Translation elongation factor 1-alpha-1

cluster 10 size=228 NKL=124 Jurkat=104 no significance

- 1.) p-value=8.662e-15 probe hits=26 /ug=Hs.78978 /gb=AA146788
/gene= Human MHC protein homologous to chicken B complex protein mRNA,
- 2.) p-value=3.348e-14 probe hits=27 /ug=Hs.78978 /gb=N39998
/gene= Human MHC protein homologous to chicken B complex protein mRNA,
- 3.) p-value=1.178e-13 probe hits=28 /ug=Hs.78978 /gb=AA316423
/gene= Human MHC protein homologous to chicken B complex protein mRNA,
- 4.) p-value=1.090e-11 probe hits=37 /ug=Hs.78978 /gb=AA477677
/gene= Human MHC protein homologous to chicken B complex protein mRNA,
- 5.) p-value=1.012e-10 probe hits=24 /ug=Hs.78978 /gb=AA677074
/gene= Human MHC protein homologous to chicken B complex protein mRNA

Das Cluster 5 enthält 1.114 Klone. Davon stammen 541 aus der NKL-Bibliothek und 573 aus der Jurkat-Bibliothek. Dieses Co-Cluster wurde als Translations-Elongationsfaktor 1-alpha-1 identifiziert und wird mit einer Signifikanz (p-value) von 1.00e-02 differentiell exprimiert. Angegeben sind die fünf wahrscheinlichsten Identitäten, wobei in diesem Fall das Co-Cluster eindeutig identifiziert werden konnte. Das Cluster 10 enthält 228 Klone, von denen 124 aus der NKL-Bibliothek und 104 aus der Jurkat-Bibliothek stammen. Auch dieses Co-Cluster konnte wieder eindeutig identifiziert werden, jedoch sind die Klone beider Bibliotheken (bezogen auf die Gesamtgröße der jeweiligen Bibliotheken) etwa gleich häufig im Co-Cluster repräsentiert und weisen somit keine Signifikanz hinsichtlich einer differentiellen Expression auf.

Es ist jedoch zu beachten, daß die Identifizierung der Klone (Cluster) nicht anhand der Sequenz der Klone erfolgt, sondern anhand der Consensus-Fingerprints. Diese wiederum ergeben sich aus der Analyse der Hybridisierungssignale der Oligonucleotid-Hybridisierungen, die z.T. relativ anfällig äußeren Einflüssen gegenüber sind (s. Kapitel 4.2.4), die die Interaktion zwischen der Probe (Oligo) und dem Target (cDNA-Sequenz) verfälschen und so eine gewisse Ungenauigkeit beinhalten können. Daraus kann sich eine mögliche Fehlerquelle bei der Identifizierung der Klone

ergeben. Um die Zuordnung zu den bekannten Sequenzen direkt bestätigen zu können, wurden einige geeignete Klone verschiedener Co-Cluster, die einen signifikanten Match in der Datenbank aufwiesen, selektiert und sequenziert. Anschließend wurden die Sequenzen mittels BLAST-Programmen (*basic local alignment tool*) (Altschul, S. F. *et al.*, 1990) in der Datenbank mit bereits bekannten Sequenzen abgeglichen. Dazu wurden insgesamt 72 Klone der Jurkat-Bibliothek sowie 172 Klone der NKL-Bibliothek mit Hilfe eines Rearranging-Roboters aus den Platten der jeweiligen Bibliotheken ausgewählt und in neue Microtiterplatten inokuliert. Nach Inkubation der Zellen ÜN bei 37°C wurden die Inserts mittels PCR amplifiziert und aufgereinigt. Anschließend wurden diese Klone vom 5'-Ende her ansequenziert, mittels BLAST-Analysen in der GenBank/EMBL Datenbank identifiziert und mit den aus der Co-Clustering-Analyse angegebenen Identitäten verglichen. Die mittels Datenbankabgleich der experimentellen Fingerprints der Klone mit den theoretischen Fingerprints aus der Datenbank vorausgesagten Identitäten konnten nach Abgleich der Sequenzen in den Datenbanken im Allgemeinen bestätigt werden.

Je häufiger ein Gen in der Zelle exprimiert wird, desto häufiger wird auch die das Gen translatierende mRNA (auch spielt die Stabilität der entsprechenden mRNA eine gewisse Rolle) in der gesamten mRNA-Population vorkommen. So sind die am häufigsten vorkommenden mRNAs (Gene) in den größten Clustern bzw. Co-Clustern vertreten. Die Cluster wurden ihrer Größe nach geordnet. So enthält Cluster 1 die meisten Klone. Je mehr Klone ein Cluster enthält, desto mehr Transkripte liegen dann in der Bibliothek -und in der Tendenz auch in der Zelle- vor. So kann die Clustering-Analyse genutzt werden, um einen groben Überblick über die in der Zelle vorkommenden Arten und Mengen an exprimierten Genen -in Form der mRNA- zu erhalten. Es kann jedoch keine absolute Voraussage über die genaue Anzahl gegeben werden, die Clustergröße zeigt lediglich eine Tendenz an. Die Erkenntnisse des Co-Clusterings werden später noch in Northern-Hybridisierungen überprüft.

Einen Überblick über die ersten 100 Co-Cluster der Bibliotheken Jurkat und NKL gibt die Tab. 4-5. Darin dargestellt ist die Gesamtgröße des entsprechenden Co-Clusters sowie das Vorkommen des entsprechenden Gens in jeder der beiden Bibliotheken, das p-Value, die Genbank ID und die Bezeichnung des codierenden Gens.

Cluster	Größ	NKL	Jurkat	p-Value	Genbank ID	Datenbank-Match
1	1981	1981	0	1.00E-02	X91906	Dents Disease candidate gene (chloride channel 5 (nephrolithiasis 2, X-linked, Dent disease))
2	1343	4	1339	1.00E-02	U26742	dvtrobrevin (DTN)
3	1293	1293	0	1.00E-02	U71207	eyes absent homolog (Fab1) clone 24534
4	1184	1184	0	1.00E-02	U71207	eyes absent homolog (Fab1) clone 24534
5	1114	541	573	1.00E-02	AA324189	Translation elongation factor 1-alpha-1
6	914	0	914	1.00E-02	X62949	H+-ATPase Mr.56.000 subunit (HO57) _vacuolar
7	349	218	131	no	H66207	actin_cytosolic (from cultured epidermal cells grown from foreskin)
8	316	145	171	8.61E-05	AA662591	23 kD highly basic protein
9	281	129	152	2.30E-04	AA536218	L1.Ren3 (40S Ribosomal protein S2 (S4))
10	228	124	104	no	AA146788	MHC protein homologous to chicken B complex protein
11	222	206	16	3.44E-28	AA347254	KTAA0220 gene
12	203	96	107	6.95E-03	AA291731	ribosomal protein L4
13	195	52	143	1.19E-17	AA381294	ribosomal protein S3A
14	192	89	103	3.51E-03	AA214725	ribosomal protein, large, P0
15	183	182	1	5.46E-33	U40622	DNA repair protein XRCC4
16	178	178	0	2.65E-33	AA102697	MHC class I HL A-B51, complete cds, hanlotvne A2, B27
17	178	174	4	9.07E-30	AJ005371	DNA from chromosome 19, specific cosmid R28461, genomic sequence
18	172	171	1	1.72E-31	U26742	dvtrobrevin (DTN)
19	144	144	0	1.30E-27	AA102034	KTAA0226 gene
20	133	124	9	6.70E-18	AJ014823	PMS4 (yeast mismatch repair gene, PMS1 homologue)
21	133	133	0	3.38E-25	R59536	CTA-1 (membrane glycoprotein)
22	132	132	0	6.79E-25	AJ076923	Glyceroldehyde-3-phosphate dehydrogenase
23	130	130	0	3.27E-25	###	###
24	125	99	26	4.73E-07	X04098	Actin, gamma 1
25	124	122	2	1.27E-21	Y09980	HOXD3
26	120	50	70	8.64E-04	X67309	ribosomal protein S6 gene, complete cds and flanking regions
27	112	44	68	1.86E-04	T59399	ribosomal protein S4, X-linked
28	112	112	0	1.37E-21	###	###
29	111	0	111	2.80E-41	AJ076923	Glyceroldehyde-3-phosphate dehydrogenase
30	102	102	0	1.77E-19	AA832103	RNA helicase (Myc-regulated dead box protein)
31	101	40	61	5.75E-04	AA284925	BBC1 (60S Ribosomal protein L13)
32	92	48	44	no	I20688	GDP-dissociation inhibitor protein (Lv-GDI)
33	92	18	74	2.57E-13	W52893	prothymosin-alpha
34	91	83	8	1.25E-11	H08910	non-histone chromosomal protein HMG-17
35	87	87	0	1.08E-17	###	###
36	86	1	85	4.36E-30	T94056	Laminin receptor (2H5 epitone)
37	86	85	1	1.84E-16	R98265	alpha-tubulin
38	85	32	53	4.10E-04	###	novel gene
39	84	8	76	8.89E-20	C74996	pancreatic tumor-related protein
40	84	84	0	9.02E-17	###	###
41	84	83	1	7.34E-16	H44597	Lamin B2 (LAMB2) gene and ntvy1 gene
42	81	81	0	7.38E-16	AC004030	DNA from chromosome 19
43	80	27	53	2.92E-05	AA653025	ribosomal protein L29
44	79	79	0	3.54E-16	H92267	hepatitis C-associated microtubular aggregate protein n44
45	78	76	2	4.45E-14	AA633954	HLA-DR alpha-chain
46	76	18	58	4.15E-09	AA523353	ribosomal protein S5
47	76	0	76	1.71E-28	K01571	T-cell receptor active beta-chain, from cell line MOLT-3
48	75	14	61	1.30E-11	W28616	heat shock protein 90-kDa
49	73	0	73	2.14E-27	AA377555	non-histone chromosomal protein HMG-17
50	73	42	31	no	R35310	homeobox protein MEIS2 (MEIS2)
51	73	72	1	1.84E-13	AA045819	GT198
52	70	67	3	9.47E-12	R55882	protein phosphatase 2A beta subunit
53	69	69	0	4.77E-14	D44497	actin binding protein n57
54	68	64	4	2.06E-10	AA136279	KTAA0224 gene
55	67	0	67	3.33E-25	X12671	Heterogeneous nuclear ribonucleonprotein A1
56	67	1	66	2.98E-23	I76259	6-nrvuvyl-tetrahydropterin synthase
57	66	38	28	no	T82139	actin_cytosolic (fragment encoding, from cultured epidermal cells grown from foreskin)
58	66	50	16	2.58E-03	AA599380	chromosome 17a21 mRNA
59	64	19	45	1.26E-05	AA313198	Enolase 1 (alpha)
60	64	12	52	4.60E-10	X06705	ribosomal protein L7a
61	64	0	64	4.15E-24	AA664625	ribosomal protein L6
62	64	64	0	1.89E-13	AA100556	pancreatic tumor-related protein
63	64	0	64	4.15E-24	AA173039	alpha-tubulin
64	63	41	22	no	AA853315	Vimentin
65	63	29	34	no	AA327830	DAP-3 (ionizing radiation resistance-conferring protein: death-associated protein 3)
66	61	21	40	4.40E-04	AA853840	profilin
67	61	26	35	no	N84665	TP1 gene for triosephosphate isomerase
68	61	19	42	5.24E-05	AA147786	T-cell cyclonhlin
69	60	13	47	2.88E-08	AA565966	ribosomal protein S8 (rS8)
70	59	59	0	6.40E-12	R12300	Nucleosomin (nucleolar phosphonrotein B23, numatrin)
71	59	15	44	9.75E-07	AA186565	ribosomal protein L5
72	58	58	0	1.51E-12	X12671	Heterogeneous nuclear ribonucleonprotein A1
73	57	16	41	1.26E-05	R99400	KTAA0097 gene
74	57	57	0	3.08E-12	###	###
75	56	9	47	3.71E-10	R22840	mRNA from chromosome 5a21-22, clone-FBR35
76	56	49	7	2.05E-06	AA047017	KTAA0309 gene
77	56	55	1	5.24E-11	M59488	S100 protein beta-subunit
78	56	42	14	6.55E-03	D12485	plasma cell membrane glycoprotein (PC-1)
79	56	54	2	3.88E-10	R86103	Signal recognition particle receptor (dockine)
80	55	55	0	1.28E-11	M58051	human novel growth factor receptor
81	54	54	0	2.59E-11	X00700	MHC class II HL A-DRw10-beta
82	53	52	1	4.09E-10	X62534	High-mobility group (nonhistone chromosomal) protein 2
83	53	0	53	4.34E-20	W68765	beta-tubulin gene, clone m40
84	53	53	0	5.24E-11	AA094834	Inhibitor of DNA binding 2, dominant negative helix-loop-helix protein
85	52	46	6	3.91E-06	J03258	Vitamin D (1,25-dihydroxyvitamin D3) receptor
86	52	13	39	3.21E-06	AA135686	KTAA0391 gene
87	51	51	0	2.13E-10	R06498	Ubiquitin-like protein
88	50	39	11	2.44E-03	R23949	human Bruton's tyrosine kinase-associated protein-135
89	50	49	1	4.48E-10	AA278921	Hematopoietic proteoglycan core protein
90	50	40	10	8.55E-04	U27193	protein-tyrosine phosphatase
91	50	50	0	5.08E-11	U32519	GAP, SH3 binding protein
92	50	50	0	5.08E-11	I27841	autoantigen pericentriol material 1 (PCM-1)
93	49	48	1	8.90E-10	AA243313	Laminin receptor (2H5 epitone)
94	49	49	0	1.04E-10	R60497	ubiquitin-cytochrome c reductase, smallest subunit
95	49	47	2	6.74E-09	N43864	HLA-A2 class I antigen
96	49	21	28	no	D78361	ornithine decarboxylase antizyme, ORF 1 and ORF 2
97	48	47	1	1.76E-09	AA129206	hsc70 gene for 71 kd heat shock cognate protein
98	48	48	0	2.12E-10	U43292	MDS1B (MDS1)
99	47	0	47	6.75E-18	AA657885	ribosomal protein S3 (XP1PO) (rS3)
100	47	19	28	no	D84276	CD38 antigen (n45)

Tab. 4-5 Die ersten 100 Co-Cluster

Dargestellt sind die ersten 100 Co-Cluster der Bibliotheken Jurkat und NKL. Fett gedruckte Gen-Bezeichnungen stellen einen signifikanten Datenbank-Match dar, für die mit ### gekennzeichneten Co-Cluster konnte kein Datenbank-Match gefunden werden und stellen somit eventuelle Kandidatengene dar.

4.3.2.1 Cluster-Splitting

Während der reversen Transkription werden die mRNAs in cDNAs umgeschrieben. Dabei kann es bei einigen Klonen zum vorzeitigen Syntheseabbruch kommen, wodurch kürzere Fragmente ein und derselben mRNA entstehen können. Die durchschnittliche Größe der Klone beider Bibliotheken beträgt ungefähr 1.5 kb, wobei die einzelnen Fragmente aber in einem Bereich von etwa 0.5 - 5 kb variieren. Aufgrund der unterschiedlichen Länge dieser Klone eines gleichen Gens kann es zu fehlenden oder zusätzlichen Hybridisierungs-Signalen bestimmter Klone kommen. So weist ein längerer Klon mehr positive Signale in den Oligonucleotid-Hybridisierungen und folglich einen anderen Fingerprint auf als der kürzere Klon und kann daher in ein anderes Cluster geordnet werden, obwohl das selbe Gen repräsentiert wird. So kann es vorkommen, daß die positiven Klone einer Longprobe-Hybridisierung nicht nur in einem Cluster, sondern in zwei oder mehreren Clustern wiederzufinden sind, was als Cluster-Splitting bezeichnet wird. Um eben dieses weitgehend einschränken zu können, werden die Longprobe-Hybridisierungen durchgeführt und anhand der Variierung einiger Parameter des Clusterings die besten Zuordnungen ermittelt. Eine Übersicht über die Entstehung des Cluster-Splittings gibt die Abb. 4-14.

Als Beispiele für das Aufsplitten eigentlich zusammengehörender Cluster können u.a. GAPDH und α -Tubulin genannt werden. GAPDH splittet in mindestens 3 größere Cluster (Co-Cluster 22, 29, 199). Dabei fällt auf, daß die einzelnen Cluster jeweils nur aus entweder Jurkat- bzw. NKL-Klonen zusammengesetzt sind (NKL:Jurkat jeweils 132:0, 0:111, 0:25). Ähnlich verhält es sich auch mit dem α -Tubulin, das in mindestens 5 Cluster (Co-Cluster 37 (85:1), 63 (0:64), 125 (0:39), 610 (8:1) und 978 (0:6)) splittet. Bei einer PCR und anschließender Gelelektrophorese zur Größenbestimmung der Fragmente von jeweils 30 Klonen der Co-Cluster 22 und 29, die für GAPDH codieren, konnten jedoch in diesem Fall keine wesentlichen Größenunterschiede festgestellt werden. So weisen beide Co-Cluster im Durchschnitt eine recht einheitliche Größe mit einem geringen Größenunterschied von nur etwa 20 - 50 bp auf. Mit diesem geringen Unterschied allein jedoch läßt sich das Cluster-Splitting in dem Fall dieser beiden Co-Cluster nicht erklären. Die Abb. 4-15 veranschaulicht die Größen der amplifizierten Fragmente beider Co-Cluster.

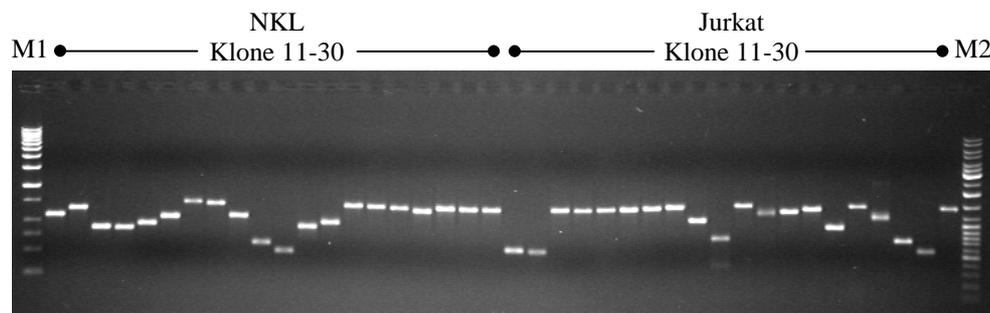


Abb. 4-15 Gelelektrophorese amplifizierter Klone der Co-Cluster 22 und 29 (GAPDH)

Abgebildet sind jeweils 20 Klone der Co-Clusters 22 bzw. 29, die beide für die GAPDH codieren, jedoch in zwei verschiedene Co-Cluster splitten. Die Klone beider Co-Cluster weisen im Durchschnitt gleiche Größen auf.

Es lassen sich aber im direkten Vergleich beider Visualisierungen der Fingerprints (siehe Abb. 4-16) trotz der sichtlichen Übereinstimmung einer Vielzahl von Signalen einige Unterschiede hinsichtlich der Hybridisierung mit einigen Oligos erkennen.

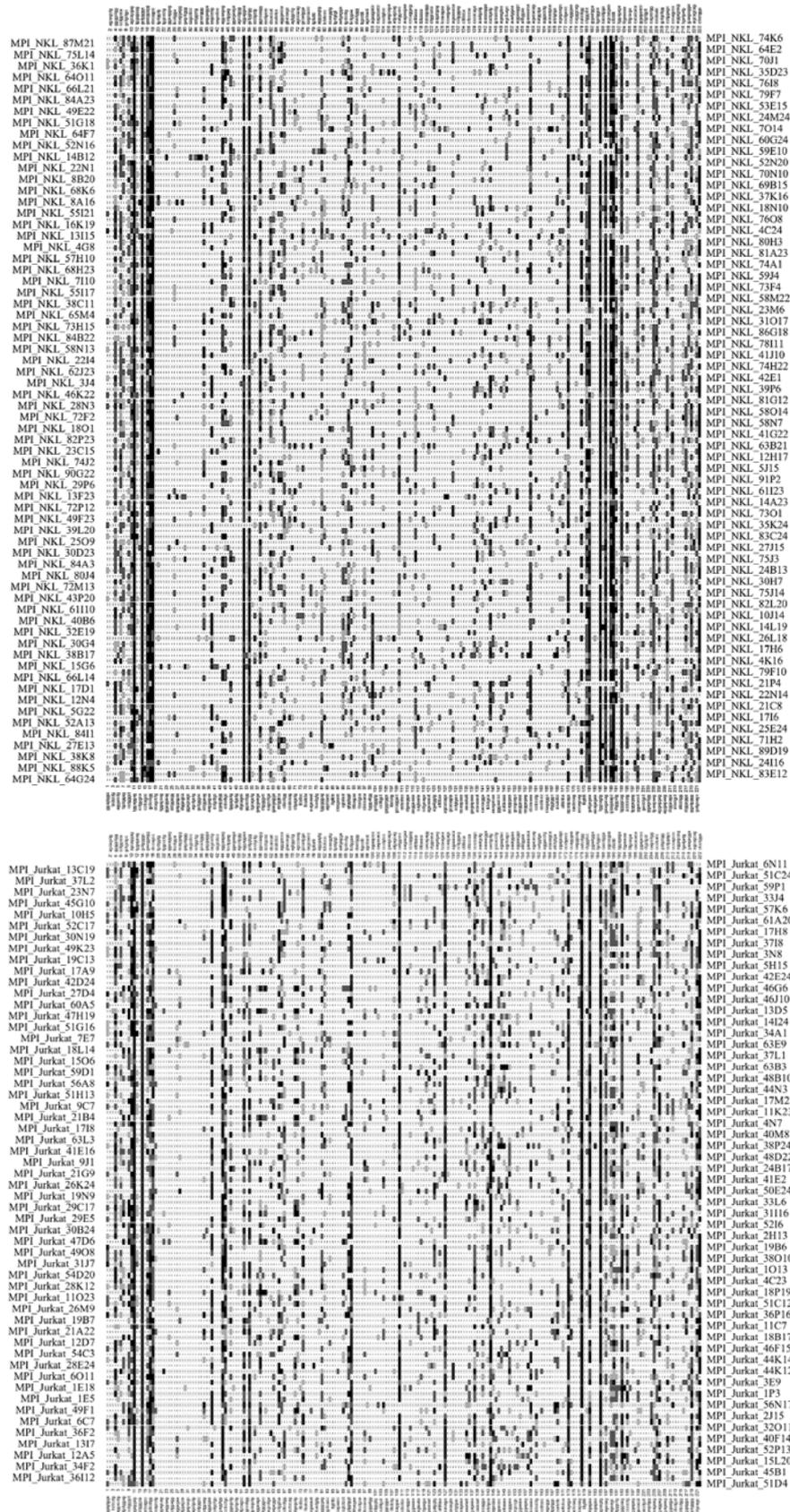


Abb. 4-16 Visualisierungen der Fingerprints der Co-Cluster 22 und 29 (GAPDH)

Vergleich der Visualisierungen der Fingerprints der beiden Co-Cluster 22 und 29, in die die GAPDH splittet. Dabei lassen sich deutlich Unterschiede bezüglich der Bindung einiger Oligos erkennen.

Auch eine Sequenzierung einiger Klone beider Co-Cluster ließ keine signifikanten Unterschiede in dem sequenzierten Bereich erkennen, die ein Aufsplitten in verschiedene Cluster begründen würden (Abb. 4-17).

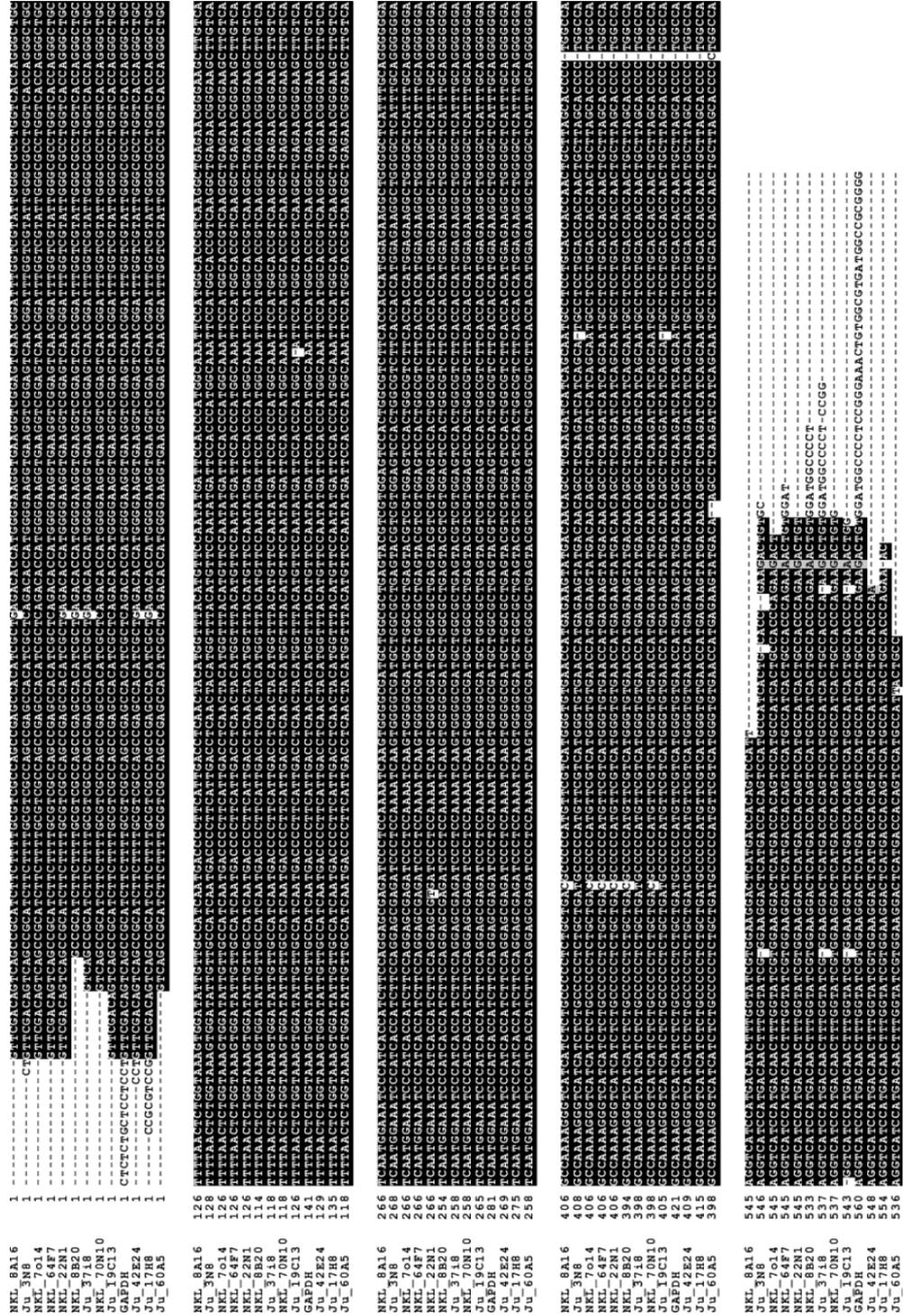


Abb. 4-17 Alignment von Klonen der Co-Cluster 22 und 29

Alignment und Boxshade von jeweils sechs Klonen der Co-Cluster 22 und 29 sowie der Sequenz der GAPDH zum Vergleich. Multiple Sequence Alignments: (<http://dot.imgen.bcm.tmc.edu:9331/multi-align/multi-align.html>); Boxshade: (http://www.ch.embnet.org/software/BOX_form.html).

Das Aufsplitten dieser eigentlich zusammengehörenden Co-Cluster in zwei oder mehrere verschiedene Co-Cluster, die jeweils nur Klone einer der beiden Bibliotheken beinhalten, ist demnach eher darauf zurückzuführen, daß beide zu vergleichenden Bibliotheken unter unterschiedlichen Bedingungen erstellt und hybridisiert wurden. Es gibt eine Reihe von Faktoren, die Einfluß auf die Qualität der Ergebnisse haben können, wie speziell in diesem Fall auf die Signalintensitäten. Solche Faktoren können beispielsweise die Verwendung verschiedener Chargen von Membranen und Radioaktivität sein oder aber auch unterschiedliche Aktivitäten der markierten Oligonucleotide. Infolgedessen lassen sich zumindest zum Teil unterschiedliche Signalintensitäten erklären. Auf weitere Einzelheiten zum Thema des Cluster-Splittings wird noch genauer in der Diskussion eingegangen.

4.3.2.2 Co-Clustering-Ergebnisse

Das Co-Clustering ist das Zusammenfassen von Clustern gleicher Identität zweier oder mehrerer zu vergleichender Bibliotheken. Dies geschieht anhand der Ähnlichkeit der Consensus-Fingerprints der korrespondierenden Cluster. Das Co-Clustering dient dem direkten Vergleich der Genexpression der zu untersuchenden Bibliotheken und man erhält einen Überblick über gemeinsam sowie differentiell exprimierte Gene.

Insgesamt konnten 24.057 Co-Cluster ermittelt werden. Davon haben 4.223 Co-Cluster eine Mindestgröße von zwei Klonen, 19.832 Klone verbleiben als Singletons. Das bedeutet, diese Cluster enthalten nur einen Klon, der keinem anderen Cluster zugeordnet werden konnte. Diese verbleiben also als Unikate. Die Co-Cluster 1 - 1.409 sind Cluster mit einer Größe von mindestens 5 Mitgliedern, d.h., diesen Clustern konnten jeweils mindestens 5 Klone zugeordnet werden. Die Co-Cluster 1.410 - 4.224 sind Cluster mit einer Anzahl von weniger als 5 Klonen. Die Co-Cluster 4.225 - 24.057 letztlich umfassen Singletons. Die Tab. 4-6 gibt einen Überblick über die mittels der Clustering-Analyse erhaltenen Cluster der einzelnen Bibliotheken Jurkat und NKL sowie die in Co-Cluster zusammengefaßten Cluster beider Bibliotheken.

	Singletons	Clustergröße					total	>/=5
		2-4	5-10	11-100	101-1000	>1000		
NKL-Cluster	10926	1671	279	201	16	1	2168	497
Jurkat-Cluster	8276	1690	308	223	9	0	2230	540
Co-Cluster	19832	2814	873	505	26	5	4223	1409

Tab. 4-6 Clustergrößen und deren Verteilung

Darstellung der Co-Clustergrößen und deren Verteilung sowie der Clusterverteilung in den einzelnen Bibliotheken und Jurkat und NKL zum Vergleich.

So konnten die Sequenzen zu 1.409 Co-Clustern mit mindestens 5 Klonen je Cluster geordnet werden. 2.815 Co-Cluster enthalten weniger als 5 Klone je Cluster. 19.833 Sequenzen verbleiben als Singletons. Dem größten Co-Cluster konnten dabei 1.981 Klone zugeordnet werden. Eine Übersicht über die genaue Verteilung der Cluster und Co-Clustergrößen ist in Abb. 4-18 gegeben, in der die Verteilung der Co-Cluster und ihrer Größen graphisch dargestellt ist. Es wird deutlich, daß es eine größere Anzahl an Clustern gibt, denen nur eine mittlere oder geringe Zahl an Mitgliedern zugeordnet werden konnten. Dahingegen gibt es nur wenige Cluster mit einer großen Anzahl an Klonen.

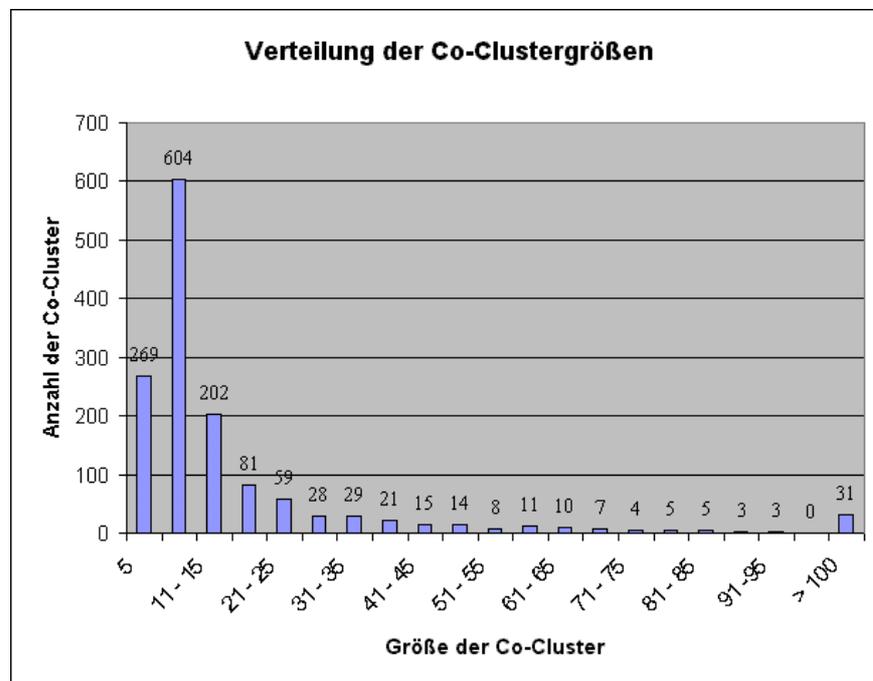


Abb. 4-18 Verteilung der Co-Clustergrößen

Graphische Darstellung der Clustergrößen der Co-Cluster mit mindestens 5 Klonen je Cluster und ihrer Verteilung. Am häufigsten sind solche Cluster, die nur relativ wenige Klone beinhalten. Seltener kommen solche Cluster vor, denen viele Klone zugeordnet wurden.

Von den 1.409 Co-Clustern mit mehr als 5 Mitgliedern sind 538 differentiell exprimiert, d.h., sie werden entweder in beiden Zell-Linien signifikant unterschiedlich stark oder aber ausschließlich nur in einer der beiden Zell-Linien exprimiert. Von diesen 538 differentiell exprimierten Genen kommen 324 NKL-spezifisch und 214 Jurkat-spezifisch vor.

Um die Ergebnisse des Co-Clusterings hinsichtlich der gemeinsamen bzw. differentiellen Expression der Gene graphisch zu veranschaulichen, wurde ein logarithmischer Plot der Ergebnisse des Co-Clusterings erstellt. Jedes Cluster ist repräsentiert durch einen Punkt im Plot mit Koordinaten, der der Frequenz des entsprechenden Clusters in der jeweiligen Bibliothek entspricht. Die Linien stellen die Grenzen des 99%-igen Konfidenzintervalls des Binomialtests dar. Alle diejenigen Cluster, die gemeinsam in beiden Bibliotheken etwa gleich stark exprimiert vorkommen, liegen innerhalb dieser Linien, alle differentiell exprimierten liegen außerhalb dieses Bereiches. Die Abb. 4-19 gibt einen Überblick über die Verteilung der gemeinsam und der differentiell exprimierten Co-Cluster.

Dabei ist es jedoch auch möglich, daß einige Cluster, die tatsächlich gemeinsam exprimierte Cluster darstellen, aufgrund des Phänomens des Cluster-Splittings in kleinere Cluster aufgesplittet sind, die nun ausschließlich nur Klone einer der beiden Bibliotheken beinhalten und somit außerhalb des Intervalls dargestellt werden. Darauf wurde bereits im Kapitel 4.3.2.1 näher eingegangen.

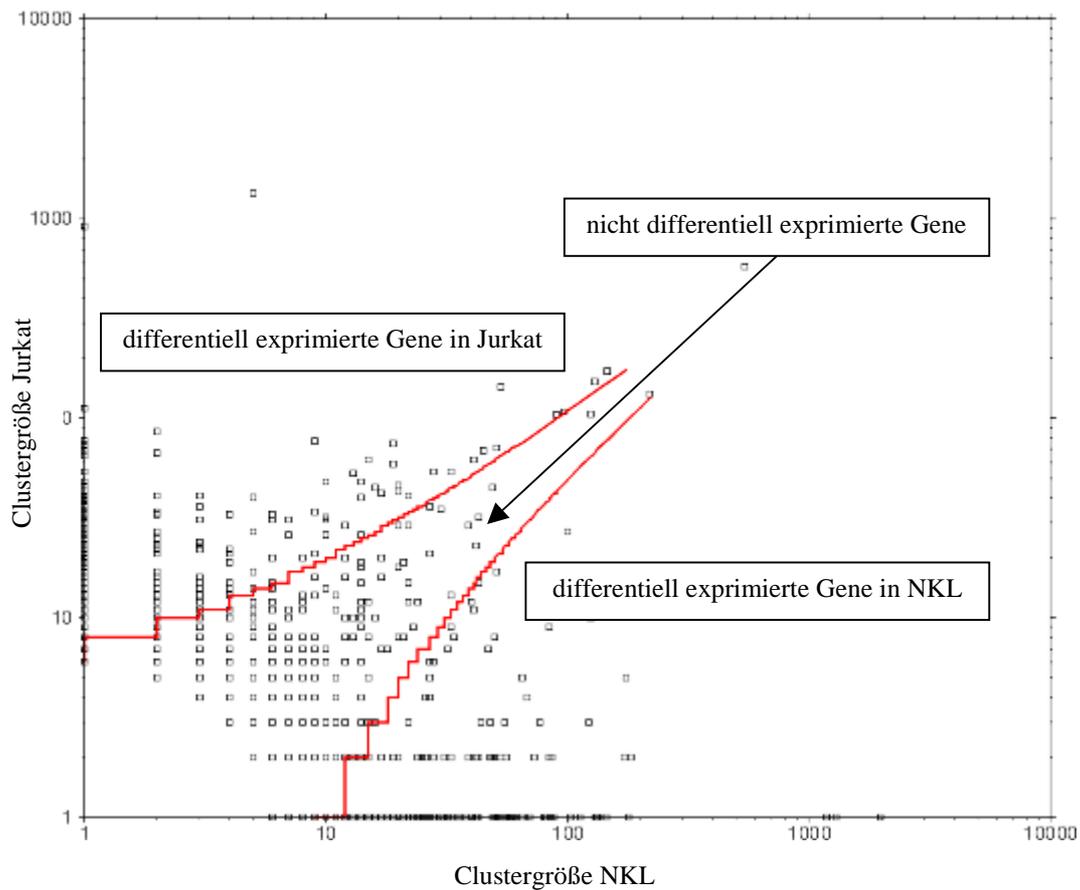


Abb. 4-19 Co-Clusterverteilung

Clusterverteilung im Co-Clustering der Bibliotheken Jurkat und NKL, jedes Cluster ist dargestellt anhand eines kleinen Quadrates. Die x-Achse zeigt die Anzahl der Klone in der NKL-Bibliothek, die y-Achse der Jurkat-Bibliothek. Für ein jedes Cluster wurde ein Binomialtest berechnet, die roten Linien stellen die Grenzen des 99%-igen Konfidenzintervalls des Tests dar.

Differentiell in Jurkat exprimierte Gene lt. Co-Clustering-Analyse sind beispielsweise die T-Zell-Rezeptoren und das CD3E-Gen -beide schon bekannte T-Zell spezifische Gene-, sowie die Elongationsfaktoren α und β , der Laminin-Rezeptor (2H5-Epitop) oder das *translationally controlled tumor protein*. Differentiell in der Zell-Linie NKL exprimierte Gene sind zum Beispiel das NKG2-D-Gen, das ein bekanntes NK-Zell spezifisches Gen ist, bzw. das HLA-A2-Gen. Die Tab. 4-7 und Tab. 4-8 geben eine Übersicht über eine Auswahl weiterer differentiell in Jurkat bzw. in NKL exprimierter Gene.

NKL	Jurkat	DB-Match
0	96	T-Zell Rezeptor β -Kette
19	42	T-Zell Cyclophilin
11	41	IL-2 Rezeptor γ -Kette
0	12	TALLA-1
0	7	Clathrin assembly Protein50 AP50
0	35	T-Zell Rezeptor α -Kette
1	10	NM23 Protein
0	6	JM26 Protein
0	12	CD3E epsilon polypeptide (TiT3 complex)
2	30	<i>insulinoma rig-analog mRNA encoding DNA binding protein</i>
1	32	<i>high-mobility group (nonhistone chromosomal) protein1</i>
3	21	Ferritin schwere Kette
3	13	CyclinD3
0	26	L-Plastin (Lymphocyten cytosolisches Protein 1)
0	26	PHAPI2b Protein
5	19	<i>proliferation-associated gene A (Natural Killer-enhancing factor A)</i>

Tab. 4-7 **Differentiell exprimierte Gene in der Zell-Linie Jurkat**

Dargestellt sind solche Gene, die laut Co-Clustering-Analyse stärker oder ausschließlich differentiell in der Zell-Linie Jurkat exprimiert werden. TALLA-1 codiert für das T-Zell akute lymphoblastische Leukämie assoziierte Antigen 1; NM23 für die Nucleosid-Diphosphat-Kinase A (auch bezeichnet als *tumor metastatic process-associated protein* oder *metastasis inhibition factor nm23*); JM26 für das Polyglutamin bindende Protein 1.

NKL	Jurkat	DB-Match
1981	0	<i>Dents Disease candidate gene (chloride channel 5)</i>
2477	0	<i>eyes absent homolog (Eab1)</i>
15	0	NKG2-D
55	3	HLA-A2 Klasse I-Antigen
14	0	RbAp48, <i>Retinoblastoma binding protein4</i>

Tab. 4-8 **Differentiell exprimierte Gene in der Zell-Linie NKL**

Dargestellt sind einige der Gene, die laut Co-Clustering-Analyse stärker oder ausschließlich differentiell in der Zell-Linie NKL exprimiert sind.

Von den 1.409 Co-Clustern mit mehr als 5 Mitgliedern sind 135 Co-Cluster ohne signifikanten Treffer in der Datenbank. Das entspricht etwa 9.6% unter diesen Clustern. Diese Gene sind eventuell noch unbekannte Gene und stellen mögliche Kandidatengene, beispielsweise neue Targetgene für die Arzneimittelforschung dar. Von diesen 135 nicht identifizierten Genen sind 43 differentiell exprimiert, 19 davon differentiell in Jurkat, 24 differentiell in NKL. Im Rahmen dieser Arbeit konnten jedoch keine näheren Untersuchungen dieser Gene vorgenommen werden.

Bei all diesen Angaben, besonders bei den nicht näher definierten, muß allerdings auch das Vorkommen des Cluster-Splittings berücksichtigt werden, welches die Angaben etwas relativiert.

Gene können bestimmten funktionellen Klassen zugeordnet werden, je nachdem, welche Rolle sie in den Stoffwechsel- und Lebensvorgängen der Zelle übernehmen. So konnten einige der identifizierten Gene als Vertreter den folgenden funktionellen Klassen zugeordnet werden:

House keeping Gene

Translations-Elongationsfaktor 1-alpha-1; verschiedene ribosomale Proteine
Glycerinaldehyd-3-phosphate-dehydrogenase, Enolase, Aldolase A, Lactatdehydrogenase
Actin gamma 1, α -Tubulin, β -Tubulin, Ubiquitin

Zellzyklus-Kontrollproteine

Cyclin D3, Cyclin A, Cyclin B1
Zellzyklus-Protein p38-2G4 Homolog (hG4-1)

Apoptosis related Proteine und proteolytische Proteine

DAP-3 (*ionizing radiation resistance-conferring protein; death-associated protein 3*)
Glutathione S-Transferase pi; *Apoptosis-related protein* TFAR15 (TFAR15)

DNA-Synthese, Reparatur- und Rekombinations-Proteine

Ku Protein-Untereinheit; RAG (*recombination activating gene*) Cohort 1
nucleäre RNA-Helicase; APEX Nuclease (*apurinic; multifunctional DNA repair enzyme*)

DNA-Bindung, Transkription, Transkriptionsfaktoren

non-histone chromosomal protein HMG-17
nucleärer Faktor NF45; Histon (H3.3) Klasse B, Histon H1 (H1.2)
Transkriptionsfaktor 3 (*E2A immunoglobulin enhancer binding factor E12*)
RNA-Polymerase II elongation factor-like protein
DNA-bindendes Protein (CROC-1A)
TATA-bindendes Protein assoziierter Faktor 30 kDa Untereinheit (tafII30)

Zellrezeptoren

T-Zell-Rezeptor α Kette, T-Zell-Rezeptor aktive β -Kette
Interleukin 2 Rezeptor γ -Kette
NKG2-D

Zelloberflächen-Antigene, Adhäsion und Migration

Clathrin leichte Kette A, E-Cadherin, Annexin II (Lipocortin II), Fibronectin 1
Macrophage migration inhibitory factor

extrazelluläres Zell-Signalling und Kommunikationsproteine / Cytokine, Wachstumsfaktoren sowie assoziierte Proteine

Basigin; TGF- β Rezeptor interagierendes Protein 1
Granulocyte colony-stimulating factor induced gene

HLA related

MHC Klasse I Antigen HLA-A2, HLA-E
MHC Klasse II Antigen HLA-DC1-alpha (DRw6,w6)
MHC Protein *homologous to chicken B complex protein*

Ionenkanäle und Transportproteine

voltage-dependent anion channel 1, voltage-dependent anion channel 2, voltage dependent anion channel 3
Phosphat-Carrier, mitochondrial; 2-Oxoglutarate-Carrier Protein
ATPase, H^+ transporting, lysosomal (*vacuolar proton pump*) 31kD

Modulatoren / Effektoren / intrazelluläre Transducer

GTP-bindendes Protein (RAB1); 14.3.3 protein, Protein-Kinase-Regulator
Lymphocyten-spezifisches Protein-Tyrosine-Kinase (*lck*)

stress response Proteine

Hitzeschock-Protein 90kD, Hitzeschock-Protein 4 70kD
hsc70 Gen (71 kd *heat shock cognate protein*)
Chaperonin (Cctg mRNA), Chaperonin-like Protein (HTR3)

Onkogene und Tumorsuppressoren / Krebsgene (www.genomsolutions.com)

(dieser Gruppe können auch weitere Gene aus den anderen Klassen zugeordnet werden)

Pancreatic tumor-related protein
translationally controlled tumor protein
Lymphocyte cytosolic protein 1 (L-plastin)
RbAp48 encoding retinoblastoma binding protein 4
Breakpoint Cluster Region Protein 1 (BCRG1)
RHOA proto-oncogene multi-drug-resistance protein

Auffällig -und natürlich- bei der Betrachtung aller Cluster hinsichtlich der Häufigkeit des Vorkommens der Genfamilien ist, daß die größten Cluster, die am stärksten in der Zelle exprimierte Gene, für *house keeping genes* kodieren, die vorrangig den Grundstoffwechsel der Zelle aufrecht erhalten. Weitere recht große Cluster sind solche, die für Gene codieren, die an der DNA-Bindung und der Transkription beteiligt sind, sowie für Transkriptionsfaktoren und für Gene, die in die Kontrolle und Regulation des Zellzyklus involviert sind. Desweiteren kommen solche Gene vor, die für Zell-Oberflächenmoleküle und Zell-Adhäsionsmoleküle codieren. So gehören zu den größten Clustern stark exprimierte Gene wie z.B. der Translations-Elongationsfaktor 1-alpha-1, Actin, verschiedene ribosomale Proteine, Tubuline, Hitzeschock-Proteine aber auch verschiedene Enzyme, wie GAPDH oder Enolase1.

Die Co-Cluster 1 bis 4 sind sehr große Cluster. Mit jeweils 1.981 bzw. mehr als 1.000 Mitgliedern in der einen sowie keinen oder nur sehr wenigen Mitgliedern in der anderen Bibliothek weisen diese Co-Cluster eine starke differentielle Expression auf. Jedoch weisen sie nach Datenbankabgleich der Fingerprints keine signifikanten Homologien zu bekannten Sequenzen der Datenbanken auf. Es ließen sich jedoch gewisse Ähnlichkeiten (e-Value etwa 10^5) zu *Dents Disease candidate gene* (auch *chloride channel 5* oder Nephrolithiasis 2, X-linked; Co-Cluster 1), Dystrobrevin (DTN; Co-Cluster 2) bzw. *eyes absent homolog* (Eab1; Co-Cluster 3 und 4) erkennen.

Auch in den sich anschließenden Sequenzierungen einiger Klone der Co-Cluster 1 und 2 konnten diese Cluster nicht eindeutig identifiziert werden. Die isolierten Sequenzen stellten sich als sehr verschieden heraus und konnten so keine einheitlichen Identitäten hervorbringen. Zwei der sequenzierten Klone des Co-Clusters 1 ergaben Treffer mit sehr hohen Wahrscheinlichkeiten und wurden identifiziert als eukaryotischer Translations-Elongationsfaktor EEF2 (e-Value = 0) bzw. als TRAF *family member-associated NFkB activator* (e-Value = 0). Im Fall des Co-Clusters 2 konnten zwei Sequenzen als MORF *related gene p15* (e-Value = $1e-139$) bzw. *als melanoma associated antigen recognised by cytotoxic T lymphocytes* (e-Value = $4e-56$) identifiziert werden.

MORF	<i>monocytic leukemia zinc finger protein related factor</i>
TRAF	TNF-Rezeptor assoziierter Faktor

Bei einigen, als differentiell exprimiert identifizierten Clustern, kann es sich auch um gesplittete Cluster handeln, besonders wenn es sich dabei um solche Cluster handelt, in denen nur Klone in einer größeren Anzahl der einen Bibliothek vorkommen und keine Klone der jeweils anderen Bibliothek vertreten sind. So ist immer Validierung mit direkten Nachweismethoden notwendig.

Im Falle der Klone dieser Co-Cluster 1 und 2 ist es möglich, daß es sich um Sequenzen handelt, die nur in bestimmten Bereichen hohe Homologie aufweisen, jedoch kein Gen, sondern eher eine Genfamilie repräsentieren. Solche Sequenzen würden auch einen ähnlichen Fingerprint ergeben und so in ein Cluster geordnet werden. Dies kann in weiteren Untersuchungen (Sequenzierung) geklärt werden.

Die Klone der Co-Cluster 3 und 4 ließen sich auch in Wiederholungen der PCR nicht amplifizieren. In einer sich anschließenden Plasmidpräparation und Restriktion mit den beiden Enzymen *NotI* und *SalI* ließ sich nur eine Vektorbande bei etwa 3.5 kb nachweisen, jedoch kein enthaltenes Insert (siehe Abb. 4-20).

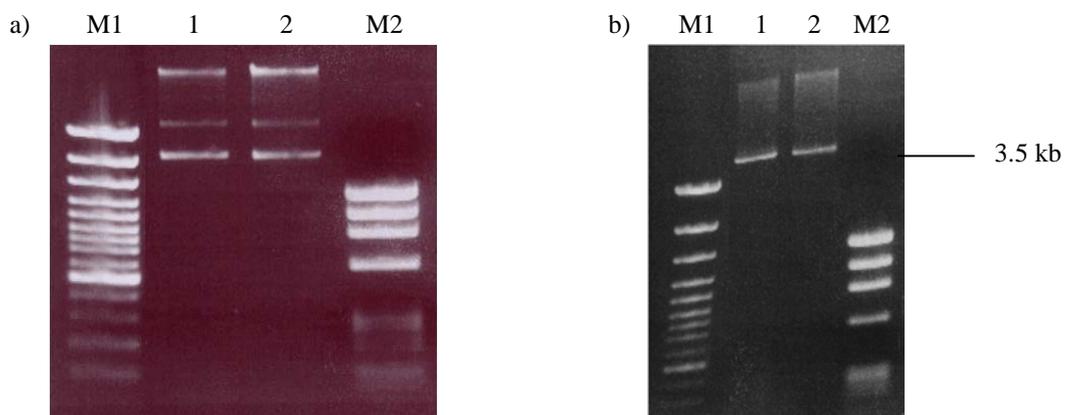


Abb. 4-20 Plasmidpräparation Co-Cluster 3 und 4

Plasmidpräparation der 1.Wahl-Klone aus den Co-Clustern 3 und 4.

- a) vor Restriktion
- b) nach Restriktion mit den Enzymen *NotI/SalI*
- 1 NKL 40o8 (Co-Cluster 1)
- 2 NKL 92E22 (Co-Cluster 3)
- M1 Gene Ruler
- M2 Φ X 174

Diese Co-Cluster enthalten somit keine amplifizierbaren Inserts, weisen aber spezifische Fingerprints auf, so daß sie in Cluster geordnet werden konnten. Beide Co-Cluster weisen zudem einen identischen bzw. sehr ähnlichen Fingerprint auf, bilden möglicherweise ein gemeinsames Co-Cluster. In der Abb. 4-21 sind als Beispiel die Fingerprints dieser beiden Co-Cluster im Vergleich dargestellt.

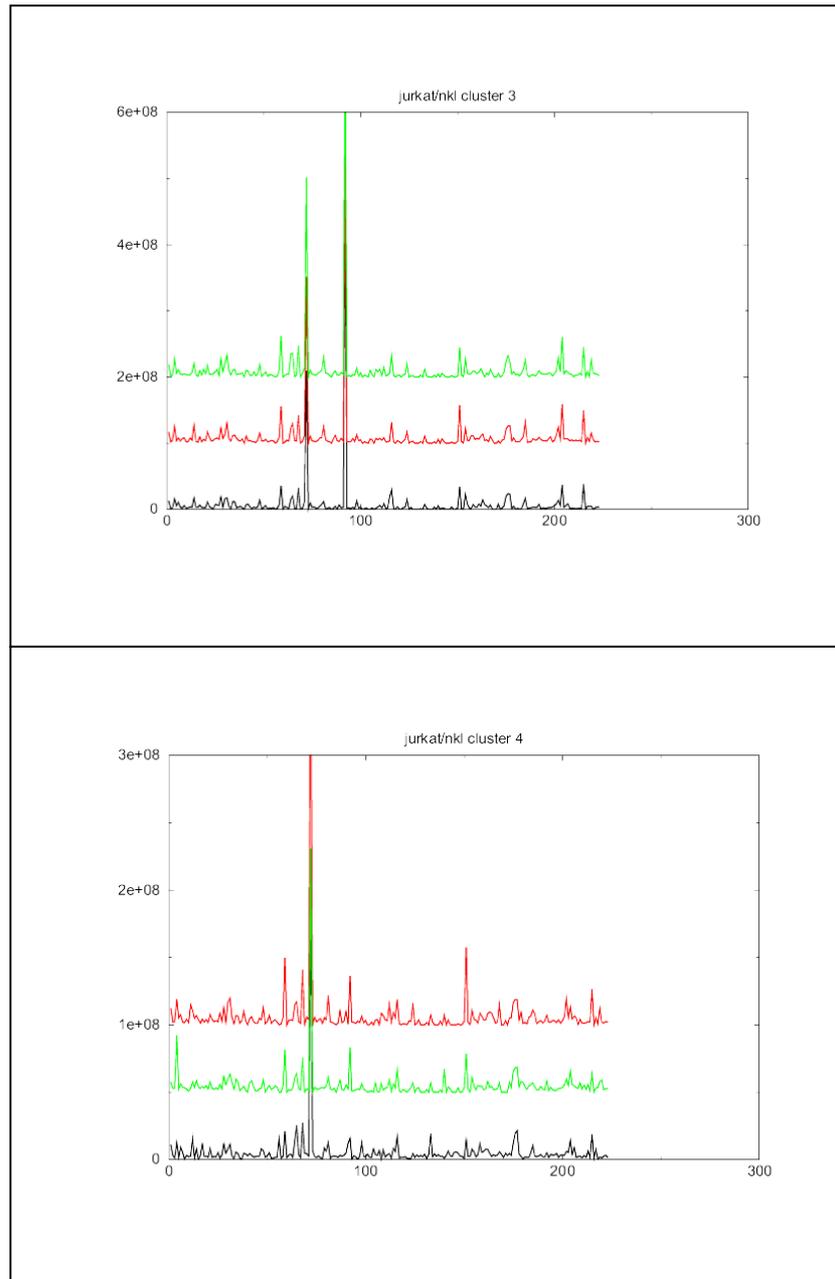


Abb. 4-21 Vergleich der Fingerprints der Co-Cluster 3 und 4

In dieser Abbildung sind die Fingerprints jeweils dreier Klone der Cluster 3 und 4 graphisch dargestellt, um ihre Ähnlichkeit zu verdeutlichen. Beide weisen in bestimmten Bereichen starke Ähnlichkeiten auf, wurden jedoch in verschiedene Cluster geordnet. Auch in wiederholten Amplifikationsreaktionen konnten keine Inserts nachgewiesen werden.

Im Falle dieser beiden Co-Cluster 3 und 4 ist es auch möglich, daß es sich dabei um sogenannte „leere“ Klone, also Klone ohne Inserts handelt. Auch solche Klone geben in Hybridisierungen Signale, wenngleich auch nur sehr schwache, die als das sogenannte Hintergrundrauschen (*background noise*) bezeichnet werden. Bei der Normalisierung wird jedem Signal, auch diesen sehr schwachen Signalen, ein numerischer Wert zugeordnet, so daß letztlich ein Fingerprint daraus

errechnet werden kann. Zur Analyse werden die 15% schwächsten Signale ausgeschlossen, jedoch ist es möglich, daß einige dieser Klone mit zur Analyse kommen, wenn mehr als nur 15% schwache Signale existieren. Aus den numerischen Werten dieser Signale läßt sich dann auch ein Fingerprint errechnen. So ließe sich erklären, warum im Falle der Co-Cluster 3 und 4 keine Amplifikation von Fragmenten möglich war, jedoch Fingerprints errechnet werden konnten. Die Anzahl von 2.477 Klonen in diesen Clustern entspräche auch in etwa dem 8%-igen Anteil an in der PCR festgestellten Klonen ohne Insert. Auf das Problem wird in der Diskussion noch näher eingegangen.

4.3.3 Northern-Hybridisierung

Das Oligonucleotid Fingerprinting ist keine absolute quantitative Methode. Bei kleinen Clustern mit einer Größe von weniger als 20 Klonen ist die Grenze der Sensitivität dieser Methode erreicht, d.h., es können keine verlässlichen Angaben mehr über die Clusterverteilung gemacht werden. Das ONF kann daher nur als Indikator dienen und Tendenzen beschreiben. Es ist jedoch eine sehr gute Möglichkeit zur Präselektion interessierender Klone für weitere Anwendungen wie der Sequenzierung, der Komplex-Hybridisierung oder zum Finden neuer Gene. Das Oligonucleotid Fingerprinting stellt eine (indirekte) Methode der Expressionsanalyse einer großen Zahl von Genen gleichzeitig dar, die im großen Maßstab einen groben Überblick über die Genexpression in der Zelle geben kann. Auch stellt das ONF nur eine der möglichen Methoden zur Quantifizierung der Expression dar und ist sehr gut zur Erstellung nicht oder nur gering redundanter Unigene-Sets geeignet, die dann mit anderen Methoden weiter untersucht werden können. Die Ergebnisse der Clustering-Analysen müssen noch mit detaillierteren Methoden untersucht und evaluiert werden. Die genaue Expression mehrerer Zellpopulationen im Vergleich zueinander, das detaillierte Expressions-Profiling interessierender Gene muß mit einer direkten Methode wie der Northern-Hybridisierung bestätigt werden.

Die Northern-Analysen sollen dazu dienen, die Zellspezifität einiger im Co-Clustering gefundener differentiell exprimierter Gene direkt nachzuweisen. Dabei werden RNA-Fractionen der beiden Zell-Linien Jurkat und NKL in Hybridisierungsexperimenten auf Sequenzhomologien zu ausgesuchten Proben untersucht.

4.3.3.1 Northern-Hybridisierungen

In den folgenden Northern-Hybridisierungen sollen nun die aus dem Co-Clustering erhaltenen Angaben über die differentielle Expression einiger Gene nachgewiesen werden, besonders solcher Gene, die von immunologischer und onkologischer Relevanz sein können. Ist ein Gen stärker in der einen Zellpopulation als in der anderen exprimiert, so läßt sich das auch in einer Hybridisierung der auf einer Membran immobilisierten RNA mit der entsprechenden DNA-Sonde nachweisen. Als Proben wurden solche Gene ausgesucht, die laut Co-Clustering-Analyse besonders signifikante Unterschiede in der Genexpression der beiden Zell-Linien aufwiesen. Einige dieser in Jurkat- bzw. NKL-spezifischen Gene sind in den Tab. 4-9 und Tab. 4-10 aufgeführt. Dazu gehören auch solche Gene, von denen bekannt ist, daß sie T-Zell-spezifisch exprimiert werden, beispielsweise TALLA-1, der T-Zell-Rezeptor mit seinen verschiedenen Ketten, CD3E, oder daß sie für NK-Zellen spezifische Gene sind, wie NKG2-D und andere Gene von immunologischer Relevanz.

NKL	Jurkat	DB-Match	Co-Cluster
0	96	T-Zell Rezeptor β -Kette	47, 260
19	42	T-Zell Cyclophilin	68
11	41	IL-2 Rezeptor γ -Kette	153, 259
0	12	TALLA-1	469
0	7	Clathrin assembly Protein50 AP50	832
0	35	T-Zell Rezeptor α -Kette	139
1	10	NM23 Protein	528
0	6	JM26 Protein	1132
0	12	CD3E epsilon Polypeptid (TiT3-Komplex)	481
2	30	<i>insulinoma rig-analog mRNA encoding DNA binding protein</i>	156
1	32	<i>high-mobility group (nonhistone chromosomal) protein1</i>	152
3	21	Ferritin schwere Kette	214
3	13	CyclinD3	308
0	26	L-Plastin (Lymphocyten cytosolisches Protein 1)	187
0	26	PHAPI2b Protein	191
5	19	<i>proliferation-associated gene A (Natural Killer-enhancing factor A)</i>	206

Tab. 4-9 Differenziell exprimierte Gene der Zell-Linie Jurkat

Dargestellt sind einige Gene, die laut Co-Clustering-Analyse stärker oder ausschließlich differentiell in der Zell-Linie Jurkat exprimiert werden sowie die diese Gene repräsentierenden Cluster. TALLA-1: T-Zell akute lymphoblastische Leukämie assoziiertes Antigen 1; NM23: Nucleosid-Diphosphat-Kinase A (NDKA), auch als *tumor metastatic process-associated protein* oder *metastasis inhibition factor nm23* bezeichnet; JM26: *similar to polyglutamine binding protein 1*; PHAPI2b: *acidic protein rich in leucines*.

NKL	Jurkat	DB-Match	Co-Cluster
1981	0	<i>Dents Disease candidate gene (chloride channel 5)</i>	1
1293	0	<i>eyes absent homolog (Eab1)</i>	3
1184	0	<i>eyes absent homolog (Eab1)</i>	4
15	0	NKG2-D	357
55	3	HLA-A2 Klasse I-Antigen	95
14	0	RbAp48, <i>Retinoblastoma binding protein4</i>	368

Tab. 4-10 Differenziell exprimierte Gene der Zell-Linie NKL

Dargestellt sind einige Gene, die laut Co-Clustering-Analyse stärker oder ausschließlich differentiell in der Zell-Linie NKL exprimiert werden sowie die Cluster, die diese Gene repräsentieren.

Die Fragmente der ausgewählten Klone, die als DNA-Sonden in der Hybridisierung zum Einsatz kommen sollten, wurden aus den entsprechenden Bibliotheken isoliert, in einer PCR amplifiziert, aufgereinigt und anschließend sequenziert, um die Identität der Proben, die anhand des Datenbank-Matches der experimentellen Fingerprints der Klone ermittelt wurden, zu bestätigen. In der Regel wurden für diese Hybridisierungen die 1.Wahl-Klone des entsprechenden Co-Clusters verwendet. Das sind diejenigen Klone, die dem Consensus-Fingerprint am ähnlichsten sind.

4.3.3.2 Northern-Gelelektrophorese und Vacuum-Blotting

Für die nachfolgenden Northern-Hybridisierungen wurde es nun vorerst notwendig, die beiden Zell-Linien in Zellkultur zu vermehren, um die erforderliche Anzahl an Zellen für die Präparation einer adäquaten Menge an RNA zur Verfügung stellen zu können. Die Zell-Linien wurden aus Kryokulturen angeimpft und bei etwa 5×10^5 Zellen/ml kultiviert. Bei einer ausreichenden Zellzahl von 10^7 bis 10^8 wurde die Gesamt-RNA isoliert (Abb. 4-22).

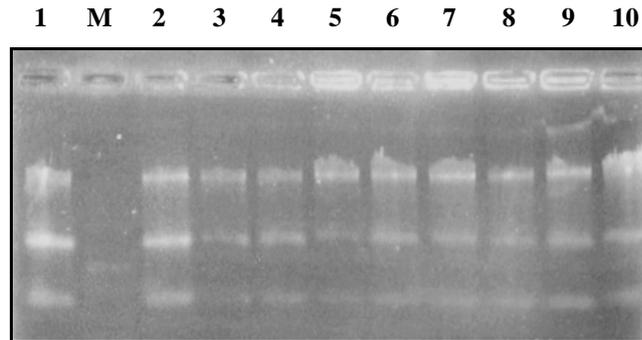


Abb. 4-22 Gelelektrophoretische Auftrennung der Gesamt-RNA für Northern-Blots

Isolierte Gesamt-RNA der Zell-Linie Jurkat für die Northern-Analysen gelelektrophoretisch aufgetrennt.

1 - 10 verschiedene Isolate der Gesamt-RNA

M RNA-Marker

Die RNA-Isolate wurden mit Hilfe des BioAnalyzers 2100 (Agilent) auf ihre Qualität hin untersucht und ihre Konzentrationen bestimmt. Diese Methode der Qualitäts- und Konzentrationsbestimmung mittels eines Chips beruht auf dem Prinzip der Kapillarelektrophorese, wobei die Detektion der RNA-Fragmente anhand eines in die RNA eingelagerten Fluoreszenzfarbstoffes erfolgt (s. Kapitel 3.10.2). Die Trennung der Fragmente erfolgt aufgrund eines angelegten elektrischen Feldes durch Mikrokapillar-Kanäle. Die RNA-Fragmente (ribosomale RNA und RNA-Transkripte verschiedener Größe) werden ihrer Masse entsprechend separiert und durch Fluoreszenz detektiert. Die Ergebnisse werden aufgezeichnet und mit der Agilent 2100 BioAnalyzer Software analysiert. Eine Übersicht über eine Analyse am Beispiel der isolierten RNA der Zell-Linie Jurkat gibt die Abb. 4-23. Der erste, kleinere Peak läuft bei etwa 35 sec und entspricht der 18S rRNA, der zweite, höhere Peak mit einer Retentionszeit von etwa 40 sec entspricht der 28S rRNA. Die Fläche unterhalb der Peaks wird zur Berechnung der RNA-Mengen herangezogen. Das Verhältnis der Konzentrationen der 28S rRNA zur 18S rRNA sollte bei etwa 2 liegen.

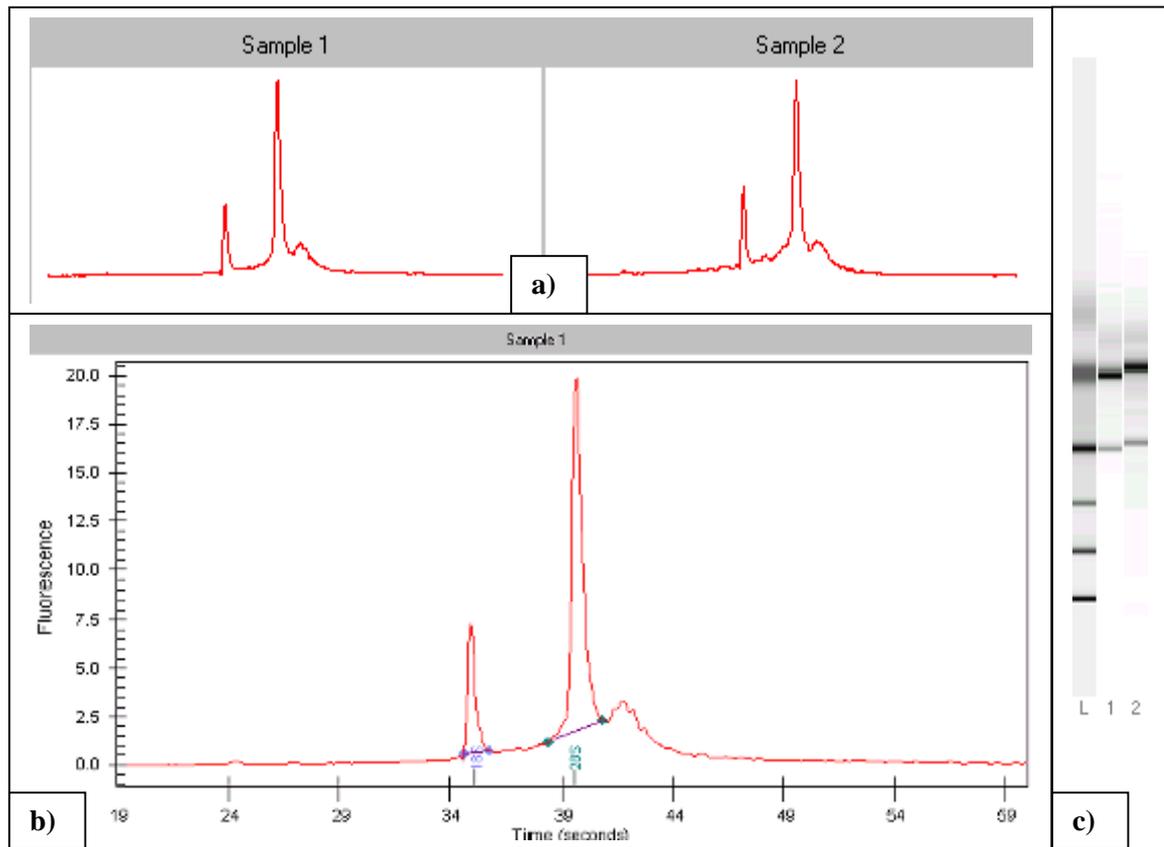


Abb. 4-23 Analyse der RNA am BioAnalyzer 2100

Die Auswertung der aus der Zell-Linie Jurkat isolierten Gesamt-RNA erfolgte am BioAnalyzer mit dem RNA 6000 LabChip® (Agilent).

- a) Übersicht der Isolate 1 und 2
- b) Detailansicht des Isolates 1 mit Retentionszeit (Time (seconds)) und Fluoreszenzintensität
- c) Gelbild beider Isolate im Vergleich zu einem Größenstandard
- L Größenstandard
- 1 Isolat 1
- 2 Isolat 2

Die RNA wurde anschließend in einem denaturierenden Agarosegel elektrophoretisch aufgetrennt. Dazu wurden paarweise 5 µg Gesamt-RNA per Slot aufgetragen. Um für die Hybridisierungen genügend Filter zur Verfügung zu stellen, wurden jeweils mehrere Aliquots der RNA der Zell-Linien Jurkat und NKL auf das Gel aufgetragen. Anschließend wurde die RNA aus dem Gel durch Vacuum-Blotting auf eine Nylon-Membran transferiert (Modell 785 VacuumBlotter, Bio-Rad) und durch UV-cross linking kovalent an die Membran gebunden. Insgesamt wurden vier Filter dieser Art erstellt. Der Transfer der RNA aus dem Gel auf die Membran durch Vacuum-Blotting hat den Vorteil, daß die relativ instabile RNA in kürzerer Zeit (90 min) und mit höherer Ausbeute als beim Kapillarblotting aus dem Gel auf eine Membran transferiert werden kann (Kroczeck, R. A. *et al.*, 1990).

Um Unterschiede in der Signalintensität nicht fälschlicherweise zurückzuführen auf Unterschiede in der Genexpression und Fehler auf Unterschiede nur allein aufgrund einer unterschiedlichen RNA-Konzentration auszuschließen, war es notwendig, eine Hybridisierung zur Normalisierung der Filter durchzuführen. Glycerinaldehyd-3-phosphat-dehydrogenase (GAPDH) ist ein *house keeping gene*, eines jener Gene, die in jeder Zelle aktiv sind und für grundlegende Proteine des Stoffwechsels und der Physiologie kodieren. Diese Gene werden in allen Zellen in etwa gleichem Maße exprimiert und sollten so auch bei Hybridisierungen gleich starke Signale geben. Deshalb wurde GAPDH als interner Standard zur Normalisierung der Filter untereinander eingesetzt. Unter bestimmten experimentellen Umständen, wenn ein Einfluß auf die Expression dieser *house keeping* Gene nicht ausgeschlossen werden kann, kann es sich auch als sinnvoll erweisen, zur Normalisierung von Northern-Blots oder cDNA-Arrays synthetische poly(A)-RNAs als externe Standards zu verwenden (Eickhoff, B. *et al.*, 1999). Da hier dagegen nicht davon ausgegangen werden muß, wurde sich für den Einsatz der GAPDH entschieden.

Bei der Hybridisierung mit der GAPDH als Sonde wurde jedoch festgestellt, daß nur auf einem der vier verwendeten Filter (Filter 4) die gleiche Menge an Jurkat- bzw. NKL-RNA zur Northern-Gelelektrophorese aufgetragen wurde. Die anderen drei Filter beinhalteten in der Jurkat-Spur nur etwa 30% der RNA-Menge im Vergleich zur NKL-Linie, was auf das Verwenden zweier unterschiedlicher Isolate der Jurkat-RNA sowie das Anwenden zweier verschiedener Konzentrations-Meßmethoden zurückzuführen ist. So wurde die NKL-RNA spektrophotometrisch, die Jurkat-RNA mittels BioAnalyzer analysiert. Anhand der Hybridisierung mit der GAPDH als Sonde läßt sich nun ein Faktor errechnen, der sich aus der unterschiedlichen Signalintensität ergibt und bei der Auswertung der Signalintensitäten der verschiedenen Northern-Hybridisierungen berücksichtigt werden muß. Die Hybridisierung mit dem *house keeping gene* GAPDH zur Normalisierung der Signalwerte ist in der Abb. 4-24 dargestellt.

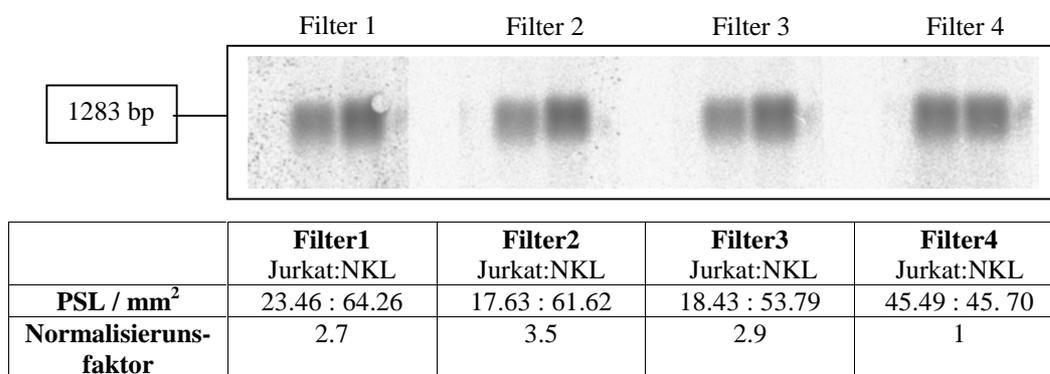
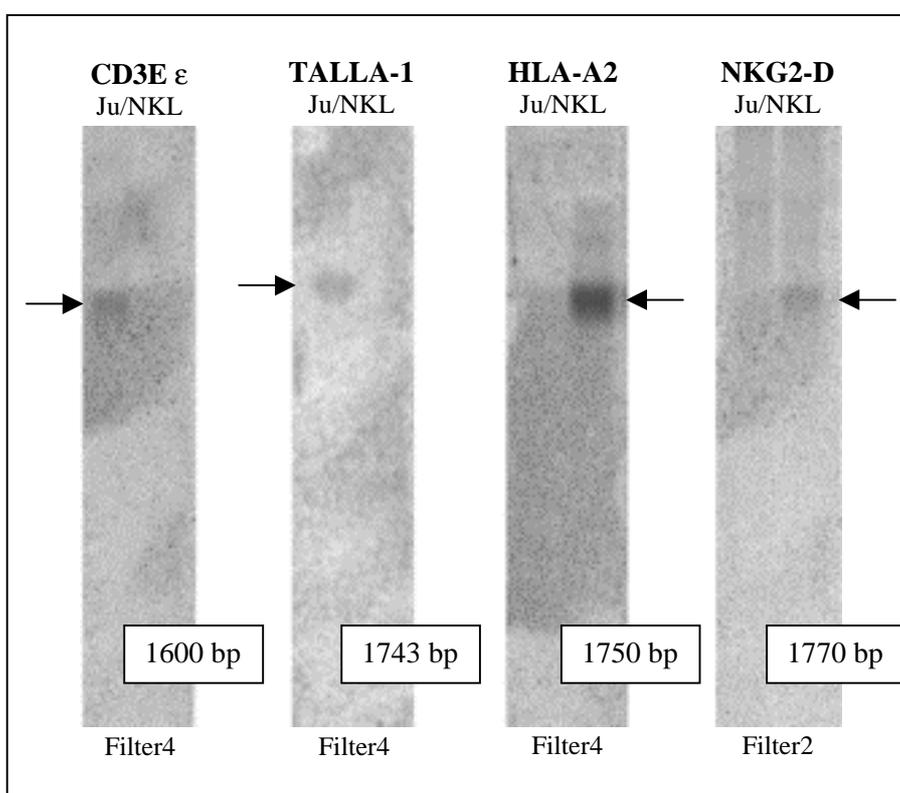


Abb. 4-24 Northern-Hybridisierung mit GAPDH zur Normalisierung

Hybridisierung mit GAPDH zur internen Normalisierung der Filter mit den jeweils entsprechenden Signalintensitäten und den sich daraus ergebenden Faktoren. Die jeweils linke Spur beinhaltet die Jurkat-RNA, die rechte Spur die NKL-RNA.

Die zu untersuchenden Proben wurden mittels *random primed labeling* radioaktiv markiert, über Sephadex-Säulen aufgereinigt und zur Hybridisierung eingesetzt. Die Hybridisierung erfolgte in Hybridisierungsflaschen über Nacht bei 42°C. Die Detektion der radioaktiven Signale erfolgte am BAS-1800 mittels Fuji-Screens (BAS-IP MS 2325, Kassette BAS 2325) mit der Ansteuersoftware BASReader 2.13h. Alle Messungen erfolgten jeweils nach 1-, 3- bzw. 16-stündiger Exposition der hybridisierten Membranen auf einem Fuji-Screen. Die Signalintensitäten wurden mit dem Bildanalyseprogramm Tina 2.10g ausgewertet. Es wurden die Mittelwerte dieser Messungen gebildet und zur Auswertung verwendet. In der Abb. 4-25 sind einige Beispiele für die Co-Clustering-Analysen, deren Angaben in Northern-Hybridisierungen bestätigt werden konnten, dargestellt.



	CD3E ε	NKG2-D	HLA-A2	TALLA-1
PSL / mm²	13.77 : 6.74	14.40 : 24.83	5.27 : 55.65	3.13 : 13.03
Normalisierungsfaktor	1	1	1	3.5

Abb. 4-25 Northern-Hybridisierung mit den Sonden für CD3E ε, TALLA-1, HLA-A2, NKG2-D

Die jeweils linke Spur beinhaltet die Jurkat-RNA, die rechte Spur die NKL-RNA. Die Werte für die Signalintensitäten sind jeweils nach der Normalisierung angegeben. Filter4 enthält einen relativ hohen lokalen Background gerade an den Stellen der Signale, so daß auch die angegebenen Hintergrundsignalwerte gegenüber etwas schwächeren Signalen relativ hoch, die eigentlichen Signale jedoch etwas geringer ausfallen und somit das Verhältnis beider zueinander verfälscht wird.

Die meisten Vorhersagen der Clustering-Analysen konnten in den Northern-Hybridisierungen bestätigt werden, vier von ihnen seien an dieser Stelle näher aufgeführt. Dabei handelt es sich um die Jurkat-spezifischen Gene CD3E epsilon Polypeptid (TiT3-Komplex) und TALLA-1 sowie die NKL-spezifischen Gene NKG2-D und HLA-A2.

CD3E epsilon Polypeptid (TiT3-Komplex)

Es ist bekannt, daß das CD3E-Gen (das durch das Co-Cluster 481 repräsentiert wird) nur spezifisch in Thymocyten und T-Zellen exprimiert wird, nicht dagegen in anderen Zellen, wie z.B. der NK-Zell-Linie NKL. Dieses Gen codiert für einen Oberflächenmarker, mit einem Molekulargewicht von 20 kD. Dieser ist mit dem T-Zell-Rezeptor (TCR) assoziiert und wird für die Oberflächenexpression des TCR und die Signal-Transduktion nach Antigen-Erkennung durch den TCR benötigt. Im Co-Clustering wurde das Verhältnis der Transkripte von Jurkat zu NKL mit 12 : 0 bestimmt. D.h., es erfolgt die Expression in Jurkat, nicht aber in NKL. Diese ausschließliche Expression in der T-Zell-Linie Jurkat konnte auch anhand der Hybridisierung bestätigt werden.

NKG2-D

Ein Natural Killer-Zell spezifisches Gen ist NKG2-D (Co-Cluster 357). Es spielt als integrales Membranprotein eine Rolle als ein Rezeptor in der Erkennung der MHC I HLA-E Moleküle durch die NK-Zellen. Das Vorkommen dieses Gens in den beiden Zell-Linien wurde mit einem Verhältnis von Jurkat zu NKL mit 0 : 15 festgestellt. Bei der Hybridisierung dieses Gens konnte diese NK-Zell spezifische Expression auch hier bestätigt werden.

TALLA-1

Ein weiteres T-Zell spezifisches Gen ist TALLA-1 (Cluster 469), das *T-cell acute lymphoblastic leukemia associated antigen 1*, ein integrales Transmembranprotein, das einen hoch spezifischen Oberflächenmarker darstellt. Dieses Gen wird in der akuten T-Zell-Leukämie sowie in Neuroblastomen exprimiert. Auch hier konnte eine Expression dieses Gens nur in Jurkat, nicht jedoch in NKL nachgewiesen werden, was die Voraussage durch die Co-Clustering-Analyse mit einer Expression von 12 : 0 bestätigt.

HLA-A2

HLA-A2 (Co-Cluster 95) ist weder ein T- noch NK-Zell spezifischer Marker. Es ist ein Histokompatibilitäts-Antigen der Lymphocyten (MHC Klasse I), der auf allen kernhaltigen Zellen exprimiert wird und an Abstoßungs- und Abwehrreaktion von fremden oder nicht eigenem Gewebe und virusinfizierten Zellen beteiligt ist. T-Zellen exprimieren auf ihrer Zell-Oberfläche entweder MHC der Klasse I oder II. In der Northern-Hybridisierung konnte das Ergebnis des Co-Clusterings,

wonach HLA-A2 (in einem Verhältnis von 3 : 55) in NKL und nicht in Jurkat exprimiert wird, auch nachgewiesen werden.

Neben den eben erwähnten Beispielen für differentiell exprimierte Cluster, die in den Northern-Hybridisierungen bestätigt werden konnten, gibt jedoch auch Beispiele für einige Gene, deren differentielle Expression in den Northern-Hybridisierungen nicht bestätigt werden konnten. Dazu zählen beispielsweise L-Plastin, *insulinoma rig analog mRNA (encoding DNA binding protein)* sowie das T-Zell Cyclophilin.

L-Plastin

Ein Beispiel für die Fälle, in denen das Co-Clustering-Ergebnis nicht betätigt werden konnte, ist z.B. das L-Plastin (Co-Cluster 187), einem Actin-bindenden Protein, welches in zwei Isoformen (L und T) vorkommt. Die L-Form, das L-Plastin, wird nur exprimiert in Zellen der hämatopoetischen Linie. Auch wurde es gefunden in vielen Typen maligner humaner Zellen nicht hämatopoetischen Ursprungs, was darauf hinweist, daß seine Expression auch während der Tumorentstehung solider Gewebe induziert ist. Laut Co-Clustering sollte dieses Gen mit einer Tendenz von 26 : 0 ausschließlich in Jurkat exprimiert werden. Festgestellt hingegen wurde, daß das L-Plastin in beiden Zell-Linien fast in gleichem Maße (etwa 0.85 : 1) exprimiert wird.

insulinoma rig analog mRNA, encoding DNA binding protein

Die *insulinoma rig analog mRNA* (Co-Cluster 156) codiert für ein ribosomales Protein, welches eine Komponente der 40S Untereinheit ist. Es gehört der S19P-Familie der ribosomalen Proteine an und ist in verschiedenen Tumoren wie dem Insulinom, dem Ösophagus- oder dem Colonkarzinomen aktiviert. In den hier durchgeführten Northern-Hybridisierungen konnte die Voraussage / Tendenz von 30 : 2, also einer stärkeren Expression in Jurkat, nicht oder nur gering in NKL, jedoch nicht bestätigt werden. Es wurde lediglich eine in etwa gleich starke Expression dieses Gens in beiden Zell-Linien gefunden.

Insgesamt wurden 18 lt. Co-Clustering-Analyse differentiell exprimierte Gene als Sonde hybridisiert. Davon konnte die Vorhersage bei 3 Genen nicht bestätigt werden, 2 von diesen zeigten sogar ein eher umgekehrtes Verhältnis der Expressions-Häufigkeiten, d.h., jeweils die andere Zell-Linie exprimiert das jeweilige Gen höher als eigentlich durch das Co-Clustering vorhergesagt. Im Allgemeinen entsprechen jedoch die Angaben des Co-Clusterings in etwa dem wirklichen Expressionsmuster in der Zelle, zumindest in der Tendenz. Jedoch das Verhältnis, welches angegeben wird, entspricht nicht unbedingt den in den Hybridisierungen erhaltenen Signalintensitäten. Bei einer Clustergröße von kleiner als 20 ist die Grenze der Sensitivität dieser ONF-Methode erreicht. Bei kleineren Clustern kann es zu Ungenauigkeiten und damit zu

Falschaussagen bezüglich der Expressionshöhe eines Gens oder der Signifikanz der differentiellen Expression kommen. Jedoch ist es möglich, Aussagen darüber zu treffen, ob ein Gen in der jeweiligen Bibliothek vorhanden ist oder nicht repräsentiert wird. Um Voraussagen treffen zu können über die reale Expression eines Gens in einer Zelle, ist es immer notwendig, diese auch in einer entsprechenden anderen Methode, wie z.B. der hier genutzten Northern-Hybridisierung, zu bestätigen.

4.3.4 Rearranging

Es gibt mehrere Anwendungen, die das neue Anordnen einiger Klone aus der großen Bibliothek in kleinere Subsets notwendig werden lassen, wenn beispielsweise das Interesse nur auf einige Klone der Bibliothek fällt, so zur Sequenzanalyse, zur Überprüfung der Bibliotheksqualität und der Validierung des Datenbank-Matches oder auch zum Erstellen eines Unigene-Sets. Dafür werden die Klone zufällig ausgewählt oder nach bestimmten Parametern selektiert und mit Hilfe eines Roboters aus der Kulturplatte gepickt und in eine neue MTP inokuliert. Dieser Vorgang wird als *rearranging* bezeichnet.

Eine weitere Möglichkeit der Anwendung des Rearranging besteht darin, all die Klone/Cluster zu selektieren und neu anzuordnen, die nach den Datenbankanalysen anhand des Consensus-Fingerprints hohe bzw. keine Homologie zu schon bekannten Sequenzen aufweisen. So können die schon bekannten Klone/Cluster weiteren Experimenten zugeführt werden. Bei den Klonen, die anhand des Consensus-Fingerprints keine Homologien aufwiesen, kann die genaue Sequenz durch die herkömmliche Sequenzierung ermittelt werden und, sofern sie in der Datenbank bekannt ist, mittels BLAST-Analysen auch identifiziert werden. Sind Klone in der Datenbank noch nicht bekannt, können sie als neue Gene in weiteren Experimenten näher charakterisiert werden.

4.3.4.1 Erstellen eines Unigene-Sets

In einer allgemeinen cDNA-Bibliothek ist das gesamte Spektrum aller in der Zelle oder einem Gewebe vorkommenden mRNAs etwa im Verhältnis ihrer Expression enthalten. Stark exprimierte Gene sind demnach häufiger in der Bibliothek vertreten (und bilden größere Cluster) als nur schwach exprimierte. Um nun aber für zukünftige Experimente, wie der Komplex-Hybridisierung, den Analyseaufwand so gering wie möglich zu halten, ist es ausreichend, daß jeweils nur ein Vertreter dieser Klone -jedes ein Gen repräsentierendes cDNA-Fragment- im Idealfall nur einmal vorkommt. Solch eine Bibliothek nicht redundanter cDNA nennt man ein Unigene-Set. Bei einer Hybridisierung wird hier nun (nicht mehr wie bei der Longprobe-Hybridisierung der cDNA-Filter)

die Häufigkeit der Expression eines Gens an der Häufigkeit der Signale auf dem Filter gemessen, sondern an der Signalintensität des betreffenden Spots.

Bezugnehmend auf die Ergebnisse des Clusterings einer Bibliothek wird im Allgemeinen der Consensus-Klon (1.Wahl-Klon) eines interessierenden Clusters zur Erstellung des Unigene-Sets selektiert. Der Consensus-Klon ist derjenige Klon eines Clusters, der dem Consensus-Fingerprint des entsprechenden Cluster am ähnlichsten ist. Die Consensus-Klone aller Cluster werden mit Hilfe eines Rearranging-Roboters aus frisch replizierten und über Nacht gewachsenen Kulturen der Bibliothek gepickt und neu in 384-well MTP angeordnet. Auf diese Art wird die ursprüngliche Bibliothek normalisiert und damit verkleinert. Nur ein repräsentativer Klon eines Clusters wird nun als Vertreter seiner RNA-Species gespottet und in weiteren Experimenten untersucht. Zu weiteren Analysen werden die Klone in einer PCR amplifiziert und für Hybridisierungen auf Nylon-Membranen, oder in neueren Verfahren auch auf Glas-Slides, auf denen eine noch höhere Dichte der Spots erreicht werden kann, gespottet (Eickhoff, H. *et al.*, 2000).

4.3.4.2 Erstellen und Spotten eines Jurkat-Unigene-Sets

Das Unigene-Set der Zell-Linie Jurkat mit jeweils dem repräsentativen Consensus-Klon eines jeden Clusters, den sogenannten 1.Wahl-Klonen der Cluster (2.230) und eines jeden Singletons besteht aus 10.506 Klone (in 28 MTP) aus ursprünglich insgesamt über 23.200 Klone. Das entspricht etwa 44.4% der Original-Bibliothek, die somit 2.25-fach normalisiert ist. Gespottet wurden jedoch nicht alle Klone des eigentlichen Unigene-Sets, sondern nur 3.396 Klone (in 10 MTP). Darin enthalten sind Vertreter aller Cluster (2.230 in 6 MTP) sowie Singletons mit einem signifikanten Datenbank-Match (e-Value $<10^{-6}$ beim Datenbankabgleich der experimentellen mit den theoretischen Fingerprints) (1.166 Klone in 4 MTP). Das hat den Vorteil, daß all diese Klone (Set) in ein Feld gespottet werden können und je Block noch 2 Positionen als Negativ-Kontrolle frei bleiben. Gespottet wurde das Unigene-Set nach den gleichen Parametern wie die Jurkat- bzw. NKL-Bibliotheken, wiederum in einem Pattern von 5x5 in Duplikaten mittels eines 250 µm Pin-Tools (10x per Spot). Die Position des Guide Dots wurde diesmal jedoch nicht gespottet. Um einen Überblick über die Quantität der amplifizierten Targets sowie die Qualität des Transfers auf die Membran zu erhalten, wurden als erstes Background-Hybridisierungen mit einem Primer durchgeführt, der zuvor zur Amplifizierung der Bibliothek verwendet wurde (NST-rev), so daß alle auf dem Filter gespotteten Klone sichtbar wurden. Diese Hybridisierung ist in der Abb. 4-26 am Beispiel eines Filters dargestellt.

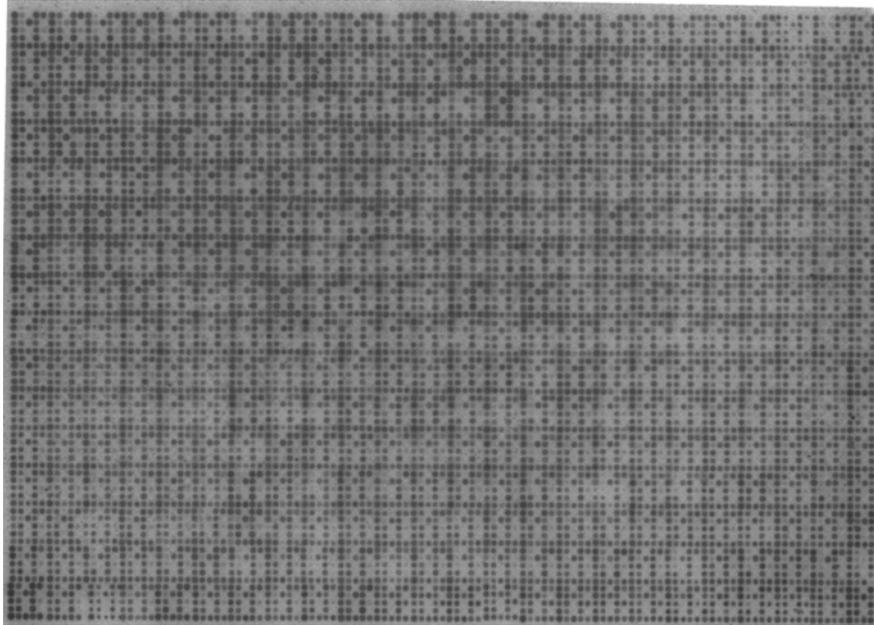


Abb. 4-26 Background-Hybridisierung des Jurkat Unigene-Sets

Background-Hybridisierung mit einem zur Amplifizierung genutzten Primer zur Qualitätskontrolle des Filters des Jurkat-Unigene-Sets. Alle auf dem Filter gespotteten Positionen werden dadurch sichtbar.

Das Jurkat-Unigene-Set wurde aus einer T-Lymphocyten-Linie erstellt und stellt damit ein T-Zell spezifisches Unigene-Set dar. Somit können auch solche Gene, die sonst in anderen Zellen, Geweben oder Organen nicht oder nur in geringem Ausmaß zur Expression gelangen, besonders Gene immunologischer Relevanz in dieser Bibliothek repräsentiert sein. Da die mRNA aus einer etablierten Zell-Linie isoliert wurde, die ursprünglich von einem Leukämiepatienten stammt, sind in dieser Bibliothek auch solche Gene exprimiert -oder auch stärker exprimiert bzw. inhibiert-, die in die Pathogenese von Krebs, insbesondere der Leukämien, involviert sein können.

Die entsprechenden Daten zur Erstellung eines Unigene-Sets der NKL-Zell-Linie liegen ebenfalls vor, so daß auch dieses Set jederzeit gespottet werden kann. Das Unigene-Set der Bibliothek NKL besteht aus insgesamt 13.094 Klonen, Repräsentativen von 2.169 Clustern einer Mindestgröße von 2 sowie 10.925 Singletons. Das entspricht einer 2.7-fachen Normalisierung der ursprünglichen Bibliothek.

Mit diesen beiden Unigene-Sets stehen Tools zweier Leukämie-Zell-Linien, die Zellen der hämatopoetischen Zellreihe sind, zur Verfügung, die in weiteren Untersuchungen detailliertere Einblicke in Krebsentstehung und Proliferation geben können. Besonders im Fall der NKL-Bibliothek könnten Untersuchungen in näherer Zukunft dazu beitragen, genauere Einblicke in die Natur der NK-Zellen zu erhalten, über die bisher weit weniger bekannt ist, als über ihre nah verwandten B- oder T-Lymphocyten.

4.3.5 Komplex-Hybridisierung

Die Komplex-Hybridisierung ist eine Art der Hybridisierung, bei der keine spezifischen Nucleinsäuren, sondern ein Gemisch aus vielen, nicht definierten Gesamt-RNA- oder mRNA-Fragmenten (RNA-Pool in cDNA revers transkribiert) auf Filter hybridisiert wird (daher komplex). Diese Art der Hybridisierung verwendet man, um z.B. verschiedene Entwicklungsstadien eines Organismus, gesunde oder erkrankte Gewebe hinsichtlich ihrer Genexpression direkt miteinander vergleichen und so gewebe- oder krankheitsspezifische Gene identifizieren zu können. Dabei erwartet man nicht nur einzelne positive Signale, sondern Signale in fast jedem, ein Gen repräsentierenden Spot auf dem Filter, jedoch beim Vergleich der Hybridisierungsmuster der einzelnen Gewebe unterschiedlich starke Signalintensitäten. Die Voraussetzung für diese Hybridisierung ist eine möglichst nicht redundante Bibliothek, das sogenannte Unigene-Set, in dem jedes Gen nur einmal präsent und möglichst auch identifiziert ist.

Im Rahmen dieser Arbeit wurden bislang einige Vorexperimente durchgeführt. So wurde das Jurkat Unigene-Set auf seine Qualität und Funktionsfähigkeit hin überprüft. Es ist bisher im Rahmen dieser Arbeit jedoch keine wissenschaftliche Anwendung damit verbunden. Diese Untersuchungen mit dem Jurkat Unigene-Set dienen exemplarisch als Modell für beide erstellten Sets und sollen ihre Anwendbarkeit und Funktionsfähigkeit demonstrieren.

Als erste Komplex-Hybridisierung wurde eine Hybridisierung mit radioaktiv markierter Jurkat-RNA durchgeführt, die gleichzeitig auch als Positiv-Kontrolle dient. Da alle Klone dieses Unigene-Sets aus der Jurkat-Bibliothek stammen, müssen bei einer Hybridisierung mit einer RNA-Probe dieser Zell-Linie auch alle Klone ein Signal geben, in einem in der Intensität der Signale für die Zelle spezifischen Muster. Die Detektion der radioaktiven Signale erfolgte am BAS-1800 mittels Fuji-Screens (BAS-IP MS 2325) mit der Ansteuerungssoftware BASReader 2.13h. Die Signale wurden ausgewertet mit Visual Grid[®], Version 3.4.1.921.

Der Nutzen dieser Komplex-Hybridisierungen liegt klar darin begründet, Veränderungen nicht nur einer bestimmten DNA oder RNA, sondern eines breiten Spektrums an exprimierten, nicht genau definierten RNAs in einem einzigen Experiment nachweisen zu können.

In weiteren Untersuchungen können nun beispielsweise aus verschiedenen Patientengruppen isolierte mRNAs als Proben hybridisiert und die Genexpressionsmuster direkt miteinander verglichen werden. Solche Patientengruppen können beispielsweise solche unterschiedlicher Diagnose sein oder solche, die unterschiedliche Therapien erhalten haben. So können einige Experimente wichtige Hinweise auf neue Therapieansätze liefern und dazu dienen, Targets für neue Medikamente zu finden. Andere Analysen, wie der Vergleich der Lymphocyten-Populationen vor und nach Therapien und Anwendungen, können helfen, Einblicke in die Wirkungsweise der

Medikamente zu erhalten und das Problem der Medikamentenresistenz molekulargenetisch näher zu charakterisieren.

Der Nachweis minimal residualer Zellen nach einer Therapie anhand ihrer spezifischen Oberflächenmarker könnte einen wesentlichen Beitrag zur Remissionserhaltung und -überwachung leisten, die vorhandenen Therapien verbessern und damit die Heilungs- und Überlebenschancen der Patienten erhöhen.