

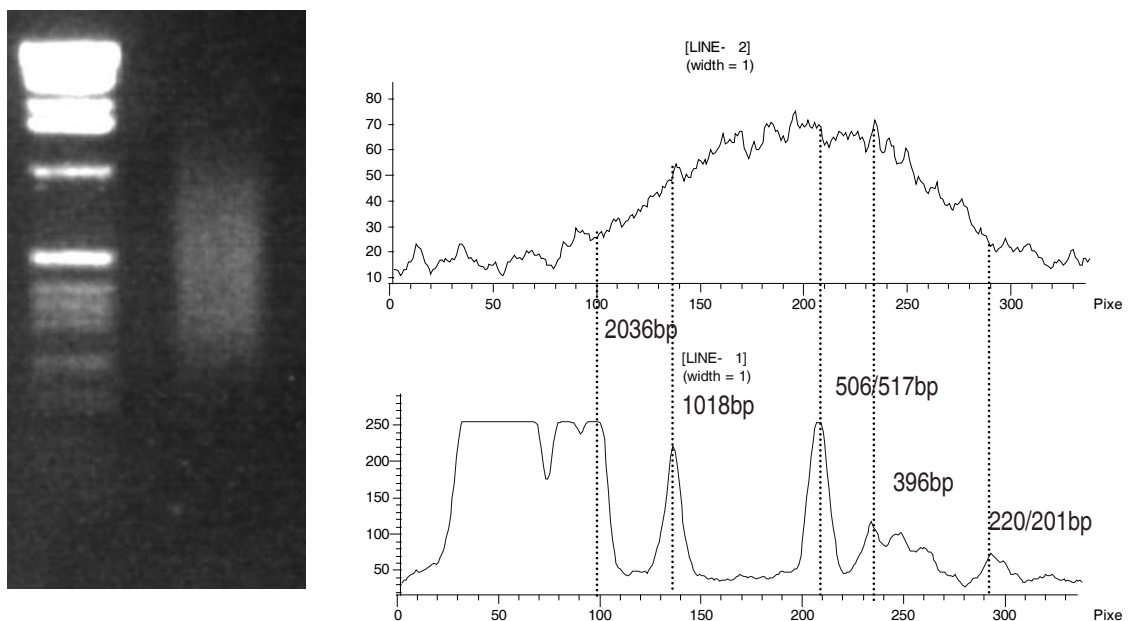
## 4 Ergebnisse

### 4.1 Grundlagen

#### 4.1.1 Zusammensetzung der Probe

##### 4.1.1.1 Längenverteilung der IA- PCR- Produkte

Das Inter- Alu- Amplikon ist das PCR- Produkt von genomischer DNS mit dem CL-2- Primer. Es umfaßt eine große Anzahl verschieden langer PCR- Produkte. Um das Größenspektrum experimentell zu erfassen, wurde ein Aliquot der PCR- Reaktion in einem Agarosegel aufgetrennt und das erhaltene Bild mittels ImageQuant quantifiziert. Das Ergebnis war keine distinkte, sondern eine kontinuierliche Bande über einen weiten Größenbereich.



**Abbildung 4-1:** Darstellung der Längenverteilung eines IA- Amplikons: links Agarosegel mit Größenstandard und dem Inter- Alu- Schmier; rechts oben Intensitätsverteilung des Amplikons nach der Digitalisierung, unten Bandenmuster des zugehörigen Größenstandards, die entsprechenden Längen sind angegeben, das Größenspektrum des IA- Amplikons beginnt bei etwa 200bp und endet bei etwa 2.500bp, das Maximum liegt bei etwa 500bp, längere Fragmente sind überrepräsentiert, da sie pro Molekül mehr Ethidiumbromid binden können

#### 4.1.1.2 Klonierung von YAC- IA- Amplikons – Bestimmung des Anreicherungs-faktors für die Inter- Alu- PCR

In den Experimenten zu Beginn der Arbeit wurden die Sonden für den Chip durch Inter- Alu- PCR auf bereits kartierten YACs erzeugt, um somit deren bekannte Lokalisation auf die IA- PCR- Klone übertragen zu können (siehe 3.4.4). Dazu wurde Gesamt- DNS aus Hefeklonen isoliert und als Ausgangsmaterial für eine Inter- Alu- PCR eingesetzt. Das PCR- Produkt wurde in einen Vektor kloniert und das Konstrukt in *E.coli* transformiert. Nach dem Ausplattieren wurden pro YAC zwölf Klone selektiert, mit denen Kolonie- PCR mit dem CL-2- Primer durchgeführt wurde. Die erzeugten PCR- Produkte wurden auf ein Agarosegel aufgetragen und durch Vergleich der Fragmentlängen wurde die Anzahl verschiedener Inter- Alu- PCR- Produkte bestimmt. Die durchschnittliche Länge dieser YACs betrug etwa ein Mbp. Insgesamt wurden auf diese Art 116 YACs ausgewertet. Zusammen ergaben sich 556 verschiedene Klone, was etwa fünf IA- PCR- Produkten pro YAC entspricht. Dieser empirisch ermittelte Wert diente als Grundlage für die folgenden Berechnungen. Wenn die Größe des humanen Genoms mit 3.000 Mbp angenommen wird, so ergibt sich daraus eine Zahl von 15.000 Inter- Alu- Fragmenten im amplifizierbaren Bereich. Unter der Annahme, daß die Länge von distinktes IA- PCR- Produkten im genomischen Amplikon im Mittel 500 bp beträgt (siehe 4.1.1.1), ergibt sich eine Verringerung der Komplexität von 3.000 Mbp zu 7,5 Mbp, d.h. um den Faktor 400.

#### 4.1.1.3 Erstellung einer gesamtgenomischen IA- Amplikonbank – Bestimmung der Komplexität des genomischen IA- PCR- Produktes

Aufgrund der Wichtigkeit zuverlässiger Angaben zur Komplexität genomischer IA- PCR- Amplikons wurde das PCR- Produkt zur Herstellung einer Plasmidbank herangezogen. Genomische DNS von fünf verschiedenen Spendern (3x männlich, 2x weiblich) wurde zur Generierung von fünf Inter- Alu- Amplikons verwendet. Die PCR- Produkte wurden zunächst in einem Agarosegel aufgetrennt, extrahiert und anschließend zu gleichen Teilen vereinigt, um eine gleichmäßige Repräsentation der Spender- DNS im Amplikon zu gewährleisten. Danach wurde das IA- PCR- Produkt kloniert und anschließend transformiert. Der Ansatz wurde auf 15 Agarplatten (22x 22cm) verteilt und die erhaltenen Klone dann durch einen Roboter in 81 384-er

Mikrotiterplatten mit LB- Medium überführt. Alle Klone wurden danach durch IA-PCR reamplifiziert.

Die PCR- Produkte einer Platte wurden ansequenziert. Dazu wurde die Plasmid- DNS präpariert und mit dem zum Vektor passenden sp6- Primer nach dem Dye- Primer-Verfahren sequenziert. Die durchschnittliche Länge der durch dieses Verfahren erzielbaren auswertbaren Sequenzen lag bei ungefähr 500bp, also deutlich unter den maximal möglichen Insertlängen. Anschließend wurde die Komplexität der Bank bestimmt. Dazu wurden in den erhaltenen Sequenz- Rohdaten zunächst die noch vorhandene Vektorsequenz ermittelt und diese maskiert. Klone ohne Vektorsequenz wurden von der weiteren Auswertung ausgeschlossen, um zu verhindern, daß DNS-Kontaminationen in die weitere Auswertung einfließen. Weiterhin wurden Klone mit langen internen poly(A)- Sequenzen verworfen, da diese oft zu vorzeitigen Abbrüchen der Sequenzierungsreaktion führen. Die verbliebenen Sequenzen wurden mit dem Programm GAP4 aus dem Staden- Software- Paket verglichen, um Überlappungen zu finden. Identische Klone sollten zumindest teilweise überlappen. Durch repetitive Elemente verursachte Überlappungen werden von dem Programm erkannt und maskiert. Es bleibt anzumerken, daß sehr lange, doppelt auftretende, identische PCR- Produkte mit einer Wahrscheinlichkeit von 50% nicht erkannt werden, wenn die Insertlänge größer als die Länge zweier Rohsequenzen und die Orientierung des PCR- Produkts im Vektor entgegengesetzt ist. Dieser Spezialfall wurde nicht weiter untersucht. Zusammengefaßt ergaben Sequenzierung und Suche nach überlappenden Klonen folgende Ergebnisse:

sequenzierte Klone:	384
davon mit Vektorsequenz:	238
Sequenz über poly(A) Bereich:	214
einzelne auftretende Klone:	195x1
doppelt auftretende Klone:	8x2
dreifach auftretende Klone:	1x3
Anteil einzelner Klone:	91,21%
Anteil mehrfacher Klone	8,88%

Ziel dieser Untersuchung war es, abzuschätzen, inwieweit einzelne Klone präferentiell in der Inter- Alu- PCR amplifiziert werden bzw. ob die ursprüngliche Komplexität der Probe während der PCR erhalten bleibt. Dazu benötigt man einen Vergleichswert, der angibt, wieviele Mehrfach- Klone bei entsprechender Stichprobengröße bei idealen Bedingungen zu erwarten wären. Dieser Sachverhalt lässt sich aufgrund der vorliegenden Polynomialverteilung nur schwer exakt in einer Gleichung formulieren. Da die relativ kleine Stichprobengröße auch nur ein Ergebnis mit einem eher weiten Konfidenzintervall liefern würde, wurde die erwartete Häufigkeit mehrfach auftretender Klone stark vereinfacht entsprechend der folgenden Gleichung abgeschätzt:

$$\begin{aligned}
 & \text{Erwartete Häufigkeit identischer Klone in \%} = \\
 = & \frac{\text{Anzahl der Klone in der Bank}}{\text{Anzahl möglicher Klone im Genom}} \cdot \frac{\text{Anzahl der ausgewerteten Klone}}{\text{Anzahl der Klone in der Bank}} \cdot 100\% \\
 = & \frac{\text{Anzahl der ausgewerteten Klone}}{\text{Anzahl möglicher Klone im Genom}} \cdot 100\% \\
 = & \frac{214}{15.000} \cdot 100\% \\
 = & 1,43 \%
 \end{aligned}$$

Die wiederholte Simulation des Vorgangs per Computer bestätigte das erhaltene Ergebnis. Unter den gemachten Annahmen ergibt sich somit ein idealer Wert von 1,43% für den Anteil an mehrfach auftretenden Klonen. Der tatsächlich gefundene

Anteil ist mit 8,88% etwa 6x so hoch. Das bedeutet, es gibt präferentielle Amplifikation einzelner Sequenzen, die davon betroffenen Klone machen zusammen aber weniger als 10% des Amplikons aus. Dieser Wert läßt darauf schließen, daß die Komplexität des Inter- Alu- PCR- Produktes während der Amplifikation fast vollständig erhalten bleibt und somit ausreichend repräsentativ für die Ausgangs- DNS ist.

#### 4.1.1.4 Nachweis der Sequenzspezifität des CL-2- Primers

Die Sequenzen der Alu- Elemente sind nicht absolut identisch. Um möglichst viele verschiedene IA- PCR- Klone zu erhalten, müssen die PCR- Bedingungen so gewählt werden, daß der verwendete Primer auch bei nicht vollständig exakter Basenpaarung an die Alu- Sequenz bindet. Die Inter- Alu- PCR wurde deshalb bei 55°C *annealing*-Temperatur durchgeführt. Da der verwendete CL-2- Primer eine optimale *annealing*-Temperatur von 60°C hat, waren die gewählten PCR- Bedingungen somit nicht stringent. Für die Charakterisierung des Amplikons wurde die Anzahl der Fehlpaarungen zwischen Primer und Erkennungssequenz bestimmt, die eine PCR- Amplifikation noch zuließen. Zunächst wurden die experimentell ermittelten Sequenzen der IA- PCR- Klone mit bekannten Datenbanksequenzen verglichen. Dazu wurde der BLAST- Algorithmus (Altschul *et al.*, 1990) verwendet. Für 22 Klone wurden die entsprechenden Datenbanksequenzen ermittelt. Die in diesen Sequenzen enthaltenen Alu- Elemente stellen die Erkennungssequenz für den CL-2- Primer dar. Die Sequenz des Primers wurde dann Base für Base mit der Erkennungssequenz aus der Datenbanksequenz verglichen und erfaßt (siehe Tabelle 4.1). Für alle aufgefundenen Alu- Elemente konnte in der Datenbanksequenz ein zweites Alu- Element in 3'-3' Orientierung in maximal 2.000 bp Entfernung nachgewiesen werden. Auch diese Sequenz wurde wieder Base für Base mit der revers- komplementären Sequenz des CL-2- Primers verglichen. Die folgende Tabelle enthält die Zusammenstellung der Ergebnisse. Die Abweichungen von der CL-2- Sequenz sind in Großbuchstaben hervorgehoben. Insertionen/ Deletionen sind durch Klammern gekennzeichnet.

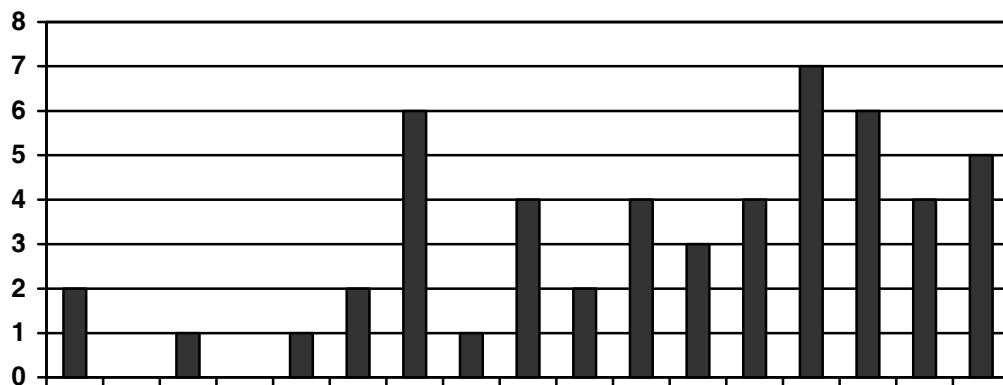
**Tabelle 4-1:** Zusammenstellung der experimentell gefundenen Primerbindungssequenzen, BAC bezeichnet die korrespondierende Datenbanksequenz.

Klon	BAC	5'-Primer	3'-Primer
a1d02	Ac007637	Ctgcactccagcctggg	cccaggctggagtgcag
a1c06	Ac005598	Ctgcactccagcctggg	cccaggctggagtgcag
a4h05	Af064860	Ctgcactccagcctggg	cccaggctggagtgcaA
a1a09	Ac004883	Ctgcactccagcctggg	cccaggGtggagtgAGg
a4h03	Al109627	Ctgcactccagcctggg	cccaggctggagtgcag
A4g03	Ac006531	CtgcacCtccagcctAgg	cccaggctggGTtgcag
A4c08	Al035079	CtgcactccagTctggg	cccaggctggagtCcag
A3h01	Ac006449	Ctgcagtccagcctggg	cccaggctggagtgcag
A3g03	Ac006123	CtgTGactccTgcctggg	cccaggTtggagtgcag
A3d05	Al022101	Ctgcactccagcctggg	cccaggctggagtgcag
A3a03	u91323	CtgcacCcTagcctggg	cccaggTtggagtgcag
A2h08	Af156545	CtCAaTctccagcctggg	CccaggctggagtgTaA
A2h07	Af135026	ctGact (C) cagcctggg	cccagAActCTCTCtg (C) a
A2e09	m27149	<sup>2</sup> ?tggg	Cccaggctggagtgcag
A2c09	Ac007786	Ctgcactccagcctggg	Tccaggctggagtgcag
A2b12	Ap000189	CGgcGttccagcctggg	CccaAgctggagtgcag
A2c02	Ac004134	CtgcactccagcctggA	CccaggctAgagtgcag
A2b03	Al021938	AtgGaGAccagcctggg	Cccagactggagtgcag
A1h09	Ac002997	Ctgcactccagcctggg	CccaggctAgagtgcaA
A1g12	Ac007193	Ctgcactccagcctggg	CccaggctGgaTGAAct
A1f02	Ac004083	Ctgcactccagcctggg	Cccaggctggagtgcag
A1d05	Ac005067	CtgcactccagccGggg	CccaggAtggagtAcag

---

<sup>2</sup> Ende der Datenbanksequenz

**Abbildung 4-2:** Verteilung der Häufigkeit der tolerierten Fehlpaarungen im CL-2- Primer in Abhängigkeit von der Position (3'→5')



Pos.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Anzahl:	2	0	1	0	1	2	6	1	4	2	4	3	4	7	6	4	5

19 von 44 gefundenen Primersequenzen zeigen keinerlei Fehlpaarungen, andererseits wurden aber auch PCR- Produkte gefunden, die eine Fehlpaarung am 3'-Ende des Primers enthalten. Insgesamt wurden 52 Mutationen in 44 Sequenzen gefunden. Eine Aufsummierung der Fehlpaarungen nach der Position bestätigt die Annahme, daß Fehlpaarungen am 5'-Ende des Primers sehr viel häufiger toleriert werden als solche am 3'- Ende. Bei beiden tolerierten 3'- Abweichungen wurde ein T an Stelle eines C zugelassen.

## 4.1.2 Filterhybridisierungen

### 4.1.2.1 Hybridisierung der Filter mit Inter- Alu- PCR- Produkten von monochromosomalen Hybridzelllinien

Monochromosomale Zelllinien sind ein wichtiges Hilfsmittel zur Kartierung von genomischen Klonen. Sie enthalten neben einem kompletten Maus- oder Hamstergenom noch ein humanes Chromosom, das einen Selektionsmarker trägt. Auch als Ausgangsmaterial zur Generierung von IA- PCR- Produkten sind monochromosomale Hybridzelllinien seit längerem bekannt (Ledbetter, S., *et al*, 1990). Man kann mit ihnen so auf einfache Weise chromosomenspezifische Sonden für Hybridisierungen erzeugen.

Als erster Schritt zur Kartierung der Klone des IA- Amplikons sollte deshalb auch eine Hybridisierung der Filter mit einem radioaktiv markierten Inter- Alu- PCR- Produkt von einer solchen Zelllinien- DNS erfolgen, um die durch die Spots repräsentierten Klone einem Chromosom zuordnen zu können. Für die Hybridisierung wurden 200 ng IA- PCR- Produkt (Chromosomen X, 10, 13, 22) radioaktiv markiert und auf die Filter unter immer stringenteren Bedingungen hybridisiert. Die Ergebnisse dieser Hybridisierungen ergaben immer wieder das gleiche Bild: alle humanen genomischen Klone, inklusive der Kontrollen, ergaben sehr starke Signale. Dies ist bemerkenswert, da vergleichsweise drastische Hybridisierungs- und Waschbedingungen verwendet wurden: Präassoziierung mit 20 µg c<sub>0</sub>t1- DNS, 2µg CL-2- Primer und/oder 8 µg eines PCR- Produktes, welches spezifisch das 3'- Ende der Alu- Sequenz abdeckt. Auch noch nach Hybridisierungen bei 70°C und Waschttemperaturen von 80°C ergaben alle genomischen Alu- Klone ein Signal. Die Vermutung, daß dies durch Hybridisierungen der verbliebenen Teile des repetitiven Alu- Elementes an den jeweiligen Enden von Sonde und Probe hervorgerufen wird, wurde durch die Hybridisierung eines einzelnen Klones auf die Filter überprüft. Auch das einzelne Inter- Alu- PCR- Produkt ergab Signale bei allen Spots.

#### 4.1.2.2 Versuche zur Entfernung des repetitiven Anteils in den IA- PCR- Produkten

Um die vielen falschpositiven Signale bei den Filterhybridisierungen zu reduzieren, wurden Versuche unternommen, aus dem zu hybridisierenden IA- PCR- Produkt den verbliebenen Anteil des repetitiven Alu- Elementes zu entfernen und so dadurch induzierte falschpositive Signale zu unterdrücken. Da der entsprechende Bereich nicht konserviert ist, schied einfache Verdauung mit bestimmten Typ- II –Restriktionsendonukleasen aus. Zur Umgehung dieses Problems wurden zwei unterschiedliche Strategien verfolgt:

- a.) Reamplifizierung des Inter- Alu- Amplikons mit einem modifizierten Primer, der eine Schnittstelle für die Typ- II- Restriktionsendonuklease *GsuI* beinhaltet. Anders als bei der Mehrheit der Typ- II- Endonukleasen ist die Erkennungssequenz nicht palindromisch und die eigentliche Schnittstelle liegt außerhalb der Erkennungssequenz. *GsuI* zeichnet sich dadurch aus, daß der Abstand mit 14/ 16 Nukleotiden sehr groß ist. Somit ist es möglich, die Schnittstelle gerichtet in eine unbekannt Sequenz zu setzen. Es wurden zwei Primer verwendet, die die Schnittstellen



möglichst weit in den polyA- Bereich setzen.

b.) Verdauung des Amplikons mit der Exonuclease BAL31

Exonukleasen dauern zunächst die Enden von DNS- Molekülen ab. Sie sind somit nicht sequenzspezifisch. Die Schwierigkeit im Umgang mit Exonukleasen besteht in ihrer exakten Dosierung. Die Anzahl und die verbleibende Länge der prozessierten DNS- Moleküle hängen von der verwendeten Enzymdosis ab. Dabei muß ein Optimum gefunden werden, damit möglichst alle Moleküle die repetitiven Enden verlieren, ohne daß sich deren Gesamtlänge zu stark reduziert und die anschließende Hybridisierung behindert wird.

Die experimentelle Überprüfung der modifizierten Inter- Alu- Amplikons erfolgte mittels *Dotblot*- Hybridisierung auf Membranen. Auf den Filtern wurden 96 Inter- Alu Amplikons von YAC- Klonen als Sonden transferiert. Durch die leicht erhöhte Komplexität (mehr als ein IA- PCR- Produkt pro Spot, im Durchschnitt werden fünf pro YAC erwartet – siehe 4.1.1.2) sollte die Chance für ein positives Signal erhöht werden. Als Proben wurden IA- Amplikons von monochromosomalen Hybridzelllinien verwendet. Die chromosomale Zuordnung der als Sonden verwendeten Amplikons war vorher bekannt.

Mit den mit *GsuI*- Primern reamplifizierten Amplikons konnten nur vereinzelte positive Signale erhalten werden. Die wahrscheinlichste Erklärung ist, daß im Zuge der Reamplifizierung mit den degenerierten bzw. nicht perfekt passenden Primer ein Großteil der Komplexität des Amplikons verloren geht und somit keine spezifischen Signale mehr erhalten werden konnten.

Für die BAL31- modifizierten Produkte konnten dagegen Signale erhalten werden. Etwa 30% der erwarteten Signale konnten detektiert werden. Nur ein Spot lieferte falsch positive Signale. Bei allen Spots waren aber noch Hintergrundsignale zu erkennen. Jedoch lieferten diese Experimente erste Hinweise darauf, daß vor allem die sich an den Enden der IA- PCR- Produkte befindenden repetitiven Alu- Sequenzen für die falsch positiven Signale aus den ersten Hybridisierungen (siehe 4.1.2.1) ursächlich sind.

Um diese Vermutungen weiter zu bestätigen, wurden neue *Dotblot*- Filter erstellt, die diesmal einzelne, reamplifizierte IA- PCR- Klone aus der gesamtgenomischen Bank (siehe 4.1.1.3) als Sonden aufwiesen. Es wurden willkürlich 48 Klone aus Bank ausgewählt. Diese wurden zur Hälfte durch Restriktion mit BAL31 von den repetitiven

Enden befreit. Auf den Filter wurden die 48 Klone in zwei Formen, vor und nach dem BAL31- Verdau, aufgetragen. Das Experiment sollte zeigen, daß mit Proben ohne repetitiven Enden keine falsch positiven Signale mehr auftreten. Das konnte für vier von fünf ausgewählten Klonen tatsächlich nachgewiesen werden. Dieses Ergebnis wies erneut darauf hin, daß es für den Erfolg der Hybridisierung entscheidend ist, den Einfluß der hochrepetitiven Enden der Alu- Elemente vollständig zu unterdrücken. Aufgrund dieser Ergebnisse wurde dann ein anderer Ansatz verfolgt, der von vornherein den störenden Einfluß der repetitiven Sequenzen auf das Ergebnis der Hybridisierung ausschließen sollte.

### 4.1.3 Generierung von Sonden

#### 4.1.3.1 IA- PCR Produkte von YACs

Die ursprüngliche Strategie sah vor, von kartierten YACs IA- PCR- Klone zu generieren und diese als Sonden für die Matrix- CGH zu benutzen. Wie in 4.1.2 beschrieben, führte dieser Weg jedoch nicht zu konsistenten Signalen. YAC- IA- PCR- Klone wurden darüber hinaus dafür benutzt, die Komplexität von IA- Amplikons abzuschätzen (siehe 4.1.1.2). Um die durch die endständigen, hochrepetitiven Alu- Sequenzen oder andere repetitive Elemente hervorgerufenen falsch positiven Signale zu unterdrücken, wurden in anschließenden Experimenten für eine große Zahl von IA- PCR- Klonen spezifische Primerpaare definiert. Dazu wurden Inter- Alu Klone verschiedener YACs sequenziert. Spezifische Primerpaare wurden mit dem Programm Primer3 so ausgewählt, daß das eingeschlossene PCR- Produkt keine repetitiven Elemente enthielt und möglichst lang war. Die ermittelten Primer mußten eine einheitliche *annealing*- Temperatur aufweisen, um später alle PCRs parallel durchführen zu können. Ein Abstand von 60 bp zwischen Alu- Element und internem Primer stellte sicher, daß die sich anschließenden Sequenzen niedriger Komplexität (z.B. poly(A)- Bereich) nicht in den internen spezifischen PCR- Produkten vorkamen. Diese Bereiche werden erfahrungsgemäß von Programmen wie RepatMasker oder entsprechenden Routinen von Primerauswahlprogrammen (PRIME, Primer3) nur unvollständig maskiert. Auf diese Weise konnten Sonden für den Chip hergestellt werden, deren genomische Lokalisation aufgrund der FISH- Kartierung der betreffenden YACs bekannt war und die einzelnen Inter- Alu- Fragmente im in IA-

PCR- Amplikons entsprachen. Zusätzlich wurde versucht, die Sequenzen der spezifischen PCR- Produkte durch Vergleich mit Datenbanksequenzen (BLAST) genauer zu lokalisieren. Mit diesen Klonen konnten bei der Hybridisierung von Amplikons monochromosomaler Zelllinien auf Mikroarrays konsistente Signale erhalten werden. Der Aufwand, mit dieser Methode Sonden zu generieren, erwies sich aber als unverhältnismäßig hoch.

#### 4.1.3.2 *In Silico*- generierte IA- PCR- Produkte

Da sich die Auswahl von Primerpaaren durch IA- PCR auf YACs als sehr aufwendig erwies und in der gleichen Zeit der Anteil der schon sequenzierten Bereiche des humanen Genoms sehr stark zunahm, wurde bald auch eine andere Methode zur Erzeugung spezifischer, interner Primerpaare verwendet. Sequenzen genomischer Klone verschiedener Datenbanken (Genbank, Sanger Centre, GoldenPath Assembly der HGP-Sequenzen) wurden benutzt, um nach putativen Bindungsstellen für den CL-2- Primer zu suchen. Dabei wurde nach solchen gesucht, die in richtiger Orientierung (3'<->3') (siehe 2.4.4) und im passenden Abstand (von 500-2500 bp) zueinander liegen. Weiterhin mußten die nicht stringenten PCR- Bedingungen berücksichtigt werden, so daß Fehlpaarungen in der Primer-Bindungssequenz toleriert wurden. Nachdem geeignete IA- PCR- Fragmente bestimmt wurden, wurde mit ihnen wie beschrieben (siehe 4.1.3.1) verfahren: repetitive Elemente wurden maskiert und spezifische interne Primerpaare wurden unter Berücksichtigung der oben aufgeführten Bedingungen ausgewählt.

Der oben dargestellte Algorithmus wurde in TurboPascal v6.0 (Borland) implementiert. Dabei wurden die Ergebnisse aus 4.1.1.1 und 4.1.1.4 berücksichtigt. In den Datenbank-Sequenzen wurde zunächst nach den ersten sechs 3'- Basen des CL-2- Primers gesucht. Ergab sich für dieses Hexamer ein Treffer, so wurden die sich in 5'- Richtung anschließenden elf Basen auf Übereinstimmung mit der entsprechenden Primersequenz überprüft. Weniger als drei Fehlpaarungen in diesem Bereich wurden als eine putative Primerbindungsstelle bewertet. Wenn zwei solche Primerbindungsstellen im richtigen Abstand und der richtigen Orientierung zueinander (siehe oben) gefunden werden konnten, dann wurde die davon eingeschlossene Sequenz extrahiert. Die Auflösung von ineinander verschachtelten Alu- Sequenzen wurde ebenfalls von dem Programm übernommen. Die extrahierten Sequenzen wurden dann als Datei ausgegeben und

anstelle der durch Sequenzierung erhaltenen Daten, wie in 4.1.3.1 beschrieben, als Grundlage zur Auswahl von spezifischen, internen Primerpaaren verwendet. Aufgrund der relativen Kürze einer Einheit von Alu- Element, Inter- Alu- Sequenz und Alu- Element auch Rohsequenzen der Phase0/Phase1 des Humanen Genom- Projektes zur Bestimmung von IA- PCR- Klonen herangezogen werden konnten.

#### 4.1.3.3 *In Silico* - Charakterisierung des Amplikons

Um zu überprüfen, ob die durch das Programm vorhergesagten IA- PCR- Klone die aus den einleitenden Experimenten (siehe 4.1.1) gewonnenen Daten hinsichtlich Verteilung und Häufigkeit bestätigen, wurden größere genomische Abschnitte ausgewertet, ohne daß anschließend Primer aus diesen Sequenzen ausgewählt wurden.

Im folgenden werden die Ergebnisse von Suchläufen dargestellt. Als Testsequenz wurden zunächst 62,4 Mbp des X- Chromosoms verwendet (HGP- Sequenzen, GoldenPath Assembly, Stand September 2000). Diese Daten enthielten nur fertig assemblierte und vollständig kartierte Contigs. Mit diesen sehr langen Teilsequenzen wurde der Algorithmus zunächst getestet und optimiert. In einem weiteren Schritt wurde die Verteilung von möglichen Inter- Alu- PCR- Produkten in Abhängigkeit ihrer genomischen Umgebung untersucht. Dazu wurden alle verfügbaren Sequenzdaten vom X- Chromosom verwendet. Diese beinhalteten auch kürzere Fragmente aus vorläufigen Assemblierungen des Humanen Genomprojektes. Die X- chromosomalen Sequenzdaten (172,4 Mbp) wurden zunächst in Blöcke zu je zehn Mbp (Blöcke 1-18) zerlegt und der Anteil der noch unbekannt Sequenzen bestimmt. Er lag zwischen 17% und 29%. Die Blöcke unterschieden sich auch stark in ihrem Anteil von Giemsa- R- und Giemsa- D- Banden:

Block 1-	32%	D- Banden,
Block 4-	49%	D- Banden,
Block 8-	71%	D- Banden,.
Block 11-	75%	D- Banden,
Block 14-	9%	D- Banden,
Block 17-	2%	D- Banden,
Block 18-	0%	D- Banden

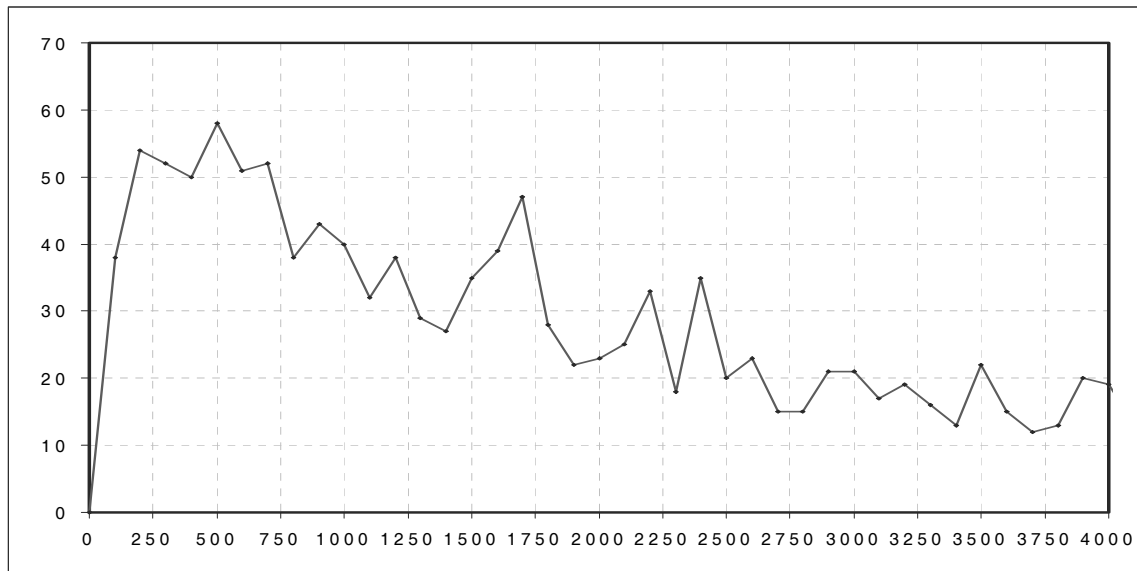
Genomische Regionen von Metaphasechromosomen, die sich von dem Farbstoff Giemsa besonders stark anfärben lassen, werden als direkte Giemsa, oder D- Banden bezeichnet. Die von Giemsa nicht anfärbaren Banden können dagegen spezifisch mit Chromomycin- Farbstoffen markiert werden. Wegen dieser reversen Färbung spricht man auch von R- Banden. Es ist bekannt, daß dem unterschiedlichen Färbeverhalten auch spezifische Eigenschaften der betroffenen DNS zugeordnet werden können: R- Banden haben einen hohen lokalen GC- Gehalt, eine hohe Gendichte, hohe transkriptionelle Aktivität und eine erhöhte Konzentration von Alu- Sequenzen. Alle diese Eigenschaften sind genau negativ mit dem Auftreten der D- (oder G- Banden) Banden korreliert (Korenberg und Engels, 1978; Korenberg und Rykowski, 1988).

**Tabelle 4-2:** Ergebnisse der Computersuche nach Inter- Alu- PCR- Produkten (Min, Max, N erlaubt: Parameter zur Steuerung des Programms, Suche nach einem PCR- Produkt von Min bis Max Basenpaaren mit N erlaubt Fehlpaarungen im Primer toleriert), Angabe der gefundenen Treffer für Alu- Elemente in Vorwärts- und Rückwärtsrichtung und die daraus resultierenden Inter- Alu- PCR- Produkte

	Min Max N erlaubt	500, 3000, 4	4	4	500, 2000, 4	500, 2000, 2	2	2
Block- Nr.	Sequenz in Mbp	Alu Elemente 5'→3'	Alu Elemente 3'→5'	Inter- Alu Produkte	Inter- Alu Produkte	Alu Elemente 5'→3'	Alu Elemente 3'→5'	Inter- Alu Produkte
1	7,75	881	1080	208	151	630	862	113
4	7,09	567	741	122	95	462	627	77
8	8,27	500	596	91	72	429	506	60
11	8,44	545	664	111	78	416	535	48
14	7,52	465	689	109	79	346	559	57
17	7,59	500	571	109	89	337	408	57
18	1,20	127	144	29	23	91	97	13

	Min	500,			500,		
	Max	2000,			2000,		
	N erlaubt	1	1	1	0	0	0
Block-Nr.	Sequenz in Mbp	Alu Elemente 5'→3'	Alu Elemente 3'→5'	Inter-Alu Produkte	Alu Elemente 5'→3'	Alu Elemente 3'→5'	Inter-Alu Produkte
1	7,75	509	689	75	331	473	31
4	7,09	390	542	62	259	367	33
8	8,27	369	432	44	256	285	20
11	8,44	378	491	44	274	331	19
14	7,52	308	498	48	211	358	25
17	7,59	296	365	44	212	240	27
18	1,20	82	87	11	56	61	6

Mit den im Ergebnis aufgelisteten Suchläufen sollte überprüft werden, ob die experimentell ermittelten Werte für die Anzahl möglicher Inter- Alu- Klone und die Toleranz des Primers unter den gewählten PCR- Bedingungen durch das Programm simuliert werden können. Experimentell wurde gefunden, daß fast die Hälfte der gefundenen Primerbindungsstellen (19 von 44) perfekt mit der CL-2- Primersequenz übereinstimmt. Insgesamt wurden 52 tolerierte Abweichungen gefunden, was etwa 1,2 pro Sequenz entspricht. Um diese gefundenen Werte am besten abzubilden, sollten die Parameter für die Länge von 500- 2000 bp und eine erlaubte Fehlpaarung pro Primersequenz benutzt werden. Man erhält dann rund ein IA- PCR- Produkt alle 200 kbp, was den Resultaten aus 4.1.1.2 entspricht. Wie erwartet werden die wenigsten Alu-Sequenzen in den Blöcken mit hohem Anteil an Giemsa- D- Banden (11 und 14) gefunden. Dagegen werden sehr hohe lokale Konzentrationen in den terminalen Sequenzen (Blöcke 1 und 18, beide Giemsa- R) gefunden.



**Abbildung 4-3:** Simulation der Längenverteilung der PCR- Produkte eines theoretischen Inter- Alu- Amplikons, ein Datenpunkt repräsentiert eine Klasse von 100 aufeinanderfolgenden Längen (Punkt 1: Längen von 1-100bp, Punkt 2: Längen von 101-200bp, etc.), Abzisse Länge der Produkte, Ordinate Häufigkeit in 100Mbp

In einer weiteren Simulation wurden 100Mbp des X- Chromosoms auf das Vorhandensein von putativen Inter- Alu- Klonen untersucht. Bei der Auswahl der Primerbindungssequenzen wurde eine Fehlpaarung zugelassen und keine Restriktion gemacht bezüglich der erwarteten Produktlänge.

Aus dem Diagramm ist zu entnehmen, daß Inter- Alu- Fragmentlängen unterhalb 100bp stark unterrepräsentiert sind. Das Maximum liegt bei Längen zwischen ca. 250bp und etwa 750bp. Danach verringert sich die Häufigkeit von putativen Inter- Alu- Klonen zunehmend mit steigender Länge der Fragmente. Erkennbar sind auch einzelne lokale Maxima, die wahrscheinlich keine biologische Relevanz besitzen. Wenn man die Beschränkungen der PCR bei der Länge der erzeugbaren Produkte in Betracht zieht, dann läßt sich dieses theoretisch erwartete Längenspektrum mit dem experimentell erhaltenen (siehe 4.1.1.1) gut korrelieren.

#### 4.1.4 Mikroarray- Technik: Grundlagen

##### 4.1.4.1 Nachweis der Selektivität von Hybridisierungen auf Mikroarrays

Der Begriff Selektivität bedeutet in der Mikroarray- Technik das möglichst vollständige Vermeiden von falsch positiven und falsch negativen Signalen. In einem Experiment wurde überprüft, ob die gewählten Hybridisierungsbedingungen stringent genug sind, eine hohe Selektivität zu gewährleisten. Dazu wurden willkürlich 16 spezifische, interne IA- PCR- Klone ausgewählt, die auch als Sonde auf dem Mikroarray repräsentiert sind. Die PCR- Produkte wurden vereinigt, fluoreszent markiert und auf den Chip hybridisiert. Dabei wurden unspezifische Bindungen nicht durch Zugabe von unmarkierter DNS unterdrückt. Die Auswertung der Mikroarrays ergab nur für die 16 ausgewählten Klone und die Positivkontrollen Signale. Die als Hintergrund eingestuften Signale anderer Klone erreichten weniger als 2% der Intensität des schwächsten ausgewählten Inter- Alu- Klons.

Es wurde somit gezeigt, daß man durch geeignete Wahl der Primer die Spezifität der Klone gewährleisten kann.

##### 4.1.4.2 Bestimmung der Sensitivität

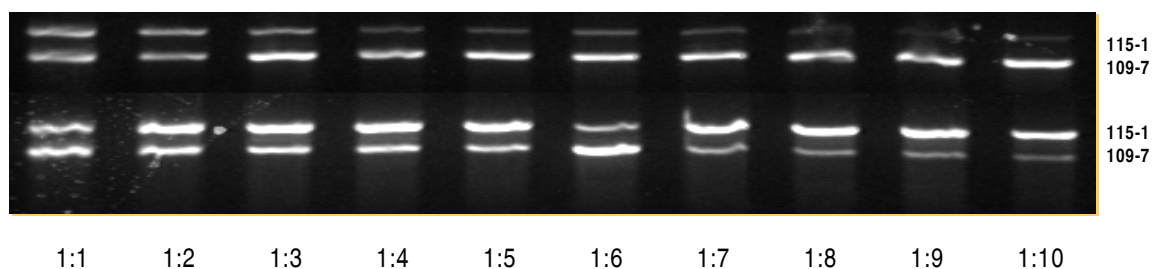
Zur Bestimmung der Sensitivität unseres Mikroarray- Systems wurde in einem Experiment untersucht, bei dem eine Probe aus 16 internen IA- PCR- Klonen (siehe oben) in Verdünnungen bis 1:100.000 auf den Chip hybridisiert wurde. Auch in der höchsten Verdünnungsstufe konnten die Spots noch ausgewertet werden. Wenn man davon ausgeht, daß die unterschiedlichen IA- PCR- Klone in ihrer Anzahl nahezu gleich verteilt sind (siehe 4.1.1.3), so kann man daraus ableiten, daß ungefähr  $16 \cdot 100.000 = 1,6$  Mio. verschiedene IA- PCR- Klone unterschieden werden können. Dieser Wert liegt deutlich über den für ein genomisches IA- PCR- Amplikon angenommenen 15.000 verschiedenen Fragmenten (siehe 4.1.1.2). Um zu gewährleisten, daß nicht Abweichungen der verwendeten Chips für unterschiedliche Signalintensitäten ursächlich sind, wurde jeweils eine identische Probe mit einem anderen Farbstoff markiert, co- hybridisiert und ausgewertet. Es wurden keine Unterschiede der Signalintensitäten der Kontrollprobe beobachtet, so daß die Intensitätsunterschiede der Proben allein auf die unterschiedlichen Verdünnungen zurückzuführen sind.



Somit wurde gezeigt, daß nach einer Reduktion der Komplexität der Proben es technisch möglich ist, die Signale zu erfassen.

#### 4.1.4.3 Simulation der Matrix- CGH in einem Minimalsystem

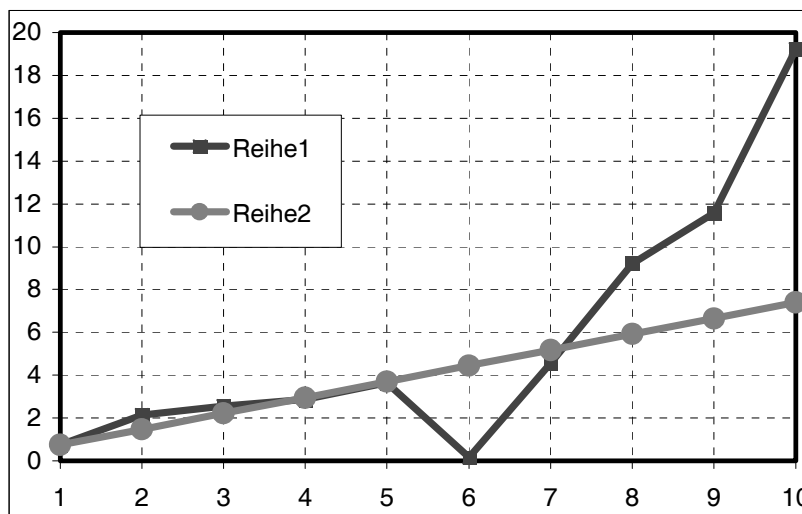
Das komplette Verfahren der Matrix- CGH wurde in einem Minimalsystem untersucht, um so die technisch erreichbaren Grenzen bestimmen und beurteilen zu können. In diesem Versuch wurde das komplexe genomische IA- Amplikon (15.000 verschiedene IA- Klone) durch ein Amplikon aus nur zwei IA- Klonen ersetzt. Dazu standen die durch IA- PCR auf YACs erzeugten Inter- Alu- PCR- Klone zur Verfügung, die noch die 3'- Enden der Alu- Elemente enthielten (siehe 4.1.1.2). Die beiden ausgewählten Klone (Bezeichnungen 109-7 und 115-1, jeweils als Plasmid- DNS im Vektor *pGemTeasy*) waren aufgrund ihrer ungleichen Länge auf dem Agarosegel unterscheidbar. Die Patienten- DNS als Ausgangsmaterial für die Inter- Alu- PCR wurde in diesem Experiment durch Ansätze ersetzt, die die beiden Klone in unterschiedlichen Mengenverhältnissen enthielten. Dazu wurden Verdünnungsreihen erstellt. Ein Klon wurde in konstanter Menge eingesetzt, der andere wurde in den Verhältnissen 1:1, 1:2 bis 1:10 dazupipettiert. Somit diente der erste Klon als Referenz und der zweite als variable Meßgröße. Zur Kontrolle wurden beide Klone in jeweils umgekehrten Verhältnissen in einer zweiten Reihe untersucht. Die so erstellten 20 Ansätze wurden durch Inter- Alu- PCR amplifiziert und auf ein Agarosegel aufgetragen (siehe Abbildung 4.4). Die Bandenmuster wurden mittels ImageQuant quantifiziert. Die



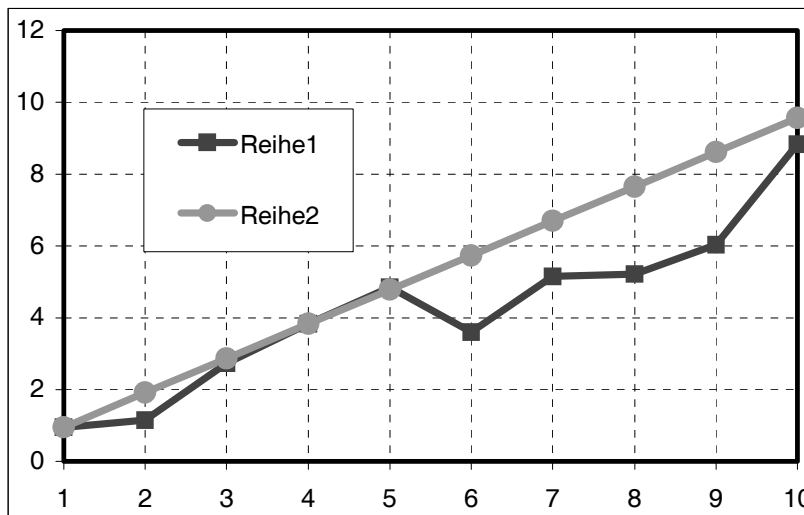
**Abbildung 4-4:** Inter- Alu- PCR auf jeweils zwei IA- Klone (109-7 und 115-1), oben: 109-7 in konstanter Menge und 115-1 in einer Verdünnungsreihe von 1:1 bis 1:10, unten 115-1 in konstanter Menge und 109-7 in einer Verdünnungsreihe von 1:1 bis 1:10

Mengenverhältnisse der Bandenpaare wurden zueinander ins Verhältnis gesetzt. Die erhaltenen Zahlenwerte wurden jeweils auf den 1:1- Ansatz normiert. Die in diesen Experimenten gewonnenen Meßwerte sind in den Abbildungen 4-6 und 4-6 zusammengefaßt. Aus dem Vergleich der Daten ist zu ersehen, daß im Bereich von 1:1 bis 1:5 für die Mengenverhältnisse an eingesetzter DNS die Kurven für die theoretisch erwarteten und experimentell ermittelten Daten nahezu identisch verlaufen. Ab einem Verhältnis von 1:6 divergieren die Kurven dagegen. Daraus ist zu entnehmen, daß unter den verwendeten Bedingungen für die Inter- Alu- PCR aus dem Verhältnis der IA- PCR- Produkte das Verhältnis der Mengen an Ausgangs- DNS direkt abzuleiten ist, sofern der relative Unterschied der eingesetzten DNS- Mengen den Faktor fünf nicht überschreitet.

In einem nächsten Schritt wurden dann die PCR- Produkte fluoreszent markiert und auf ein Mikroarray hybridisiert. Auf dem Array befanden sich als Sonden die entsprechenden internen Inter- Alu- PCR- Produkte (109-7, 115-1) in jeweils vier verschiedenen Verdünnungen, sowie Negativ- und Positivkontrollen. Für die Proben wurde jeweils ein Referenzamplikon (beide IA- Klone im Verhältnis 1:1) gegen unterschiedliche Testamplikons (beide Klone im Verhältnis 1:1, 3:1, 5:1 und 9:1) vergleichend hybridisiert. Zunächst wurden die Signalintensitäten der beiden IA- Klone aus einem Amplikon ermittelt. Diese wurden zueinander ins Verhältnis gesetzt. Die Zahlenwerte für diese Verhältnisse ergaben als Quotient dann den Meßwert, der im Diagramm 4.7 als Ordinate angegeben ist. Dieser Meßwert wurde mit dem Verhältnis

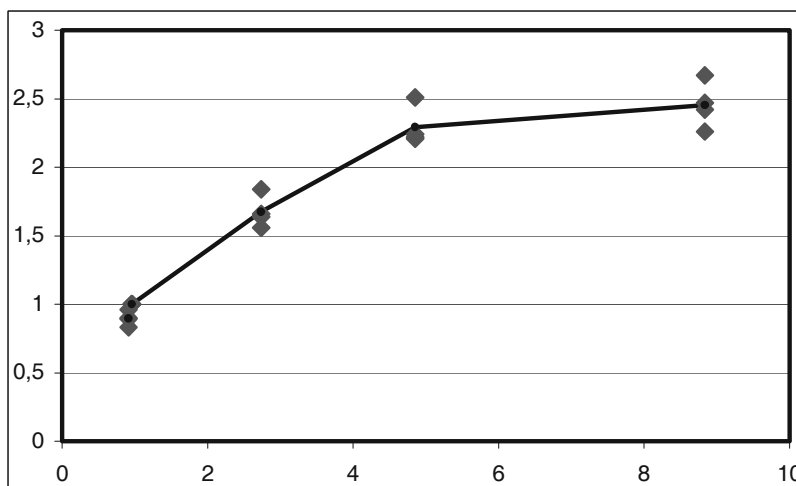


**Abbildung 4-5:** Experiment für die Verdünnungsreihe von Klon 115-1: Abzisse: eingesetzte (erwartete) Mengenverhältnisse, Ordinate: gemessene Mengenverhältnisse der PCR- Produkte, Reihe 1: tatsächlich ermittelte Werte, Reihe 2: idealer Verlauf wenn Mengenverhältnisse konserviert werden



**Abbildung 4-6:** Experiment für die Verdünnungsreihe von Klon 109-7: Abzisse: eingesetzte (erwartete) Mengenverhältnisse, Ordinate: gemessene Mengenverhältnisse der PCR-Produkte, Reihe 1: tatsächlich ermittelte Werte, Reihe 2: idealer Verlauf wenn Mengenverhältnisse konserviert werden

der DNS- Mengen verglichen, die tatsächlich für die Markierung eingesetzt wurden.



**Abbildung 4-7:** Vergleich der Verhältnisse der Mikroarray-Signale mit den Mengen der eingesetzten PCR-Produkte, Abzisse: Verhältnis der eingesetzten Mengen an PCR-Produkt, Ordinate experimentell ermittelte Signalverhältnisse, die Punkte stellen die Ergebnisse von verschiedenen Sondenmengen dar, die Kurve verbindet die Durchschnitte der einzelnen Meßwerte

Aus der sich ergebenden Kurve ist abzuleiten, daß die auf dem Chip ermittelten Meßwerte nicht direkt die DNS- Mengenverhältnisse wiedergeben. Für Mengenverhältnisse von mehr als fünf gehen die Meßwerte in einen Sättigungsbereich über. Für den linearen Bereich darunter ergibt sich folgende Gleichung:

$$(4.1) \quad y = 0,33x + 0,63 \quad | \quad y = \text{Meßwert}, x = \text{tatsächlicher Wert}$$

oder

$$(4.2) \quad x = (y - 0,63) / 0,33$$

Damit ergeben sich für die Auswertung von Matrix- CGH- Experimenten folgende Richtlinien: Für gemessene Verhältnisse von Signalintensitäten größer als 2,5 ist das tatsächliche Verhältnis der Kopienzahlen der DNS größer als fünf, genauere Aussagen können mit der verwendeten Methodik nicht erzielt werden. Meßwerte kleiner als 2,5 müssen mit der Formel (4.2) auf das tatsächliche Kopienzahlverhältnis umgerechnet werden.

## 4.2 Lokalisierung von YAC- Klonen mittels Mikroarrays

Mit Hilfe von IA- PCR- Mikroarrays sollten sich große genomische Klone, wie zum Beispiel YACs, lokalisieren lassen. Voraussetzung ist, daß mögliche Inter- Alu- Sequenzen des YACs in entsprechenden internen, spezifischen IA- Klonen auf dem Mikroarray repräsentiert sind.

Zwei YACs, die der Xq21.1- Region zugeordnet wurden, wurden ausgewählt und als Ausgangsmaterial für eine Inter- Alu- PCR benutzt. Die erhaltenen Produkte wurden aufgereinigt, markiert und auf den Chip hybridisiert. Für jede Probe wurde jeweils ein positives Signal erhalten:

YAC- Bezeichnung	IA- PCR- Klon auf dem Chip
106	Z84487
2924	A1121869

Da bis auf die Positivkontrollen keine weiteren Signale erhalten wurden, konnte eine eindeutige Zuordnung der YAC- Klone zu bekannten korrespondierenden Sonden auf dem Mikroarray erfolgen. Die Ergebnisse wurden durch PCR mit den internen Primerpaaren auf der isolierten YAC- DNS bestätigt.

### 4.3 Kartierung einer hemizygoten Deletion

#### 4.3.1 Einführung

Die vollständige Deletion von genomischen Abschnitten ist in zwei Varianten möglich. Homozygote Deletionen umfassen identische Abschnitte von homologen Autosomen. Auch eine Deletion auf dem einen X- Chromosom in männlichen Zellen (hemizygot) führt zum kompletten Funktionsverlust der betroffenen Gene, da keine zweite Kopie wie bei weiblichen Zellen zur Kompensation vorhanden ist. Die Untersuchung von solchen vollständigen Verlusten von genomischen Abschnitten in der humanen DNS war ein erstes Hauptziel der vorliegenden Arbeit. Da für jede durch einen auf dem Chip befindlichen Spot repräsentierte Sonde das erhaltene Signalintensitätsverhältnis nur den Kategorien ‚vorhanden‘ oder ‚nicht vorhanden‘ zugeordnet werden kann, wurde eine vergleichsweise einfache Auswertung solcher Experimente erwartet.

#### 4.3.2 Beschreibung des klinischen Phänotyps der betroffenen Patienten

Es wurde die DNS von drei Zelllinien von Patienten mit Mentaler Retardierung (MR), Choroideremie, Lippen-, Kiefer- Gaumenspalte und Taubheit untersucht. Bei allen drei Patienten lagen große Deletionen in der Xq21- Region vor (van der Maarel *et al.*, 1995). Die für die Taubheit und die Choroideremie, eine Form von erblich bedingter Blindheit, ursächlichen Gene sind bekannt: X- chromosomal vererbte Taubheit – POU3F4, OMIM 304400 (de Kok *et al.*, 1995); Choroideremie – CHM, OMIM 303100 (van Bokhoven *et al.*, 1994).

**Tabelle 4-3:** Verzeichnis der untersuchten Zelllinien, klinische Befunde nach van der Maarel *et al.*, 1995

Interne Bezeichnung	Alias	Klinische Befunde der Patienten:
303/99	NP	CHM, MR, Lippen-, Kiefer- Gaumenspalte
525/00	RvD	MR, Hörbeeinträchtigung
526/00	XL62	CHM, MR, Hörbeeinträchtigung

Das Gen, dessen Mutation für die Mentale Retardierung der Patienten verantwortlich ist, konnte noch nicht bestimmt werden. Aufgrund ihrer Lage zwischen den beiden zuvor beschriebenen Genen werden vor allem die folgenden Gene als Kandidatengene für Mentale Retardierung angesehen:

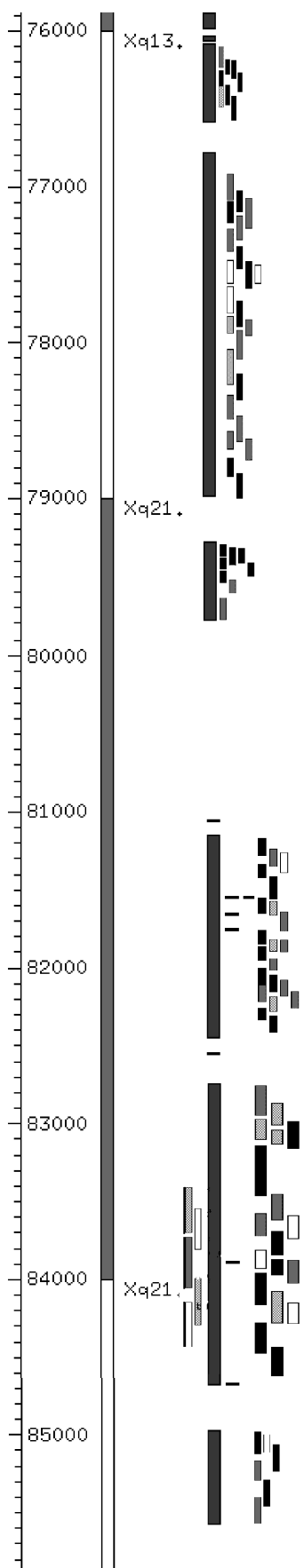
**Tabelle 4-4:** Liste von möglichen MR- Kandidatengenen in der Xq21- Region, die angegebenen Positionen wurden aus dem GoldenPath Assembly der HGP- Sequenzen, Stand Februar 2001 entnommen

Kurzform	alias	5'-Position	Referenz	OMIM	Beschreibung
<b>RPS6KA6</b>		78551267	Yntema <i>et al.</i> , 1999	300303	ribosomal protein S6 kinase, 90kD, polypeptide 6
<b>ZNF6</b>		79724387	Colleaux <i>et al.</i> , 1996.	314990	zinc finger protein 6 (CMPX1)
<b>SH3BGRL</b>		80214389	Egeo <i>et al.</i> , 1998	300190	SH3 domain binding glutamic acid-rich protein like
<b>STIP1</b>	STUB1	81098806	Honore <i>et al.</i> , 1992	605063	STIP1 stress induced phosphoprotein 1

Für die genetische Ursache der Lippen-, Kiefer- Gaumenspalte bei Patient NP (303/99) bestanden bis jetzt noch keine Vermutungen.

#### 4.3.3 Erstellung eines geordneten Rasters von spezifischen, internen IA- PCR-Klonen

Zur Abgrenzung der betreffenden Region wurden genomische Marker (van der Maarel *et al.*, 1995) herangezogen. Anschließend wurden aus bereits vorliegenden Datenbanksequenzen des Humanen Genomprojektes (HGP, Sanger Centre, [www.sanger.ac.uk/](http://www.sanger.ac.uk/), Stand Juli 2000) BAC- Klone ausgewählt (siehe nächste Seite), um nach putativen IA- PCR- Klonen in dieser Region zu suchen.



Wie aus der Abbildung zu entnehmen ist, war die Xq21-Region zu diesem Zeitpunkt noch nicht vollständig durch genomische Klone abgedeckt. Etwa 60-70% der vorhandenen Klone enthielten zumindest teil-assemblierte Sequenzen, die für die Suche nach Inter-Alu-Sequenzen geeignet waren. Aus diesen verfügbaren Daten wurden die Sequenzen für 21 spezifische, interne Inter-Alu-Primerpaare ausgewählt. Zusammen mit den schon vorher vorhandenen Sonden standen somit 28 interne IA-PCR-Klone für den Bereich zwischen 70-90Mbp auf dem X-Chromosom zur Verfügung. Diese vergleichsweise geringe Anzahl von verwertbaren IA-PCR-Klonen spiegelte den damaligen Stand der Sequenzierung aber auch die geringere lokale Konzentration von Alu-Elementen in dieser Region wieder.

**Abbildung 4-8:** Sequenzierungsplan des Humanen Genomprojektes (Sanger Centre, Hinxton, UK) der Xq21-Region im Juli 2000 - ganz links: Abstand der genomischen Region von Xpter in kbp, Mitte: cytogenetisches Bandenmuster, rechts: Abdeckung der Region durch die vorhandenen genomischen Klone (breiter Balken generelle Abdeckung, daneben der *minimal tiling path* mit dem durch den Farbton dargestellten Stand des Sequenzierfortschrittes: weiß: lokalisierter Klon ohne verfügbare Sequenzinformation, schwarz: fertig sequenzierter und assemblierter Klon, Grautöne geben die verschiedenen Zwischenschritte wieder

#### 4.3.4 Auswertung

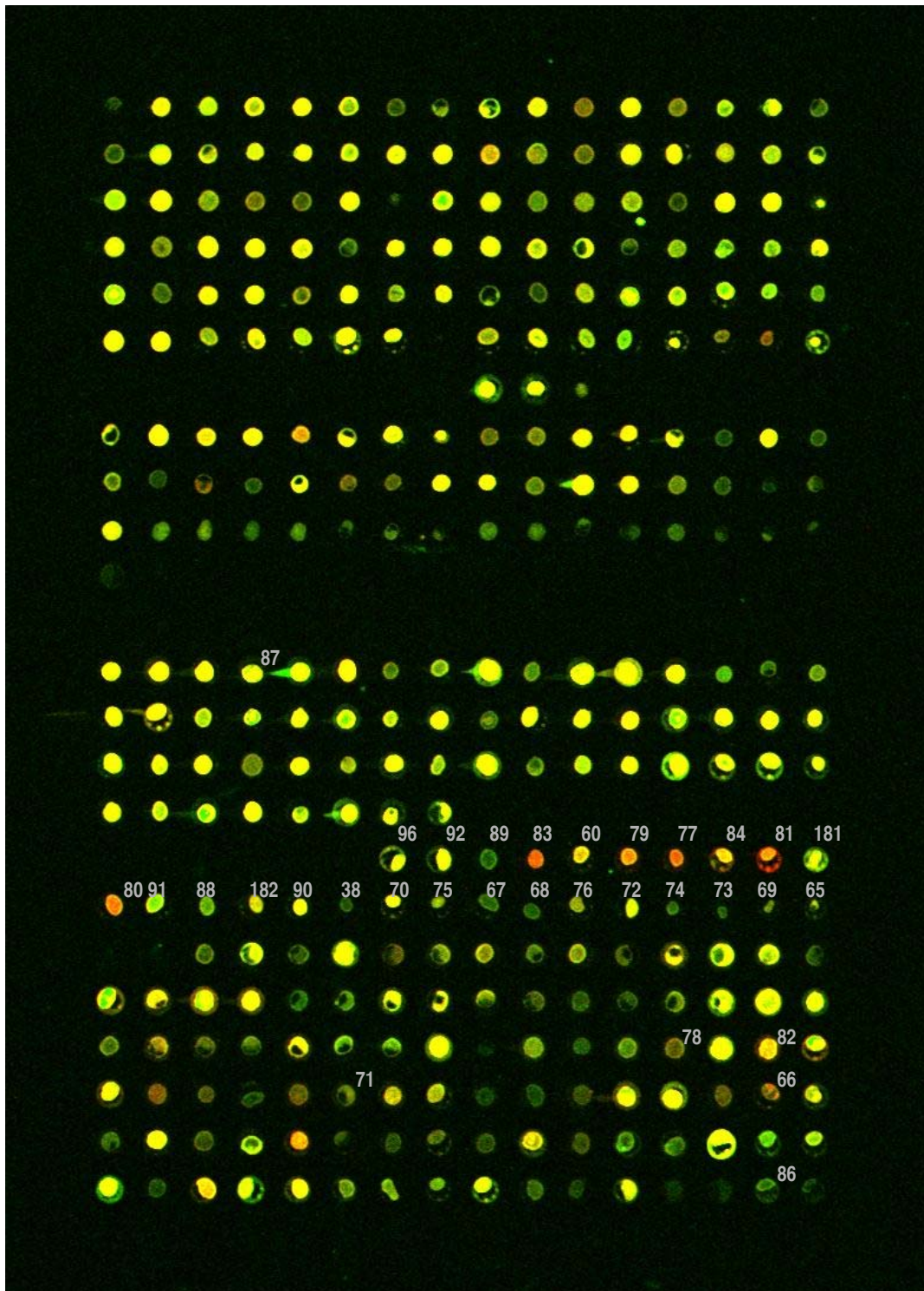
Von den drei Patienten- DNS- Proben und einer männlichen Kontroll- DNS- Probe wurden Inter- Alu- Amplikons erzeugt und mit Cy3 (Patienten) bzw. Cy5 (Kontrolle) markiert. Die markierten Produkte wurden danach auf IA- PCR- Mikroraster co-hybridisiert. Anschließend wurden die Cy3- und Cy5- spezifischen Signale getrennt aufgenommen und falschfarbenkodiert übereinander projiziert (siehe nächste Seiten). Deletierte IA- PCR- Klone erscheinen in dieser Darstellung rot. Für eine genaue Lokalisierung der Deletion wurden die Daten statistisch ausgewertet. Zunächst wurde für jede Messung der Ausgleichs- oder Normierungsfaktor zwischen den Kanälen (Cy3 und Cy5) bestimmt. Dazu wurde jeweils die durchschnittlichen Signalintensität der 96 vom Chromosom 22 gewonnenen IA- PCR- Klone ermittelt. Der erhaltene Quotient der beiden Mediane ergab den Normierungsfaktor. Zur Normierung wurden dann die einzelnen Signalintensitätsverhältnisse (*ratios*) mit diesem Faktor multipliziert. Die so auf diese Weise korrigierten Intensitätsverhältnisse wurden dann den einzelnen IA- PCR- Klonen zugeordnet. Die Lage aller IA- PCR- Produkte auf dem X-Chromosom wurde durch Sequenzvergleich mit den assemblierten Daten des Humanen Genomprojektes (BLAST- Algorithmus, GoldenPath Assembly, Stand März 2001) bestimmt. Für die weitere Auswertung wurden nur IA- PCR- Klone verwendet, die eine zweifelsfreie, eindeutige Zuordnung zuließen (siehe 7.2.1).

Die folgenden Abbildungen zeigen die bei der vergleichenden Hybridisierung der IA- Amplikons von Patienten- und Kontroll- DNS gewonnenen Ergebnisse. In den falschfarbenkodierten Überlagerungen erscheinen deletierte IA- PCR- Klone rötlich, die sowohl im Kontroll- wie in den Patientengenomen vorhandenen Sequenzen dagegen gelb bis grünlich. Die normalisierten Verhältnisse der ermittelten Signalintensitäten wurden für die IA- PCR Klone aus der Xq21- Region tabellarisch zusammenfaßt (siehe Tabellen 4.5, 4.6 und 4.7). Mit den spezifischen internen Primerpaaren wurde anschließend eine Kontroll- PCR auf der Patienten- DNS durchgeführt. Das Ausbleiben eines spezifischen PCR- Produktes ermöglichte somit ebenfalls die Zuordnung eines Klones zum deletierten Bereich. Zur Visualisierung der ermittelten Intensitätsverhältnisse wurden die Werte für 171 X- chromosomale IA- PCR- Klone in der Reihenfolge ihrer Position schematisch dargestellt. Die Anwendung eines gleitenden



Durchschnitts über drei bzw. fünf Elemente ermöglichte eine klarere Darstellung der erhaltenen Ergebnisse durch Unterdrückung des Signalrauschens einzelner Meßwerte.

#### 4.3.4.1 Kartierung der Deletion in der DNS von der Zelllinie 526/00

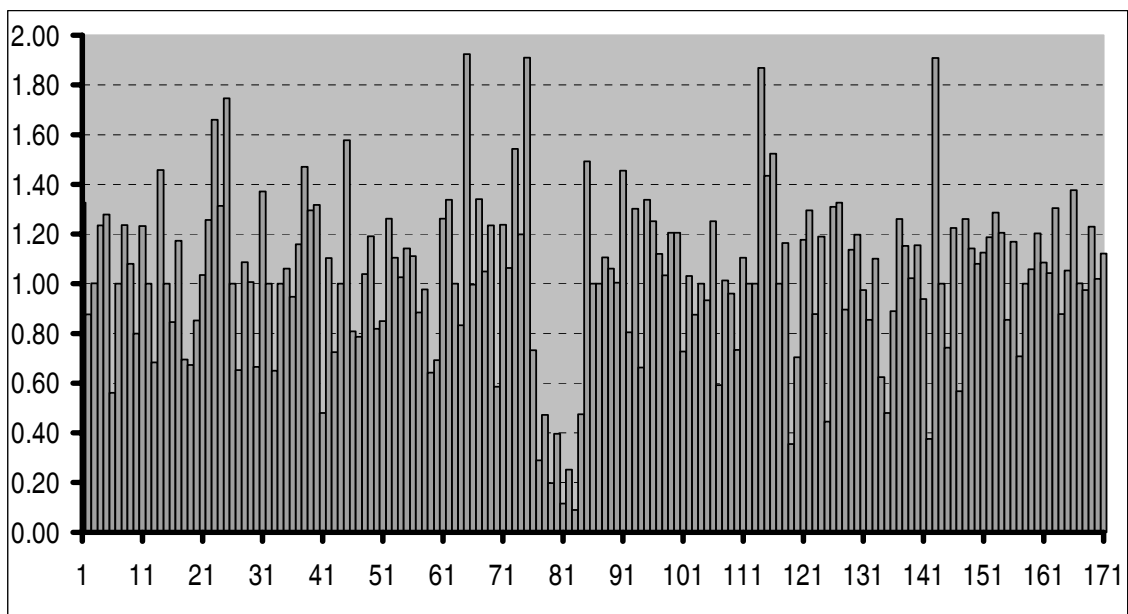


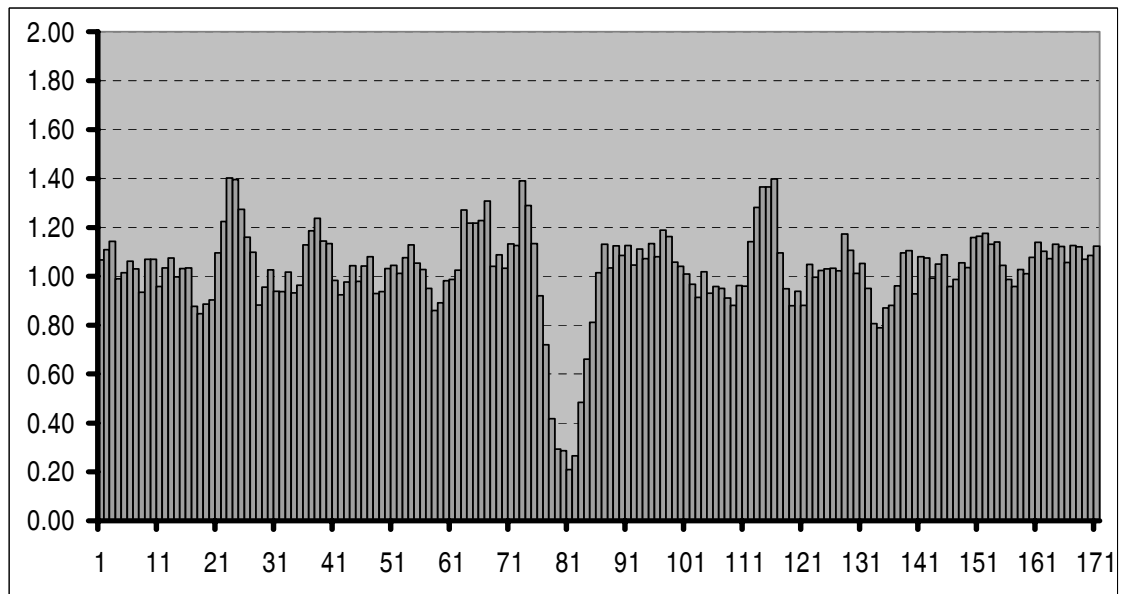
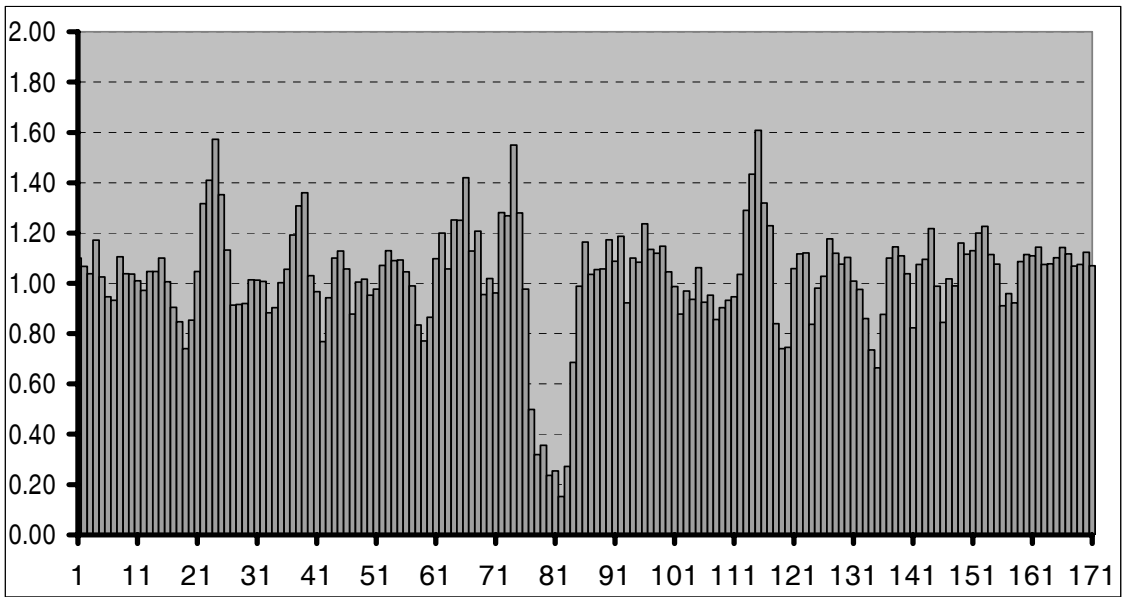
**Abbildung 4-9:** Falschfarbenkodierte Überlagerung der Signale der Hybridisierung der DNS- Probe 526/00 (rot) gegen eine Referenz- Probe (grün), die oberen sechs Reihen bestehen aus den autosomalen Kontrollklonen, die numerierten Spots entsprechen den Indizes der IA- PCR- Klone aus der Xq21- Region

**Tabelle 4-5:** Auswertung der Mikroarray- Hybridisierung mit IA- PCR- Sonden für den Bereich zwischen 70 und 80 Mbp (von Xpter aus) für die Zelllinie 526/00: aufgeführt sind die Klonbezeichnung, die Lage des Klon, das ermittelte normierte Verhältnis der Signalintensitäten und das Ergebnis der Kontroll- PCR- Ansätze mit den spezifischen, internen IA- PCR- Primern

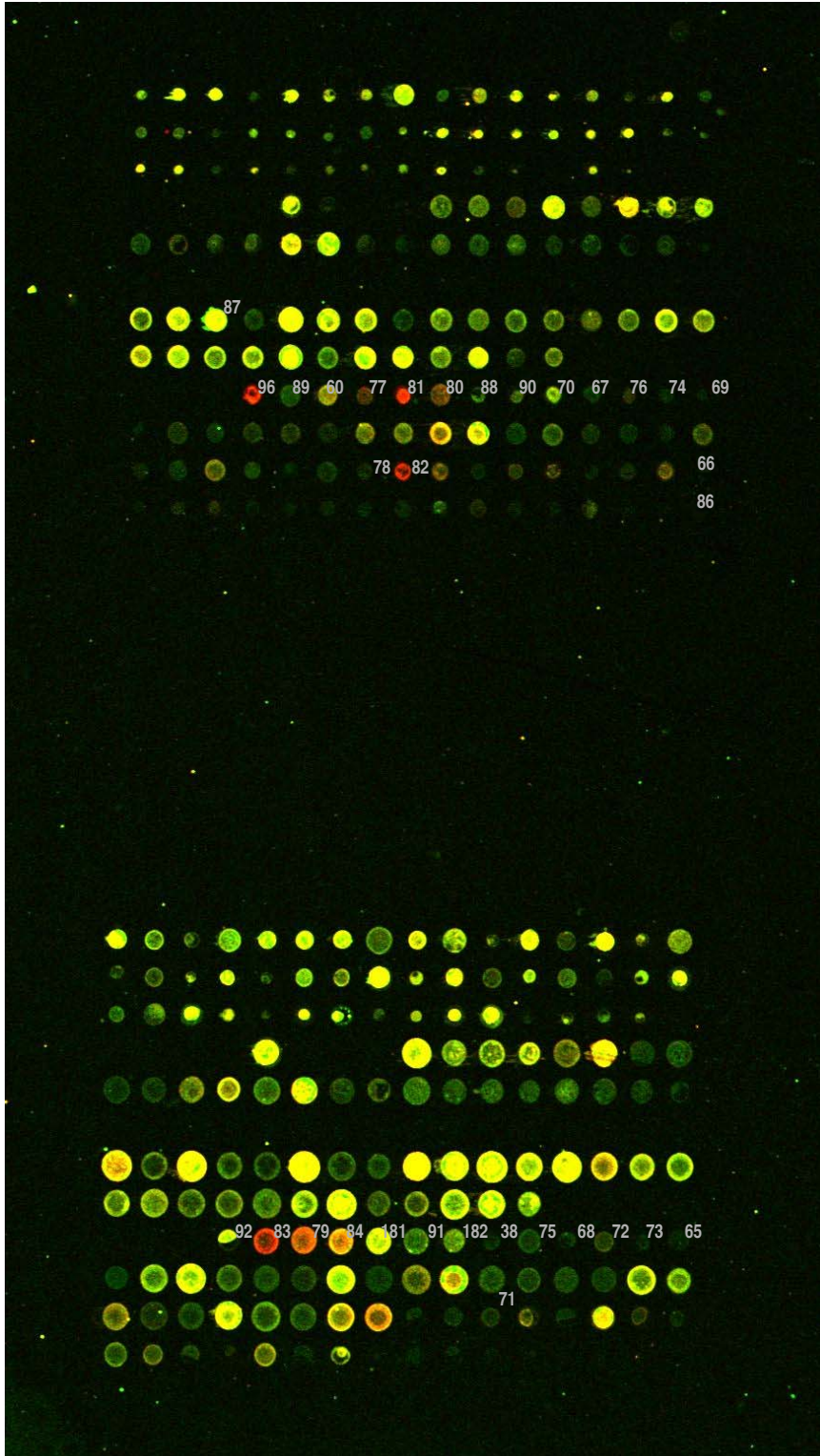
Index	Klon- Bezeichnung	Position	Signal- verhältnis	PCR- Ergebnis	Index	Klon- Bezeichnung	Position	Signal- verhältnis	PCR- Ergebnis
65	AC004677	70.756.612	1,92	+	79	AL160266	77.423.756	0,20	-
66	hs333e23	73.007.918	1,00	-	80	AL158820	79.074.886	0,40	-
67	Z84487_2	73.618.669	1,34	+	81	AL359885	80.154.487	0,12	-
68	Z84487_11	73.634.878	1,05	+	82	da145b8	81.150.346	0,25	-
69	AL049589	73.742.680	1,23	+	83	AL035451	81.150.396	0,09	-
70	Z94801	73.816.374	0,58	+	84	AC025202	82.876.366	0,48	-
71	hs465g10b	73.816.445	1,24	+	85	hs393p23	83.169.128	1,49	+
72	AL138743_6	74.041.011	1,06	+	86	652-12	87.114.284	1,00	+
73	AL138743	74.119.096	1,54	+	87	3019_1	88.073.039	1,00	+
74	AL138743_23	74.136.279	1,20	+	88	AL136362	88.323.926	1,11	+
75	AL109753	74.364.174	1,91	+	89	AL160266_2	88.358.928	1,06	+
76	AL162912	74.780.589	0,73	+	90	AL136362_2	88.358.928	1,00	+
77	AL031313	76.409.633	0,29	+	91	AL133274	88.723.181	1,45	+
78	dj1111n9	76.769.657	0,47	-	92	AL121869	88.867.588	0,80	+

**Abbildung 4-10:** Graphische Darstellung der für die Zelllinie 526/00 ermittelten, normierten Signalintensitätsverhältnisse für die entsprechend ihrer Lage geordneten IA- PCR- Klone, von Xpter (links) bis Xqter (rechts), die Zahlen geben den Positionsindex an und entsprechen dem Eintrag in der Tabelle, die erste Abbildung beinhaltet die tatsächlichen Meßwerte, in der mittleren Abbildung sind die Meßwerte durch Verwendung eines gleitenden Durchschnitts über drei Elemente, in der dritten Abbildung über fünf Meßwerte, geglättet





## 4.3.4.2 Kartierung der Deletion in der DNS von der Zelllinie 525/00

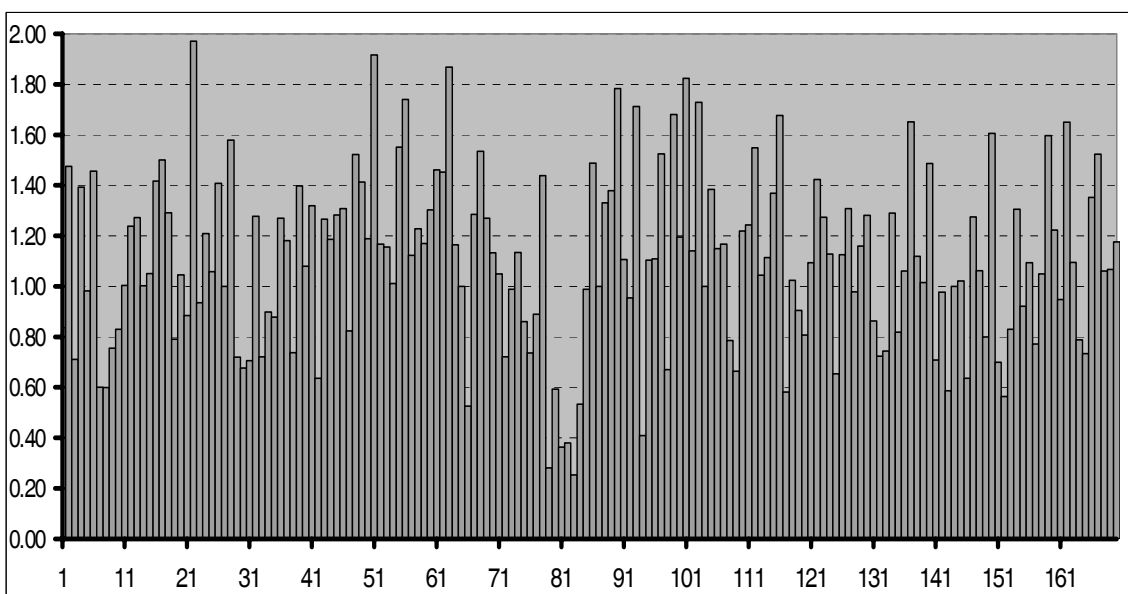


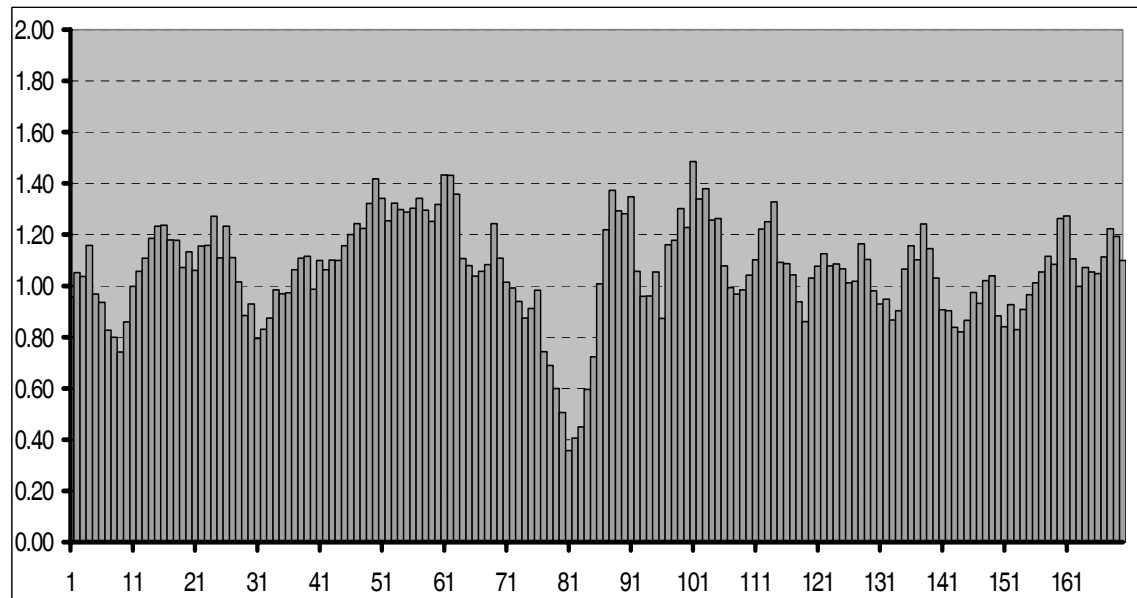
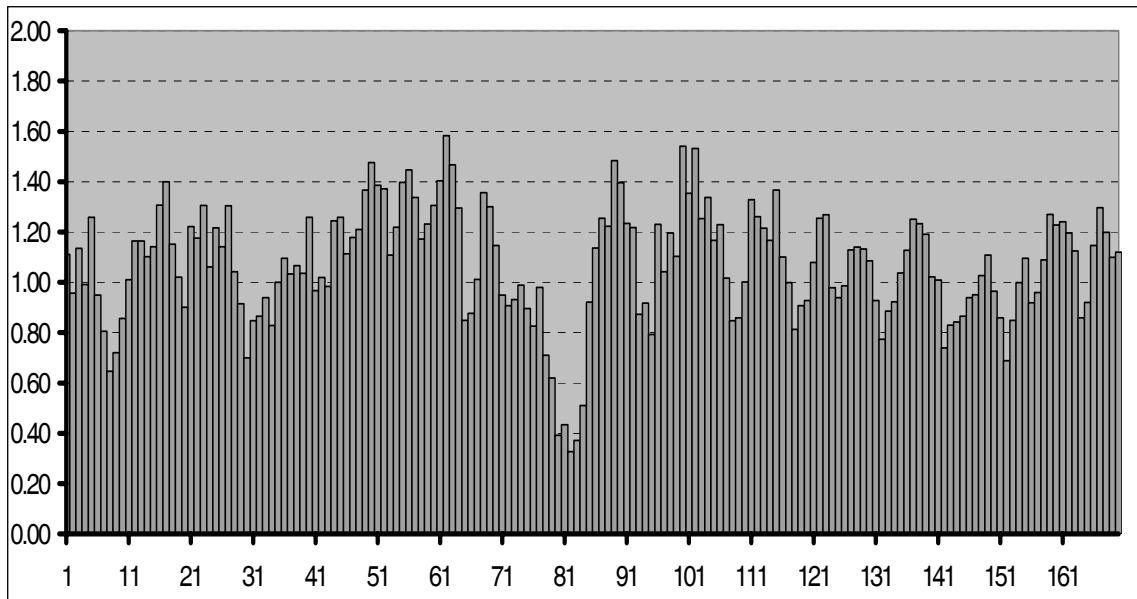
**Abbildung 4-11:** Falschfarbenkodierte Überlagerung der Hybridisierung der DNS- Probe 525/00 (rot) gegen eine Referenz- Probe (grün), die numerierten Spots entsprechen IA- PCR- Klonen aus der Xq21- Region

**Tabelle 4-6:** Auswertung der Mikroarray- Hybridisierung mit IA- PCR- Sonden für den Bereich zwischen 70 und 80 Mbp (von Xpter aus) für die Zelllinie 525/00: aufgeführt sind die Klonbezeichnung, die Lage des Klons, das ermittelte normierte Verhältnis der Signalintensitäten und das Ergebnis der Kontroll- PCR- Ansätze mit den spezifischen, internen IA- PCR- Primern

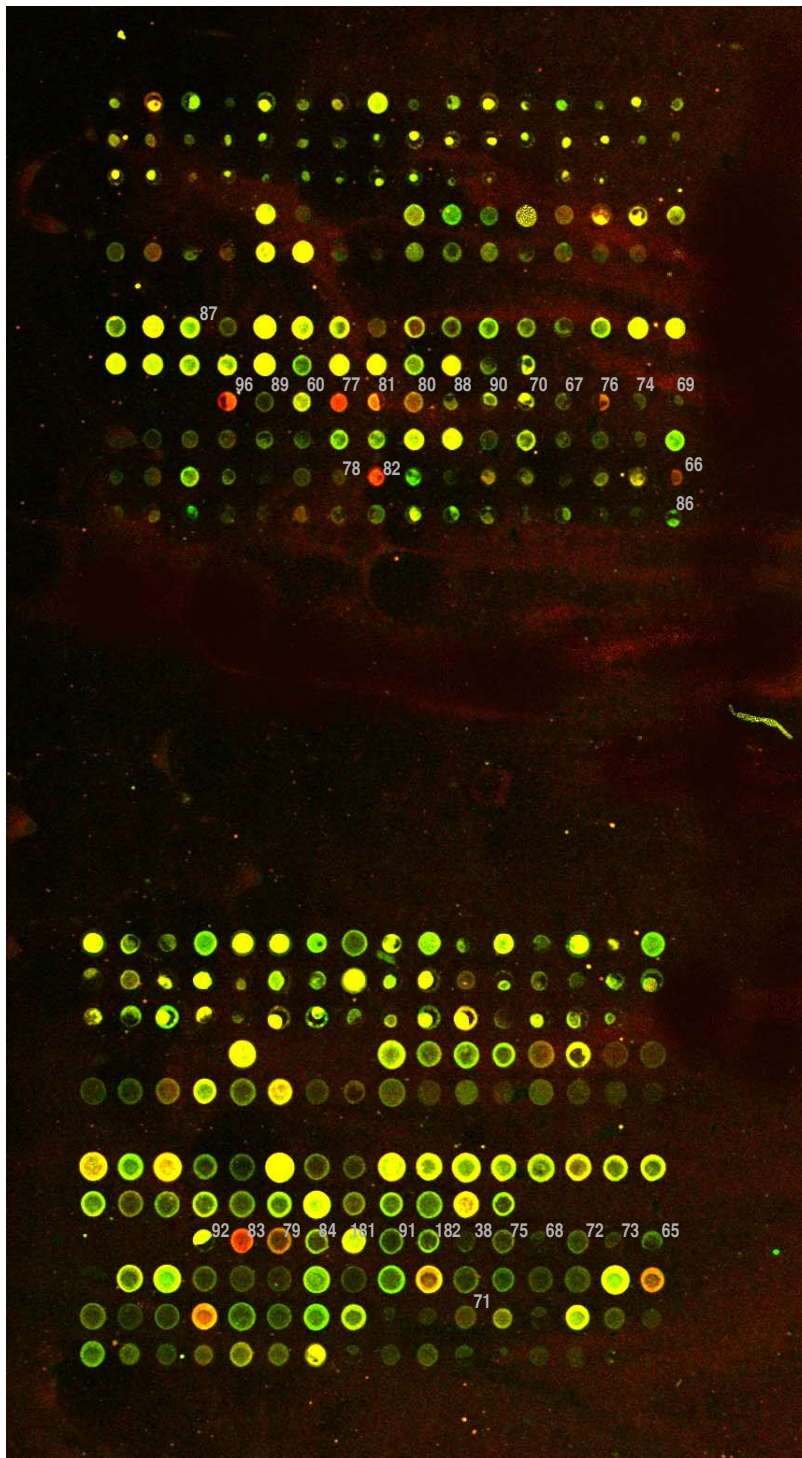
Index	Klon- Bezeichnung	Position	Signal- verhältnis	PCR- Ergebnis	Index	Klon- Bezeichnung	Position	Signal- verhältnis	PCR- Ergebnis
65	AC004677	70.756.612	1.00	-	79	AL160266	77.423.756	0.28	-
66	hs333e23	73.007.918	0.53	+	80	AL158820	79.074.886	0.59	-
67	Z84487_2	73.618.669	1.28	+	81	AL359885	80.154.487	0.36	-
68	Z84487_11	73.634.878	1.53	+	82	da145b8	81.150.346	0.38	-
69	AL049589	73.742.680	1.27	+	83	AL035451	81.150.396	0.25	-
70	Z94801	73.816.374	1.13	+	84	AC025202	82.876.366	0.53	-
71	hs465g10b	73.816.445	1.05	+	85	hs393p23	83.169.128	0.99	-
72	AL138743_6	74.041.011	0.72	+	86	652-12	87.114.284	1.49	+
73	AL138743	74.119.096	0.99	+	87	3019_1	88.073.039	1.00	+
74	AL138743_23	74.136.279	1.13	+	88	AL136362	88.323.926	1.33	+
75	AL109753	74.364.174	0.86	+	89	AL160266_2	88.358.928	1.38	+
76	AL162912	74.780.589	0.74	-	90	AL136362_2	88.358.928	1.78	+
77	AL031313	76.409.633	0.89	+	91	AL133274	88.723.181	1.11	+
78	dj1111n9	76.769.657	1.44	-	92	AL121869	88.867.588	0.95	+

**Abbildung 4-12:** Graphische Darstellung der für die Zelllinie 525/00 ermittelten, normierten Signalintensitätsverhältnisse für die entsprechend ihrer Lage geordneten IA- PCR- Klone, von Xpter (links) bis Xqter (rechts), die Zahlen geben den Positionsindex an und entsprechen dem Eintrag in der Tabelle, die erste Abbildung beinhaltet die tatsächlichen Meßwerte, in der mittleren Abbildung sind die Meßwerte durch Verwendung eines gleitenden Durchschnitts über drei Elemente, in der dritten Abbildung über fünf Meßwerte, geglättet





#### 4.3.4.3 Kartierung der Deletion in der DNS von der Zelllinie 303/99

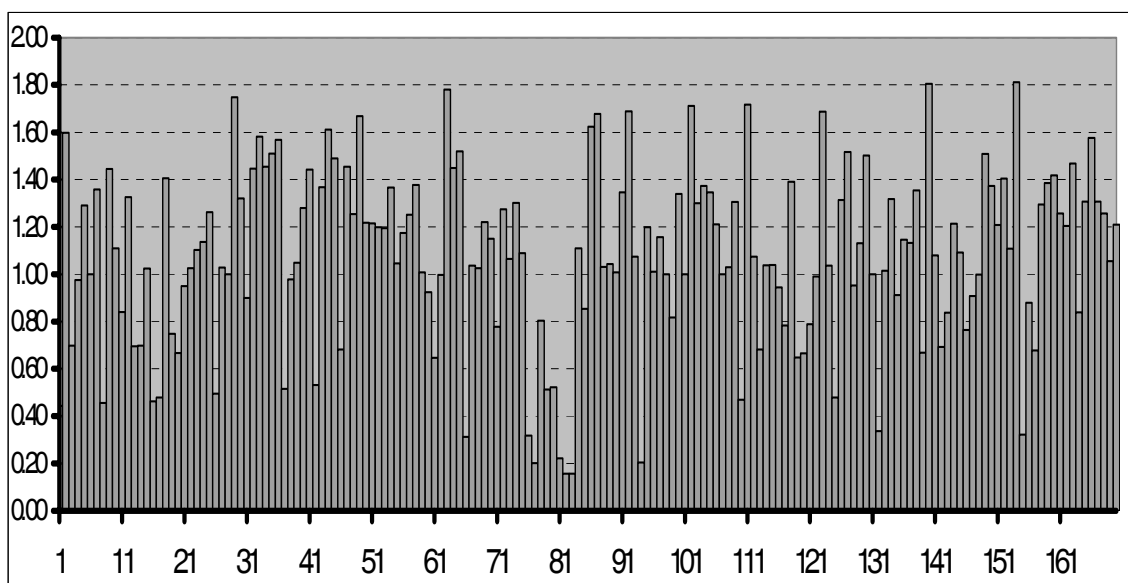


**Abbildung 4-13:** Falschfarbenkodierte Überlagerung der Hybridisierung der DNS- Probe 303/99 (rot) gegen eine Referenz- Probe (grün), die numerierten Spots entsprechen IA- PCR- Klonen aus der Xq21- Region

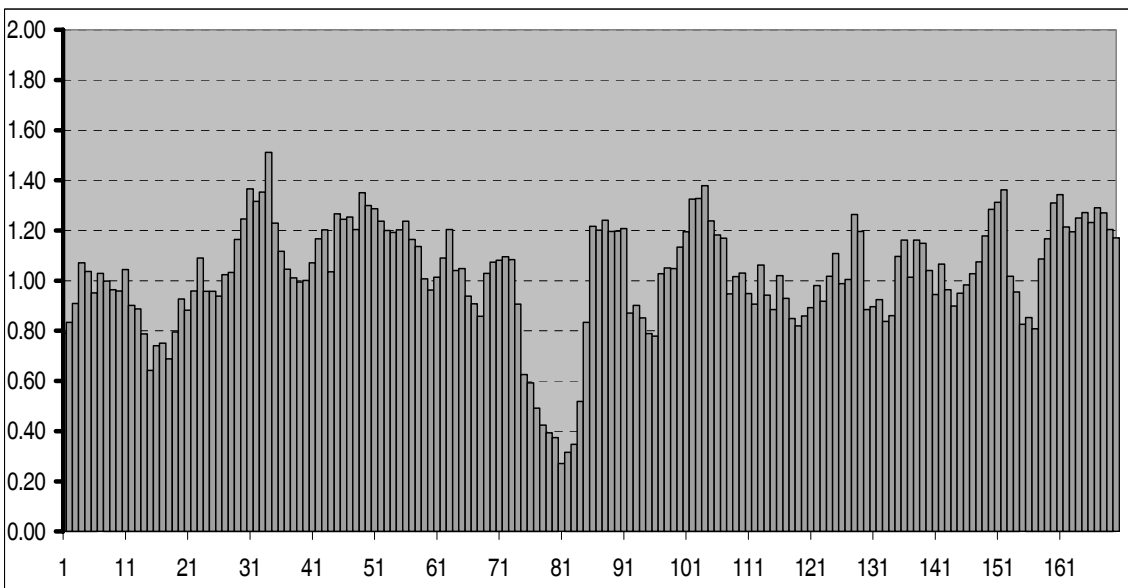
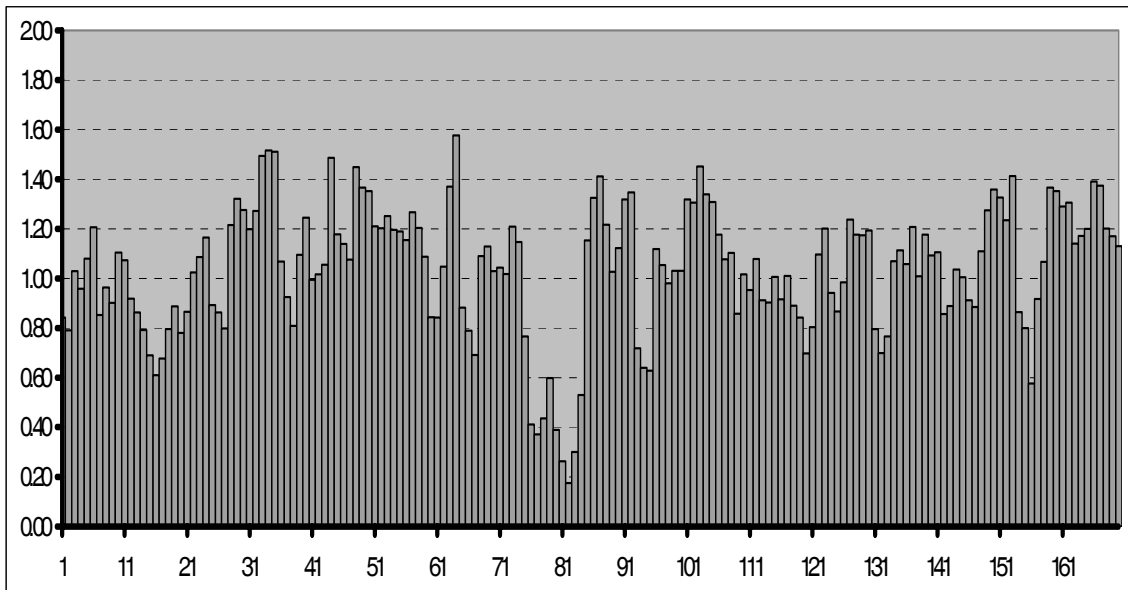
**Tabelle 4-7:** Auswertung der Mikroarray- Hybridisierung mit IA- PCR- Sonden für den Bereich zwischen 70 und 80 Mbp (von Xpter aus) für die Zelllinie 303/99, aufgeführt sind die Klonbezeichnung, die Lage des Klons, das ermittelte normierte Verhältnis der Signalintensitäten und das Ergebnis der Kontroll- PCR- Ansätze mit den spezifischen, internen IA- PCR- Primern

Index	Klon- Bezeichnung	Position	Signal- verhältnis	PCR- Ergebnis	Index	Klon- Bezeichnung	Position	Signal- verhältnis	PCR- Ergebnis
65	AC004677	70.756.612	1.52	+	79	AL160266	77.423.756	0.51	-
66	hs333e23	73.007.918	0.31	-	80	AL158820	79.074.886	0.52	-
67	Z84487_2	73.618.669	1.04	+	81	AL359885	80.154.487	0.22	-
68	Z84487_11	73.634.878	1.02	+	82	da145b8	81.150.346	0.16	-
69	AL049589	73.742.680	1.22	+	83	AL035451	81.150.396	0.16	-
70	Z94801	73.816.374	1.15	+	84	AC025202	82.876.366	1.11	-
71	hs465g10b	73.816.445	0.78	+	85	hs393p23	83.169.128	0.85	-
72	AL138743_6	74.041.011	1.27	+	86	652-12	87.114.284	1.62	+
73	AL138743	74.119.096	1.06	+	87	3019_1	88.073.039	1.68	+
74	AL138743_23	74.136.279	1.30	+	88	AL136362	88.323.926	1.03	+
75	AL109753	74.364.174	1.09	+	89	AL160266_2	88.358.928	1.04	+
76	AL162912	74.780.589	0.32	-	90	AL136362_2	88.358.928	1.01	+
77	AL031313	76.409.633	0.20	+	91	AL133274	88.723.181	1.35	+
78	Dj1111n9	76.769.657	0.80	-	92	AL121869	88.867.588	1.69	+

**Abbildung 4-14:** Graphische Darstellung der für die Zelllinie 303/99 ermittelten, normierten Signalintensitätsverhältnisse für die entsprechend ihrer Lage geordneten IA- PCR- Klone, von Xpter (links) bis Xqter (rechts), die Zahlen geben den Positionsindex an und entsprechen dem Eintrag in der Tabelle, die erste Abbildung beinhaltet die tatsächlichen Meßwerte, in der mittleren Abbildung sind die Meßwerte durch Verwendung eines gleitenden Durchschnitts über drei Elemente, in der dritten Abbildung über fünf Meßwerte, geglättet







#### 4.3.5 Zusammenfassung der Meßergebnisse

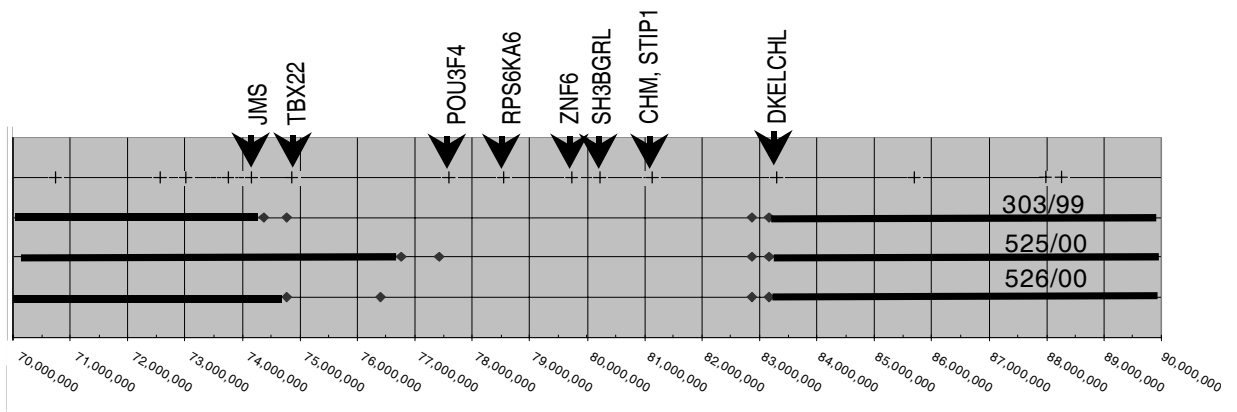
Für die Bestimmung der Lage und der Größe der Deletion mußten die vorhandenen Meßwerte den Kategorien ‚vorhanden‘ oder ‚deletiert‘ zugeordnet werden. Dazu wurden zunächst die Standardabweichungen aller Meßwerte bestimmt:

Zelllinie:	Standardabweichung:
526/00	0,32
525/00	0,34
303/99	0,37

Klone, deren relative Signalintensitäten um mindestens eine Standardabweichung kleiner war als eins (der normalisierten relativen Signalintensität für nicht deletierte IA-PCR- Produkte), wurden als deletiert eingestuft, wenn auch die unmittelbar benachbarten Klone ähnliche Werte aufwiesen. Die Aussagekraft dieses Verfahrens wird durch die Einführung des gleitenden Durchschnitts über drei oder fünf benachbarte Meßwerte erhöht. Für die Identifizierung einer Deletion wurde dieser Durchschnitt herangezogen, jedoch wurden die Grenzen des deletierten Bereichs wurden zusätzlich anhand der nicht geglätteten Meßwerten abgeleitet. Die Ausdehnung dieser Region konnten dann auf ein Intervall eingeeengt werden, das durch die Positionen des letzten nicht deletierten und des ersten deletierten Klons bestimmt wird. Auf diese Weise wurden die Grenzen der Deletionen in der Patienten- DNS wie folgt kartiert:

**Tabelle 4-8:** Kartierung der Deletionen in der Xq21- Region nach den Matrix- CGH- Ergebnissen, die Bestimmung ergibt sich aus den den Indices zugeordneten Positionen

Zelllinie	Klone in der Deletion [Index]	Klone außerhalb der Deletion [Index]	Minimale Größe der Deletion [Mbp]	Maximale Größe der Deletion [Mbp]
526/00	77-84	1-76, 85-171	6,467	8,389
525/00	79-84	1-78, 85-171	5,453	6,400
303/99	76-84	1-75, 85-171	8,096	8,805



**Abbildung 4-15:** Graphische Darstellung der ermittelten Daten zur Ausdehnung der Deletionen in der Xq21-Region entlang dem Positionsindex (Abstand Xpter in bp), die Deletionen sind als Unterbrechungen der durchgezogenen Linien dargestellt, die Positionen der in Tabelle 4-8 aufgeführten IA-PCR-Klone sind als Rauten gekennzeichnet (jeweils paarweise für den letzten noch vorhandenen bzw. den ersten deletierten IA-PCR-Klon, in der obersten Reihe ist die Position der bekannten Gene in der Xq21-Region markiert, unmittelbar in oder an der Deletion gelegene Gene sind namentlich aufgeführt (Stand März 2001)

#### 4.3.6 Bewertung und genetische Implikationen

Bei allen drei Patienten konnte wie erwartet eine große Deletion in der Xq21-Region durch Untersuchung mittels Matrix-CGH nachgewiesen werden. Bis auf eine Ausnahme (Klon hs333e23) bildeten die als deletiert ermittelten IA-Sequenzen eine Reihe von topographisch aufeinanderfolgenden IA-PCR-Klonen. Die durch die Matrix-CGH ermittelten Zuordnungen zu den Kategorien ‚vorhanden‘ oder ‚nicht vorhanden‘ wurden durch PCR mit den spezifischen internen Primerpaaren auf der Patienten-DNS verifiziert. Wie aus den Tabellen 4-5 bis 4-7 zu entnehmen ist, stimmen die Ergebnisse aus Matrix-CGH und PCR zu über 90% überein. Die erreichte Auflösung von durchschnittlich 590 kB ermöglichte die Zuordnung eines einzelnen Gens (TBX22, Laugier-Anfossi und Villard, 2000) zu einem genomischen Abschnitt, der im Patient 303/99 deletiert in den beiden anderen Patienten aber (sehr wahrscheinlich) vorhanden ist. Patient 303/99 hat zusätzlich zu den auch bei den anderen beiden Patienten auftretenden Symptomen eine Kiefern-Lippen-Gaumenspalte (siehe Tabelle 4-3). Aufgrund seiner Zugehörigkeit zu den T-Box Genen, einer Klasse von entwicklungsbiologisch relevanten Transkriptionsfaktoren, erschien TBX22 als

---

geeignetes Kandidatengen, diesen zusätzlichen Phänotyp zu erklären. Tatsächlich konnten nach Abschluß der vorliegenden Arbeit Mutationen in TBX22 als Ursache für die nichtsyndromische, X- chromosomal vererbte Form der Kiefer-, Lippen-, Gaumenspalte (CPX) nachgewiesen werden (Baybrook *et al.*, 2001, Herr *et al.*, 2003)

## 4.4 Bestimmung des Auflösungsvermögens für kleine Kopienzahlunterschiede anhand von multiplen X- Zelllinien

### 4.4.1 Grundlagen

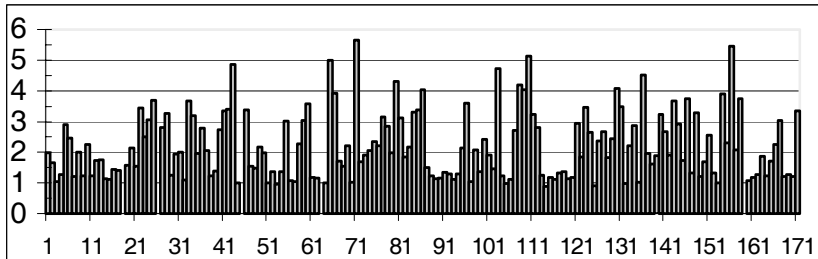
Zur Bestimmung der experimentellen Auflösungsgrenze für kleine Kopienzahlunterschiede wurde eine Reihe von vergleichenden Hybridisierungen durchgeführt. Dabei wurden die IA- Amplikons von Patienten mit mehreren X- Chromosomen mit denen von normal männlicher (XY) und normal weiblicher (XX) Referenz- DNS verglichen. Im einzelnen wurden dazu folgende Konstellationen untersucht:

Probe 1	Probe 2	erwartetes Verhältnis
XXX	XX	1,5
XXXX oder XX	XX oder XY	2
XXXXX	XX	2,5
XXX	XY	3
XXXX	XY	4
XXXXX	XY	5

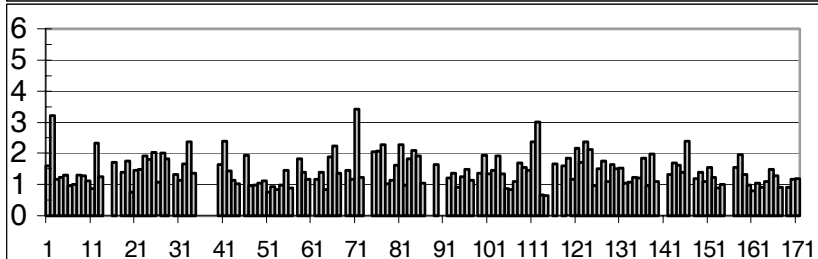
Die erhaltenen Meßwerte wurden wieder über die Signale der 96 autosomalen IA- PCR- Klone vom Chromosom 22 normiert.

## Darstellung der Meßwerte

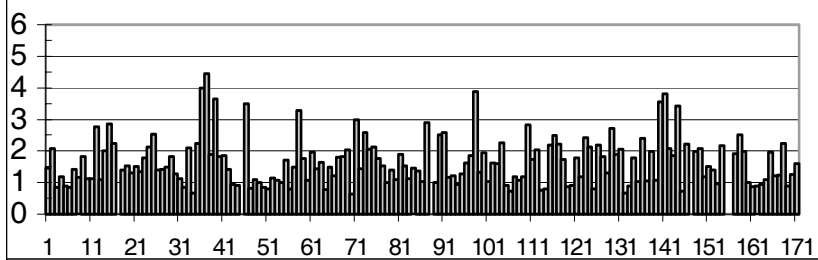
Im folgenden sind die ermittelten Signalintensitätsverhältnisse aus den zuvor auf-



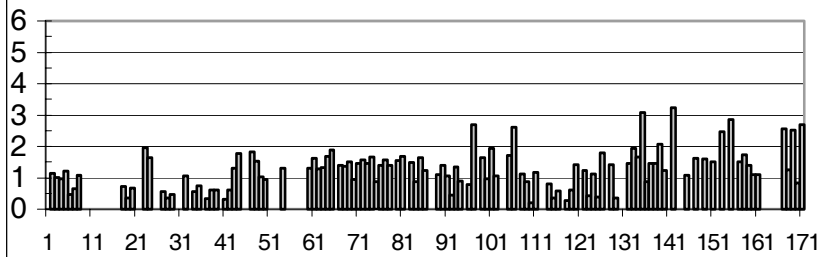
**Abbildung 4-16:** Ergebnisse für die Hybridisierung der Zelllinien 377/99 und 290/99, entsprechend 5:1



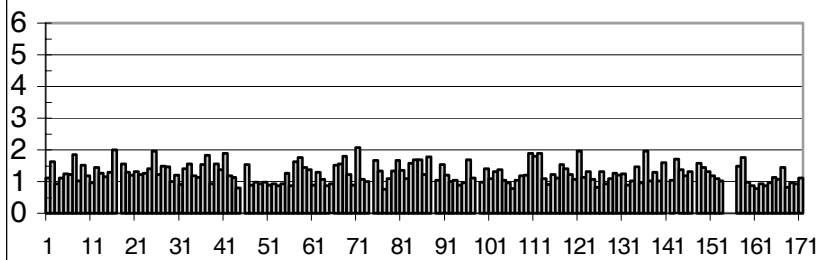
**Abbildung 4-17:** Ergebnisse für die Hybridisierung der Zelllinien 372/99 und 290/99, entsprechend 4:1



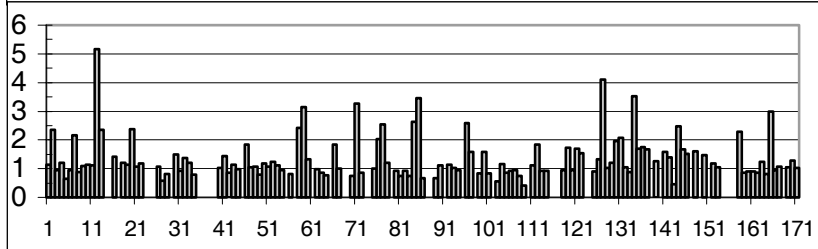
**Abbildung 4-18:** Ergebnisse für die Hybridisierung der Zelllinien 451/99 und 290/99, entsprechend 3:1



**Abbildung 4-19:** Ergebnisse für die Hybridisierung der Zelllinien 364/99 und 453/99, entsprechend 5:2



**Abbildung 4-20:** Ergebnisse für die Hybridisierung der Zelllinien 363/99 und 453/99, entsprechend 4:2



**Abbildung 4-21:** Ergebnisse für die Hybridisierung der Zelllinien 451/99 und 453/99, entsprechend 3:2

geführten Experimenten graphisch dargestellt. Die Ordinate zeigt die normalisierten Signalintensitätsverhältnisse, die Abzisse ermöglicht die Zuordnung zum Index der untersuchten IA- PCR- Klone. Das Kopienzahlverhältnis eins zu eins wurde nicht gesondert untersucht, da die autosomalen Kontrollklone diesen Fall wiedergeben. Die Simulation einer unvollständigen Deletion (eins zu zwei) ergibt sich aus den Werten für das Verhältnis zwei zu eins, da das gleiche Verhältnis nur reziprok ausgelesen werden muß. Spots, die weniger als die doppelte Intensität des lokalen Hintergrunds aufwiesen, wurden von der weiteren Auswertung ausgeschlossen, da diese Werte als nicht signifikant betrachtet wurden.