

TECHNISCHE UNIVERSITÄT MÜNCHEN

Institut für Experimentelle Genetik

Kopplungsanalysen und Bestimmung von
Kopienzahlvariationen mit Oligonukleotid-Mikroarrays

Janine Wagenstaller

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors für Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:

Univ.-Prof. Dr. A. Gierl

Prüfer der Dissertation:

1.

Univ.-Prof. Dr. M. Hrabé de Angelis

2.

Univ.-Prof. Dr. Chr.-C. Schön

Die Dissertation wurde am 06.12.2007 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 04.03.2008 angenommen.

Inhaltsverzeichnis

Inhaltsverzeichnis.....	1
Abkürzungsverzeichnis.....	5
Zusammenfassung.....	7
A Einleitung	9
1 Mikroarrays und ihre Anwendung in der Wissenschaft.....	9
1.1 <i>Expressionsarrays</i>	9
1.2 <i>Vergleichende genomische Hybridisierung</i>	10
1.3 <i>Oligonukleotid-Arrays</i>	10
2 Untersuchungen mit SNP-Oligonukleotid-Mikroarrays.....	12
2.1 Kopplungsanalysen mit SNP-Oligonukleotid-Mikroarrays.....	13
2.1.1 <i>Grundbegriffe der Kopplungsanalysen</i>	13
2.1.2 <i>Vergleich von Mikrosatelliten versus SNPs in Kopplungsanalysen</i>	15
2.1.3 <i>Krankheitsgeschichten der untersuchten Familien</i>	16
2.1.3.1 <i>Der Phosphatstoffwechsel im Zusammenhang mit bisher bekannten Formen der Hypophosphatämien</i>	16
2.1.3.2 <i>Erbliche hypophosphatämische Rachitis mit Hyperkalziurie (HHRH)</i>	17
2.1.3.3 <i>Autosomal-rezessive Hypophosphatämie (ARHP)</i>	18
2.1.3.4 <i>Spondylo-oculares Syndrom (SOS)</i>	18
2.2 Kopienzahlvariationen (CNVs).....	19
2.2.1 <i>Mechanismen zur Entstehung von CNVs</i>	20
2.2.1.1 <i>Nicht-allelele homologe Rekombination (NAHR)</i>	21
2.2.1.2 <i>Nicht-homologe Endverknüpfung (NHEJ)</i>	22
2.2.2 <i>Pathologien durch CNVs</i>	25
2.2.3 <i>Mentale Retardierung</i>	26
3 Aufgabenstellung.....	28
B Material und Methoden	29
1 Patientenmaterial.....	29
2 Verwendete Materialien und Geräte.....	30
2.1 Chemikalien.....	30
2.2 Enzyme.....	30

2.3	Kits.....	31
2.4	Technische Geräte.....	31
2.5	Datenbanken und Computerprogramme.....	32
3	DNA-Isolierung.....	32
4	Polymerase-Kettenreaktion (PCR).....	33
4.1	Standard-PCR.....	33
4.2	PCR für das Affymetrix 10K- und 100K-Protokoll.....	34
4.3	RT-PCR und <i>nested</i> -PCR.....	35
4.4	RACE-PCR.....	36
4.5	PCR für das Idaho-Protokoll.....	36
4.6	PCR für den iPLEX Gold Assay.....	37
4.7	Quantitative PCR.....	37
4.8	PCR für Mikrosatellitenanalyse.....	38
5	Reinigung von PCR-Produkten.....	38
5.1	Reinigung durch Ultrazentrifugation.....	38
5.2	Reinigung durch Gelextraktion.....	39
6	DNA-Sequenzierung.....	39
6.1	Sequenzreaktion.....	39
6.2	Fällung der Sequenzreaktion.....	40
6.3	Sequenzanalyse.....	40
7	Klonierung von PCR-Produkten.....	40
7.1	Ligation.....	41
7.2	Transformation.....	41
7.3	Präparation von Plasmid-DNA.....	42
7.4	Restriktionsspaltung.....	42
8	Mutationsscreening mit DNA-Schmelzkurvenanalyse (Idaho).....	43
9	Inaktivierung von X-Chromosomen.....	43
10	Affymetrix GeneChip Mapping Assay.....	44
10.1	GeneChip Mapping 10K 2.0 Assay.....	44
10.1.1	<i>Xba</i> I-Restriktionsenzymverdau.....	45
10.1.2	<i>Ligation</i>	45

10.1.3	PCR.....	46
10.1.4	Fragmentierung.....	46
10.1.5	Biotin-Endmarkierung der Fragmente.....	46
10.1.6	Hybridisierung.....	47
10.1.7	Waschen, Färben und Scannen.....	48
10.2	GeneChip Mapping 100K Assay.....	49
10.2.1	XbaI-Restriktionsenzymverdau.....	49
10.2.2	Ligation.....	49
10.2.3	PCR.....	50
10.2.4	Fragmentierung.....	50
10.2.5	Biotin-Endmarkierung der Fragmente.....	51
10.2.6	Hybridisierung.....	51
10.2.7	Waschen, Färben und Scannen.....	52
11	iPLEX Gold Assay.....	52
11.1	iPLEX Gold PCR.....	52
11.2	SAP-Behandlung.....	53
11.3	iPLEX Reaktion.....	53
11.4	Reinigung.....	53
11.5	MALDI TOF Massenspektrometeranalyse.....	54
12	Feinkartierung mit Mikrosatelliten.....	54
13	Analyse der Daten.....	54
13.1	Analyse der 10K-Affymetrix-Daten.....	54
13.2	Analyse der 100K-Affymetrix-Daten.....	56
C	Ergebnisse.....	59
1	Kopplungsanalysen mit SNP-Oligonukleotid-Mikroarrays.....	59
1.1	Kopplungsanalyse in 5 Familien mit HHRH.....	59
1.1.1	Identifizierung einer Kandidatenregion in den Familien mit HHRH.....	59
1.1.2	Mutationsanalyse in der Kandidatenregion 9q34.3 in den Familien mit HHRH...	62
1.2	Kopplungsanalyse in 3 Familien mit ARHP.....	64
1.2.1	Identifizierung einer Kandidatenregion in den Familien mit ARHP.....	64
1.2.2	Mutationsanalyse in der Kandidatenregion 4q21 in den Familien mit ARHP.....	66

1.3	Kopplungsanalyse in einer Familie mit SOS.....	67
1.3.1	<i>Identifizierung einer Kandidatenregion in der Familie mit SOS.....</i>	67
1.3.2	<i>Mutationsanalyse in der Kandidatenregion 17 in der Familie mit SOS.....</i>	70
2	CNVs.....	71
2.1	Datenanalyse der 100K-Arrays.....	71
2.2	Kandidatenregionen.....	73
2.3	CNVs, die von low-copy repeats flankiert sind.....	76
2.4	CNVs, die nicht von low-copy repeats flankiert sind.....	82
2.5	CNV mit einem deletierten Gen.....	87
2.6	Potentiell pathologische CNV mit einem deletierten Gen.....	89
2.7	CNVs in Patienten mit Translokationen.....	91
D	Diskussion	93
1	Kopplungsanalysen.....	93
2	Kopienzahlvariationen.....	97
2.1	Unterscheidung zwischen Polymorphismen und pathogenen CNVs.....	98
2.2	Pathogene CNVs.....	99
2.2.1	<i>CNVs, die von low-copy repeats (LCRs) flankiert werden.....</i>	99
2.2.2	<i>CNVs, ohne flankierende LCRs.....</i>	101
2.2.3	<i>Bestätigung der Bruchpunkte.....</i>	102
2.2.4	<i>CNVs mit nur einem deletierten Gen.....</i>	102
2.3	CNVs an Bruchpunkten von Translokationen.....	104
2.4	Eignung von SNP-Arrays zur Identifizierung von CNVs.....	104
E	Literaturverzeichnis	105
F	Anhang	
	Primertabelle.....	I
	Mikrosatellitentabelle.....	IX
	Tabellenverzeichnis.....	X
	Abbildungsverzeichnis.....	XI
	Danksagung.....	XII
	Lebenslauf.....	XIII
	Erklärung.....	XV

Abkürzungen

A	Adenin
ADHR	autosomal-dominant hypophosphatemic rickets
ARHP	autosomal-recessive hypophosphatemic rickets
BAC	Bacterial artificial chromosome
bp	Basenpaar
BSA	bovine serum albumin
C	Cytosin
cDNA	komplementäre Desoxyribonukleotidsäure
CGH	comparative genomic hybridization
cM	centi-Morgan
CNV	Kopienzahlvariation
dATP	2'-Desoxyadenosin-5'-triphosphat
DBS	Doppelstrangbruch
dCTP	2'-Desoxycytidin-5'-triphosphat
ddNTP	2'-Didesoxynukleosid-5'-triphosphat
dGTP	2'-Desoxyguanin-5'-triphosphat
DGV	Database of genomic variants
DMSO	Dimethylsulfoxid
DNA	Desoxyribonukleotidsäure
dNTP	2'-Desoxynukleotid-5'-triphosphat
dTTP	2'-Desoxythymidin-5'-triphosphat
EDTA	Ethylendiamintetraacetat
EST	expressed sequence tag
FISH	Fluoreszenz in situ Hybridisierung
F-Primer	forward primer
G	Guanin
HHRH	hereditary hypophosphatemic rickets with hypercalciuria
HS-DNA	herring sperm DNA
kb	Kilobasen
LCR	Low copy repeats
LTR	long tandem repeat
M	Molar
MAD	mittlere absolute Abweichung
Mb	Megabasen
MES	2-(N-Morpholino)-ethansulfonsäure
min	Minute
MM	mismatch Probe
MR	mentale Retardierung
mRNA	messenger RNA
NaCl	Natriumchlorid
NAHR	nicht-allelische homologe Rekombination
NHEJ	nicht-homologe Endverknüpfung
OMIM	Online Mendelian Inheritance in Man
PCR	Polymerasekettenreaktion
PM	perfect match Probe
PTH	Parathormon
qPCR	quantitative PCR
RACE	rapid amplification of cDNA ends
RNA	Ribonukleinsäure
rpm	rounds per minute
R-Primer	reverse primer
RT	Raumtemperatur
RT-PCR	reverse Transkriptase Polymerasekettenreaktion
SAPE	Streptavidin, R-phycoerythrin Konjugat - 1
SD	Standardabweichung
sec	Sekunde
SNP	Single nucleotide polymorphism
SNR	Siganl-Rausch-Verhältnis

SOS	spondylo-oculares Syndrom
T	Thymidin
TIO	Tumor-induced osteomalacia
TMACL	tetramethylammoniumchlorid
U	Unit (Enzymeinheit)
ÜN	über Nacht
XLH	X-linked hypophosphatemia

Zusammenfassung

Kopplungsanalysen im menschlichen Genom wurden in den letzten 15 Jahren mit Mikrosatellitenmarkern durchgeführt. In den Jahren 2003/2004 standen die ersten kommerziellen SNP-Oligonukleotid-Mikroarrays zur Verfügung, deren Entwicklung im Rahmen des Internationalen HapMap-Projektes vorangetrieben wurde. Im ersten Teil der Arbeit wurde untersucht, ob und wie gut sich diese Arrays für Kopplungsuntersuchungen eignen. Im zweiten Teil der Arbeit wurde getestet, inwieweit und mit welcher Auflösung sich Veränderungen in der Kopienzahl genomischer DNA mit SNP-Oligonukleotid-Mikroarrays feststellen lassen.

Im ersten Teil der Arbeit wurde die Methode der 10K-Arrays von Affymetrix im Labor etabliert. Mit diesen 10K-Arrays, war es möglich, ca. 10000 Genotypen pro Patienten in wenigen Tagen zu erhalten. In der vorliegenden Arbeit werden die Ergebnisse der Kopplungsanalysen von 3 verschiedenen Erkrankungen vorgestellt. Bei 2 Erkrankungen, der hypophosphatämischen Rachitis mit Hyperkalziurie (HHRH) und der autosomal-rezessiven Form der Hypophosphatämie (ARHP) konnten Mutationen in einem Gen identifiziert werden, die als Ursache der Erkrankung angesehen werden. Bei dem spondylo-ocularem Syndrom (SOS) konnte eine Kandidatenregion identifiziert werden.

SNP-Oligonukleotid-Mikroarrays haben sich inzwischen gegenüber den Mikrosatelliten für Kopplungsanalysen durchgesetzt. Die Vorteile der Mikroarrays sind der geringe zeitliche Aufwand, die geringe Ausgangsmenge an DNA und die Möglichkeit der industriellen Herstellung der Arrays. Es gilt inzwischen als gesichert, dass SNP-Oligonukleotid-Mikroarrays mit mehr als 10.000 Markern einen größeren Informationsgehalt besitzen als 300 bis 400 Mikrosatellitenmarker, die früher für Kopplungsuntersuchungen verwendet wurden. In der Arbeit wurde gezeigt, dass der Informationsgehalt einzelner Arraytypen nicht für jede Untersuchung ausreichend ist. Bei der Untersuchung der Familien mit HHRH konnte mit einem Arraytyp (Linkage IV Panel von Illumina) eine signifikante Kopplung nachgewiesen werden, während sie mit einem anderen Arraytyp (10K-Array von Affymetrix) nicht detektierbar war. Dieser Array enthielt in der kritischen Region nur einen SNP. Weiterhin zeigte sich, dass Mikrosatellitenmarker benutzt werden können, um die Kandidatenregion weiter einzugrenzen. In den Familien mit SOS konnte die Region von 10 Mb auf 5 Mb eingegrenzt werden.

Die SNP-Oligonukletid-Mikroarrays wurden inzwischen verbessert: zum einen wurde die Auswahl der SNPs für die dedizierten Linkage-Arrays optimiert, zum anderen können für Kopplungsuntersuchungen auch Arrays höherer Dichte verwendet werden und aus diesen selbst eine optimale Auswahl von SNPs vorgenommen werden.

Mit der herkömmlichen zytogenetischen Diagnostik, der GTG-Bänderung, können Aberrationen im Genom ab einer Größe von 5-10 Mb detektiert werden. Im Laufe der Arbeit zeigte sich, dass sich SNP-Oligonukleotid-Mikroarrays auch zur Identifizierung von Veränderungen in der Kopienzahl genomischer DNA eignen. In dieser Arbeit wurden 67 Kinder mit ungeklärter mentaler Retardierung mit 100K-Arrays von Affymetrix untersucht. In die Studie wurden nur Kinder mit unauffälligen Karyogrammen nach GTG-Bänderung aufgenommen. Es wurden 11 pathogene CNVs identifiziert, die entweder *de novo* entstanden sind und/oder mit bereits bekannten Deletions- bzw. Duplikationssyndromen übereinstimmen. Die 11 CNVs (8 Deletionen und 3 Duplikationen) variieren in der Größe zwischen 0.2 Mb und 7.5 Mb. 5 der CNV sind von *low-copy repeats* flankiert. Zwei davon, auf Chromosom 15q25.2 und Xp22.31 wurden bisher in der Literatur nicht beschrieben und könnten neue Deletions- bzw. Duplikationssyndrome darstellen. In zwei Patienten konnten Deletionen identifiziert werden, die nur ein einzelnes Gen betreffen. Zusätzlich zu den 67 Patienten mit unauffälligen Karyogrammen wurden noch 5 Patienten mit scheinbar balancierten *de novo* Translokationen untersucht. In 4 Fällen wurden Deletionen mit 1.1 Mb bis 14.3 Mb in den Bruchpunktregionen identifiziert.

Das Auflösungsvermögen von SNP-Oligonukleotid-Mikroarrays, übertrifft das von herkömmlichen Diagnostikmethoden der Zytogenetik. Es zeigte sich das das Signal-Rausch-Verhältnis stark von den experimentellen Bedingung abhängt. Mit der Entwicklung von neuen Arrays und Protokollen, die experimentelle Fehler reduzieren und ein besseres Signal-Rausch-Verhältnis erreichen, wird sich die Auflösung weiter verbessern. Es ist zu erwarten, dass diese Technik schon bald Eingang in die Routinediagnostik finden wird.

A Einleitung

1. Mikroarrays und ihre Anwendung in der Wissenschaft

Im Rahmen des *Human Genome Projects* und des *International HapMap Projects* wurde die Entwicklung neuer Technologien vorangetrieben, die es ermöglichen eine große Anzahl von Genen gleichzeitig zu untersuchen.

Eine dieser Techniken sind Mikroarrays. Sie bieten verschiedene Optionen zur Untersuchung des Genoms von Menschen und Tieren. Die häufigsten Anwendungen sind Untersuchungen der mRNA-Expression und Bestimmung der Genotypen und Kopienzahlvariationen (*Copy Number Variations*, CNVs) von genomischer DNA.

Für diese 3 Anwendungsgebiete ist die Technik gleich. Auf der Matrix des Arrays, einer Glasoberfläche, werden einzelsträngige Oligonukleotide an festgelegten Orten fixiert. Die zu untersuchende Probe wird mit einem Fluoreszenzfarbstoff markiert, als Einzelstrang auf den Array hybridisiert und bildet mit den komplementären, immobilisierten Oligonukleotiden einen DNA-Doppelstrang. Über Bildanalyseprogramme wird anschließend die zu untersuchende Probe auf dem Array lokalisiert und anhand der Fluoreszenz quantifiziert.

1.1 Expressionsarrays

Eine mögliche Anwendung von Mikroarrays ist die Untersuchung von Genexpressionsmustern. Um Gene auf ihre Expression hin untersuchen zu können, werden Arrays mit cDNA-Klonen oder genspezifischen Oligonukleotiden verwendet, die den zu untersuchenden Genen entsprechen. Aus dem zu untersuchenden Gewebe wird RNA gewonnen, mit der reversen Transkriptase in cDNA umgeschrieben, mit einem Fluoreszenzfarbstoff markiert und auf den Array hybridisiert. Die Fluoreszenzintensitäten geben Aufschluss darüber, in welcher Menge die markierte cDNA vorhanden ist und wie hoch bzw. niedrig dieses Gen exprimiert ist.

1.2 Vergleichende genomische Hybridisierung (CGH)

Die Techniken zur Identifizierung von Kopienzahlvariationen der DNA entwickelten sich aus der Karyotypisierung, über die Bandenfärbung von Metaphasepräparaten, der *Fluoreszenz in situ Hybridization* (FISH) zu der vergleichenden genomischen Hybridisierung (*Comparative Genomic Hybridization*, CGH).

Die vergleichende genomische Hybridisierung bedient sich einer kompetitiven Hybridisierung von verschiedenfarbig markierten Fragmenten einer zu untersuchenden DNA und einer Referenz-DNA auf denselben Array, um Unterschiede in der Kopienzahl festzustellen. Als Zielsequenzen zur Bindung der DNA-Fragmente werden genomische Klone (BACs) verwendet. Das Auflösungsvermögen der CGH ist abhängig von dem Abstand zwischen den Klonen und der Größe der verwendeten Klone. In der Regel variiert die Auflösung zwischen 50 kb und mehreren Mb (Vissers et al. 2003, Veltman et al. 2002, Feuk et al. 2006).

1.3 Oligonukleotid-Arrays

Die array-CGH war die erste Methode mit Mikroarrays, die entwickelt wurde, um genomweite Untersuchungen auf Kopienzahlvariationen durchzuführen.

Die array-CGH ist eine Weiterentwicklung der konventionellen CGH, bei der die Technik der CGH mit der Technik der Expressionsarrays kombiniert wurde. Als Hybridisierungstargets werden nicht mehr BAC-Klone verwendet, sondern definierte DNA-Fragmente, die auf dem Array immobilisiert werden. Für die Analyse werden Test- und Referenz-DNA mit verschiedenen Farbstoffen markiert und auf einem Array kohybridisiert. Numerische Veränderungen in der Test-DNA können durch Farbunterschiede identifiziert werden. Der erste Array in diesem Format wurde von Lucito et al. als *Representational Oligonucleotide Microarray Analysis* (ROMA) vorgestellt. Diese Methode reduziert die Komplexität des Genoms vor der Hybridisierung durch den Verdau der Test-DNA und der Referenz-DNA mit einem Restriktionsenzym und anschließender Amplifikation (Lucito et al. 2003).

Zu Beginn der Arbeit kamen die ersten kommerziellen Oligonukleotidarrays auf den Markt, die mit SNPs arbeiteten. *Single Nucleotide Polymorphisms* (SNPs) sind polymorphe Variationen in der DNA auf dem Niveau von einzelnen Nukleotiden. Man erhoffte sich durch die Entwicklung von SNP-Oligonukleotid-Mikroarrays, dass sie, im Vergleich zu der array-CGH, eine industrielle Herstellung ermöglichen, schneller in der Anwendung und universeller im Einsatz sein würden. Ein großer Vorteil der SNP-Mikroarrays ist, dass man bei der Analyse sowohl die Genotypen der SNPs erhält, als auch Kopienzahlvariationen identifizieren kann.

Ursprünglich wurden die SNP-Oligonukleotid-Mikroarrays für Assoziationsstudien oder für Kopplungsuntersuchungen entwickelt (Zhao et al. 2004, Slater et al. 2005, Huang et al. 2004, Bignell et al. 2004). Erst in den letzten 3 Jahren wurden SNP-Oligonukleotid-Mikroarrays auch zur Identifizierung von Variationen in der Kopienzahl genomischer DNA verwendet. Es wurden Studien zur Untersuchung von submikroskopischen Kopienzahlvariationen mit array-CGH und SNP-Oligonukleotid-Mikroarrays durchgeführt. Dabei wurden Deletionen und Duplikationen von einigen Kilobasen (kb) bis hin zu mehreren Megabasen (Mb) detektiert (Iafate et al. 2004; Sebat et al. 2004; Sharp et al. 2005; Tuzun et al. 2005; Conrad et al. 2006; Hinds et al. 2006; McCarroll et al. 2006; Locke et al. 2006, Redon et al. 2006).

Tabelle 1 Vergleich verschiedener Methoden zur Identifizierung von CNVs.

Methode	Translokationen	Inversionen	große CNVs (>50 kb)	CNVs (1-50 kb)	kleine Sequenzvarianten (<1 kb)
Karyotypisierung	ja (>3 Mb)	ja (>3 Mb)	ja (>3 Mb)	nein	nein
BAC-Array	nein	nein	ja (>50 kb)	nein	nein
ROMA	nein	nein	ja (>35 kb)	ja (>35 kb)	nein
SNP-Array	nein	nein	ja	ja	ja (SNPs)
Genotypisierung mit Mikrosatelliten	nein	nein	ja (Deletionen)	ja (Deletionen)	ja

(Tabelle modifiziert nach Feuk et al. 2006)

Die oben beschriebenen oder erwähnten Methoden analysieren die DNA entweder genomweit oder fokussieren sich auf definierte Bereiche des Genoms. Die Auflösung, die dabei erreicht werden kann ist sehr variabel. Auch die Art der detektierbaren CNVs ist bei den einzelnen Methoden unterschiedlich. So ist z. B. die Karyotypisierung wegen ihrer geringen Auflösung nur zur Identifizierung sehr großer Deletionen oder Duplikationen geeignet, hat aber den Vorteil, dass Inversionen und balancierte Translokationen gesehen werden können. Im Gegensatz dazu haben SNP-Oligonukleotid-Mikroarrays eine wesentlich bessere Auflösung und können auch sehr kleine

Variationen aufdecken, sind aber nicht zur Identifizierung von Inversionen oder balancierten Translokationen geeignet (Tabelle 1).

Alle hier beschriebenen Methoden vergleichen eine Test-DNA mit einer Referenz-DNA, um Unterschiede in der Kopienzahl festzustellen. Bei der array-CGH werden Test- und Referenz-DNA mit verschiedenen Farbstoffen markiert, um neue Variationen identifizieren zu können. Im Gegensatz dazu wird bei den SNP-Oligonukleotid-Mikroarrays eine Einfarbentechnik verwendet. Um Unterschiede in der Kopienzahl zu erkennen, werden die Signalintensitäten auf dem Array der Test-DNA mit Signalintensitäten von Arrays mit Kontroll-DNA verglichen. Die neu identifizierten Variationen werden im Anschluss mit der *Human Genome Reference Assembly* (NCBI) verglichen.

In der physikalischen Karte des Genoms existieren immer noch über 300 Lücken. Untersuchungen mit array-CGH zeigten eine signifikante Häufung von solchen Lücken in der Umgebung von CNVs (Cheung et al. 2003, Eichler et al. 2004, lafrate et al 2004, Sharp et al. 2005, Redon et al. 2006).

Ein Vergleich zwischen dem NCBI Referenz Genom und dem Referenzgenom von Celera zeigte, dass einige Sequenzen nur im Celera-Genom vorkommen und einige nur im NCBI-Genom (Feuk et al. 2006). Diese Unterschiede lassen sich dadurch erklären, dass einige Sequenzen nicht klonierbar sind, Fehler beim Assemblieren der Sequenzen entstehen oder Unterschiede in der Struktur der verwendeten DNA-Quellen vorkommen können (Feuk et al. 2006).

2. Untersuchungen mit SNP-Oligonukleotid-Arrays

In der vorliegenden Arbeit sollte die Anwendung von SNP-Oligonukleotid-Mikroarrays für Kopplungsanalysen und Bestimmungen der Kopienzahlvariationen getestet werden. Zu Beginn der Arbeit wurden Kopplungsanalysen noch mit Mikrosatelliten durchgeführt. Eine Aufgabe der vorliegenden Arbeit bestand darin, die Anwendung der 10K-SNP-Mikroarrays von Affymetrix für Kopplungsanalysen zu testen und die Technik dafür im Labor zu etablieren. Dafür wurden in einigen Familien Kopplungsanalysen durchgeführt. Die Prinzipien einer Kopplungsanalyse und die Krankheiten der untersuchten Familien werden in den kommenden Abschnitten erläutert. Auf die

Bestimmung der Kopienzahlvariationen wird im Abschnitt 2.2 Kopienzahlvariationen (s.S. 19) näher eingegangen.

2.1 Kopplungsanalysen mit SNP-Oligonukleotid-Mikroarrays

2.1.1 Grundbegriffe der Kopplungsanalysen

Die Loci von mendelisch vererbten Krankheiten lassen sich mit parametrischer Kopplung identifizieren. Die Grundlage der Kopplungsanalyse beruht auf der Tatsache, dass zwei benachbarte Loci mit einer umso höheren Wahrscheinlichkeit (Rekombinationsfrequenz θ) gemeinsam vererbt werden, je näher sie auf einem Chromosom zusammen liegen (Strachan et al. 2004). Je weiter die Loci auseinander liegen, umso häufiger werden sie während der Meiose durch Crossing over voneinander getrennt. Dabei trennen sich Loci auf verschiedenen Chromosomen nach der dritten mendelschen Regel (Unabhängigkeitsregel) mit einer Wahrscheinlichkeit von 50% ($\theta = 0.5$) (Strachan et al. 2004). Erfolgt die Trennung zweier Loci auf einem Chromosom seltener, liegt eine Kopplung vor ($\theta < 0.5$). Allele verschiedener gekoppelter Loci, die gemeinsam vererbt werden, werden als Haplotypen bezeichnet. Die genetische Distanz zwischen zwei Loci ergibt sich aus der Rekombinationsfrequenz und wird in der Einheit „Morgan“ (M) gemessen. Werden zwei Gene in einer von 100 Meiosen getrennt, so entspricht das dem Abstand von 1 centi-Morgan (cM). Im menschlichen Genom entspricht 1 cM ungefähr einem Abstand von 1 Mb. Im Mausgenom ist der Abstand größer und 1 cM entspricht ungefähr 2 Mb.

Für Kopplungsanalysen sind Meiosen informativ, wenn bestimmt werden kann, ob ein Gamet rekombinant ist, oder nicht. Um das in einem Stammbaum nachvollziehen zu können, wird die genetische Kartierung von Genen mit Hilfe von Markern durchgeführt. Als Marker wurden anfangs Blutgruppen und Polymorphismen in Serumproteinen verwendet. Mehr Informationen liefern die jetzt verwendeten DNA-Mikrosatelliten und SNPs.

Die Aufgabe der Kopplungsanalyse ist es θ zu schätzen und zu testen, ob $\theta < 0.5$ und damit statistisch signifikant ist (Terwillinger et al.1994). Mit bestimmten Züchtungsschemata ist es bei Labortieren möglich, Träger rekombinanter und nicht rekombinanter Chromosomen zu zählen und mit Hilfe des

Chi-Quadrat-Tests zu testen. Beim Menschen sind die Rekombinanten nicht einfach zu zählen und man muss sich auf vorhandene Familiendaten stützen. Die Familien müssen eine ausreichende Anzahl von betroffenen und nicht betroffenen Familienmitgliedern haben, um mit entsprechenden Programmen eine Kopplungsanalyse durchführen zu können. Es besteht auch die Möglichkeit mehrere kleinere Familien einzeln zu analysieren und anschließend die Ergebnisse zu addieren, unter der Voraussetzung, dass keine genetische Heterogenität vorliegt und die Erkrankung durch Mutationen am gleichen Locus verursacht wird (Ott 1991). Die Ergebnisse der Berechnung werden dabei umso aussagekräftiger, je mehr Meiosen stattgefunden haben.

Die Effizienz und Validität der parametrischen Kopplungsanalyse ist dabei abhängig von den gewählten Parametern: dem Erbang, d. h. den geschätzten Penetranzen, der Allelfrequenzen der Marker und des Krankheitsallels.

Mit Analyseprogrammen, wie dem Programm LINKAGE, ist es möglich, mit unterschiedlichen Annahmen der Rekombinationsfrequenz zwischen 2 Loci die Wahrscheinlichkeit einer Kopplung ($L(\theta)$) für einen vorgegebenen Stammbaum zu schätzen. In Kopplungsanalysen wird das Verhältnis der Wahrscheinlichkeiten durch $L(\theta) / L(\theta = 0.5)$ ausgedrückt (Terwillinger et al. 1994). Für Kopplungsanalysen wird die folgende Formel für die Berechnung des LOD Scores z verwendet:

$$z(\theta) = \log_{10} [L(\theta) / L(0.5)]$$

Ein LOD Score $z(\theta) > 3$ wird in einem Genomscan als signifikant für eine Kopplung betrachtet und entspricht einem p-Wert von kleiner 0.5. Ein LOD Score von $z(\theta) < -2$ schließt eine Kopplung aus (Ott 1991).

Es wird zwischen der Twopoint- und der Multipoint-Analyse unterschieden. Bei der Twopoint-Analyse werden zwei Loci, der Locus des Markers und der des Krankheitsallels, zur gleichen Zeit betrachtet, während bei der Multipoint-Analyse mehrere Loci in der Analyse berücksichtigt werden.

Für parametrische Analysen wurden Computerprogramme entwickelt. Die parametrische Twopoint-Analyse wird mit dem Programmpaket LINKAGE 5.2 berechnet (Lathrop et al. 1984). Kruglyak et al. entwickelten das Programm GENEHUNTER, womit eine parametrische Multipointanalyse mit dem

Lander Green Algorithmus durchgeführt werden kann (Kruglyak et al 1996). Weitere Programme, mit denen Multipointanalysen durchgeführt werden sind ALLEGRO und MERLIN (Gudbjartsson et al. 2000, Abecasis et al. 2002). Bei den Programmen für Multipointanalysen steigt die Laufzeit linear mit der Anzahl der Marker und exponentiell mit der Größe des Stammbaums, während bei den Programmen für die Twopointanalysen die Laufzeit exponentiell mit der Anzahl der Marker und linear mit der Größe des Stammbaums ansteigt.

Nichtparametrische Kopplungsanalysen werden verwendet, wenn es nicht möglich ist, ein bestimmtes genetisches Modell in der betreffenden Familie zu erkennen. Mit dieser Methode der Kopplungsanalysen werden nicht betroffene Personen aus der Berechnung herausgelassen und es wird nach Allelen oder Chromosomenabschnitten gesucht, die von den betroffenen Personen geteilt werden (Strachan et al. 2004). Sind Allele in einem oder mehreren Individuen identisch, weil sie von dem gleichen Vorfahren stammen, nennt man diese Übereinstimmung *Identity by Descent* (IBD). Im Gegensatz dazu gibt es die *Identity by State* (IBS), bei der Allele in mehreren Individuen übereinstimmen, man aber keine Aussage über die Herkunft des Allels treffen kann.

2.1.2 Vergleich von Mikrosatelliten versus SNPs in Kopplungsanalysen

Die genomweiten Kopplungsanalysen wurden bis vor ein paar Jahren mit Mikrosatelliten durchgeführt. Mikrosatelliten sind kleine, meist weniger als 0.1 kb lange Tandemrepeats einer einfachen DNA-Sequenz aus 1-4 bp, z. B. (CA)_n (Strachan et al. 2004). Die meisten Untersuchungen verwendeten ein Panel von 300 bis 400 Mikrosatelliten, die das Genom mit einem durchschnittlichen Intervall von ca. 10 cM abdeckten (Evans et al. 2004, John et al. 2004, Middleton et al. 2004, Schaid et al. 2004). Viele mendelisch vererbte Erkrankungen wurden erfolgreich mit Mikrosatelliten kartiert.

Zu Beginn dieser Arbeit wurden die ersten Versuche unternommen, Kopplungsanalysen mit SNP-Oligonukleotid-Mikroarrays durchzuführen. Ein Aufgabenteil der vorliegenden Arbeit bestand darin, die Verwendung von SNP-Oligonukleotid-Mikroarrays für Kopplungsanalysen mit Arrays zu testen, die über 10000 SNPs enthielten. SNPs (*Single Nucleotid Polymorphisms*) sind polymorphe Variationen auf dem Niveau von einzelnen Nukleotiden (Strachan et al. 2004).

2.1.3 Krankheitsgeschichten der untersuchten Familien

2.1.3.1 Der Phosphatstoffwechsel im Zusammenhang mit bisher bekannten Formen der Hypophosphatämien

Phosphat ist ein wesentlicher Bestandteil der Knochensubstanz, der Nukleinsäuren und, als Phospholipide, der Zellmembranen. Der Phosphatspiegel liegt unter physiologischen Bedingungen zwischen 2.5 und 4.5 mg/dl (Greenberg et al. 1960). Sinkt die Phosphatkonzentration im Serum unter 2.5 mg/dl, wird dies als Hypophosphatämie bezeichnet. Liegt ein chronischer Phosphatmangel vor, kommt es zu Defekten bei der Mineralisation und zu einer Anhäufung von unmineralisiertem Osteoid. Dieses Krankheitsbild wird bei Menschen im Wachstum als Rachitis und nach dem Wachstum als Osteomalazie bezeichnet.

Die Regulierung der Phosphatkonzentration erfolgt vor allem durch renale Reabsorption, Freisetzung und Speicherung im Knochengewebe und im geringeren Maße durch die intestinale Absorption. Bis vor wenigen Jahren war lediglich bekannt, dass die Hormone Vitamin D ($1,25(\text{OH})_2\text{D}_3$) und das Parathormon (PTH), welche primär die Kalziumhomöostase regulieren, auch in die Regulation der Phosphathomöostase eingreifen (Murer et al. 2001).

Kalzium ist essentiell an der Muskelkontraktion und der Neurotransmitterabgabe beteiligt und übt eine wichtige Stützfunktion in Knochen und Zähnen aus. Es ist wichtig, dass die Konzentration an extrazellulärem, ionisiertem Kalzium für diese Prozesse zur Verfügung steht und konstant gehalten wird. Sinkt die Serumkonzentration von Kalzium unter den Normwert, wird die Bildung und Sekretion von PTH in den Nebenschilddrüsenzellen aktiviert. PTH stimuliert die 1α -Hydroxylase in der Niere und damit die Bildung der aktiven Form von Vitamin D, dem 1,25-Dihydroxyvitamin D_3 . PTH und Vitamin D wirken synergistisch in Knochen, Dünndarm und den Nieren zur Steigerung der Kalziumkonzentration im Serum. Im Knochen sorgt PTH über eine Stimulation von Osteoklasten für einen Abbau von Hydroxylapatit ($\text{Ca}_5(\text{PO}_4)_3(\text{OH})$) aus dem Knochengewebe und damit zur Freisetzung von Kalzium und Phosphat in den Extrazellulärraum. Im Dünndarm steigern PTH und Vitamin D die aktive Aufnahme von Kalzium und Phosphat in der Bürstensaummembran. In der Niere wird durch PTH die tubuläre Rückresorption von Kalzium stimuliert. Sekundär kommt es dabei zu einer Hemmung der tubulären

Phosphatreabsorption durch PTH und es folgt eine vermehrte Ausscheidung von Phosphat. Dies hat zur Folge, dass eine Ausfällung von Kalziumphosphat, bedingt durch die erhöhte Phosphatkonzentration, verhindert wird. Die Hormone PTH und Vitamin D regulieren sich dabei gegenseitig. PTH stimuliert die Bildung von Vitamin D und erhöhte Konzentrationen von Vitamin D hemmen die Bildung von PTH, so dass es zu einer Normalisierung der Kalziumkonzentration kommt. Als Resultat der Wirkungen von PTH und Vitamin D kommt es zu einer Zunahme der Kalzium- und Phosphatkonzentration im Serum, zu einer erhöhten Reabsorption von Kalzium in der Niere und gleichzeitig zu einer vermehrten renalen Ausscheidung von Phosphat.

Untersuchungen von genetisch bedingten Erkrankungen, die Hypophosphatämie verursachen, führten zur Identifizierung von Faktoren, die primär an der normalen Regulation des Phosphathaushaltes beteiligt sind. Die häufigste Ursache der Hypophosphatämien sind X-chromosomale Hypophosphatämien (*X-linked Hypophosphatemia*, XLH; OMIM 307800), die durch inaktivierende Mutationen in einer putativen Endopeptidase (PHEX) entstehen (The HYP Consortium 1995). Seltener kommt die autosomal-dominante Form der hypophosphatämischen Rachitis (*Autosomal-dominant Hypophosphatemic Rickets*, ADHR; OMIM 193100) vor, die durch Mutationen im Fibroblasten-Wachstumsfaktor FGF23 entstehen (The ADHR Consortium 2000, White et al. 2001, Benet-Pages et al. 2004). Die Wildtyp-Form von FGF23 wird in bestimmten mesenchymalen Tumoren vermehrt exprimiert und verursacht einen erhöhten FGF23-Spiegel im Serum. Durch den vermehrten FGF23-Spiegel kommt es zu einem renalen Phosphatverlust und zur Osteomalazie. Diese Form wird als tumorinduzierte Osteomalazie (*tumor-induced Osteomalacia*, TIO) bezeichnet (White et al. 2001, Shimada et al. 2001, Yamazaki et al. 2002, Jonsson et al. 2003).

Im Folgenden werden 2 weitere Formen der Hypophosphatämie beschrieben. In der vorliegenden Arbeit sollte versucht werden diese Loci zu identifizieren.

2.1.3.2 Erbliche hypophosphatämische Rachitis mit Hyperkalziurie (HHRH)

Es wurden 5 Familien untersucht, deren betroffene Familienmitglieder an der erblichen hypophosphatämischen Rachitis mit Hyperkalziurie (*Hereditary Hypophosphatemic Rickets with*

Hypercalciuria, HHRH, OMIM 241530) erkrankt sind. HHRH wird durch einen autosomal-rezessiven Erbgang vererbt.

Im Gegensatz zu den anderen Formen der Hypophosphatämie kommt es in den Patienten mit HHRH zu einer Stimulation der renalen 25-Hydroxyvitamin D 1 α -Hydroxylase, die einen erhöhten Serumspiegel des biologisch aktiven Vitamin D-Metaboliten 1,25-Dihydroxyvitamin D (1,25(OH)₂D) verursacht. Daraus resultiert eine gesteigerte intestinale Absorption von Kalzium und daraus folgend eine Hyperkalziurie (Tieder et al. 1985). Der erhöhte Vitamin D-Spiegel ist ein wichtiges Kriterium um die HHRH von den bisher beschriebenen Hypophosphatämien zu unterscheiden, in denen die Vitamin D-Konzentration trotz niedrigem Phosphatspiegel im Normbereich liegt. Bei der HHRH wird das Auftreten der Hypophosphatämie sekundär durch einen renalen Phosphatverlust verursacht (Tieder et al. 1985, Tieder et al. 1987, Gazit et al. 1991, Lorenz-Depierreux et al. 2006, Bergwitz et al. 2006).

2.1.3.3 Autosomal-rezessive Hypophosphatämie (ARHP)

Weiterhin wurden 3 Familien untersucht, deren betroffene Familienmitglieder klinische, biochemische und histomorphometrische Parameter aufweisen, die denen von Patienten mit XLH und ADHR gleichen. Die bekannten Formen XLH und ADHR wurden aufgrund einer Stammbaumanalyse und einer Mutationsanalyse ausgeschlossen. Die Hypophosphatämie wurde in den untersuchten Familien autosomal-rezessiv vererbt (*Autosomal-recessive Hypophosphatemic Rickets*, ARHP). Bei den Betroffenen wurden Mutationen in den Genen *PHEX* und *FGF23* ausgeschlossen, die für die Entstehung von XLH und ADHR verantwortlich sind (Lorenz-Depierreux et al. 2006).

2.1.3.4 Spondylo-oculares Syndrom (SOS)

Neben den Familien mit Hypophosphatämien, wurde eine Familie mit konsanguinen Eltern untersucht, in der bei 6 von 7 Kindern ein spondylo-oculares Syndrom (SOS) diagnostiziert wurde. Das Syndrom ist charakterisiert durch Katarakte, die Ablösung der Retina und den daraus resultierenden Verlust der Sehkraft, faziale Dysmorphien, faziale Hypotonie, eine normale Körpergröße mit einem unproportional kurzen Oberkörper, Platyspondylie, generalisierte Osteoporose, eine unbewegliche Wirbelsäule mit thorakaler Kyphosis und eingeschränkter lumbaler Lordosis (Schmidt et al. 2001, Rudolph et al. 2003).

Der Stammbaum und die Konsanguinität der Eltern, deuten auf eine Erkrankung mit einer autosomal-rezessiven Vererbung hin.

Die klinischen Aspekte der Erkrankung beziehen sich vor allem auf die Augen und das Skelett, ohne dass andere Organe beeinträchtigt sind. Es gibt mehrere Syndrome, die einen ähnlichen Phänotyp aufweisen, z. B. Stoffwechselerkrankungen, wie Mukopolysaccharidose und Homocystinurie, das Osteoporose-Pseudoglioma-Syndrom (OMIM 259770), und Kollagenopathien, wie das Kniest-Syndrom (OMIM 156550) oder das Stickler-Syndrom (OMIM 108300) (Sewell et al. 1991, Somer et al. 1988, Stickler et al. 1965, Spranger et al. 1997). Um die bekannten Syndrome ausschließen zu können, wurden bei den Betroffenen Stoffwechseluntersuchungen durchgeführt, die normale metabolische Parameter zeigten (Schmidt et al. 2001, Rudolph et al. 2003).

Mutationen in Genen, wie *COL11A1*, *COL11A2* und *COL2A1*, die das Stickler- oder Kniest-Syndrom verursachen, sowie Mutationen des Gens *low-density lipoprotein receptor-related protein 5 (LRP5)*, welches das Osteoporose-Pseudoglioma-Syndrom verursacht, wurden ausgeschlossen. Auch das Gen *PAX6* wurde untersucht, da es aktiv an der Entwicklung der Linse und des anterioren Auges beteiligt ist und Mutationen in *PAX6* zu Entwicklungsstörungen der Linse oder zur primären Aphakie führen. In keinem der Betroffenen wurden Mutationen in diesem Gen gefunden (Rudolph et al. 2003).

Durch eine Kopplungsanalyse sollte in dieser Familie der Krankheitslocus bestimmt werden.

2.2 Kopienzahlvariationen (CNVs)

Im Laufe der Arbeit entwickelten sich neue SNP-Oligonukleotid-Mikroarrays, die 100K Arrays von Affymetrix, mit denen über 100000 SNPs pro Genom untersucht werden konnten. Variationen in der Kopienzahl im Genom, die sowohl pathogen als auch polymorph sind, wurden in vielen Studien untersucht. Die meisten Studien zu diesem Thema wurden mit array-CGH durchgeführt und nur wenige mit SNP-Oligonukleotid-Mikroarrays. In dieser Arbeit sollte die Eignung der 100K-Arrays von Affymetrix zur Untersuchung von CNVs getestet werden.

Die genetische Variabilität im menschlichen Genom ist vielfältig und reicht von sehr großen, mikroskopisch sichtbaren Abweichungen der Chromosomen bis hin zu Veränderungen auf dem Niveau von einzelnen Nukleotiden. CNVs wurden nicht nur bei Patienten mit speziellen Krankheiten gefunden, sondern auch in der Normalbevölkerung. Eine CNV ist ein DNA-Segment, das mit einer variablen Kopienzahl im Vergleich zu einem Referenzgenom vorkommt. CNVs können in Insertionen, Deletionen und Duplikationen eingeteilt werden (Feuk et al. 2006).

Mit wachsendem Interesse an CNVs wurde 2004 eine Datenbank, die *Database of Genomic Variants* (DGV), eingerichtet, in der zurzeit bereits über 6000 Variationen erfasst sind. CNVs, die in mehr als 1% der Normalbevölkerung vorkommen und keinen pathologischen Phänotyp verursachen, werden als Polymorphismen bezeichnet.

Es besteht ein großes Interesse an der Sammlung von CNVs aus verschiedenen Studien, weil dadurch Polymorphismen bestätigt werden können und es eine Möglichkeit gibt, pathologische CNVs von Polymorphismen zu unterscheiden. Die CNVs sind ungleichmäßig über das gesamte Genom verteilt und es gibt keine längeren Abschnitte, in denen keine CNV vorkommt. Der Verteilung der CNVs auf die einzelnen Chromosomen schwankt zwischen 6% und 19% (Redon et al. 2006). Der Gesamtanteil an CNVs im Genom liegt bei etwa 12% oder ungefähr 360 Mb (Redon et al. 2006).

In der vorliegenden Arbeit lag der Schwerpunkt auf der Identifizierung von CNVs mit starken Effekten bei Kindern mit mentaler Retardierung. Im Folgenden werden zum einen die Mechanismen erläutert, die zur Entstehung von CNVs führen können, zum anderen wird auf Pathologien durch CNVs, insbesondere auf mentale Retardierung eingegangen.

2.2.1 Mechanismen zur Entstehung von CNVs

Molekulare Untersuchungen zeigten, dass Umlagerungen im Genom durch nicht-allelische homologe Rekombination (NAHR) und durch nicht-homologe Endverknüpfung (*non-homologous end joining*, NHEJ) entstehen können.

Die Voraussetzung für Umlagerungen, die durch die nicht-allelische homologe Rekombination entstehen, sind homologe Sequenzen an den Bruchpunkten. Die nicht-homologe Endverknüpfung

findet als Reparaturmechanismus nach Doppelstrangbrüchen statt und benötigt im Gegensatz zur NAHR keine homologen Sequenzen.

2.2.1.1 Nicht-allelische homologe Rekombination (NAHR)

Umlagerungen durch NAHR entstehen bevorzugt in Regionen, die von großen, homologen Repeats, so genannten Low Copy Repeats (LCRs), flankiert sind. Low Copy Repeats (LCRs) wurden definiert als gepaarte, intra- und interchromosomale Duplikationen mit über 95% Sequenzidentität, die 10 kb oder größer sind und durch Sequenzen von 50 kb bis 10 Mb getrennt sind (Stankiewicz et al. 2002). Ungefähr 5% des menschlichen Genoms bestehen aus LCRs (Sharp et al. 2006). Das Misalignment, das es nicht-allelicen LCRs ermöglicht sich in der homologen Rekombination zu paaren, scheint von einigen genomischen Strukturen abhängig zu sein. So korreliert die Größe der chromosomalen Umstrukturierung mit dem Intervall zwischen den Repeats, d. h. je länger der Repeat ist, umso größer ist das neu angeordnete genomische Fragment (Lupski 1998, Stankiewicz et al. 2002). Damit der Rekombinationsmechanismus ablaufen kann, werden als Minimum 300-500 bp homologer Sequenz benötigt (Reiter et al. 1998). LCRs haben in der Regel größere homologe Bereiche.

Verschiedene Syndrome, wie das DiGeorge/Velocardiofacial-Syndrom, das Williams-Beuren Syndrom oder das Smith-Magenis-Syndrom entstehen durch NAHR, die durch flankierende LCRs vermittelt wird.

NAHR zwischen direkt orientierten LCRs in homologen Chromosomen oder Schwesterchromatiden führt zu Deletionen und Duplikationen und innerhalb eines Chromatids zu einer Deletion und zur Bildung eines azentrischen Fragments (Abb 1 a-c). Liegen die LCRs in invertierter Orientierung auf homologen Chromosomen vor, kommt es zur Inversion (Abb 1 d-f). Manche LCRs haben auch komplexe Strukturen, die aus direkt orientierten und invertierten LCRs aufgebaut sind. Komplexe LCRs können sowohl zu Inversionen oder Deletionen / Duplikationen führen, als auch zu Rearrangierungen der Repeatsequenzen, abhängig davon, welcher Teil des LCR als Rekombinationssubstrat verwendet wird (Abb 1 g-i) (Stankiewicz et al 2002).

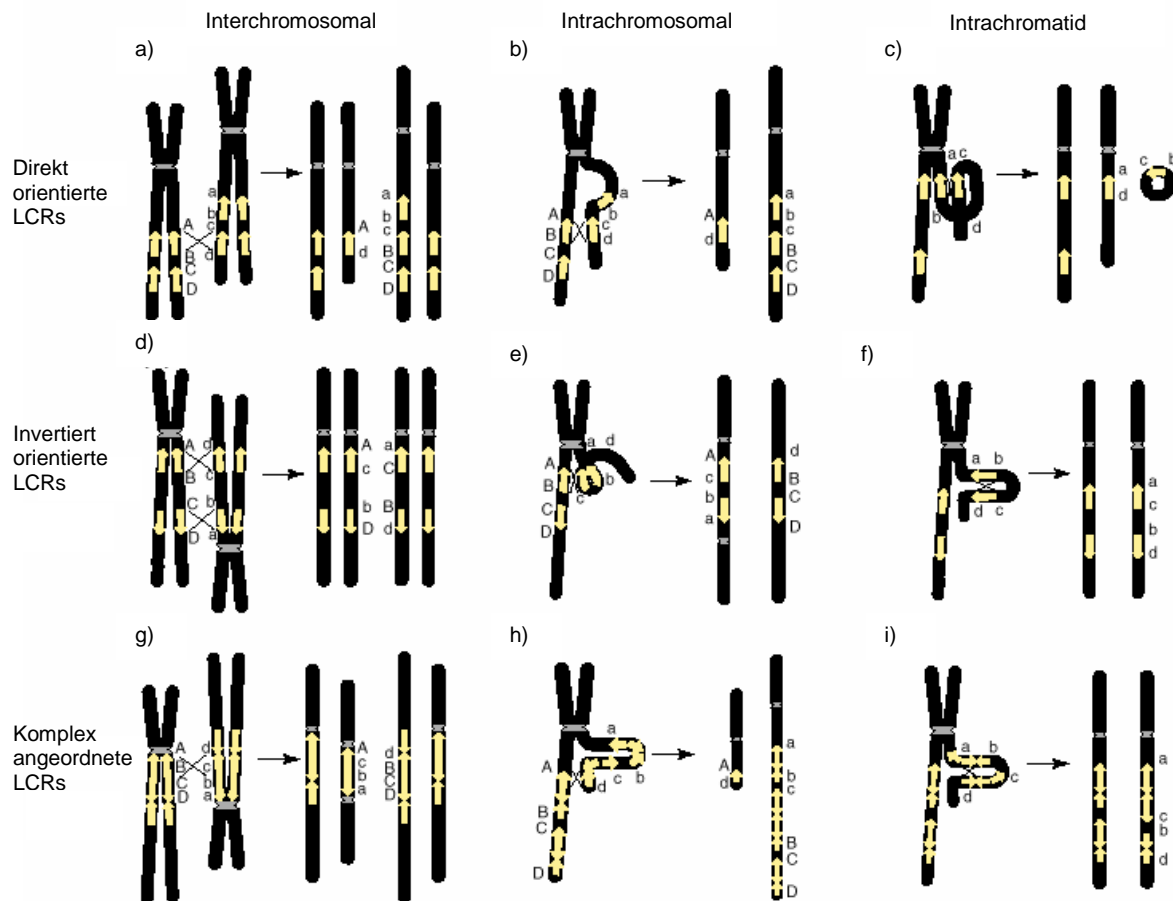


Abb. 1 Mechanismen der NAHR und ihre Auswirkungen auf die genomische Struktur. Die Chromosomen sind in schwarz dargestellt, die Zentromere in grau und LCRs als gelbe Pfeile. Direkt orientierte LCRs führen zwischen zwei homologen Chromosomen zu Deletionen / Duplikationen a), innerhalb eines Chromosoms zwischen den zwei Chromatiden zu Deletionen / Duplikationen b) oder innerhalb eines Chromatids zur Bildung einer Deletion und eines azentrischen Fragments c). Invertierte LCRs dagegen führen zu Inversionen d-f). Komplex arrangierte LCRs können zu Inversionen oder Deletionen und Duplikationen führen g- i). (Abbildung modifiziert nach Stankiewicz et al. 2002)

2.2.1.2 Nicht-homologe Endverknüpfung (NHEJ)

Ein Mechanismus, der zur Bildung von Rearrangierungen ohne homologe Repeats führen kann, ist die nicht-homologe Endverknüpfung (*Non-Homologous End Joining*, NHEJ).

Eine Untersuchung der Bruchpunkte des *Dystrophin*-Gens zeigte, dass 30% der Bruchpunkte in Regionen mit *Alu*- oder LTR-Sequenzen lagen, während die restlichen 70% der Bruchpunkte in Regionen ohne Repeats vorkamen (Nobile et al. 2002, Shaw et al. 2004).

Kenntnisse über die die Reparatur von Doppelstrangbrüchen stammen aus Untersuchungen an Patienten mit Gendefekten, wie dem Njimegen Breakage Syndrom, SCID oder Xeroderma pigmentosum, die zur Identifizierung beteiligter Gene geführt haben.

Der Mechanismus der NHEJ läuft folgendermaßen ab. Initialisiert wird die NHEJ durch die Bindung der Proteine Ku70 und Ku80 an die Bruchstellen der DNA (Abb 2 b). Anschließend wird die katalytische Untereinheit der DNA-abhängigen Proteinkinase (DNA-PKcs) rekrutiert, die zusammen mit den Ku-Proteinen die DNA-abhängige Kinase (DNA-PK) bildet. Zusätzlich werden noch der Proteinkomplex MRE11-RAD50-NBS1 und ein weiteres Protein angelagert, Artemis, das von der DNA-PK phosphoryliert wird. Artemis ist eine 5'→3' DNA-Einzelstrangexonuklease, die als DNA-PKcs/Artemis-Komplex zu einer Endonuklease für 5'- und 3'-Überhänge wird. Durch Protein-Protein-Interaktionen zwischen den an die DNA gebundenen DNA-PKcs-Molekülen werden die DNA-Enden überbrückt (Abb 2 c und d). Durch die Verbindung der DNA-PKcs-Moleküle wird die DNA-Ligase IV mit ihrem assoziierten Faktor XRCC4 und die Polynukleotidkinase (PNK) angelagert (Abb 2 e). Zusammen mit Mitgliedern der Familie der Pol X-Polymerasen, Pol μ und Pol λ , ist der gesamte Proteinkomplex an der Auffüllung der restlichen Lücken und der Ligierung der Bruchstücke beteiligt und der DSB wird geschlossen (Abb 2 f und g) (Hefferin et al. 2005, Sonoda et al. 2006)

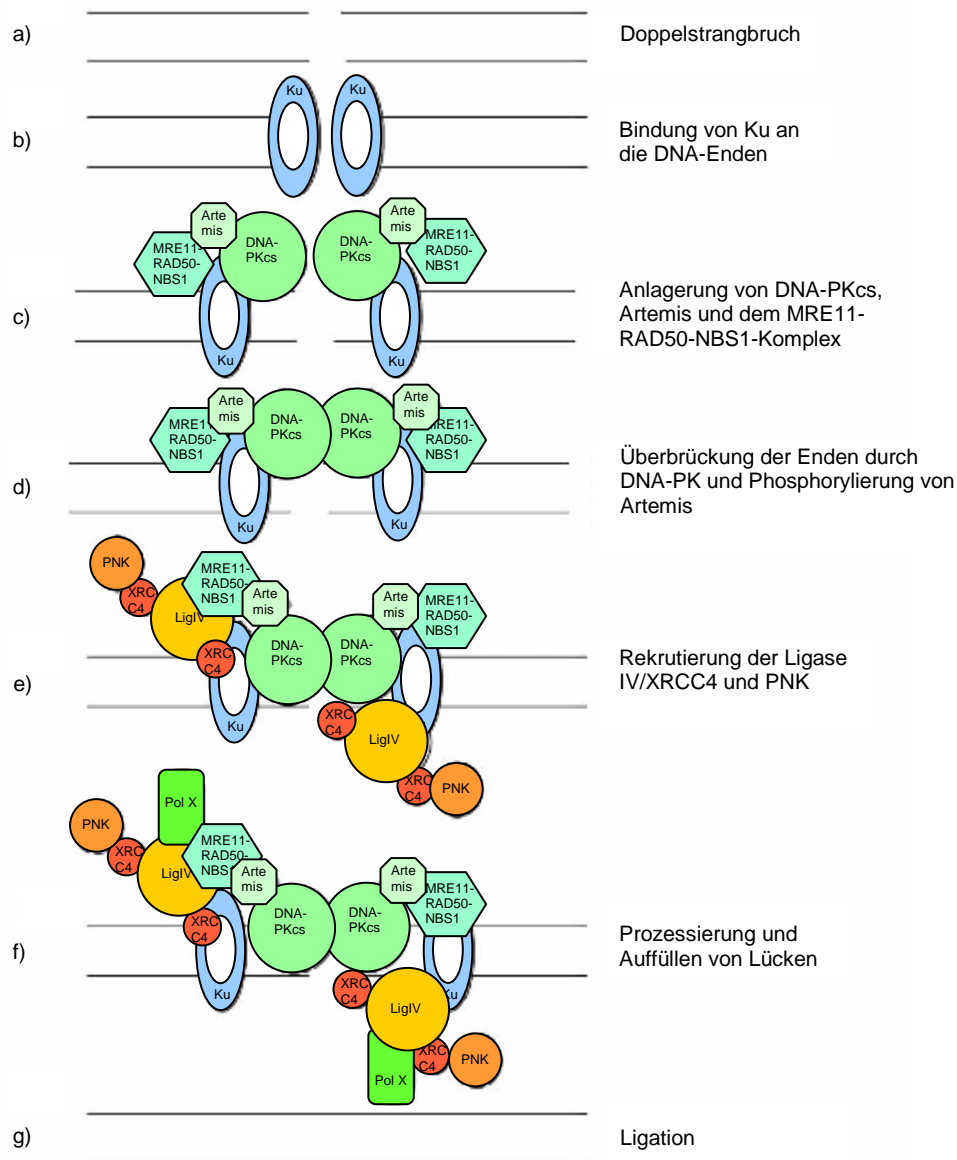


Abb. 2 Modell der Reparatur von DSBs durch NHEJ. In Säugetierzellen führt die Bildung von DSBs a) zur Rekrutierung des Ku-Heterodimers und dessen Bindung an die freien DNA-Enden b). Anschließend folgen die Bindung der katalytischen Untereinheit der DNA Proteinkinase (DNA-PKcs), des MRE11-RAD50-NBS1-Komplexes und des Proteins Artemis, und die Phosphorylierung von Artemis durch die DNA-PK c). Die Protein-Protein-Interaktion zwischen den DNA-PKcs-Molekülen verbindet die DNA-Enden d) und resultiert in der Anlagerung des DNA-Ligase IV / XRCC4-Komplexes e). Die Polynukleotidkinase (PNK) interagiert mit XRCC4 und wird ebenfalls an den Bruchstellen gebunden e). Durch die Prozessierung durch Artemis und anderen Faktoren, zusammen mit dem Auffüllen von Lücken durch DNA-Polymerasen der X Familie f), kann die Ligase IV den DSB reparieren g). (Abb modifiziert nach Hefferin et al. 2005)

2.2.2 Pathologien durch CNVs

Eine mögliche Folge von Variationen der Kopienzahl sind Krankheiten, die durch eine Veränderung der Gendosis entstehen. CNVs können zum einen starke Effekte haben, wenn z.B. die CNV *de novo* entstanden ist. Eine Folge sind dann Deletions- oder Duplikationssyndrome, wie das Charcot-Marie-Tooth Syndrom Typ 1a oder das Smith-Magenis Syndrom (Inoue et al. 2002; Lupski et al. 2005; Shaw-Smith et al. 2004). Zum anderen kann der Einfluss einer CNV auch schwach sein, wenn z.B. das betroffene Gen oder die betroffenen Gene eine Rolle bei komplexen Erkrankungen spielen (McCarroll et al. 2006, Buckland et al. 2003; Nguyen et al. 2006; Repping et al. 2006).

Im Folgenden werden 4 verschiedene Mechanismen beschrieben, wie sich CNVs auf die Gendosis auswirken können (Inoue et al. 2002, Feuk et al. 2006, Freeman et al. 2006, Lupski et al. 2005, Feuk et al. 2006).

- a) Dosis-sensitive Gene, die in einer CNV liegen, können durch eine Duplikation oder Deletion pathogen werden. Eine Krankheit kann auch ausgelöst werden, wenn durch eine Deletion auf dem einen Allel, eine rezessive Mutation auf dem zweiten Allel aufgedeckt wird (Abb 3 a).
- b) Gene, die in Bruchpunktregionen von CNVs liegen, werden durch Inversion, Translokation oder Deletion auseinander gerissen und es kommt zu einer reduzierten Expression der dosis-sensitiven Gene. Bruchpunkte von CNVs, die in Genen liegen, können auch zur Bildung von neuen Transkripten führen, die durch Genfusion oder *exon-shuffling* entstehen (Abb 3 b).
- c) Durch Positionseffekte können CNVs auch dosis-sensitive Gene beeinflussen, die nicht in unmittelbarer Nähe zu der Variation liegen. Dieser Effekt kann entstehen, wenn durch die CNV regulatorische Elemente verändert werden, entweder direkt durch Deletion, Translokation oder Inversion oder, wenn ein Allel deletiert ist und dadurch im zweiten Allel ein funktioneller Polymorphismus in einem Effektor wirksam wird, der die Genfunktion beeinflusst (Abb 3 c).
- d) CNVs können auch einen Einfluss auf komplexen Krankheiten haben (Abb 3 d).

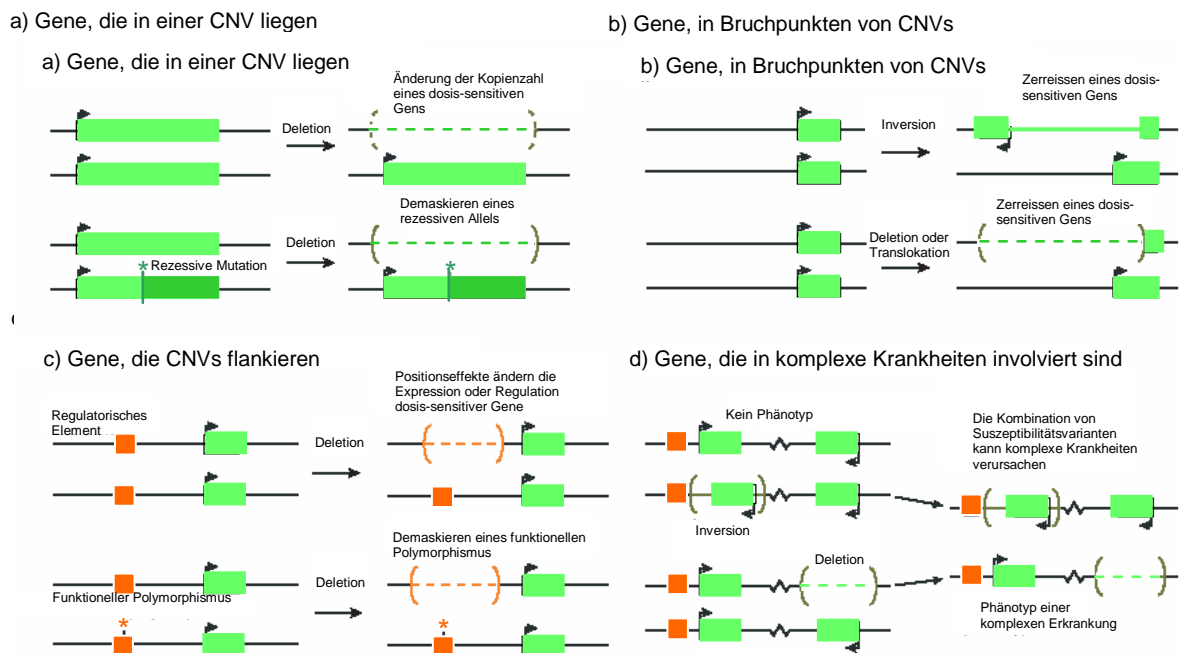


Abb. 3 Der Einfluss von CNVs auf den Phänotyp. Dosis-sensitive Gene, die in einer CNV liegen, werden pathogen durch eine Deletion (obere Zeile) oder durch das Aufdecken einer rezessiven Mutation auf dem homologen Chromosom (untere Zeile) a). Gene, die mit Bruchpunkten einer CNV überlappen, werden durch eine Inversion (obere Zeile) oder eine Deletion oder Translokation (untere Zeile) getrennt und haben dadurch eine niedrigere Expression b). Gene, die von CNVs flankiert werden, können durch Positionseffekte beeinflusst werden. Eine Deletion von wichtigen regulatorischen Elementen kann die Genexpression verändern (obere Zeile). Wird durch eine Deletion ein funktioneller Polymorphismus in einem regulatorischen Element aufgedeckt, kann das ebenfalls Auswirkungen auf die Genexpression haben c). CNVs können als Suszeptibilitätsallele agieren, wenn eine Kombination verschiedener genetischer Faktoren nötig ist, um einen Phänotyp zu produzieren. So haben z. B. zwei CNVs einzeln betrachtet keinen Einfluss auf den Phänotyp, aber in Kombination können sie zu einer komplexen Erkrankung beitragen d). (Abbildung modifiziert nach Feuk et al. 2006)

2.2.3 Mentale Retardierung

In der vorliegenden Arbeit sollten Kinder mit mentaler Retardierung auf CNVs untersucht werden. Mentale Retardierung oder Entwicklungsverzögerung wird von der *American Association on Intellectual and Developmental Disabilities* definiert als Erkrankung, die charakterisiert ist durch signifikante Einschränkungen sowohl der intellektuellen Fähigkeiten, als auch der Anpassungsfähigkeit bei sozialen, praktischen und konzeptionellen Fertigkeiten. Mentale Retardierung (MR) wird nach der *International Classification of Diseases (ICD-10)* der *World Health Organization* (WHO) in 5 Schweregrade eingeteilt, die leichteste Form ist die Lernbehinderung (Intelligenzquotient

(IQ) 71-85), gefolgt von der leichten MR (IQ von 51-70), der mäßigen MR (IQ 36-50), der schweren MR (IQ 21-35) und der schwersten MR (IQ <20). Die MR ist häufig noch mit weiteren Symptomen verbunden, z. B. mit Autismus, Fehlbildungen des Gehirns und motorische und/oder sprachliche Beeinträchtigungen. Viele Syndrome mit MR sind mit schwerwiegenden Beeinträchtigungen der Organe assoziiert und wirken sich negativ auf die Lebenserwartung aus.

Die Ursachen der MR sind zum einen endogene Faktoren, wie Chromosomenaberrationen, monogene und polygene Defekte oder molekulargenetisch definierte Syndrome. Aber auch exogene Faktoren, die cerebrale Schäden verursachen, wie Unfälle, Sauerstoffmangel bei der Geburt, Hirnhautentzündungen, Alkoholkonsum der Mutter während der Schwangerschaft oder Strahlung, verursachen MR.

MR kommt mit einer Prävalenz von 2-3% in der Population vor (Roeleveld et al. 1997). Die Häufigkeit der mäßigen bis schweren MR mit einem IQ von <50, wird von den meisten Autoren mit 0.3-0.4% angegeben, während die Häufigkeit der leichten Form der MR zwischen den einzelnen Studien stark variiert (Roeleveld et al. 1997, Yeargin-Allsopp et al. 1997, Leonard et al. 2002).

Genetische Untersuchungen ermöglichen eine Diagnose einer Vielzahl von Erkrankungen, die mit einer MR einhergehen. Damit kann auch eine Aussage zu einem eventuellen Wiederholungsrisiko in den betroffenen Familien getroffen werden und gegebenenfalls eine gezielte Frühförderung der erkrankten Kinder einsetzen.

Die Ätiologie von MR kann mit klinischen Untersuchungen (z.B. dysmorphologische, metabolische und neurologische Untersuchungen), zytogenetischen Untersuchungen (z.B. Karyotypisierung und Untersuchungen der Subtelomerbereiche) oder Untersuchungen auf spezifische Mutationen (z.B. Fragiles X-Syndrom) erfolgen (van Karnebeek et al. 2005). Aber trotz der vielfältigen Untersuchungsmöglichkeiten bleibt in vielen Fällen die Ursache der MR unklar. Neben der Trisomie 21, die in ungefähr 5-15% die Ursache der genetisch verursachten MR ist, werden Chromosomenanomalien in nicht mehr als 5% der Fälle entdeckt, die zytogenetisch untersucht werden (Leonard et al. 2002, de Vries et al. 1997, van Karnebeek et al. 2005). Das Auflösungsvermögen der mikroskopischen Techniken zur Entdeckung von anormalen Karyotypen liegt

zwischen 5-10 Mb. Kleinere Rearrangierungen im Subtelomerbereich können in ungefähr 5% der betroffenen Kinder mit *Flourescence in situ Hybridization* (FISH) oder *Multiplex ligation-dependend Probe Amplification* (MLPA) nachgewiesen werden (Flint et al. 1995, Knight et al. 1999, de Vries et al. 2003, Koolen et al. 2004).

Erst in den letzten Jahren konnte die Auflösung mit Hilfe von array-CGH bis auf einige 100 kb erhöht werden und Chromosomenaberrationen in weiteren 10-20% der Patienten gefunden werden (de Vries et al. 2005, Shaw-Smith et al. 2004, Vissers et al. 2003, Menten et al. 2006, Tyson et al. 2005). Neben der array-CGH wurden SNP-Oligonukleotid-Mikroarrays entwickelt, um das Genom auf Aberrationen zu untersuchen. Nur wenige Arbeiten zu diesem Thema wurden mit SNP-Oligonukleotid-Mikroarrays durchgeführt (Slater et al. 2005, Friedman et al. 2006, Tyson et al. 2005). In dieser Arbeit sollte die Eignung von SNP-Oligonukleotid-Mikroarrays bei der Identifizierung von CNVs untersucht werden.

3. Aufgabenstellung

Im ersten Teil der Arbeit sollte die Technik der SNP-Oligonukleotid-Mikroarrays von Affymetrix etabliert werden und die Eignung von SNP-Oligonukleotid-Mikroarrays für Kopplungsanalysen untersucht werden. Die ersten Untersuchungen wurden mit den 10K-Arrays durchgeführt und sollten in Familien mit unterschiedlichen Krankheiten zur Identifizierung einer Kandidatenregion führen. Die Kopplungsanalysen sollten an 5 Familien mit HHRH, 3 Familien mit ARHP und einer Familie mit SOS durchgeführt werden. Aus den gefundenen Kandidatenregionen sollte im Anschluss eine Auswahl an Kandidatengenen getroffen werden und diese auf Mutationen überprüft werden.

Im zweiten Teil der Arbeit sollte versucht werden, mit den 100K-Arrays Deletionen und Duplikationen im submikroskopischen Bereich zu identifizieren. Zur Untersuchung standen dafür DNAs von Kindern mit mentaler Retardierung zur Verfügung, die in vorausgegangenen zytogenetischen Untersuchungen unauffällige Karyogramme hatten. Es sollte untersucht werden, ob und bis zu welcher Auflösungsgrenze CNVs detektiert werden können. Zusätzlich wurden noch 5 Patienten mit scheinbar balancierten Translokationen auf CNVs in den Bruchpunktregionen mit den 100K-Arrays analysiert.

B Material und Methoden

1 Patientenmaterial

Das in der vorliegenden Arbeit verwendete Patientenmaterial wurde im Rahmen genetischer Diagnostik und Beratung an das Institut für Humangenetik gesandt. Die beteiligten Institute und Praxen sind im Folgenden aufgeführt:

Material für HHRH:

Dr. Ze'ev Hochberg, Meyer Children's Hospital and Technion-Israel Institute of Technology, Haifa, Israel

Dr. Norbert Albers, Kinderkrankenhaus, Osnabrück

Dr. Dirk Schnabel, Department for Pediatric Endocrinology, Charité, Berlin

Material für DMP1:

Dr. Ursula Müller-Barth, Medizinische Genetik, Hanau

Dr. Feliciano Ramos, Department of Pediatrics, University Hospital, Faculty of Medicine, University of Zaragoza, Zaragoza, Spanien

Dr. Harald Jüppner, Endocrine Unit and Pediatric Nephrology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, USA

Material für SOS:

Dr. Heinrich Schmidt, Kinderkrankenhaus, Ludwig-Maximilians-Universität, München

Dr. Gunther Rudolph, Augenklinik, Ludwig-Maximilians-Universität, München

Material für die Untersuchung der Kopienzahlvariationen:

Dr. Stephanie Spranger und Dr. Bernd Kazmierczak, Praxis für Humangenetik, Bremen

Dr. Michaela Nathrath, Abteilung für Pädiatrie, Technische Universität München

Dr. Dagmar Wahl, Praxis für humangenetische Beratung, Augsburg

Dr. Babett Heye, Institut für Humangenetik, München

2 Verwendete Materialien und Geräte

2.1 Chemikalien

Die organischen und anorganischen Chemikalien entsprechen, wenn möglich dem Reinheitsgrad „pro analysis“.

Produkt	Firma
Ampicillin	Roth
anti-Streptavidin Antikörper (biotinyliert, aus Ziege)	Vector Laboratories
β-Mercaptoethanol	Stratagene
Denhardt's Lösung	Sigma
DMSO	Sigma
dNTP-Mix (2,5 mM each)	Takara
EDTA	Sigma
Ethanol	Merck
Glycerin	Merck
H ₂ O _{reinst} (HPLC-Wasser)	Merck
H ₂ O _{Affymetrix} (Ultra Pure Water)	Invitrogen
HSDNA	Promega
IPTG (dioxanfrei)	Roth
Isopropanol	Merck
Menschliche Cot-1	Invitrogen Corporation
MES Hydrat	Sigma
MES Salz	Sigma
Mineralöl	Sigma
NaCl	Merck
R-Phycoerythrin Streptavidin SAPE	Molecular Probes
SSPE 20x	BioWhittaker Molecular Applications/ Cambrex
TMACI 5M	Sigma
Tween-20 (10 %)	Pierce
X-Gal	Sigma

2.2 Enzyme

Enzym	Firma
<i>fast digest</i> -NotI	New England Biolabs
<i>fast digest</i> -PstI	New England Biolabs
HhaI (20U)	New England Biolabs
HindIII (20000 U/ml)	New England Biolabs
HpaII (20U)	New England Biolabs
Proteinase K	Merck
T4 DNA Ligase	New England Biolabs
XbaI (20000 U/ml)	New England Biolabs

2.3 Kits

Kit	Firma
Advantage cDNA PCR Kit	Clontech
AmpliTaq Gold PCR Kit (AmpliTaq Gold Polymerase (5 U/μl), 10x PCR Puffer II, MgCl ₂ (25 mM))	Applied Biosystems
BigDye Terminator Cycle Sequencing Ready Reaction v.3.1 Kit	Applied Biosystems
E.coli XL10-Gold Ultracompetent Cells	Stratagene
iPLEX Gold Assay	Sequenom
iPLEX Termination Mix	Sequenom
LCGreen-Lightscanner-PCR	Applied Biosystems
Marathon Ready cDNA	Clontech
Millipore MultiScreen PCR μ 96	Millipore
NucleoFast 96 PCR clean up Kit	Macherey-Nagel
Platinum Pfx DNA Polymerase Kit (Pfx Polymerase (2,5 U/μl), 10x Pfx Amplification Puffer, 10x PCR Enhancer, 50 mM)	Invitrogen Corporation
Power SYBR Green Master Mix	Applied Biosystems
Promega PCR Mastermix	Promega
QIAprep Spin Plasmid Miniprep Kit	Qiagen
Rapid-Ligation Kit (pGEM-T-Vektorsystem II)	Promega
SAP Behandlung	Sequenom
Thermo-Start DNA Polymerase	Qiagen
Affymetrix 10K Assay	
Adaptor Xba (5 μM)	Affymetrix
PCR-Primer (10 μM)	Affymetrix
Fragmentation Reagent (DNase I) (3 U/μl)	Affymetrix
10x Fragmentation Puffer	Affymetrix
GeneChip DNA Labeling Reagenz (7,5 mM)	Affymetrix
Terminale Desoxynukleotid Transferase (30 U/μl)	Affymetrix
5x Terminale Desoxynukleotid Transferase Puffer	Affymetrix
Oligokontroll-Reagenz	Affymetrix
Affymetrix 50K Assay	
Adaptor Xba (5 μM)	Affymetrix
Adaptor Hind (5 μM)	Affymetrix
PCR-Primer (10 μM)	Affymetrix
Fragmentation Reagent (DNase I) (3 U/μl)	Affymetrix
10x Fragmentation Puffer	Affymetrix
GeneChip DNA Labeling Reagenz (7,5 mM)	Affymetrix
Terminale Desoxynukleotid Transferase (30 U/μl)	Affymetrix
5x Terminale Desoxynukleotid Transferase Puffer	Affymetrix
Oligokontroll-Reagenz 0100	Affymetrix

2.4 Technische Geräte

Gerät	Hersteller
7900HT real-time PCR System	Applied Bioystems
ABI Prism 3730 DNA Sequenzer	Applied Bioystems
DNA Engine Tetrad PTC-225	MJ Research
Fluidics Wash Station 450	Affymetrix
GeneAmp PCR System 9700	Applied Bioystems
GeneChip Hybridization Ofen 640	Affymetrix
GeneChip Scanner 300 7G	Affymetrix
LightScanner HR I 384 Instrument der Firma Idaho	Idaho Technologies

2.5 Datenbanken und Computerprogramme

Bezeichnung	Webseite/Firma
Allegro	http://www.decode.com/software/
Alohomora	http://gmc.mdc-berlin.de/alohomora/
Auswerteskripts für die Analyse der CNV-Daten	http://ihg.gsf.de/cnv-scripts
BLAT	http://genome.ucsc.edu/cgi-bin/hgBlat?command=start&org=Human&db=hg18&hgsid=98847378
Call-IT Software	Idaho Technology
Database of Genomic Variants (DGV)	http://projects.tcag.ca/variation/
DECIPHER Database	http://projects.tcag.ca/variation/
Exon Primer	http://ihg.gsf.de/ihg/ExonPrimer.html
GeneBank	http://www.ncbi.nlm.nih.gov/Genbank/
GeneChip Operating Software (GCOS)	Affymetrix
GeneMapper Software	Applied Biosystems
GTYPE 4.0 Software	Affymetrix
Hi-Res Melting Software	Idaho Technology
Merlin	http://www.sph.umich.edu/csg/abecasis/Merlin/
NCBI	http://www.ncbi.nlm.nih.gov
OMIM	http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM
PedCheck 1.1	http://watson.hgen.pitt.edu/register/docs/pedcheck.html
Pub Med	http://www.ncbi.nlm.nih.gov/sites/entrez
Staden	MCR Laboratory of Molecular Biology, Cambridge, UK
UCSC Genome Browser	http://genome.ucsc.edu/

3 DNA-Isolierung

Die DNA-Extraktion erfolgt aus 5-10 ml EDTA-Vollblut nach einem modifizierten Protokoll von Miller *et al.* aus peripheren Blutleukozyten (Miller *et al.* 1988).

Durch Zugabe von 30 ml RBC-Lysis-Puffer werden die Erythrozyten lysiert und in einem anschließenden Zentrifugationsschritt (2500 rpm, 10 min, 4°C) von den Leukozyten getrennt. Das Pellet mit den kernhaltigen Zellen wird in 5 ml SE-Puffer resuspendiert und mit 25 µl Proteinase K (10 mg/ml) und 250 µl SDS (20%) bei RT über Nacht inkubiert. Nach der Proteolyse werden zur Proteinfällung 5 ml SE-Puffer zugegeben und für 5 min bei 55°C inkubiert. Danach werden 3 ml 6 M NaCl hinzugefügt, gut gemischt und bei 3500 rpm für 15 min bei RT zentrifugiert. Nach der Proteinpräzipitation wird der Überstand, der die DNA enthält, in ein sauberes Röhrchen überführt und die DNA durch die Zugabe von einem Volumenanteil Isopropanol gefällt. Nach einem Zentrifugationsschritt (4500 rpm, 5 min, RT) wird der Überstand abgekippt und die zu einem Knäuel aggregierte DNA bei RT für 30 min getrocknet. Anschließend wird der DNA-Faden in 100 - 1000 µl TE-Puffer aufgenommen und gelöst. Die Konzentration der DNA wird mit einem NanoDrop

Spectrophotometer (ND-1000 V.3.1.2) gemessen und mit H₂O_{reinst} auf die gewünschte Konzentration eingestellt.

4 Polymerase-Kettenreaktion (PCR)

4.1 Standard-PCR

Die DNA-Amplifikation mit Hilfe der PCR wird nach der Standard-Methode von Saiki (Saiki et al., 1988) durchgeführt. Abhängig von der verwendeten Ausgangs-DNA (genomische DNA oder cDNA) variieren die Reaktionsansätze und PCR-Bedingungen. Spezifische Oligonukleotidprimer für die PCR werden mit dem Programm *Exonprimer* konstruiert und die Spezifität anschließend über elektronische PCR (*NCBI*) und BLAT-Suche (*NCBI*) überprüft. Die Synthese der Primer wird von der Firma Metabion durchgeführt. Detaillierte Beschreibungen der Primersequenzen, Anlagerungstemperaturen und amplifizierten Produktgrößen finden sich im Anhang. Es wird immer eine Negativkontrolle und, wenn erforderlich, eine Positivkontrolle mitgeführt. Als Ausgangsmenge für die Amplifikation werden 50 - 100 ng genomische DNA in einem Reaktionsvolumen von 25 µl verwendet.

Als Standard-PCR-Ansatz wurde das Protokoll des Promega-Mastermixes verwendet:

<u>Ansatz Promega-PCR:</u>	<u>Volumen in µl</u>
Mastermix	12,5 µl
Primer F (20 pmol/µl)	1,0 µl
Primer R (20 pmol/µl)	1,0 µl
Genomische DNA (50-100 ng/µl)	1,0 µl
H ₂ O _{reinst}	ad 25,0 µl

Der Promega-MasterMix enthält 50 U/ml Taq DNA Polymerase, Puffer, je 400 µM dATP, dTTP, dGTP, dCTP und 3mM MgCl₂.

Die PCR-Reaktion erfolgte, wenn nicht anders erwähnt, in programmierbaren Thermocyclern (DNA Engine Tetrad PTC-225, MJ Research und GeneAmp PCR System 9700, Applied Biosystems) in 0,5 ml Reaktionsgefäßen bzw. 96-well-Microtiterplatten (ABgene) unter folgenden Bedingungen:

1.	95 °C	5 min	Initiale Denaturierung
2.	95 °C	30 sec	Denaturierung
3.	(*) °C	30 sec	Primeranlagerung
4.	72 °C	(**) sec	Synthese
5.	Wiederholung der Schritte 2-4 in 29-34 Zyklen		
6.	72 °C	5 min	Finale Elongation

Dabei sind die Temperatur der Primeranlagerung (*) und die DNA-Synthesezeit (**) variabel und von der Basenzusammensetzung des jeweiligen Primers bzw. der Größe des zu amplifizierenden PCR-Produkts abhängig.

4.2 PCR für das Affymetrix 10K- und 100K-Protokoll

Die Amplifikation für die Affymetrix-Arrays erfolgt nach dem Restriktionsenzymverdau (siehe Material und Methoden 10.1.1) und der Ligierung des Adaptors (siehe Material und Methoden 10.1.2) mit einem universellen Primer (PCR Primer Xba bzw. PCR Primer Hind; *Affymetrix*), der eine Sequenz auf dem jeweiligen Adaptor (Xba- oder Hind-Adaptor, *Affymetrix*) erkennt und daran bindet. Dabei werden alle Fragmente mit gebundenem Adaptor amplifiziert und die Größe der Produkte wird nur durch die PCR-Bedingungen festgelegt. Für die Amplifikation der Affymetrix-PCR wird ausschließlich der Thermocycler GeneAmp PCR Systems 9700 der Firma Applied Biosystems benutzt.

<u>Ansatz Affymetrix 10K PCR:</u>	<u>Volumen in µl</u>
AmpliTaq Gold PCR-Puffer II (10x)	40 µl
dNTP-Mix (2,5 mM, Takara)	40 µl
MgCl ₂ (25 mM)	40 µl
PCR Primer Xba (10 µM)	30 µl
AmpliTaq Gold Polymerase (5 U/µl)	8 µl
Ansatz Affymetrix-Ligation 10K	40 µl
H ₂ O _{reinst}	ad 400 µl

Bedingungen Affymetrix 10K-PCR:

1.	95 °C	3 min
2.	95 °C	30 sec
3.	59 °C	30 sec
4.	72 °C	30 sec
5.	Wiederholung der Schritte 2-4 in 34 Zyklen	
6.	72 °C	7 min

<u>Ansatz Affymetrix 100K PCR:</u>	<u>Volumen in μl</u>
Pfx Amplification Puffer (10x)	30 μ l
dNTP-Mix (2,5 mM, Takara)	36 μ l
PCR Enhancer (10x)	30 μ l
MgSO ₄ (50 mM)	40 μ l
PCR Primer (10 μ M)	30 μ l
Pfx Polymerase (2.5 U/ μ l)	6 μ l
Ansatz Affymetrix-Ligation 100K	30 μ l
H ₂ O _{reinst}	ad 300 μ l

Bedingungen Affymetrix 100K-PCR:

1. 94 °C 3 min
2. 94 °C 30 sec
3. 60 °C 45 sec
4. 68 °C 60 sec
5. Wiederholung der Schritte 2-4 in 29 Zyklen
6. 68 °C 7 min

4.3 RT-PCR und *nested*-PCR

Die *nested*-PCR bietet die Möglichkeit auch geringe Ausgangsmengen eines PCR-Templates oder ein zu untersuchendes Gen, das im jeweiligen Gewebe nur sehr niedrig exprimiert ist, nachzuweisen. Bei der *nested*-PCR werden zwei Amplifikationsschritte durchgeführt. Nach der ersten PCR wird das entstandene Produkt als Template für eine zweite PCR genutzt. Für die zweite Runde liegen entweder beide Primer (*nested*-PCR) oder nur einer (*semಿನested*-PCR) innerhalb des PCR-Produkts der ersten Runde. Für die RT-PCR und die *nested*-PCR wird der Advantage cDNA PCR Kit (Clontech) verwendet.

<u>Ansatz RT-PCR und <i>nested</i>-PCR:</u>	<u>Volumen in μl</u>
Advantage Puffer (10x)	2,5 μ l
Advantage dNTP-Mix (50x)	0,5 μ l
Primer F (10 μ M)	0,5 μ l
Primer R (10 μ M)	0,5 μ l
Advantage Polymerase (2.5 U/ μ l)	0,5 μ l
cDNA bzw. PCR-Produkt	1,0 μ l
H ₂ O _{reinst}	ad 25,0 μ l

PCR-Bedingungen RT-PCR und *nested*-PCR:

1. 94 °C 1 min
2. 94 °C 30 sec
3. 60 °C 30 sec
4. 68 °C 90 sec
5. Wiederholung der Schritte 2-4 in 29 Zyklen
6. 68 °C 3 min

4.4 RACE-PCR

Die RACE-PCR („Rapid Amplification of cDNA Ends“) dient zur Vervollständigung des 5'- bzw. des 3'-Endes von cDNA. Als Ausgangsmaterial wird cDNA mit synthetischen Sequenz-Adaptoren verwendet (*Marathon Ready cDNA, human brain, human fetal brain, human testis; Clontech*). Im ersten Amplifikationsschritt wird mit Hilfe eines genspezifischen Primers, der komplementär revers in der codierenden Sequenz liegt und einem Adaptor-Primer (AP1) ein spezifisches PCR-Produkt generiert. Es folgt eine *nested-* oder *semಿನested-*PCR, bei der das PCR-Produkt der ersten Amplifikation als Template dient. Für die Amplifikation der RACE-Produkte wird der Promega PCR MasterMix (Promega) verwendet.

<u>Ansatz RACE-PCR 1.Runde</u>	<u>Volumen in μl</u>
Promega MasterMix (10x)	12,5 μ l
Primer AP1	1,0 μ l
5racePrimer R (10 μ M)	1,0 μ l
cDNA	1,0 μ l
H ₂ O _{reinst}	ad 25,0 μ l

<u>Ansatz RACE-PCR 2.Runde</u>	<u>Volumen in μl</u>
Promega MasterMix (10x)	12,5 μ l
Primer AP2	1,0 μ l
5racePrimer R (10 μ M)	1,0 μ l
PCR-Produkt aus 1. Runde	1,0 μ l
H ₂ O _{reinst}	ad 25,0 μ l

Bedingungen RACE-PCR:

1. 95 °C 5 min
2. 95 °C 30 sec
3. 62 °C 30 sec
4. 72 °C 90 sec
5. Wiederholung der Schritte 2-4 in 29 Zyklen
6. 72 °C 5 min

4.5 PCR für das Idaho-Protokoll

<u>Ansatz LCGreen-Lightscanner-PCR:</u>	<u>Volumen in μl</u>
Thermo-Start High Performance Puffer (10x)	0,50 μ l
dNTP-Mix (2 mM)	0,50 μ l
MgCl ₂ (25 mM)	0,50 μ l
Primer F (0,4 μ M)	0,20 μ l
Primer R (0,4 μ M)	0,20 μ l
LCGreen PLUS (1x)	0,25 μ l
Thermo-Start DNA Polymerase (0,25 U)	0,05 μ l
H ₂ O _{reinst}	ad 5,00 μ l

Bedingungen LCGreen-Lightscanner-PCR:

Die PCR wurde mit 8 µl Mineralöl (Sigma) überschichtet.

1. 95 °C 15 min
2. 94 °C 30 sec
3. 61 °C 30 sec
4. 72 °C 30 sec
5. Wiederholung der Schritte 2-4 in 39 Zyklen
6. 94 °C 30 sec
7. 20 °C 30 sec

4.6 PCR für iPLEX Gold Assay

Ansatz iPLEX Gold Assay PCR:

	<u>Volumen in µl</u>
PCR Puffer mit MgCl ₂ (10x)	0,625 µl
dNTP-Mix (25 mM)	0,100 µl
MgCl ₂ (25 mM)	0,325 µl
Primermix (50 µM)	1,000 µl
Hotstar Taq (5 U/µl)	0,010 µl
Genomische DNA (5-10 ng/µl)	1,000 µl
H ₂ O _{reinst}	ad 5,00 µl

Bedingungen iPLEX Gold Assay PCR:

1. 94 °C 15 min
2. 94 °C 20 sec
3. 56 °C 30 sec
4. 72 °C 60 sec
5. Wiederholung der Schritte 2-4 in 44 Zyklen
6. 72 °C 3 min

4.7 Quantitative PCR

Die quantitative PCR (qPCR) ist ein Verfahren, bei dem die Menge des entstandenen PCR-Produktes während der PCR-Reaktion in Echtzeit gemessen wird. Diese Methode wird in der vorliegenden Arbeit verwendet, um die DNA-Kopienanzahl zu quantifizieren und um Bruchpunkte von Deletionen näher einzugrenzen. Dazu wird dem PCR-Ansatz SYBR-Green zugesetzt, ein fluoreszierender Farbstoff, der sich unspezifisch in die große Grube der doppelsträngigen DNA einlagert. Die qPCR wird mit dem 7900HT real-time PCR System (Applied Biosystems) und dem Power SYBR Green Master Mix (Applied Biosystems) durchgeführt. Bei jedem Assay werden eine Negativkontrolle (no-template control, NTC) und vier Kontroll-DNAs (2 männliche und 2 weibliche DNA-Proben) und ein Referenzgen mitgeführt. Alle Proben, mit Ausnahme der NTC werden als Duplikate angesetzt.

<u>Ansatz Power SYBR Green-PCR:</u>	<u>Volumen in μl</u>
Power SYBR Green Master Mix (2x)	10,0 μ l
Primer F (0.2 μ M)	0,4 μ l
Primer R (0.2 μ M)	0,4 μ l
DNA (10ng/ μ l)	5,0 μ l
H ₂ O _{reinst}	ad 20 μ l

Bedingungen Power SYBR Green-PCR:

1. 50 °C 2 min
2. 95 °C 10 min
3. 95 °C 15 sec
4. 58 °C 60 sec
5. Wiederholung der Schritte 3-4 in 39 Zyklen
6. 95 °C 15 sec
7. 60 °C 15 sec
8. 95 °C 15 sec

Die Schritte 6-8 sind die Bedingungen für eine Dissoziationskurve, die dem Ausschluss unspezifischer Produkte dient.

4.8 PCR für Mikrosatellitenanalyse

<u>Ansatz PCR:</u>	<u>Volumen in μl</u>
10x PCR-Puffer (Qiagen, incl. MgCl ₂)	1,0 μ l
Qiagen-Taq Polymerase (5U/ μ l)	0,1 μ l
dNTP-Mix (je 2 mM)	1,0 μ l
Primer F (10 pmol/ μ l)	0,1 μ l
Primer R (10 pmol/ μ l)	0,1 μ l
DNA (1ng/ μ l)	5,0 μ l
H ₂ O _{reinst}	ad 10,0 μ l

Bedingungen für den Qiagen-PCR-Ansatz:

1. 95 °C 15 min
2. 95 °C 20 sec
3. 55 °C 30 sec
4. 72 °C 45 sec
5. Wiederholung der Schritte 2-4 in 35 Zyklen
6. 72 °C 15 min

5 Reinigung von PCR-Produkten

5.1 Reinigung durch Ultrafiltration

Die in den jeweiligen PCR-Ansätzen erhaltenen Produkte müssen für die weitere Verwendung (Sequenzierung, Klonierung, Fragmentierung) aufgereinigt werden, um Puffer, Nukleotide und ungebundene Primer zu entfernen. Für die Reinigung wurden der NucleoFast 96 PCR clean up Kit (Macherey-Nagel) oder die MultiScreen PCR _{μ 96} Filter Plates (Millipore) verwendet. Beide Kits

unterliegen dem gleichen Ultrafiltrations-Prinzip. Das PCR-Produkt wird mit Wasser auf ein Volumen von mindestens 100 µl eingestellt, auf eine Ultrafiltrationsmembran geladen und unter Vakuum (600 mbar) für 15 bis 45 min gesetzt. Die Porengröße der Filtermembran ist so gewählt, dass Fragmente, die größer sind als 150 bp auf der Membran haften bleiben, während kleinere Fragmente (Primer), Salze und Nukleotide durch die Membran gesaugt werden. Optional wird ein Waschschrift mit 100 µl Wasser (H_2O_{reinst}) durchgeführt. Um das gereinigte PCR-Produkt zu erhalten, werden mindestens 25 µl Wasser (H_2O_{reinst}) oder EB-Puffer zugefügt und durch Mischen auf einem Schüttler (200 rpm, 15 min) von der Membran gelöst. Anschließend kann das PCR-Produkt in ein sauberes Röhrchen überführt und für weitere Schritte verwendet werden.

5.2 Reinigung durch Gelextraktion

Für die Gelextraktion wurde der Qiaquick Gel Extraction Kit der Firma Qiagen verwendet. Das entsprechende PCR-Produkt wurde so exakt wie möglich aus einem präparativen Agarosegel ausgeschnitten und nach Angaben des Herstellers extrahiert. Das gereinigte PCR-Produkt wurde in 30 µl H_2O_{reinst} eluiert.

6 DNA-Sequenzierung

6.1 Sequenzreaktion

Die Sequenzreaktion basiert auf der Kettenabbruchsequenzierung von Sanger (Sanger et al., 1977) und der linearen Amplifikation von Craxton (Craxton, M., 1991). Dabei wird die DNA mit Fluoreszenzfarbstoffen an Didesoxynukleotiden markiert, die von dem Kapillarsequenziergerät ABI Prism 3730 DNA Sequenzer (Applied Biosystems) während der Elektrophorese mittels Laser angeregt und gemessen werden. Die vier Didesoxynukleotide (ddATP, ddTTP, ddGTP, ddCTP) sind mit vier verschiedenen Farbstoffen markiert und ermöglichen so die Unterscheidung der Nukleotide. Die Sequenzreaktion wird mit dem BigDye Terminator Cycle Sequencing Ready Reaction v.3.1 Kit der Firma Applied Biosystems angesetzt.

<u>Ansatz Sequenzreaktion:</u>	<u>Volumen in μl</u>
Big Dye Terminator Puffer	1,0 μ l
Primer F bzw. R (0.2 μ M)	1,0 μ l
Big Dye Terminator	0,5 μ l
PCR (50-100 ng/ μ l)	1,0-2,5 l
H ₂ O _{reinst}	ad 5,0 μ l

Bedingungen Sequenzreaktion:

1. 96 °C 10 sec
2. 50 °C 5 sec
3. 60 °C 4 min
5. Wiederholung der Schritte 1-3 in 25 Zyklen

6.2 Fällung der Sequenzreaktion

Um nicht gebundene Didesoxynukleotide aus dem Sequenzansatz zu entfernen, wird eine Ethanolfällung durchgeführt. Zum Reaktionsansatz in der Mikrotiterplatte werden 25 μ l Ethanol (100%) zugegeben, 15 min bei RT inkubiert und danach für 30 min bei 3000 x g zentrifugiert. Anschließend wird das Pellet mit 50 μ l Ethanol (70%) gewaschen und zentrifugiert (15 min, 2000 x g). Um das Pellet zu trocknen, wird die Mikrotiterplatte mit der Öffnung nach unten 1 min bei 600 x g zentrifugiert, 5 min bei RT getrocknet und schließlich in 25-50 μ l Wasser gelöst.

6.3 Sequenzanalyse

Die Sequenzanalyse erfolgt auf dem automatisierten Sequenziergerät ABI Prism 3730 DNA Sequenzer (Applied Biosystems) nach den vom Hersteller angegebenen Bedingungen. Es werden je 25 μ l der gereinigten Sequenzreaktion auf eine Sequenzierplatte aufgetragen. Die erhaltenen Daten werden durch das *Staden Package* als Elektropherogramme dargestellt und ausgewertet. Dabei werden den 4 Basen die Farben rot für T, grün für A, schwarz für G und blau für C zugeordnet.

7 Klonierung von PCR-Produkten

Die Klonierung von PCR-Produkten dient in der vorliegenden Arbeit zur Vereinzelung von DNA-Molekülen.

7.1 Ligation

Die Ligation erfolgt nach dem TA-Klonierungsprinzip, bei dem das PCR-Produkt einen A-Überhang am 3'-Ende besitzt und in einen Vektor mit einem T-Überhang kloniert wird („*sticky end*“-Ligation). Bei der Amplifikation mit Taq-Polymerasen wird den synthetisierten 3'-Strangenden eine zusätzliche Base, meistens ein dATP angehängt. Wenn bei der PCR Polymerasen verwendet werden, die keinen A-Überhang synthetisieren, muss vor der Ligation ein zusätzlicher Schritt, das sogenannte A-Tailing durchgeführt werden. Das gereinigte PCR-Produkt wird dabei mit Taq-Polymerase und dATPs für 30 min inkubiert, um so einen A-Überhang zu generieren. Die Ligation wird mit dem pGEM-T-Vektorsystem II von Promega (Rapid-Ligation Kit) durchgeführt. Für die Ligation wird ein molares Verhältnis von etwa 1:1 zwischen Insert und Vektor angestrebt. 50 ng gereinigtes PCR-Produkt werden in dem Ligationsansatz mit 50 ng pGEM-T-Vektor, 1x Rapid-Puffer und 3 U T4-Ligase in einem Reaktionsvolumen von 10 µl bei 4 °C ÜN inkubiert.

<u>Ansatz A-Tailing:</u>	<u>Volumen in µl</u>
Puffer (10x)	1,0 µl
MgCl ₂ (25 mM)	0,6 µl
dATP (2 mM)	1,0 µl
Taq-Polymerase (5 U/µl)	1,0 µl
Gereinigtes PCR-Produkt	2,4 µl
H ₂ O _{reinst}	ad 10,0 µl

Inkubation: 30 min bei 70 °C

<u>Ansatz Ligation:</u>	<u>Volumen in µl</u>
Rapid Puffer (2x)	5 µl
pGEM-T-Vektor	1 µl
T4-Ligase	1 µl
A-Tailing-Ansatz	3 µl

Inkubation: ÜN bei 4°C

7.2 Transformation

Die Ligation wird in den kompetenten Bakterienstamm *E.coli XL10-Gold Ultracompetent Cells* (Stratagene) transformiert. Ein Aliquot von 100µl kompetenter Bakterien wird auf Eis aufgetaut und mit 4 µl β-Mercaptoethanol (Stratagene) vermischt. Anschließend wird der Ligationsansatz dazugegeben

und für 30 min auf Eis inkubiert. Es folgt eine Hitzeschockbehandlung von 30 sec bei 42 °C und eine Abkühlung auf Eis (2 min). Nach der Zugabe von 900 µl Medium NZY⁺broth werden die transformierten Bakterien für 1 h im Inkubationsschüttler bei 37 °C und 250 U/min kultiviert. Die Bakterien werden bei 1000 U/min für 10 min zentrifugiert und 900 µl Überstand abpipettiert. Das Pellet wird resuspendiert. Die Bakteriensuspension wird auf einer LB/Amp-Agarplatte mit 1,25 µl 1 M IPTG und 25 µl 10% X-Gal ausgestrichen und ÜN bei 37 °C im Brutschrank inkubiert. Durch die Blau-Weiß Selektion können positive Klone (weiße Kolonien) mit dem richtigen Insert identifiziert werden. Positive Einzelkolonien werden in 5 ml NZY⁺broth-Medium mit 50 µg/µl Ampicillin aufgenommen und ÜN im Schüttler bei 37 °C und 250 U/min kultiviert.

7.3 Präparation von Plasmid-DNA

Aus der ÜN-Kultur werden 500 µl entnommen, mit 500 µl Glycerin vermischt und als Glycerinstock bei -70 °C konserviert.

Die Isolierung der Plasmid-DNA erfolgt mit dem QIAprep Spin Plasmid Miniprep Kit (Qiagen) nach den Angaben des Herstellers. Die Elution der Plasmid-DNA erfolgt mit 50 µl H₂O_{reinst}.

7.4 Restriktionsspaltung

Zur Überprüfung der isolierten Plasmide wird ein Restriktionsenzymverdau mit den *fast digest*-Enzymen PstI oder NotI (New England Biolabs) durchgeführt. Die Restriktionsspaltung wird anschließend auf ein Agarosegel aufgetragen und Produkte mit der richtigen Größe sequenziert.

<u>Ansatz Restriktionsspaltung:</u>	<u>Volumen in µl</u>
Puffer fast digest	2,0 µl
PstI / NotI fast digest enzyme	1,0 µl
Gereinigte Plasmid-DNA	0,3-1,0 µl
H ₂ O _{reinst}	ad 20,0 µl

Inkubation: 8 min bei 37°C

8 Mutationsscreening mit DNA-Schmelzkurvenanalyse (Idaho)

Um einzelne Nukleotidvariationen in Genen in einem Patientenpanel zu untersuchen, wird eine Technik verwendet, bei der Sequenzvariationen in einem Fragment amplifizierter DNA durch die Analyse der Schmelzkurven detektiert werden können. In einem diploiden Genom werden paternale und maternale Chromosomen zu gleichen Teilen amplifiziert. Die DNA-Fragmente werden analysiert und in Homoduplicae (Fragmente, die vollständig zusammen passen) und Heteroduplicae (Fragmente, die einen Mismatch enthalten) aufgeteilt. Die Mutationen werden durch Änderungen in der Gestalt der Schmelzkurve im Vergleich zu Referenzkurven identifiziert.

Zu dem PCR-Mastermix wird ein Farbstoff, LCGreen, beigemischt. LCGreen bindet an dsDNA und hat ein Fluoreszenzoptimum bei 470-520 nm. Die Fluoreszenz wird im LightScanner (Idaho) gemessen und Bilder von der Schmelzkurve mittels einer CCD-Kamera aufgenommen. Durch den Vergleich zwischen den einzelnen Proben werden die Schmelzkurven ausgewertet.

Für den Amplifikationsschritt werden Primer konstruiert, die die codierenden Exons flankieren und Produkte von maximal 300 bp amplifizieren. Der Zusatz des Farbstoffes erhöht die Schmelztemperatur der DNA um 1-3 °C. Die Anlagerungstemperatur muss deshalb angepasst werden. Genomische DNA (je 10 ng pro Well) wird auf einer 384-Mikrotiterplatte verteilt und eingetrocknet.

Die Durchführung der PCR ist weiter oben beschrieben (siehe Material und Methoden 4.5).

Anschließend folgt die Schmelzkurvenanalyse auf einem LightScanner HR I 384 Instrument der Firma Idaho (Idaho Technologies) und die Auswertung mit der *Hi-Res Melting Software* und der *Call-IT Software* (Idaho Technologies).

9 Inaktivierung von X-Chromosomen

Zur Untersuchung der X-chromosomalen Inaktivierung wird der hoch polymorphe Trinukleotidrepeat im ersten Exon des Gens für den humanen Androgenrezeptor (*HUMARA*) verwendet (Allen et al, 1992).

2 µg genomische DNA werden mit den methylierungs-sensitiven Restriktionsenzymen HhaI oder HpaII (jeweils 20U) (New England Biolabs) ÜN verdaut. Ein Aliquot von 2 µl wird durch eine PCR-Reaktion mit einem FAM-markierten F-Primer und einem unmarkiertem R-Primer amplifiziert. Die PCR-Produkte werden auf einem ABI Prism 3730 DNA Sequenzer (Applied Biosystems) aufgetrennt und mit der GeneMapper Software (Applied Biosystems) analysiert.

<u>Ansatz HhaI-Verdau</u>	<u>Volumen in µl</u>
Puffer NEB4 (10x)	4,0 µl
BSA (10x)	0,4 µl
HhaI	2,0 µl
H ₂ O _{reinst}	ad 40,0 µl

<u>Ansatz HpaII-Verdau</u>	<u>Volumen in µl</u>
Puffer NEB1 (10x)	4,0 µl
BSA (10x)	0,4 µl
HpaII	2,0 µl
H ₂ O _{reinst}	ad 40,0 µl

<u>Ansatz AmpliTaq Gold-PCR:</u>	<u>Volumen in µl</u>
AmpliTaq Gold Puffer (10x)	1,5 µl
dNTP-Mix (2,5 mM)	1,0 µl
Primer F (5 pmol)	0,5 µl
Primer R (5 pmol)	0,5 µl
AmpliTaq Gold Polymerase (1U)	0,2 µl
Ansatz Enzymverdau HhaI / HpaII	2,0 µl
H ₂ O _{reinst}	ad 15,0 µl

PCR-Bedingungen für AmpliTaq Gold-PCR:

1. 95 °C 12 min
2. 94 °C 20 sec
3. 58 °C 20 sec
4. 72 °C 30 sec
5. Wiederholung der Schritte 2-4 in 34 Zyklen
6. 72 °C 10 min

10 Affymetrix GeneChip Mapping Assay

10.1 GeneChip Mapping 10K 2.0 Assay

Der Affymetrix GeneChip Mapping 10K 2.0 Assay (Affymetrix 10K) wurde für Kopplungsanalysen von monogenen und komplexen Krankheiten entwickelt. Diese Technik verwendet ein einfaches Verfahren, um die Komplexität des menschlichen Genoms zu reduzieren. Die genomische DNA (250 ng) wird mit dem Restriktionsenzym XbaI (New England Biolabs) verdaut. Dabei entsteht ein 4 bp-Überhang, an den ein von Affymetrix bereitgestellter Adaptor ligiert wird. Dieser Adaptor besitzt eine

Primerbindestelle, mit Hilfe derer ein Amplifikationsschritt mit einem generischen Primer durchgeführt wird. Die entstandenen PCR-Produkte werden anschließend in einem DNaseI-Verdau fragmentiert und durch die terminale Transferase mit Biotin endmarkiert. Die Proben werden mit einem Hybridisierungscocktail gemischt, denaturiert und auf den Chip hybridisiert. Anschließend folgen die Stringenzwaschung und der Nachweis der Markierung mit Immunofluoreszenz.

10.1.1 XbaI-Restriktionsenzymverdau

XbaI gehört zur Familie der Typ II Restriktionsenzyme und besitzt eine palindromische Erkennungssequenz 5'-...T[^]CTAGA...-3' (bzw. 3'-...AGATC[^]T...-5'). Beim Schneiden der DNA entsteht ein spezifischer 4 bp-Überhang.

<u>Ansatz XbaI-Verdau:</u>	<u>Volumen in µl</u>
NE Puffer (10x)	2,0 µl
BSA (10x 1 mg/ml)	2,0 µl
XbaI (20 U/µl)	0,5 µl
Genomische DNA (50 ng/µl)	5,0 µl
H ₂ O _{reinst}	ad 20,0 µl

Bedingungen XbaI-Verdau:

1. 37 °C 120 min
2. 70 °C 20 min
3. 4 °C 10 min

10.1.2 Ligation

<u>Ansatz Ligation:</u>	<u>Volumen in µl</u>
T4 DNA Ligase Puffer (10x)	2,50 µl
Adaptor Xba (5 µM)	1,25 µl
T4 DNA Ligase (400 U/µl)	1,25 µl
Ansatz XbaI-Verdau	20,00 µl

Bedingungen Ligation:

1. 16 °C 120 min
2. 70 °C 20 min
3. 4 °C 10 min

Der Ligationsansatz wird anschließend mit 75 µl H₂O_{reinst} auf ein Volumen von 100 µl erhöht.

10.1.3 PCR

Für jede Probe werden jeweils 4 PCR-Reaktionen mit 100 µl angesetzt.

Die Durchführung der PCR ist weiter oben beschrieben (siehe Material und Methoden 4.2).

Zur Kontrolle wird nach der PCR-Reaktion ein Aliquot von 3 µl auf ein 2%iges Agarosegel aufgetragen (Laufzeit 45-60 min bei 120 V). Die Länge der PCR-Produkte muss zwischen 250 bp und 1000 bp liegen.

Die Aufreinigung der PCR-Produkte erfolgt nach dem Ultrafiltrationsprinzip (siehe Material und Methoden 5.1).

Um das PCR-Produkt anschließend auf die richtige Größe fragmentieren zu können, muss eine Quantifizierung durchgeführt werden. Dazu wird 1 µl der PCR mit dem NanoDrop Spectrophotometer (ND-1000 V.3.1.2) gemessen und der PCR-Ansatz auf die gewünschte Konzentration von 20 µg PCR-Produkt in 45 µl EB-Puffer eingestellt.

10.1.4 Fragmentierung

Um die Länge der PCR-Produkte auf <180 bp zu verkürzen, wird die gereinigte PCR mit dem Fragmentierungsreagenz (DNase I, 3 U/µl, Affymetrix) verdaut. Das Fragmentierungsreagenz wird auf eine Konzentration von 0,048 U/µl verdünnt und mit dem PCR-Produkt vermischt.

<u>Ansatz Fragmentierung:</u>	<u>Volumen in µl</u>
Fragmentierungspuffer (10x)	5,0 µl
verdünntes Fragmentierungsreagenz (0,048 U/µl)	5,0 µl
PCR-Produkt (20 µg)	45,0µl

Bedingungen Fragmentierung:

1. 37 °C 30 min
2. 95 °C 15 min
3. 4 °C 10 min

Zur Kontrolle der Fragmentierung werden 4 µl auf ein 4%iges TBE-Gel aufgetragen und bei 120 V aufgetrennt. Die Größe der Fragmente soll unter 180 bp liegen.

10.1.5 Biotin-Endmarkierung der Fragmente

<u>Ansatz Markierung:</u>	<u>Volumen in μl</u>
Terminaler Desoxynukleatid Transferase Puffer (5x)	14,0 μ l
GeneChip DNA Labeling Reagenz (5 mM)	2,0 μ l
Terminale Desoxynukleotid Transferase (30 U/ μ l)	3,4 μ l
Fragmentierungsansatz	50,6 μ l

Bedingungen für Markierung:

1. 37 °C 120 min
2. 95 °C 15 min
3. 4 °C 10 min

10.1.6 Hybridisierung

<u>Ansatz Hybridisierung:</u>	<u>Volumen in μl</u>
MES Stock	12 μ l
DMSO (100%)	13 μ l
Denhardt's Lösung (50x)	13 μ l
EDTA (0,5 M)	3 μ l
HSDNA (10 mg/ml)	3 μ l
Oligonukleotid-Kontroll-Reagenz	2 μ l
Menschliche Cot-1 (1mg/ml)	3 μ l
Tween-20 3%ig	1 μ l
TMACI (5M)	140 μ l
Labelingansatz	70 μ l

<u>Ansatz MES-Stock-Lösung:</u>	<u>Volumen in μl</u>
MES Hydrat	70,4 g
MES Salz	193,3 g
H ₂ O _{reinst}	ad 1000 ml

Die Lösung wird auf einen pH-Wert von 6,5-6,7 eingestellt und durch einen Filter mit einer Porengöße von 0,2 μ M filtriert.

Bedingungen Markierung:

1. 95 °C 10min
2. auf Eis 10 sec
3. 48 °C 2 min

Aus dem denaturierten Hybridisierungs-Mastermix werden 80 μ l in einen Affymetrix 10K-Array injiziert.

Die Hybridisierung erfolgt in einem Hybridisierungsofen der Firma Affymetrix (Affymetrix GeneChip

Hybridization Ofen 640) bei 48 °C und 60 rpm für 16 -18 Stunden.

10.1.7 Waschen, Färben und Scannen

Die Stringenzwaschschritte und das Färben werden automatisiert in der Affymetrix Fluidics Wash Station 450 mit dem Programm „Mini_Mapping10Kv1_450“ durchgeführt. Um eine maximale Signalintensität zu erhalten, werden die mit Biotin markierten, hybridisierten DNA-Fragmente durch die „Sandwichmethode“ nachgewiesen. Dabei bindet zuerst ein Antikörper (R-Phycoerythrin Streptavidin, SAPE) an das Biotin der endmarkierten Nukleotide, an den wiederum ein biotinylierter Antikörper (anti-Streptavidin Antikörper) bindet. Zum Schluss wird noch einmal mit dem Antikörper (SAPE) gefärbt. Die Signale werden mit dem Affymetrix GeneChip Scanner 300 7G abgelesen. Die Steuerung der Fluidics Stationen und des Scanners erfolgt über die Affymetrix GeneChip Operating Software (GCOS), die auch für die Verwaltung der Experimentinformationen und der Rohdaten verantwortlich ist.

Programm „Mini_Mapping10Kv1_450“:

- | | |
|-------------------------|---|
| 1. Post Hyb Waschen #1 | 6 Zyklen mit 5x Mischen/Zyklus mit Waschpuffer A bei 25 °C |
| 2. Post Hyb Waschen #2 | 6 Zyklen mit 5x Mischen/Zyklus mit Waschpuffer B bei 45 °C |
| 3. 1. Färbeschritt | Färben des Arrays mit SAPE für 10 min bei 25 °C |
| 4. Post Färben Waschen | 6 Zyklen mit 5x Mischen/Zyklus mit Waschpuffer A bei 25 °C |
| 5. 2. Färbeschritt | Färben des Arrays mit Antikörperlösung für 10 min bei 25 °C |
| 6. 3. Färbeschritt | Färben des Arrays mit SAPE für 10 min bei 25 °C |
| 7. Finaler Waschschrift | 10 Zyklen mit 6x Mischen/Zyklus mit Waschpuffer A bei 30 °C |
| 8. Befüllen des Arrays | Befüllen des Arrays mit Array Holding Buffer (25 °C) |

Ansatz Waschpuffer A:

	<u>Volumen in µl</u>
20x SSPE	300 ml
10 % Tween-20	1 ml
H ₂ O _{reinst}	ad 1000 ml

Der Waschpuffer A (nicht-stringent) wird durch einen Filter mit einer Porengröße von 0,2 µM filtriert und bei RT gelagert.

Ansatz Waschpuffer B:

	<u>Volumen in µl</u>
20x SSPE	30 ml
10 % Tween-20	1 ml
H ₂ O _{reinst}	ad 1000 ml

Der Waschpuffer B (stringent) wird durch einen Filter mit einer Porengröße von 0,2 µM filtriert und bei RT gelagert.

<u>Ansatz Array Holding Buffer (1x):</u>	<u>Volumen in μl</u>
12x MES Stock Puffer	8,3 ml
5 M NaCl	18,5 ml
10% Tween-20	0,1 ml
H ₂ O _{reinst}	ad 100 ml

Der Array Holding Buffer wird bei 4 °C unter Lichtabschluss aufbewahrt.

10.2 GeneChip Mapping 100K Assay

Mit dem GeneChip Mapping 100K Assay können über 100000 SNPs auf zwei Arrays (GeneChip Human Mapping 50K Array Xba 240 und GeneChip Human Mapping 50K Array Hind 240, Affymetrix) genotypisiert werden. Das Prinzip des Assays ist analog zu dem 10K Assay. Der Unterschied besteht vor allem darin, dass die Komplexität des Genoms durch den Verdau der genomischen DNA mit zwei Restriktionsenzymen, XbaI und HindIII reduziert wird.

10.2.1 Restriktionsenzymverdau

<u>Ansatz XbaI / HindIII-Verdau:</u>	<u>Volumen in μl</u>
NE Puffer (10x)	2,0 μ l
BSA (10x 1 mg/ml)	2,0 μ l
XbaI / HindIII (20 U/ μ l)	0,5 μ l
Genomische DNA (50 ng/ μ l)	5,0 μ l
H ₂ O _{reinst}	ad 20,0 μ l

Bedingungen XbaI-Verdau:

1. 37 °C 120 min
2. 70 °C 20 min
3. 4 °C 10 min

10.2.2 Ligation

<u>Ansatz Ligation:</u>	<u>Volumen in μl</u>
T4 DNA Ligase Puffer (10x)	2,500 μ l
Adaptor Xba / Hind (5 μ M)	1,250 μ l
T4 DNA Ligase (400 U/ μ l)	0,625 μ l
Ansatz XbaI / HindIII-Verdau	20,000 μ l
H ₂ O _{reinst}	ad 25,000 μ l

Bedingungen Ligation:

1. 16 °C 120 min
2. 70 °C 20 min
3. 4 °C 10 min

Der Ligationsansatz wird anschließend mit 75 µl H₂O_{reinst} auf ein Volumen von 100 µl erhöht.

10.2.3 PCR

Für jede Probe werden jeweils 3 PCR-Reaktionen mit 100 µl angesetzt.

Die Durchführung der PCR ist weiter oben beschrieben (siehe Material und Methoden 4.2).

Zur Kontrolle wird ein Aliquot von 3 µl auf ein 2%iges Agarosegel aufgetragen (Laufzeit 45-60 min bei 120V). Die Länge der PCR-Produkte muss zwischen 250 bp und 2000 bp liegen.

Vor der Aufreinigung der PCR-Produkte nach dem Ultrafiltrationsprinzip (siehe Material und Methoden 5.1) werden 8 µl EDTA (100 mM, pH 8) je PCR-Ansatz zugesetzt. Um das PCR-Produkt anschließend auf die richtige Größe fragmentieren zu können, muss eine Quantifizierung durchgeführt werden. Dazu wird 1 µl der PCR mit dem NanoDrop Spectrophotometer (ND-1000 V.3.1.2) gemessen und der PCR-Ansatz auf die gewünschte Konzentration von 40 µg PCR-Produkt in 45 µl EB-Puffer eingestellt.

10.2.4 Fragmentierung

Um die Länge der PCR-Produkte auf <180 bp zu verkürzen, wird die gereinigte PCR mit dem Fragmentierungsreagenz (DNase I, 3U/µl, Affymetrix) verdaut. Das Fragmentierungsreagenz wird auf eine Konzentration von 0,04 U/µl verdünnt und mit dem PCR-Produkt vermischt.

Ansatz Fragmentierung:

	<u>Volumen in µl</u>
Fragmentierungspuffer (10x)	5,0 µl
verdünntes Fragmentierungsreagenz (0,04 U/µl)	5,0 µl
PCR-Produkt (40 µg)	45,0µl

Bedingungen für Fragmentierung:

1. 37 °C 35 min
2. 95 °C 15 min
3. 4 °C 10 min

Zur Kontrolle der Fragmentierung werden 4 µl auf ein 4%iges TBE-Gel aufgetragen und bei 120 V aufgetrennt. Die Größe der Fragmente soll unter 180 bp liegen.

10.2.5 Biotin-Endmarkierung der Fragmente

<u>Ansatz Biotin-Markierung:</u>	<u>Volumen in μl</u>
Terminaler Desoxynukleotid Transferase Puffer (5x)	14,0 μ l
GeneChip DNA Labeling Reagenz (5 mM)	2,0 μ l
Terminale Desoxynukleotid Transferase (30 U/ μ l)	3,4 μ l
Fragmentierungsansatz	50,6 μ l

Bedingungen für Markierung:

1.	37 °C	120 min
2.	95 °C	15 min
3.	4 °C	10 min

10.2.6 Hybridisierung

<u>Ansatz Hybridisierung:</u>	<u>Volumen in μl</u>
MES Stock	12 μ l
DMSO (100%)	13 μ l
Denhardt's Lösung (50x)	13 μ l
EDTA (0,5 M)	3 μ l
HSDNA (10 mg/ml)	3 μ l
Oligonukleotid-Kontroll-Reagenz	2 μ l
Menschliche Cot-1 (1mg/ml)	3 μ l
Tween-20 3%ig	1 μ l
TMACI (5M)	140 μ l
Labelingansatz	70 μ l

<u>Ansatz MES-Stock-Lösung:</u>	<u>Volumen in μl</u>
MES Hydrat	70,4 g
MES Salz	193,3 g
H ₂ O _{reinst}	ad 1000 ml

Die Lösung wird auf einen pH-Wert von 6,5-6,7 eingestellt und durch einen Filter mit einer Porengöße von 0,2 μ M filtriert.

Bedingungen Markierung:

1.	95 °C	10min
2.	auf Eis	10 sec
3.	48 °C	2 min

Aus dem denaturierten Hybridisierungs-Mastermix werden 200 μ l in einen Affymetrix Mapping 50K Array Hind 240 bzw. Affymetrix Mapping 50K Array Xba 240 injiziert. Die Hybridisierung erfolgte in einem Hybridisierungsofen der Firma Affymetrix (Affymetrix GeneChip Hybridization Ofen 640) bei 48 °C und 60 rpm für 16-18 Stunden.

10.2.7 Waschen, Färben und Scannen

Die Stringenzwaschschritte und das Färben werden automatisiert in der Affymetrix Fluidics Wash Station 450 mit dem Programm „Mapping100Kv1_450“ durchgeführt. Das Prinzip der Antikörperfärbung und des Scannens ist mit dem des Affymetrix 10K Arrays identisch. Die Zusammensetzung der Waschpuffer A und B, sowie des Array Holding Buffers unterscheidet sich nicht zwischen den einzelnen Arraytypen (10K und 100K) und ist unter 10.1.7 beschrieben

Programm „Mapping100Kv1_450“:

- | | |
|-------------------------|---|
| 1. Post Hyb Waschen #1 | 6 Zyklen mit 5x Mischen/Zyklus mit Waschpuffer A bei 25 °C |
| 2. Post Hyb Waschen #2 | 6 Zyklen mit 5x Mischen/Zyklus mit Waschpuffer B bei 45 °C |
| 3. 1. Färbeschritt | Färben des Arrays mit SAPE für 10 min bei 25 °C |
| 4. Post Färben Waschen | 6 Zyklen mit 5x Mischen/Zyklus mit Waschpuffer A bei 25 °C |
| 5. 2. Färbeschritt | Färben des Arrays mit Antikörperlösung für 10 min bei 25 °C |
| 6. 3. Färbeschritt | Färben des Arrays mit SAPE für 10 min bei 25 °C |
| 7. Finaler Waschschrift | 10 Zyklen mit 6x Mischen/Zyklus mit Waschpuffer A bei 30 °C |
| 8. Befüllen des Arrays | Befüllen des Arrays mit Array Holding Buffer (25 °C) |

11 iPLEX Gold Assay

Diese Technik wurde im Institut für Humangenetik an der GSF in der Arbeitsgruppe von Peter Lichtner durchgeführt.

Der iPLEX Gold Assay (Sequenom) ist eine Technik, mit deren Hilfe Sequenzunterschiede auf dem Niveau von einzelnen Nukleotiden festgestellt werden können. Der iPLEX Gold Assay arbeitet auf dem Prinzip des MassARRAY Systems, das die Chemie der Primerextensionreaktion mit der Methode der MALDI-TOF Massenspektrometrie verbindet. Dieser Arbeitsschritt wurde in der Arbeitsgruppe von Peter Lichtner im Institut für Humangenetik durchgeführt.

11.1 iPLEX Gold PCR

Die Durchführung der iPLEX Gold PCR ist weiter oben beschrieben (siehe Material und Methoden 4.6).

11.2 SAP Behandlung

Dieser Schritt ist notwendig, um dNTPs zu dephosphorylieren, die während der PCR-Reaktion nicht eingebaut wurden. Phosphorylierte dNTPs erschweren die Datenanalyse, da sie Kontaminationspeaks bilden können.

<u>Ansatz SAP Behandlung:</u>	<u>Volumen in μl</u>
SAP Puffer (10x)	0,17 μ l
SAP Enzym (1 U/ μ l)	0,30 μ l
iPLEX Gold PCR	5,00 μ l
H ₂ O _{reinst}	ad 7,00 μ l

Bedingungen SAP Behandlung:

1. 37 °C 40 min
2. 85 °C 5 min
3. 4 °C 10 min

11.3 iPLEX Reaktion

<u>Ansatz iPLEX Reaktion:</u>	<u>Volumen in μl</u>
iPLEX Puffer (10x)	0,200 μ l
iPLEX Termination Mix	0,200 μ l
Primer Mix (7 μ M:14 μ M)	0,804 μ l
iPLEX Enzym	0,041 μ l
SAP-Ansatz	7,000 μ l
H ₂ O _{reinst}	ad 9,000 μ l

Bedingungen iPLEX Reaktion:

1. 94 °C 30 sec
2. 94 °C 5 sec
3. 52 °C 5 sec
4. 80 °C 5 sec
5. Wiederholung der Schritte 3-4 in 5 Zyklen
6. Wiederholung der Schritte 2-4 in 39 Zyklen
7. 72 °C 3 min
8. 4°C 10 min

11.4 Reinigung

Zur Reinigung werden 6 mg Harz (Sequenom) dazugegeben, kurz abzentrifugiert und mit 16 μ l H₂O_{reinst} verdünnt. Der Ansatz wird 10-15 min auf dem Schüttler gemischt und anschließend für 10 min bei 3000 U/min zentrifugiert.

11.5 MALDI-TOF MS Analyse

Der gereinigte iPLEX-Ansatz wird in den 384-Element SpectroCHIP Bioarray geladen und im MALDI-TOF Massenspektrometer gemessen und analysiert.

12. Feinkartierung mit Mikrosatelliten

Zwei der Kandidatenregionen, die nach der Kopplungsanalyse der 10K-Arrays identifiziert wurden, wurden mit Mikrosatellitenmarkern bestätigt oder weiter eingegrenzt. Dafür wurden in der entsprechenden Region Mikrosatellitenmarker mit einem hohen Heterozygotenindex gesucht. Die verwendeten Primersequenzen für die ausgewählten Mikrosatellitenmarker stammen von der Internetseite des UCSC Genome Browsers (<http://genome.ucsc.edu/>) (siehe Anhang Tabelle Mikrosatelliten). Der Ansatz für die PCR ist weiter oben beschrieben (siehe Material und Methoden 4.8). Anschließend wird 1 ml der PCR 1:100 verdünnt und mit 12 µl des Längenstandards Formamid-GS500LIZ vermischt. Die so vorbereiteten PCR-Produkte werden für 5 min bei 95 °C denaturiert und auf Eis abgekühlt. Die Analyse der Fragmente erfolgt im Sequenziergerät ABI Prism 3730 DNA Sequenzer (Applied Biosystems) nach den vom Hersteller angegebenen Bedingungen. Die Auswertung der Daten wurde mit dem Programm GeneMapper v3.5 (Applied Biosystems) durchgeführt.

13 Analyse der Daten

13.1 Analyse der 10K-Affymetrix-Daten

Nach dem Scannen werden die Rohdaten mit Hilfe der GCOS 1.4 Software (Affymetrix) verarbeitet und anschließend mit der GTYPE 4.0 Software (Affymetrix) die Genotypen berechnet.

Die Affymetrix Arrays bestehen aus kleinen DNA-Fragmenten, so genannte Proben, die mit Hilfe der Photolithographie auf spezifischen Positionen einer Quarzoberfläche synthetisiert werden. Die Fläche, auf der eine Probe synthetisiert ist, wird als Feature bezeichnet. Diese Proben sind Oligonukleotide mit einer Länge von 25 Nukleotiden, die eine hohe Spezifität aufweisen und die

komplementäre Sequenz in der zu untersuchenden DNA erkennen können. Für jede Probe auf dem Array, die perfekt an die Zielsequenz bindet („Perfect-Match“ Probe, PM), gibt es auch eine „Mismatch“ Probe (MM). Diese „Mismatch“ Probe enthält ein Nukleotid in der Mitte der Sequenz, das nicht zur komplementären Sequenz passt. Die „Mismatch“ Proben dienen zur Quantifizierung von Hintergrundsignalen und als interne Kontrolle für den jeweiligen „Perfect-Match“ Partner, um unspezifische Bindungen zu quantifizieren und von dem eigentlichen Signal subtrahieren zu können. Zur Erhöhung der Reproduzierbarkeit sind auf den Arrays 40 Proben für jeden Genotypen vorhanden, die beide Allele des jeweiligen SNPs abdecken.

Die Berechnung der Genotypen erfolgt mit der GTYPE 4.1 Software (Affymetrix). GTYPE verwendet den „Dynamic Model“ Algorithmus (DM), ein automatisierter, modell-basierter Algorithmus zur Berechnung der Genotypen. Dieser DM-Algorithmus setzt eine Normalisierung der Intensität der Pixel in einem bestimmten Feature voraus und berechnet die logarithmische Wahrscheinlichkeit für jedes Probenquartett (PM Allel A, MM Allel A, PM Allel B, MM Allel B) unabhängig voneinander mit 4 verschiedenen Modellen für *Null* (keine eindeutige Zuordnung zu einem Genotyp möglich oder Ausfall des SNPs), homozygot A (*AA*), homozygot B (*BB*) und heterozygot (*AB*). Somit werden für jedes Probenquartett die Wahrscheinlichkeiten zu einem Ergebnis kombiniert. Anschließend werden die Ergebnisse verschiedener Probenquartette kombiniert. Mit dem Wilcoxon-Rangsummen-Test werden schließlich die *P*-Werte für die 4 Modelle (*Null*, *AA*, *BB*, *AB*) bestimmt. Der DM-Algorithmus entscheidet, welches der 4 Modelle am wahrscheinlichsten ist, ausgehend vom minimalsten *P*-Wert aller 10 Probenquartette für jeden SNP und jeden Array. Die Berechnung der Genotypen mit dem DM-Algorithmus berücksichtigt alle Features eines Arrays und eines SNPs.

Die Genotypdaten werden mit dem Programm Alohomora (Rueschendorf et al. 2005) auf die Kopplungsanalyse vorbereitet. In dem Programm Alohomora werden die Daten auf mendelsche Fehler und Genotypisierungsfehler mit dem Programm PedCheck1.1 überprüft (O'Connell et al. 1998). Unwahrscheinliche Genotypen werden mit Hilfe des Programmes Merlin ausgeschlossen (Abecasis et al. 2002). Die Kopplungsanalyse wird danach mit dem Programm Allegro (Version 1.1d) durchgeführt (Gudbjartsson et al. 2000).

13.2 Analyse der 100K-Affymetrix-Daten

Die Auswertung der 100K-Arrays unterscheidet sich von der Auswertung der 10K-Arrays. Für Linkage-Analysen werden die Genotypen benötigt, während bei der Berechnung der Variation der Kopienzahl die Stärke der Intensitäten der einzelnen Oligonukleotide ausschlaggebend ist. Um entscheiden zu können, in welcher Dosis eine bestimmte Region im Genom der Testperson vorliegt, muss ein Vergleich mit Referenzproben durchgeführt werden. Aufgrund von Qualitätsunterschieden der DNAs und experimentellen Schwankungen variieren die Signalintensitäten zwischen den Arrays.

Die Kopienzahl wird bestimmt, indem zuerst Gruppen berechnet werden, in denen Arrays mit ähnlichen Intensitätsprofilen zu so genannten Clustern zusammengefasst werden. Mit der hierarchischen Clusteranalyse werden die untersuchten Arrays so gruppiert, dass die Unterschiede zwischen den Signalintensitäten von Arrays einer Gruppe möglichst gering und die Unterschiede zwischen den Clustern möglichst groß sind. Der binäre Logarithmus aus der Summe der median normalisierten Intensitätswerte beider Allele wird verwendet, um die euklidische Distanzmatrix zu berechnen, die wiederum die Grundlage des hierarchischen Clusters bildet. Für jede Gruppe werden anschließend die Kopienzahlen wie folgt berechnet. Zuerst werden die Rohdaten der Intensitätswerte an jedem SNP-Locus bestimmt, indem der Mittelwert der „Perfect-Match Proben“ für das A- und das B-Allel separat berechnet werden. Die Rohdaten der Intensitätswerte werden median normalisiert. Die Kopienzahl oder das \log_2 -Verhältnis der Intensitätswerte wird für die drei Genotypen (AA, AB, BB) und die „no-calls“ getrennt berechnet. Dafür werden die normalisierten Intensitätswerte des Testarrays durch die medianen Werte aller Arrays dividiert, die den gleichen Genotyp an dem jeweiligen SNP-Locus besitzen. Diese Berechnung der Kopienzahl wird verwendet, da das Hintergrundrauschen der Arrays durch die Verwendung genotyp-spezifischer Intensitätswerte niedriger ist, als bei der Verwendung der Summe der Intensitäten beider Allele für alle Genotypen. Um die Intensitätswerte für X-Chromosomen von männlichen und weiblichen Proben vergleichen zu können, wird die durchschnittliche Dosis der männlichen SNPs auf dem X-Chromosom auf die durchschnittliche Dosis der autosomalen SNPs angeglichen.

In vielen Dosisplots kann beobachtet werden, dass das durchschnittliche Dosislevel abhängig ist, von der Länge der Restriktionsfragmente. Diese Abhängigkeit wird mit Hilfe der quadratischen Regression (Nannya et al. 2005) korrigiert.

Die Qualität der Daten wird durch zwei Methoden bestimmt. Zum einen wird die Standardabweichung (SD) und die Median-Deviation (MAD) der \log_2 Intensitätswerte berechnet. Zum anderen wird das Signal-Rausch-Verhältnis (SNR) für die männlichen Proben berechnet, indem das mediane Verhältnis der \log_2 Intensitätswerte der X-chromosomalen SNPs von dem medianen Verhältnis der \log_2 Intensitätswerte der autosomalen SNPs subtrahiert wird und diese Differenz durch die Hälfte der Summe der medianen Absolutabweichung der \log_2 Intensitätswerte der autosomalen und X-chromosomalen SNPs dividiert wird.

$$\text{SNR} = \frac{\text{median}(\log_2(\text{autosomale SNPs})) - \text{median}(\log_2(\text{X-chromosomale SNPs}))}{0,5 * (\text{mad}(\log_2(\text{autosomale SNPs})) + \text{mad}(\log_2(\text{X-chromosomale SNPs})))}$$

Der Nenner in dieser Formel misst den Unterschied zwischen den \log_2 Intensitätswerten der autosomalen und X-chromosomalen SNPs in männlichen Proben. Dieser Unterschied ist aufgrund technischer Nichtlinearitäten weniger als 1. In 95 Hind III-Arrays betrug die durchschnittliche Differenz 0,63, in 98 Xba I-Arrays 0,70. Der Zähler der Formel schätzt die Größe der Variation ab, indem die autosomalen und X-chromosomalen SNPs gleichwertig behandelt werden. Alle Berechnungen werden robust durchgeführt, so dass sie gegen Abweichungen von dem Standard-Normalmodell und Ausreißern geschützt sind.

Die Mindestanzahl an aufeinander folgenden SNPs, die schwach exprimiert sind und signifikant eine Deletion anzeigen, wird berechnet mit folgendem Ausdruck:

$$\text{Min} \{n: 67 * 10^5 * P(\log_2 CN < -0.5 * \text{sep}_{1\text{copy}, 2\text{copies}})^n < 0.05\} = \text{min} \{n: \phi(-0.5 * \text{SNR})^n < 7.5 * 10^{-8}\}$$

Es steht $\text{sep}_{1\text{copy}, 2\text{copies}}$ für den medianen \log_2 Intensitätswert von SNPs mit 1 oder 2 Kopienanzahl, P steht für die Wahrscheinlichkeit einer gemessenen Kopienzahl (CN) eines SNPs mit einer realen Kopienzahl von 2 unter einem normalen Modell, ϕ ist die Verteilungsfunktion einer normalen Standardverteilung und der Faktor $67 * 10^5$ steht für die Bonferroni-Korrektur für multiples Testen.

Es wurden außerdem Methoden entwickelt, um Regionen auszuwählen, die eventuelle Deletionen, Duplikationen und den Verlust von parentalen Allelen aufweisen. Um Kopienzahlvariationen zu detektieren, wird die array-spezifische SD mit einkalkuliert. Je höher die SD ist, umso höher muss die Anzahl aufeinander folgender SNPs sein, um eine Kopienzahlvariation anzuzeigen. Als unteren Grenzwert wurde der Median von 5 aufeinander folgenden SNPs gewählt, um eine Kopienzahlvariation zu definieren. Verdächtige Regionen wurden mit bekannten Kopienzahlvariationen verglichen, die in Datenbanken, wie der Database of Genomic Variants (<http://projects.tcag.ca/variation>) und der DECIPHER Datenbank (<http://www.sanger.ac.uk/PostGenomics/decipher/>) verzeichnet sind.

C Ergebnisse

1 Kopplungsanalysen mit SNP-Oligonukleotid-Mikroarrays

1.1 Kopplungsanalyse in 5 Familien mit HHRH

1.1.1 Identifizierung einer Kandidatenregion in den Familien mit HHRH

Die erbliche hypophosphatämische Rachitis mit Hyperkalziurie (HHRH) ist eine autosomal-rezessive Form der Hypophosphatämie, die charakterisiert ist durch eine erniedrigte renale Phosphatreabsorption, Hypophosphatämie und Rachitis (Rowe 1994).

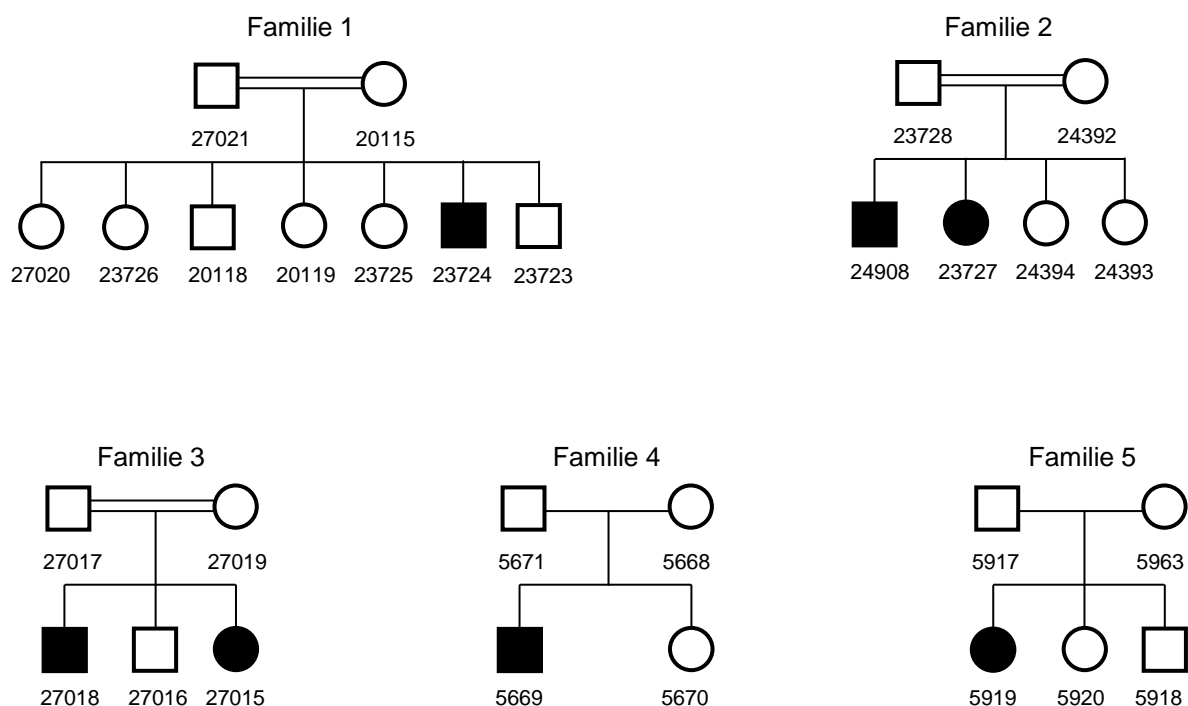


Abb. 4 Familienstammbäume der untersuchten Familien mit HHRH. Die Familien 1 bis 3 sind arabischer Abstammung, die Familien 4 und 5 kommen aus Deutschland. Betroffene Individuen sind mit schwarzen Symbolen gekennzeichnet, nicht betroffene Individuen mit weißen. DNA für die Untersuchung war von allen nummerierten Individuen vorhanden.

In dieser Arbeit wurden 3 arabische und 2 deutsche Familien mit insgesamt 7 Betroffenen untersucht (Abb 4). Die Eltern der 3 arabischen Familien sind konsanguin (Cousin und Cousine 1. Grades). Zuvor wurden bereits Mutationen in den Genen *PHEX* und *FGF23* ausgeschlossen.

Um eine Kandidatenregion für HHRH zu identifizieren, wurde eine Kopplungsanalyse mit den Affymetrix 10K 2.0 Arrays von Affymetrix durchgeführt. Die Genotypdaten, die aus den 10K-Arrays gewonnen wurden, wurden mit dem Programm Alohomora auf die Kopplungsanalyse vorbereitet (Rueschendorf et al. 2005). Alohomora benützt das Programm PedCheck1.1, um mendelsche Fehler oder Genotypisierungsfehler aufzudecken (O'Connell et al. 1998). Unwahrscheinliche Genotypen wurden durch das Programm Merlin ausgeschlossen (Abecasis et al. 2002). Im Anschluss wurde eine Multipoint-Kopplungsanalyse mit dem Programm Allegro (Version 1.1d) durchgeführt (Gudbjartsson et al. 2000). Es wurde ein autosomal-rezessives Vererbungsmodell angenommen. Die Frequenz des Krankheitsallels wurde auf 0.001 gesetzt, die Penetranz auf 99%. Nach der Auswertung konnte keine Kandidatenregion identifiziert werden.

Die Marker auf dem Array sind nicht gleichmäßig über das Genom verteilt, so dass bestimmte Teile in einer Kopplungsanalyse nicht analysiert werden. Im Institut stand eine zweite Plattform mit SNP-Arrays zur Verfügung. Da sich die Verteilung der SNPs auf dem Linkage IV Panel von Illumina und den 10K-Arrays von Affymetrix unterscheidet, wurde die Kopplung mit den Familien 1 und 2 mit dem Linkage IV Panel wiederholt. Die Durchführung des Experiments wurde in der Arbeitsgruppe von Peter Lichtner durchgeführt. Das Linkage IV Panel enthält ungefähr 5800 SNPs.

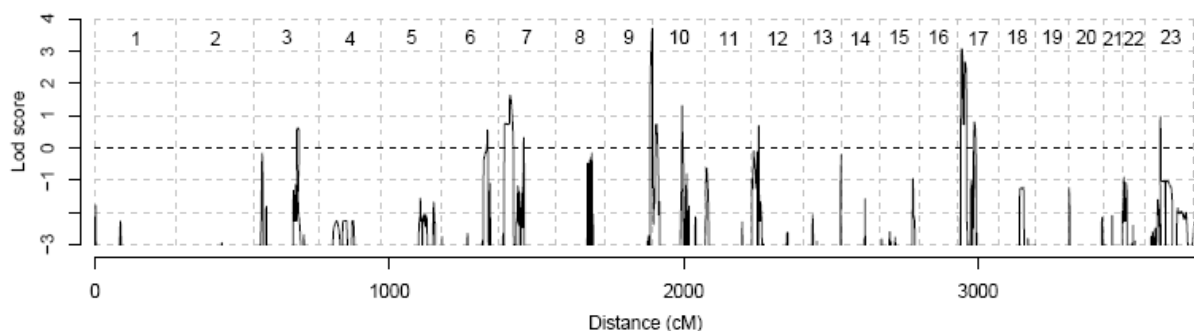


Abb. 5 Genomweite Kopplungsanalyse der Familien 1 und 2. Die Abbildung zeigt das Ergebnis der Kopplungsanalyse der Familien 1 und 2 mit dem Linkage IV Panel von Illumina. Die Analyse der Daten wurde mit dem Programm Allegro durchgeführt ($q = 0.001$; $f_1 = 0.0$; $f_2 = 0.0$; $f_3 = 0.99$). Der höchste LOD Score wurde auf Chromosom 9 mit einem Wert von 3.7 erreicht.

Die Auswertung der Daten wurde identisch zu den 10K-Arrays von Affymetrix durchgeführt. Die Analyse der Daten erbrachte eine Kandidatenregion am Ende des langen Armes von Chromosom 9 (9q34.3) mit einem signifikanten LOD Score von 3.7 (Abb 5).

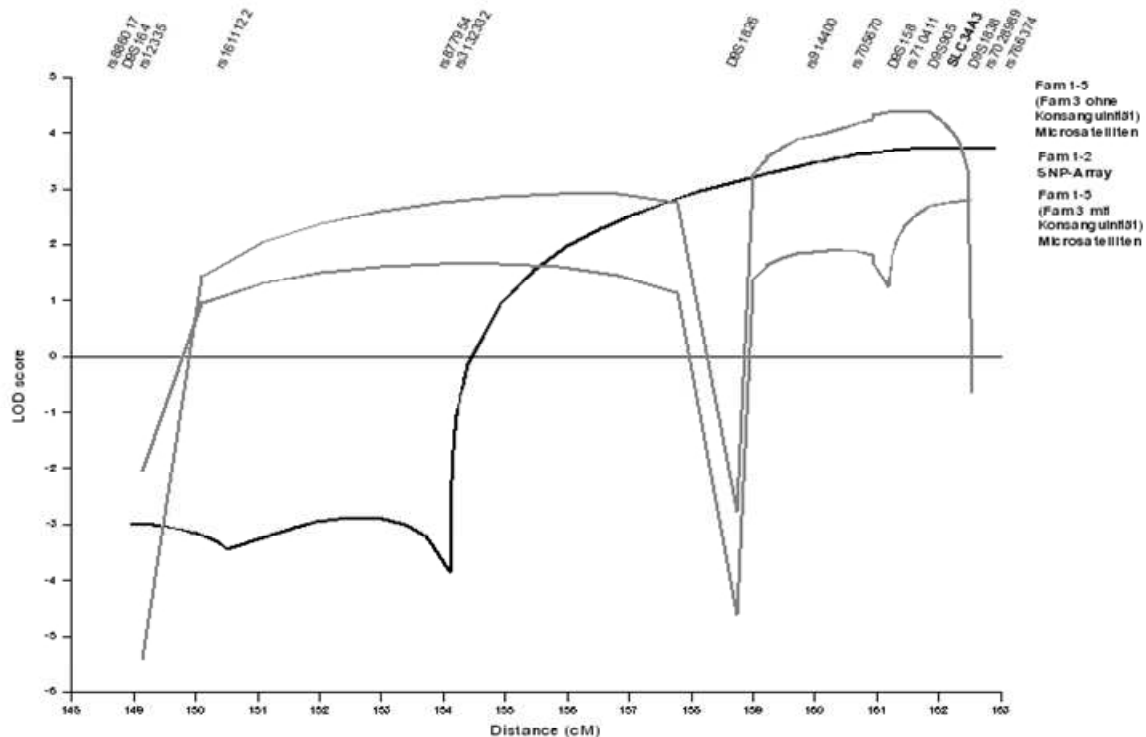


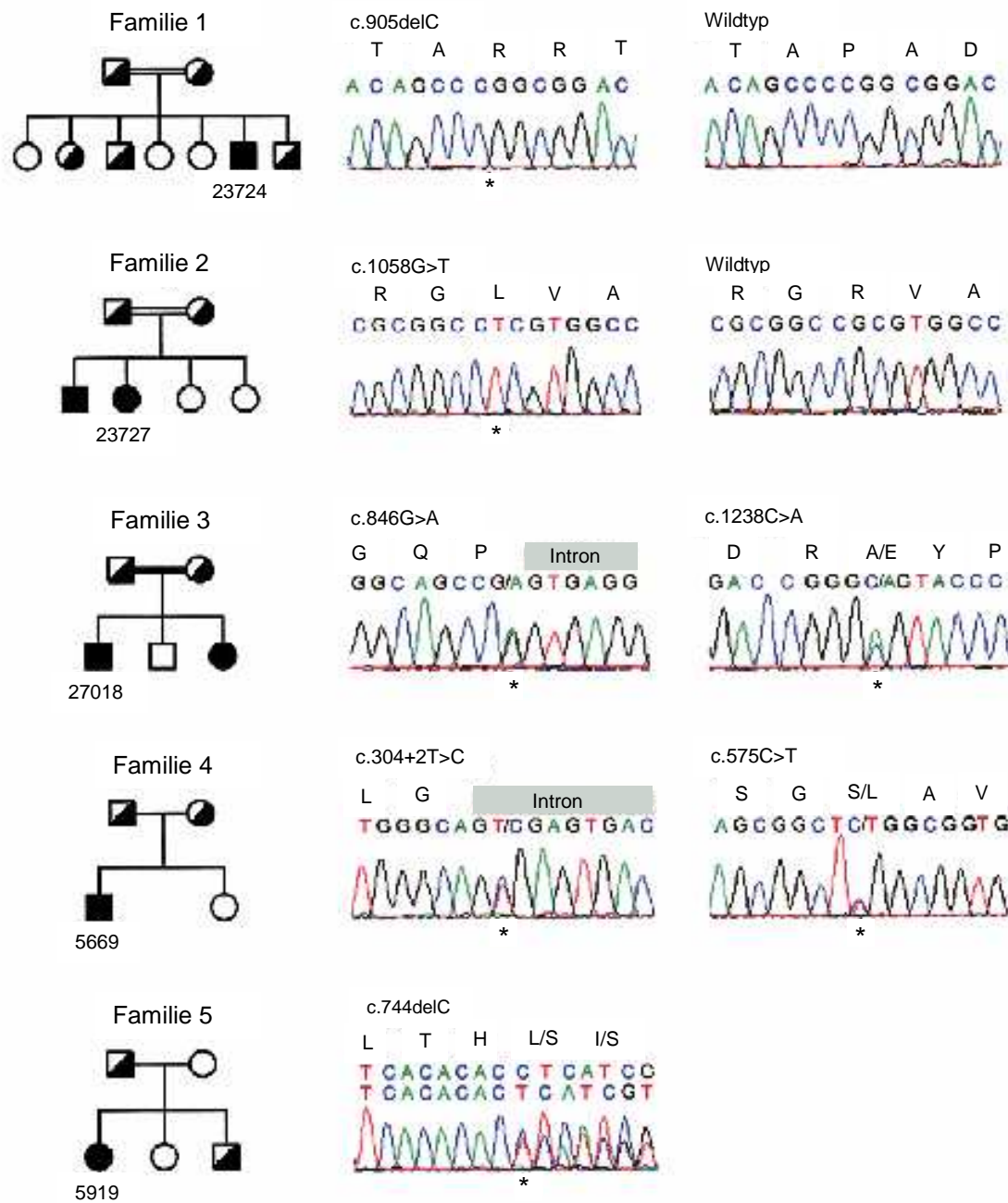
Abb. 6 Multipoint-Kopplungsanalyse am langen Arm des Chromosoms 9. Die Abbildung zeigt das Ergebnis der Kopplungsanalyse im Bereich von 15 cM des Telomers von Chromosom 9. Eine Analyse der SNP-Marker wurde in den Familien (Fam) 1 und 2 durchgeführt (schwarze Linie). Eine Analyse der Mikrosatellitenmarker wurde in allen 5 Familien zweimal durchgeführt. Einmal mit Annahme der Konsanguinität in der Familie 3 (untere graue Linie) und einmal ohne Annahme der Konsanguinität in Familie 3 (obere graue Linie). (Abbildung nach Lorenz-Depiereux et al. 2006)

Der maximale Multipoint-LOD Score lag distal des Markers *rs3132332* auf Chromosom 9q34.3 (Abb.6). Zur Unterstützung der Kopplungsanalyse mit dem SNP-Array wurde in der Region eine zusätzliche Kopplungsanalyse in allen 5 Familien mit Mikrosatellitenmarkern durchgeführt (Tabelle Mikrosatellitenmarker, siehe Anhang S. IX). Die Kandidatenregion, die anhand der Analyse der Mikrosatelliten festgelegt wurde, lag mit einem maximalen Multipoint-LOD Score von 2.8 distal des Markers *D9S158* (Abb 6). Den größten Beitrag zum LOD Score wurde von den Familien 1 und 2 geleistet (Abb 6). Die Familien 4 und 5 waren zu klein, um zum LOD Score beitragen zu können. Sie zeigten aber keinen negativen LOD Score und damit keinen Ausschluss dieser Region. Die Familie 3

zeigte einen LOD Score von -1 in dieser Region. Wie sich im Laufe des Projektes zeigte, trugen die betroffenen Individuen der Familie 3 trotz der Konsanguinität der Eltern compound-heterozygote Mutationen. Die Analyse aller Familien ohne der Annahme einer Konsanguinität der Eltern in Familie 3 brachte einen LOD Score von 4.4 distal des Markers *D9S1826* (Abb 6).

1.1.2 Mutationsanalyse in der Kandidatenregion 9q34.3 in den Familien mit HHRH

In der Kopplungsanalyse wurde eine Kandidatenregion von ungefähr 2.7 Mb auf dem Chromosom 9 zwischen dem Marker *D9S1826* und dem Ende des Chromosoms festgelegt. In der Region liegen 85 Gene. Darunter war ein offensichtliches Kandidatengen, *SLC34A3*, das für den Natrium-Phosphat-Kotransporter Typ IIc kodiert. *SLC34A3* besteht aus 13 Exons, von denen die Exons 2 bis 13 kodierend sind. Für die Mutationsanalyse wurden alle Exons von *SLC34A3* und die flankierenden Regionen der Introns sequenziert. In den betroffenen Personen wurden 1 heterozygote, 2 compound-heterozygote und 2 homozygote Mutationen gefunden (Abb 7). Die Überträger waren heterozygot für diese Mutationen. Drei der Mutationen resultieren voraussichtlich in Transkripte, die zu vorzeitig abgebrochenen Translationsprodukten führen. Der Indexpatient 23724 in Familie 1 hat eine homozygote 1 bp-Deletion in Exon 9 (c.905delC). Der Patient 5669 der Familie 4 hat eine Mutation an der Donor-Spleißstelle des Introns 4 (c.304+2T>C). Die dritte Mutation, die zu einem inkompletten Transkript führt, ist eine 1 bp-Deletion im Exon 7 (c.744delC) des Patienten 5919 der Familie 5. Die restlichen 2 Mutationen sind eine Missense-Mutation im Exon 10 (c.1058G>T) des Patienten 23727 der Familie 2 und eine stille Mutation an der letzten Position des Exons 8 (c.846G>A) im Patienten 27018 der Familie 3, die zu einem irregulärem Spleißprodukt führen könnte.

**Abb. 7**

Mutationsanalyse in dem Gen *SLC34A3* in den 5 Familien mit HHRH.

In der Abbildung sind auf der linken Seite die Stammbäume der Familien und auf der rechten Seite die Elektropherogramme des jeweiligen betroffenen Individuums dargestellt. Homozygote Mutationen wurden in den Familien 1 und 2 gefunden, compound-heterozygote Mutationen in den Familien 3 und 4 und in der Familie 5 wurde nur eine heterozygote Mutation entdeckt. Betroffene Individuen sind durch schwarze Symbole dargestellt, nicht betroffene Individuen mit heterozygoten Mutationen durch schwarz-weiße Symbole.

(Abbildung nach Lorenz-Depiereux et al. 2006)

1.2 Kopplungsanalyse in 3 Familien mit ARHP

1.2.1 Identifizierung einer Kandidatenregion in den Familien mit ARHP

In drei Familien, die aus der Türkei (Familie 1), Spanien (Familie 2) und aus dem Libanon (Familie 3) stammen, wurde eine Kopplungsuntersuchung durchgeführt. Die klinischen Symptome und die biochemischen und histomorphologischen Parameter der Betroffenen gleichen den Merkmalen, die in XLH oder ADHR vorkommen. Die Analyse der Stammbäume ließ einen autosomal-rezessiven Erbgang vermuten. Weiterhin waren Mutationsanalysen in den Genen *PHEX*, *FGF23* und *SLC34A3* ohne Befund. Die betroffenen Mitglieder der Familie 3 stammen aus einer konsanguinen Beziehung (Cousin und Cousine 1. Grades), in den Familien 1 und 2 war keine Konsanguinität bekannt (Abb 8).

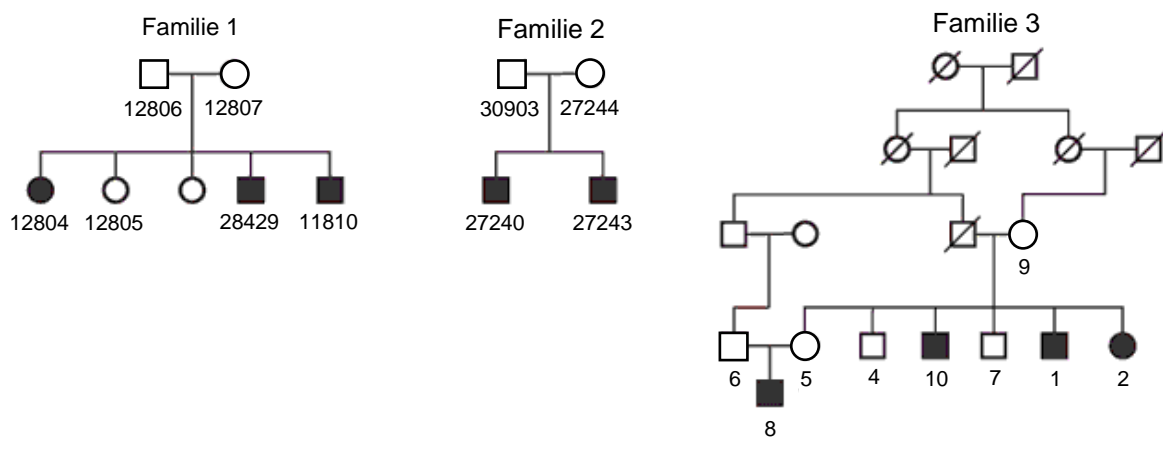
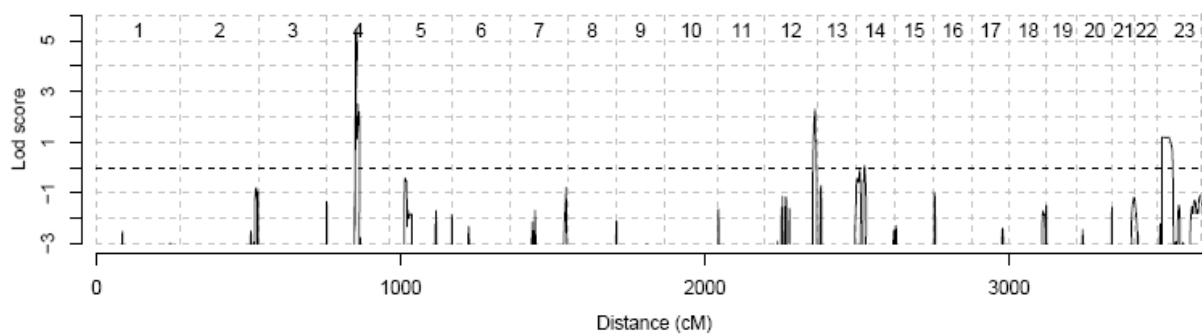


Abb. 8 Familienstammbäume der untersuchten Familien mit ARHP. Die Familien stammen aus der Türkei (Familie 1), Spanien (Familie 2) und aus dem Libanon (Familie 3). Betroffene Individuen sind mit schwarzen Symbolen gekennzeichnet, nicht betroffene Individuen mit weißen. DNA für die Untersuchung war von allen nummerierten Individuen vorhanden.

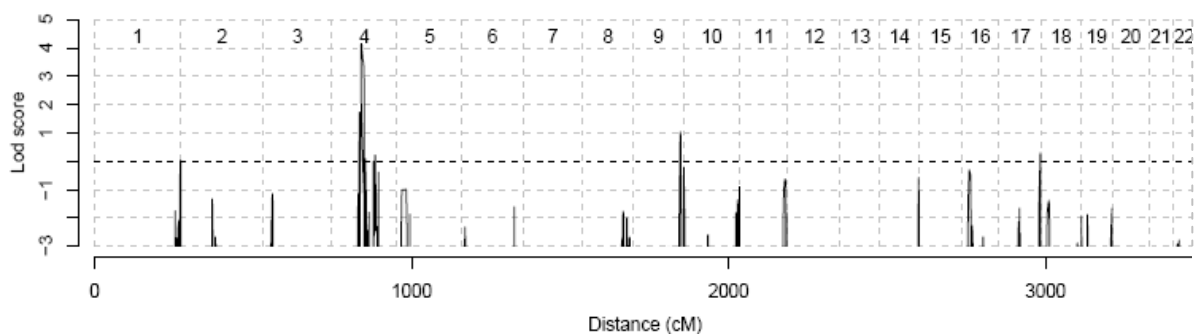
Es wurde eine genomweite Kopplungsanalyse mit den Affymetrix 10K-Arrays durchgeführt. Die Analyse der Daten erfolgte unter der Annahme, dass das betroffenen Allel *identical by descent* in allen drei Familien sein könnte. Die 4.6 Mb große Kandidatenregion auf Chromosom 4q21 wurde durch Homozygotie-Kartierung identifiziert. Die parametrische Berechnung des LOD Scores erfolgte unter der konservativen Annahme einer Beziehung zwischen Cousin und Cousine 2. Grades für die Familien 1 und 2. Für die Berechnungen wurde ein autosomal-rezessives Vererbungsmodell angenommen. Die Frequenz des betroffenen Allels wurde auf 0.001 gesetzt, die Penetranz auf 99%.

Die maximal erreichten LOD Score-Werte waren 3.1 für die Familie 1 und 2.4 für die Familie 2. Die Addition der beiden LOD Score-Werte der Familien 1 und 2 ergab einen gemeinsamen LOD Score von 5.5 (Abb 9a). Die Berechnung des LOD Scores in der Familie 3 erfolgte unter der Berücksichtigung der bekannten Konsanguinität und erreichte einen maximalen LOD Score von 4.2 (Abb 9b).

a)



b)

**Abb. 9**

Genomweite Kopplungsanalyse der Familien 1 bis 3.

Die Abbildung zeigt die Ergebnisse der Kopplungsanalyse der Familien 1 und 2 zusammen mit einem LOD Score von 5.5 (Abbildung 9a) und der Familie 3 mit einem LOD Score von 4.2 (Abbildung 9b). Die Analyse der Daten wurde mit dem Programm Allegro durchgeführt ($q = 0.001$; $f_1 = 0.0$; $f_2 = 0.0$; $f_3 = 0.99$).

Der maximale LOD Score in den 3 Familien ergab eine 4.6 Mb große Kandidatenregion auf Chromosom 4q21 zwischen den SNP-Markern *rs340204* und *rs722937*.

1.2.2 Mutationsanalyse in der Kandidatenregion 4q21 in den Familien mit ARHP

Durch die Kopplungsanalyse wurde eine Kandidatenregion mit 4.6 Mb auf dem Chromosom 4q21 identifiziert. In dieser Region liegen etwa 50 Gene, darunter eine Gengruppe, die für eine Klasse von nicht-kollagenen Dentin- und Knochenmatrix-Proteine kodiert, die auch als SIBLING-Proteine (*Small Integrin-Binding Ligand N-linked Glycoproteins*) bezeichnet werden. Zu dieser Genfamilie gehören die Gene *dentin sialophosphoprotein (DSPP)*, *dentin matrix protein (DMP1)*, *integrin-binding sialoprotein (IBSP)*, *matrix extracellular phosphoglycoprotein (MEPE)* und *osteopontin (secreted phosphoprotein 1, SPP1)*. Es wird vermutet, dass diese polyanionischen Proteine eine Schlüsselrolle in der Mineralisierung von Knochen und Dentin innehaben (Qin et al. 2004, Fisher et al. 2001, Lorenz-Depiereux et al. 2006). Aus diesem Grund wurden diese Gene als Kandidatengene für ARHP ausgewählt und auf Mutationen in den Exons und den flankierenden Intronsequenzen überprüft.

Es wurden in allen Betroffenen der 3 Familien verschiedene homozygote Mutationen in *DMP1* gefunden, die vermutlich zu einem Funktionsverlust führen. Die Eltern der betroffenen Individuen sind heterozygot für die jeweilige Mutation (Abb 10). Die Geschwister mit ARHP in der Familie 1 haben alle eine 1 bp-Deletion im Exon 6 (c.362delC), die zu einem vorzeitigen Stoppkodon nach 120 Aminosäuren führt. In der Familie 2 tragen die 2 betroffenen Brüder eine homozygote Mutation in der kanonischen Spleiß-Akzeptorsequenz des Introns 2 (c.55-1G>C). Die betroffenen Individuen in Familie 3 sind Träger einer Missense-Mutation im Exon 2, die das Startkodon von ATG nach GTG verändert (c.1A>G, p.M1V).

Es wurden keine Auffälligkeiten in den Genen *MEPE* und *SPP1* identifiziert und lediglich bekannte Sequenzvariationen in den Genen *DSPP* und *IBSP*.

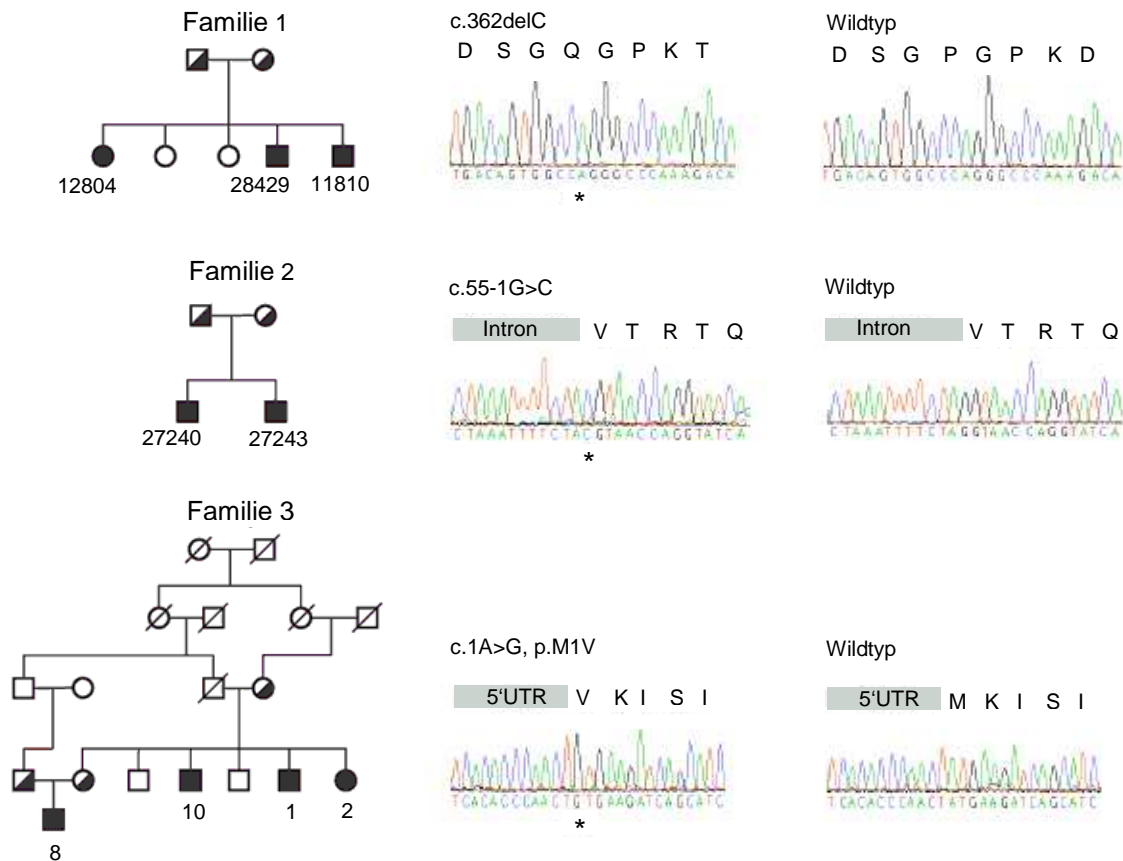


Abb. 10 Mutationsanalyse in dem Gen *DMP1* in den 3 Familien mit ARHP. In der Abbildung sind auf der linken Seite die Stammbäume der Familien und auf der rechten Seite die Elektropherogramme der betroffenen Individuen dargestellt. Homozygote *loss-of-function*-Mutationen wurden in allen betroffenen Individuen gefunden. Die Eltern der Betroffenen sind heterozygot für die jeweilige Mutation. Betroffene Individuen sind durch schwarze Symbole dargestellt, nicht betroffene Individuen mit heterozygoten Mutationen durch schwarz-weiße Symbole. (Abbildung nach Lorenz-Depiereux et al. 2006)

1.3 Kopplungsanalyse in einer Familie mit SOS

1.3.1 Identifizierung einer Kandidatenregion in der Familie mit SOS

In einer Familie aus dem Irak wurde anhand der klinischen Daten ein spondylo-oculares Syndrom diagnostiziert. In der Familie gibt es 7 Kinder, von denen 6 betroffen sind. Von der Tochter II:2, die laut Anamnese ebenfalls betroffen ist, stand kein genetisches Material zur Verfügung. (Abb 11).

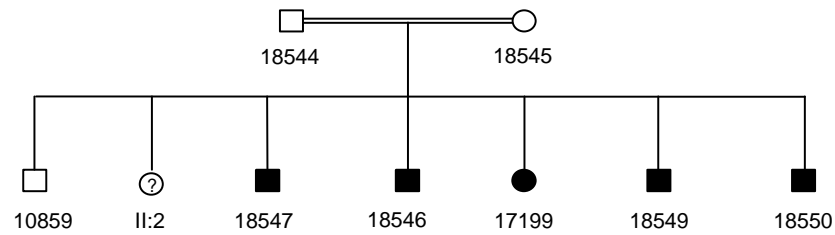


Abb. 11 Familienstammbaum der untersuchten Familie mit SOS. Die Eltern sind Cousin und Cousine 1. Grades. Betroffene Individuen sind mit schwarzen Symbolen gekennzeichnet, nicht betroffene Individuen mit weißen. DNA für die Untersuchung war von allen Individuen, mit Ausnahme von Individuum II:2, vorhanden.

Die Eltern sind Cousin und Cousine 1. Grades und phänotypisch unauffällig. Untersuchungen der Betroffenen in der Vergangenheit erbrachten normale Karyogramme und konnten Syndrome mit bekannten Gendefekten, z.B. das Osteoporose-Pseudoglioma-Syndrom, das Kniest- und das Stickler-Syndrom ausschließen (Schmidt et al. 2001, Rudolph et al. 2003).

Um die Ursache der Erkrankung identifizieren zu können, wurde mit der Familie eine Kopplungsanalyse mit den 10K-Arrays von Affymetrix durchgeführt. Die Genotypdaten, wurden mit dem Programm Alohomora auf die Kopplungsanalyse vorbereitet (Rueschendorf et al. 2005). Es wurde ein autosomal-rezessives Vererbungsmodell angenommen. Die Frequenz des geschädigten Allels wurde auf 0.001 gesetzt, die Penetranz auf 99%. Die Analyse der Daten erbrachte eine einzelne Kandidatenregion mit über 11 Mb auf dem langen Arm des Chromosoms 17 (17q21.2-q22) mit einem maximalen LOD Score von 3.7 bei Marker *rs1320283* (Abb 12).

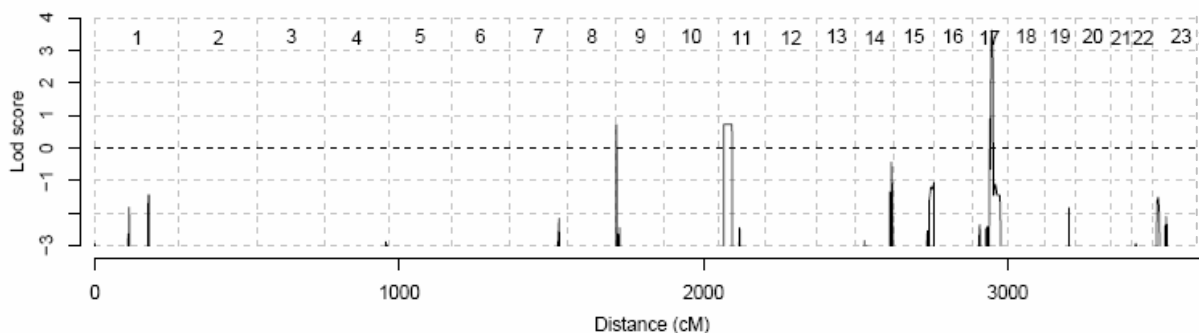


Abb. 12 Genomweite Kopplungsanalyse der Familie mit SOS. In der Abbildung ist das Ergebnis der Kopplungsanalyse dargestellt. Die Analyse der Daten wurde mit dem Programm Allegro durchgeführt ($q = 0.001$; $f_1 = 0.0$; $f_2 = 0.0$; $f_3 = 0.99$). Es wurden alle Individuen mit Ausnahme von II:2 untersucht.

Die Haplotypen der untersuchten Personen wurden mit dem Programm Allegro berechnet und es zeigt sich eine Rekombination zwischen den den SNP-Markern *rs721411* und *rs955712* (Abb 13). In der gesamten Region liegen nur 7 SNPs, mit einem Abstand von ungefähr 3 Mb zwischen den ersten beiden SNPs, *rs721411* und *rs2017257* und einem Abstand von ungefähr 2.5 Mb zwischen den letzten beiden SNPs *rs1406012* und *rs955712* (Abb 14). Da der Abstand zwischen den SNPs, in dem die Rekombination vermutet wird, sehr groß ist, wurde versucht, mit Hilfe von Mikrosatellitenmarkern die Region einzugrenzen.

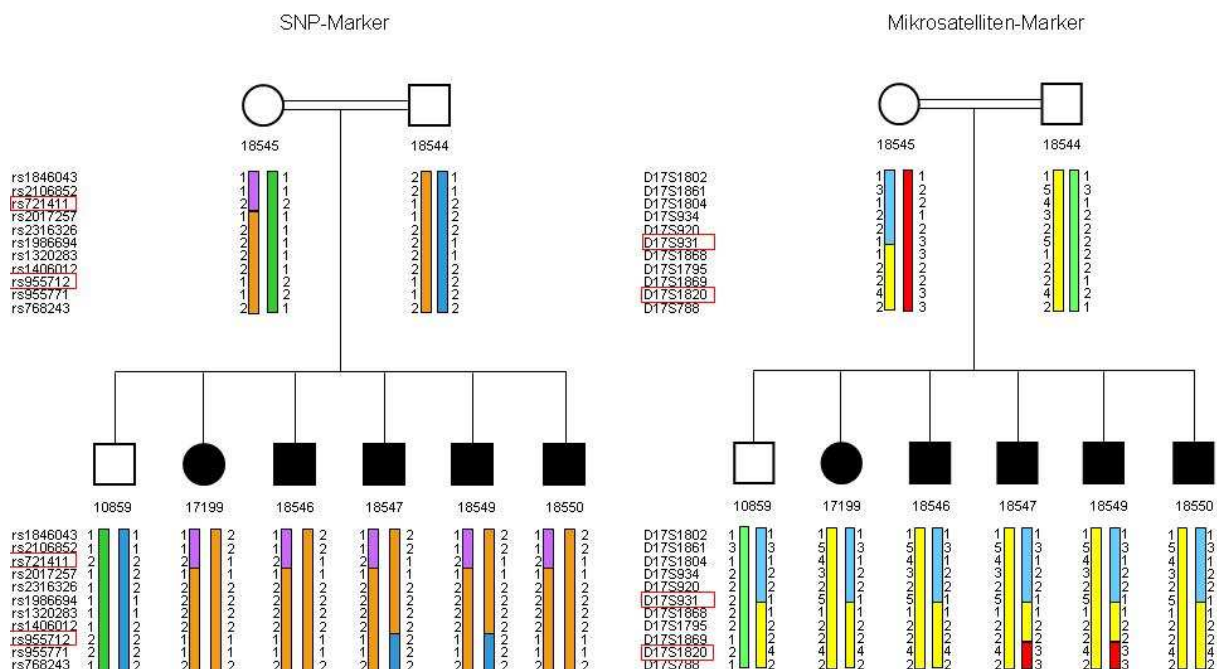


Abb. 13 Darstellung der Haplotypen nach der Analyse der SNP- und Mikrosatellitenmarker. In der Abbildung ist links der Stammbaum mit den Haplotypen nach Berechnung der SNP-Marker dargestellt, rechts der Stammbaum mit den Haplotypen mit den Mikrosatellitenmarkern. Die Marker, die die äußere Grenze der rekombinanten Region darstellen, sind rot umrandet.

Zur Feinkartierung wurde mit 11 Mikrosatellitenmarkern genotypisiert (Tabelle Mikrosatellitenmarker, siehe Anhang S. IX). Dadurch konnte die Region auf ungefähr 5.3 Mb zwischen den Markern D17S931 und D17S788 auf Chromosom 17q21.32-q22 eingengt werden. Der Marker D17S1868

ergab einen maximalen LOD Score von 3.27. Die Haplotypen der Familie nach der Analyse der Mikrosatellitenmarker ist in Abb 13 dargestellt.

1.3.2 Mutationsanalyse in der Kandidatenregion 17 in der Familie mit SOS

Die in der Kopplungsanalyse mit den SNP-Markern identifizierte Kandidatenregion enthält über 150 Gene. Für die Mutationsanalyse wurden 26 Gene aus der 11 Mb-Kandidatenregion ausgewählt, die in Knorpel, Knochen, Knochenmark, Niere, Linse oder Retina exprimiert sind, an der Embryonalentwicklung beteiligt sind oder entsprechende Mausmodelle haben (Abb 14).

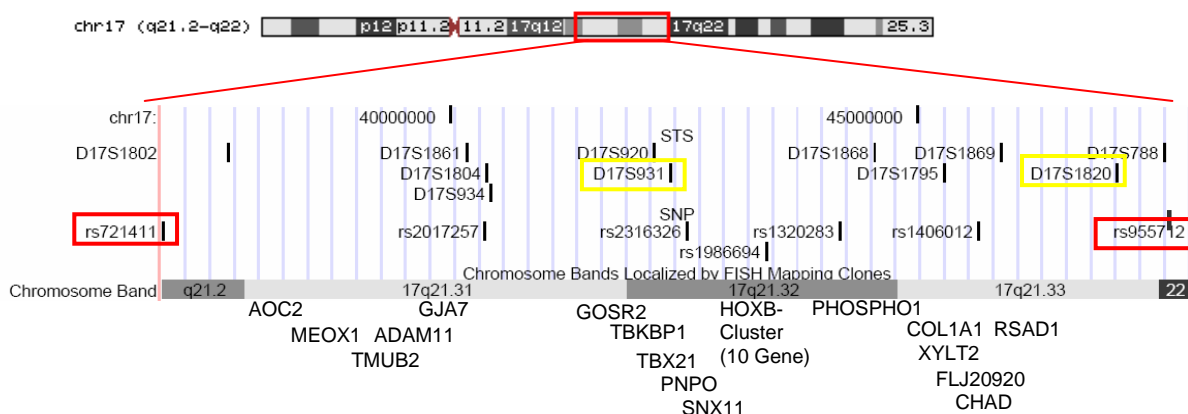


Abb. 14 Auswahl der Kandidatengene in der Region 17q21.32-q22.

In der Abbildung ist oben das gesamte Chromosom 17 dargestellt. Die Kandidatenregion ist rot umrandet. Darunter ist ein vergrößerter Ausschnitt der Kandidatenregion zu sehen mit der ungefähren Lage der zur Sequenzierung ausgewählten Gene und der Lage der Mikrosatelliten (STS) und SNPs, die bei der Kopplung verwendet wurden. Rot umrandete Marker bilden jeweils die äußere Grenze der Kandidatenregion, gelb umrandete Marker bilden die innere Grenze der Kandidatenregion.

In allen 26 untersuchten Genen wurden außer bekannten Sequenzvariationen keine Mutationen gefunden.

2 CNVs

2.1 Datenanalyse der 100K-Arrays

Es wurde genomische DNA von 67 Patienten mit Entwicklungsverzögerung untersucht, die in der Routinediagnostik keine Auffälligkeiten bei der Chromosomenuntersuchung mit GTG-Bänderung zeigten. Zusätzlich wurden noch 5 Patienten untersucht, bei denen zuvor eine balancierte *de novo* Translokation festgestellt wurde. Die DNA der Patienten wurde mit den 100K-Arrays von Affymetrix untersucht, die aus je einem *HindIII*-Array mit 57244 SNPs und einem *XbaI*-Array mit 58960 SNPs bestehen. Die 100K-Arrays bedienen sich einer Ein-Farben-Methode, bei der die Proben auf den Arrays alle mit dem gleichen Farbstoff gefärbt werden und die Auswertung der Daten über die Lokalisation der Proben und die Intensität der Färbung erfolgt. Die normalisierten Intensitätswerte einer zu untersuchenden Testperson müssen deshalb immer mit denen von Referenzarrays verglichen werden. Insgesamt wurden 169 100K-Arrays analysiert. Die Qualität der Daten wurde mit Hilfe der Standardabweichung (SD) und der mittleren absoluten Abweichung (MAD) der \log_2 Dosiswerte aller autosomaler Proben berechnet. Die Werte sind in der Tabelle 2 zusammengefasst und stimmen mit Ergebnissen überein, die von anderen Arbeitsgruppen mit vergleichbaren Techniken erhalten wurden (Redon et al. 2006; Nannya et al 2005).

Tabelle 2 Zusammenfassung der Standardabweichung und der mittleren absoluten Abweichung der \log_2 Intensitätswerte

Array	Standardabweichung	Mittlere absolute Abweichung
HindIII	0.19	0.16
XbaI	0.18	0.16

Bei der Analyse der ersten Arrays zeigte sich, dass der Vergleich mit externen Referenzwerten, die von Affymetrix zur Verfügung gestellt wurden, eine größere Standardabweichung zur Folge hatte. Unterschiede in der Qualität der DNA-Proben und experimentelle Schwankungen sind wahrscheinlich der Grund für die Unterschiede in den Intensitätswerten, da Arrays, die in einem Experiment hybridisiert wurden, ähnliche Intensitätsprofile aufweisen. Um eine optimale Standardabweichung zu erhalten, wurden die genotyp-spezifischen \log_2 Intensitätswerte hierarchisch geclustert und in Gruppen mit ähnlichen Intensitätsprofilen eingeteilt, die dann getrennt analysiert wurden. Die

Standardabweichung wurde dadurch deutlich verringert. Die *HindIII*-Arrays zeigten 4 verschiedene Gruppen, die *XbaI*-Arrays 3 Gruppen, mit deutlich unterschiedlichen Intensitätsprofilen (Abb 15).

Die Abhängigkeit der \log_2 Intensitätswerte von der Fragmentlänge der amplifizierten Produkte wurde durch quadratische Regression korrigiert, um die Standardabweichung weiter zu verringern.

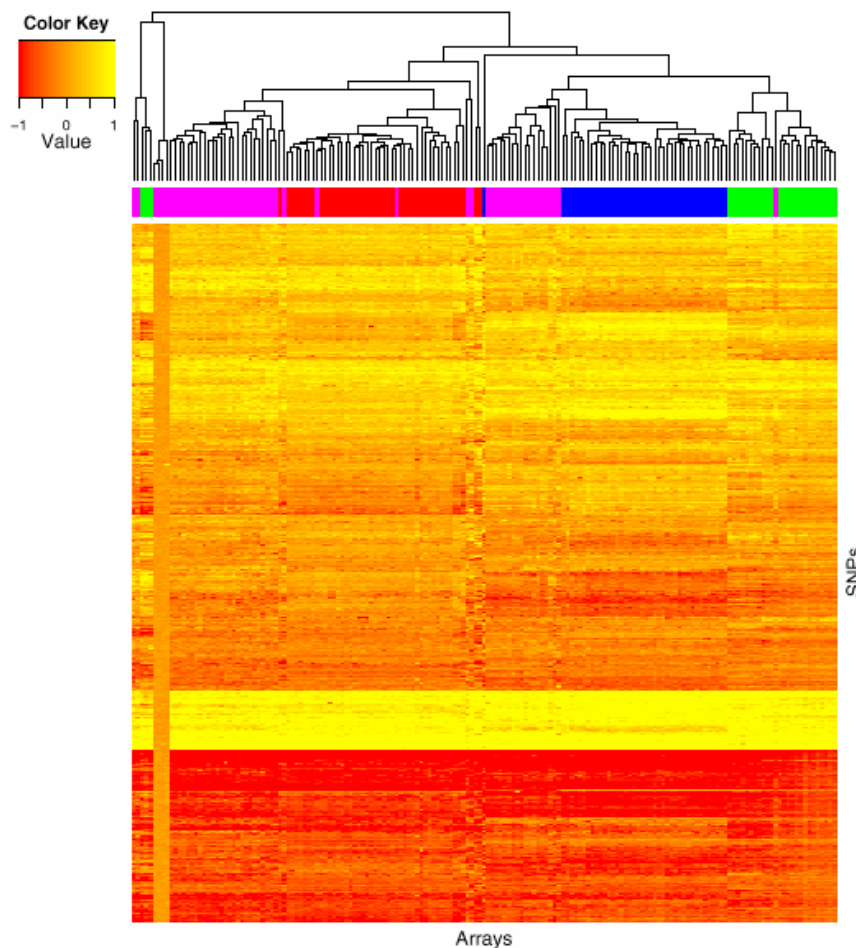


Abb. 15 Heatmap der median normalisierten Intensitätswerte von 169 *HindIII*-Arrays. Die Abbildung zeigt die \log_2 Intensitätswerte von 1000 SNPs nach dem hierarchischen Clustern. Für die Berechnung wurde die Summe der Intensitätswerte der A- und B-Allele verwendet. Die Spalten enthalten die verschiedenen *HindIII*-Arrays, die Zeilen die 1000 SNPs. Die Zeile über der Heatmap zeigt einen Farbcode für die 4 unterschiedlichen Zeitpunkte, an denen die Arrays hybridisiert wurden. Obwohl sich die Profile der Kopienanzahl aller Arrays gleichen sollten, zeigen sie 4 klar trennbare Gruppen.

Um verschiedene Arraytypen und Methoden vergleichen zu können, ist die Standardabweichung und die mediane Abweichung nicht ausreichend, da das Ausmaß der Zunahme bzw. Abnahme der \log_2 Intensitätswerte durch die Duplikationen bzw. Deletionen unterschiedlich ist. Damit die Ergebnisse vergleichbar sind, wurde die Differenz der \log_2 Intensitätswerte der autosomalen SNPs von den X-

chromosomalen SNPs der männlichen Proben verwendet, um das Signal-Rausch-Verhältnis abzuschätzen. Diese Werte sind in Tabelle 3 aufgeführt.

Tabelle 3 Zusammenfassung des durchschnittlichen Signal-Rausch-Verhältnisses (SNR)

Array	SNR
HindIII	4.6
XbaI	5.14

Aus den Werten geht hervor, dass die durchschnittliche Anzahl an SNPs, die aufeinander folgend deletiert sein müssen, um in dem Kollektiv von 67 Patienten eine signifikante Deletion oder Duplikation anzuzeigen, bei 5 SNPs für den HindIII-Array und bei 6 SNPs für den XbaI-Array liegt.

In den vorliegenden Daten hatten die einzelnen Arrays unterschiedliche Standardabweichungen und die \log_2 Intensitätswerte entsprachen nicht einer Normalverteilung. Deshalb wurde die tatsächliche Anzahl an SNPs, die auf eine Kopienzahlvariation hinweisen, der jeweiligen Standardabweichung der einzelnen Arrays so angepasst, dass maximal 8 CNVs pro untersuchter DNA-Probe erhalten wurden.

2.2 Kandidatenregionen

Die Anwendung der oben genannten Kriterien zur Bestimmung von CNVs ergab 27 Kandidatenregionen in 24 Patienten.

Ausgeschlossen wurden zuvor schon alle Stellen, die signifikant mit CNVs überlappen und die bereits in der Normalbevölkerung beschrieben wurden (Database of Genomic Variants).

Die 27 Kandidatenregionen wurden mit quantitativer PCR überprüft. In 14 Regionen, die mehr als 20 SNPs enthalten, wurde die CNV bestätigt. In 13 Regionen, die 5 bis 20 SNPs enthielten, wurden 8 Regionen bestätigt (Tabelle 4). Insgesamt wurden also 22 Kandidatenregionen (81,5 % der Regionen) als CNVs bestätigt. Die deletierten und duplizierten Stellen enthalten bis zu 49 Gene und variieren in ihrer Größe von 10 kb bis 7.5 Mb.

Es wurde auch überprüft, ob die CNVs von einem phänotypisch unauffälligen Elternteil vererbt wurden. Dafür wurden bei allen verfügbaren Eltern quantitative PCRs an den Stellen durchgeführt, die bei dem jeweiligen Kind eine CNV anzeigen.

Für die weitere Auswertung wurden folgende 10 Regionen nicht berücksichtigt:

- a) Regionen, die von einem Elternteil auf den Patienten vererbt wurden.
- b) Regionen, in denen keine kodierende Sequenz enthalten ist.
- c) Regionen, bei denen nicht geprüft werden konnte, ob sie *de novo* entstanden sind, da keine parentale DNA zur Verfügung stand. Eine Ausnahme wurde bei dem Patienten 29945 gemacht. Obwohl die paternale DNA nicht vorhanden war, wurde die Region weiter untersucht, da nur ein Gen deletiert war.

Tabelle 4 Zusammenfassung der Kandidatenregionen

Table Kandidatenregionen									
ID	Chromosom	Deletion/ Duplikation	Start rsSNP	Ende rsSNP	Anzahl der SNPs	Länge (Mb)	qPCR Bestätigung	Anzahl der qPCRs	Primer
27384	1q31.1-q31.3	Deletion	10494585	10494695	464	7.75	Ja	1	8920
30437	2p25.3-p25.1	Deletion	2313466	1964092	193	3.79	Ja	1	8930, 9069, 9070
29922	8p23.1	Deletion	2945251	2466115	177	3.9	Ja	3	8926, 9061, 9062
30375	3p25.3-p25.2	Deletion	10510400	10510422	98	2.83	Ja	3	8928, 9065, 9066
28735	12p13.33	Deletion	953385	2283285	62	2.01	Ja	3	8924, 9077, 9078
30428	17q21.31	Deletion	436667	1918798	38	0.48	Ja	1	8929, 9067, 9068
28283	7q21.13	Deletion	10487988	10488004	36	0.27	Ja	3	9651, 9652, 9653
29836	Xp22.31	Duplikation	719632	10521669	31	1.42	Ja	4	9820, 9821, 9822, 9060
27737	17p11.2	Duplikation	4073940	1373147	29	3.22	Ja	3	8921, 9054, 9055
29837	Xp22.31	Duplikation	7889947	10521669	29	1.3	Ja	4	8925, 8933, 9060
28430	15q25.2	Deletion	17158372	10520569	27	1.37	Ja	3	8923, 9058, 9059
27581	9p23	Deletion	10511570	1900218	23	0.28	Ja	1	8934
28701	13q12.11	Duplikation	9315234	4570685	23	0.53	Ja	3	8922, 9056, 9057
29945	3p13	Deletion	10511008	10511014	22	0.44	Ja	5	8939, 9063, 9064
27581	1q25.2	Duplikation	10494517	7555418	20	0.59	Nein	3	8932, 9073, 9074
27831	3q24	Deletion	2140300	10513274	19	0.19	Ja	1	8935
27526	Xp22.31	Duplikation	10521668	5934414	15	0.34	Nein	4	9820, 9821, 9822, 9060
29608	4q28.3	Duplikation	1495265	10518595	14	0.22	Ja	1	8937
27733	11q14.1	Duplikation	870066	10501436	13	0.11	Nein	1	8941
29195	2q23.1	Deletion	2890919	10497034	13	0.2	Ja	2	8936, 8948
27733	9p23	Deletion	372412	983282	11	0.14	Ja	1	8940
29996	5p15.2	Deletion	31953	26152	10	0.03	Ja	1	8944
30227	16p13.12	Deletion	190013	9452	8	0.05	Ja	1	8946
29700	13q31.2	Deletion	221022	452708	7	0.03	Ja	4	8938, 9655, 9656, 9657
29945	3p14.2	Deletion	9311853	10510892	7	0.05	Nein	1	8939
28283	8q22.1	Deletion	6996243	1378125	6	0.66	Nein	3	9648, 9649, 9650
29199	6q22.33	Deletion	3778130	265353	6	0.01	Ja	1	8942
30221	1q23.3	Duplikation	9330294	10494355	6	0.06	Ja	2	8945, 9076

Schließlich wurden 12 CNVs weiter untersucht, 9 Deletionen und 3 Duplikationen, die als Ursache der mentalen Retardierung angesehen wurden (Abb. 16 und Tabelle 5).

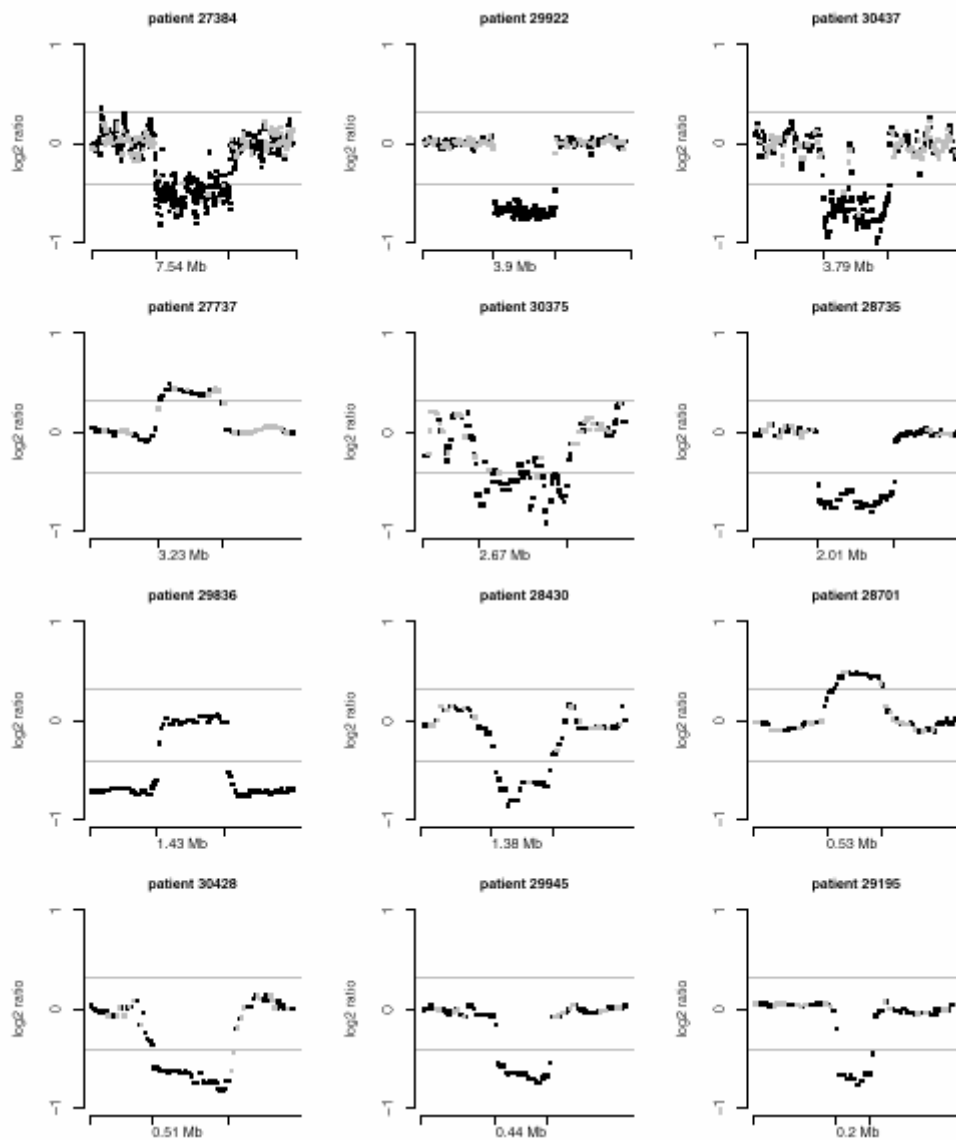


Abb. 16 Überblick über die \log_2 Intensitätsprofile der 12 ausgewählten CNVs. Die genomischen Profile zeigen die \log_2 Intensitätsverhältnisse der CNVs und der benachbarten Regionen in den Patienten mit Entwicklungsverzögerung. Auf der X-Achse ist die Länge der CNVs aufgeführt, auf der Y-Achse die \log_2 Intensitätsverhältnisse nach dem medianen Glättungsverfahren mit einer Fenstergröße von 9 SNPs. Die Dosiswerte von heterozygoten SNPs sind in grau dargestellt, die homozygoten SNPs in schwarz. Die Schwellenwerte, die eine Deletion bzw. Duplikation anzeigen, sind als graue, horizontale Linien bei $\log_2(0.75)$ und $\log_2(1.25)$ eingezeichnet.

11 CNVs wurden ausgewählt, da sie folgenden Kriterien entsprachen:

- 8 CNVs sind *de novo* entstanden.
- 2 CNVs überlappen mit bekannten Deletions- und Duplikationssyndromen (8p23.1 und 17p11.2).
- 1 maternal vererbte 1.4 Mb große Duplikation auf Chromosom Xp22.31 in einem männlichen Patienten, da Deletionen in dieser Region mentale Retardierung verursachen und die Duplikation bei der Mutter auf dem nicht zufällig inaktiviertem X-Chromosom liegt

Zusätzlich wurde als 12. CNV die Deletion in dem Patienten 29945 genauer untersucht. In der deletierten Region liegt nur ein Gen. Da nur die maternale DNA zur Verfügung stand, konnte nicht bestimmt werden, ob die Deletion *de novo* entstanden ist.

Tabelle 5 Deletionen und Duplikationen

Table Deletionen und Duplikationen										
ID	Deletion oder Duplikation	Chromosom	Länge (Mb)	Anzahl der SNPs	Anzahl der Gene	Segmentale Duplikationen	LOH paternal	LOH maternal	Anzahl der qPCRs	Bestätigung durch eine 2. Hybridisierung
27384	Deletion	1q31.1-q31.3	7.75	464	17	Nein	48	0	1	Ja
30437	Deletion	2p25.3-p25.1	3.79	193	4	Nein	0	18	1	Nein
29195	Deletion	2q23.1	0.20	13	1	Nein	0	0	2	Ja
30375	Deletion	3p25.3-p25.2	2.83	98	39	Nein	8	0	3	Nein
29945	Deletion	3p13	0.44	22	1	Nein	n.a.	n.d.	3	Ja
29922	Deletion	8p23.1	3.90	177	34	Ja	n.a.	0	3	Nein
28735	Deletion	12p13.33	2.01	62	18	Nein	4	0	3	Nein
28430	Deletion	15q25.2	1.37	27	11	Ja	0	2	3	Ja
30428	Deletion	17q21.31	0.48	38	7	Ja	26	0	1	Ja
28701	Duplikation	13q12.11	0.53	23	5	Nein			3	Ja
27737	Duplikation	17p11.2	3.22	29	50	Ja	n.a.		3	Ja
29836	Duplikation	Xp22.31	1.42	31	4	Ja			3	Ja

n.a.: nicht verfügbar, n.d.: nicht bestimmt

2.3 CNVs, die von low-copy repeats flankiert sind

Von den 12 Deletionen und Duplikationen sind 5 CNVs von low-copy repeats (LCRs) flankiert. Eine nicht allelische Rekombination wird oft begünstigt durch diese flankierenden duplizierten Regionen, die eine Größe von über 1kb bis zu mehreren 100000 kb haben und zu über 90% identische Sequenzen haben.

Die erste CNV ist eine 3.9 Mb große Deletion auf Chromosom 8p23.1 in einem männlichen Patienten (ID 29922) mit einer milder Entwicklungsverzögerung. Die Deletion wird von 177 deletierten SNPs angezeigt. Die Untersuchung der maternalen DNA zeigte, dass die Mutter eine normale Kopienzahl in dieser Region hat. Die Analyse der Genotypdaten in der Deletion zeigte außerdem, dass der Patient ein maternales Allel besitzt und daher das paternale Allel deletiert ist.

Die Lage und Größe der Deletion entspricht der des 8p23.1-Deletionssyndroms. Das 8p23.1-Deletionssyndrom ist durch einen kongenitalen Herzfehler, meist atrioventrikuläre Septumdefekte, eine Mikrozephalie, eine intrauterine Wachstumsretardierung, eine milde mentale Retardierung und ein hyperaktives, impulsives Verhalten charakterisiert. (Devriendt et al. 1999, Faivre et al. 1998). Die Deletion enthält das Gen *GATA4*. Es wurde gezeigt, dass die Deletion dieses Gens für die Herzfehler bei Patienten mit dem 8p23.1-Syndrom verantwortlich ist. In einer Studie wurden neun Patienten mit

einer 8p-Deletion untersucht. Dabei wurde festgestellt, dass alle Patienten mit einem Herzfehler eine Deletion hatten, die das Gen *GATA4* betraf (Devriendt et al. 1999). Durch ein Knockout-Mausmodell für das Gen *Gata4* wurde gezeigt, dass *Gata4* an der Morphogenese des Mausherzens beteiligt ist (Devriendt et al. 1999). Bei dem Patienten 29922 konnte mit quantitativer PCR bestätigt werden, dass das Gen *GATA4* in der deletierten Region liegt. Im Einzelnen ist der Patient von folgenden Symptomen betroffen, Mikrozephalus, Hypospadie und ein Vorhofseptum-Defekt, der auf die Haploinsuffizienz des Gens *GATA4* zurückzuführen ist. Der Patient hat keine fazialen Dysmorphien, keine Epilepsie und keine Hernie des Diaphragmas.

Die zweite CNV mit flankierenden LCRs ist die 3.2 Mb große Duplikation auf Chromosom 17p11.2. Der Patient (ID 27737) ist von einer milden Entwicklungsverzögerung betroffen. Die Anzahl an duplizierten SNPs in dieser Region betrug 29. Die Untersuchung der maternalen DNA zeigte, dass die Mutter 2 Kopien der Region besitzt, die DNA des Vaters konnte nicht untersucht werden. Die Duplikation stimmt mit dem 3.7 Mb großen, interstitiellen 17p11.2-Duplikationssyndrom überein. Dieses Syndrom ist charakterisiert durch Hypotonie, Gedeihstörung im Säuglingsalter und der frühen Kindheit, Entwicklungsretardierung, mentale Retardierung, faziale Dysmorphien, schwere Störung der kommunikativen Fähigkeiten, autistische Verhaltensmuster, kardiovaskuläre Auffälligkeiten, Schlafapnoe und in seltenen Fällen Kleinwuchs (Potocki et al. 2007). Die Lage und Größe der Duplikation beim 17p11.2-Duplikationssyndrom entspricht der Deletion, durch die das Smith-Magenis-Syndrom (MIM 182290) verursacht wird. Die Region ist am proximalen und distalen Ende von den SMS-REPs flankiert. Es wird vermutet, dass durch diese Elemente bei beiden Syndromen eine nicht allelische homologe Rekombination begünstigt wird (Potocki et al. 2000; Potocki et al. 2007).

Der Patient 27737 hatte normales Geburtsgewicht und eine normale Geburtslänge. Postnatal zeigte sich eine Gedeihstörung. Im Alter von 10 Monaten lag seine Körpergröße bei der 10. Perzentile und der Kopfumfang war unter der 3. Perzentile. Er hat eine Hypospadie Grad 1 und entwickelte eine Hypotonie. Erst im Alter von 27 Monaten begann der Junge einzelne Wörter zu sprechen. Die fazialen Dysmorphien des Patienten, wie ein Telekanthus, eine dreieckige Gesichtsform und eine breite Stirn entsprechen den Symptomen, die bei Kindern mit dem Potocki-Lupski-Syndrom beschrieben sind.

Eine weitere CNV mit LCRs ist die 500 kb große Deletion auf Chromosom 17q21.3. Sie wird durch 38 SNPs angezeigt. Die Patientin 30428 hat eine moderate Form der Entwicklungsverzögerung.

Zusätzlich sind die motorische und sprachliche Entwicklung verzögert, Das Mädchen ist sehr infektanfällig, hat eine Muskelschwäche, faziale Dysmorphien, wie tiefliegende Augen mit weitem Abstand und eine Sehschwäche mit alternierendem Schielen.

Die Deletion überspannt das erst kürzlich entdeckte Mikrodeletionssyndrom 17q21.3 (Sharp et al. 2006; Koolen et al. 2006, Shaw-Smith et al. 2006). Patienten mit dem 17q21.3-Deletionssyndrom zeigen eine moderate mentale Retardierung, schwere Hypotonie von Geburt an mit daraus resultierender motorischen Entwicklungsretardierung, eine sprachliche Retardierung, lange Finger und charakteristische faziale Dysmorphien, wie Ptosis, Blepharophimosis, große, niedrig angesetzte Ohren, breite Nase und Kinn. Das 17q21.3-Deletionssyndrom umfasst 6 Gene, den kortikotropin-ausschüttender Hormonrezeptor 1 (*CRHR1*), das mikrotubulus-assoziierte Tauprotein (*MAPT*), die intramembrane Protease 5 (*IMP5*) und 3 hypothetische Proteine (*STH*, *FLJ25168*, *LOC284058*).

Die Deletion umfasst einen bekannten, 900 kb großen Inversionspolymorphismus. Dieser Inversionspolymorphismus tritt bei ca. 20 % der Nordeuropäer auf. Die Region unterscheidet sich in zwei Haplotypen, den H1- und H2-Haplotypen. Der H2-Haplotyp hat, im Gegensatz zum H1-Haplotyp einen direkt orientierten LCR, der an den Bruchpunkten der 17q21.3-Deletionen liegt. Es wird daher vermutet, dass die Deletionen durch eine nicht-allelische homologe Rekombination verursacht werden, die durch die H2-spezifischen Haplotypen ausgelöst wird (Stefansson et al. 2005, Evans et al. 2004).

Die Analyse der Genotypdaten der 38 deletierten SNPs der Patientin und ihrer Eltern ergab, dass der Vater homozygot für den H2 Haplotyp ist und die Mutter homozygot für den H1 Haplotyp. Bei dem Mädchen trat die Deletion in einem der paternalen H2 Haplotypen auf.

Auf Chromosom 15q25.5 wurde eine vierte CNV mit LCRs detektiert. Die 1.4 Mb große Deletion umfasst 11 Gene und wurde bei einem Mädchen (ID 28430) mit einer milden Entwicklungsverzögerung gefunden. Das Mädchen wurde in der 38. Schwangerschaftswoche mit einem Gewicht von 1950 g und einer Größe von 42 cm geboren. Die Patientin hat neben der Entwicklungsverzögerung, eine Polysplenie und leichte faziale Dysmorphien. Die Patientin entwickelte im ersten Lebensjahr eine hypoproliferative makrozytische Anämie und benötigte Bluttransfusionen bis zum vierten Lebensjahr. Aufgrund der Anämie und des Kleinwuchses wurde die Diagnose Diamond-Blackfan-Anämie (MIM 105650) in Erwägung gezogen. Mit dem Alter von 11 Jahren musste

sie wegen Ösophagusvarizenblutungen behandelt werden. In der Ultraschallsonographie wurde eine schwere Leberfibrose festgestellt.

Die fünfte CNV, die von LCRs flankiert wird, ist eine 1.4 Mb große Duplikation auf Chromosom Xp22.31 bei einem Jungen mit schwerer Entwicklungsverzögerung (ID 29836). Die Duplikation wurde von der phänotypisch gesunden Mutter auf den Sohn übertragen. Die Duplikation tritt in der Region auf dem Chromosom Xp22.31 auf, in der Deletionen die X-chromosomale Ichthyose auslösen können. X-chromosomale Ichthyose ist eine genetische Erkrankung, die durch eine Haploinsuffizienz der Steroid-Sulfatase (*STS*) ausgelöst wird. Bei 80-90 % der Patienten ist das komplette *STS*-Gen deletiert, aber die Größe der Deletion ist variabel. Die meisten Patienten haben als klinisches Symptom nur die Ichthyose, aber in einigen Fällen ist auch eine mentale Retardierung beschrieben. Es wird vermutet, dass die Deletion durch nicht-allelische homologe Rekombination zwischen zwei 9 kb großen LCRs entsteht, die jeweils ein *VCX*-Gen (*VCX3A* und *VCX2*) enthalten (Van Esch et al., 2005). In zwei Studien wurden Patienten mit und ohne mentaler Retardierung bei X-chromosomaler Ichthyose untersucht. Abhängig davon, ob die homologen Sequenzen proximal oder distal der *VCX*-Sequenzen für die Rekombination verwendet wurden, ist entweder das *VCX3A*-Gen deletiert und das *VCX2*-Gen in normaler Dosis erhalten oder umgekehrt (Fukami et al., 2000, Van Esch et al., 2005). Die mentale Retardierung scheint nach diesen Studien in Zusammenhang mit der Deletion des Gens *VCX3A* zu stehen. Diese Annahme ist jedoch umstritten, da es Patienten mit einer Xp;Yq Translokation gibt, deren *VCX3A*-Gen deletiert ist und die trotzdem eine normale Intelligenz haben. Eine alternative Hypothese ist, dass das Gen *VCX3A* nicht allein für die mentale Retardierung verantwortlich ist, sondern dass es einen dosisabhängigen Mechanismus der *VCX/Y*-Genfamilie gibt, der eine normale kognitive Entwicklung gewährleistet.

Da Deletionen in dieser Region zu mentaler Retardierung führen können, wurde der Patient 29836, bei dem diese Region dupliziert ist, weiter untersucht. Die Duplikation ist lokalisiert zwischen den Genen *VCX3A* und *VCX2* und wurde von 18 duplizierten SNPs angezeigt. Mit Hilfe von quantitativer PCR wurde festgestellt, dass Amplikons, die 6 kb distal von *VCX3A* und 3 kb proximal von *VCX2* liegen in normaler einfacher Dosis vorliegen, während Amplikons 3 kb proximal von *VCX3A* und 5 kb distal von *VCX2* dupliziert sind (Abbildung 17). Abhängig davon, wo die Rekombination stattgefunden hat, ist entweder *VCX3A* oder *VCX2* dupliziert, oder es wurde ein Fusionprodukt aus beiden Genen

gebildet. Alle Gene, die innerhalb dieser beiden VCX-Gene liegen (*HDHD1A*, *STS*, *VCX*, *PNPLA4*), sind bei dem Patienten dupliziert.

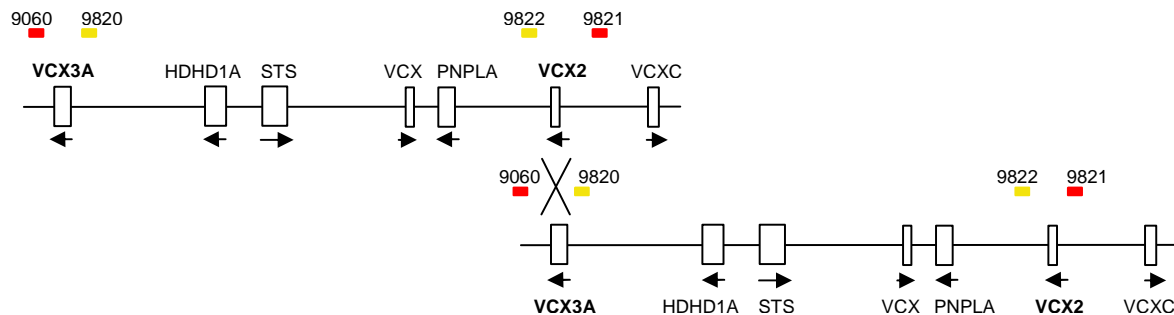


Abb. 17 Schematische Darstellung der betroffenen Region Xq22.31 des Patienten 29836. Gene sind durch weiße Blöcke dargestellt, die Pfeile darunter geben die Richtung der Transkription an. Die roten und grünen Kästchen über den Genen zeigen die Lage der Primer für die quantitative PCR mit den zugehörigen Primernummern. Die Primer 9060 und 9821 sind in einfacher Dosis vorhanden (rote Kästchen), während die Primer 9822 und 9820 in doppelter Dosis (gelbe Kästchen) vorliegen. Die nicht-allelische homologe Rekombination muss zwischen den homologen Regionen der Primer 9060 und 9820 und der Primer 9822 und 9821 stattgefunden haben.

Die genetische Untersuchung der phänotypisch gesunden Eltern zeigte, dass der Vater nur eine Kopie der Region auf seinem X-Chromosom trägt, während die Mutter drei Kopien besitzt und daher Überträgerin dieser Duplikation ist. Es wurden Untersuchungen zur X-Inaktivierung im peripheren Blut der Mutter durchgeführt. Die X-Inaktivierung wurde mit den polymorphen Trinukleotid-Repeats im Androgen-Rezeptor-Gen (*HUMARA*) untersucht. Dieser Marker ermöglicht es bei heterozygoten Frauen (Heterozygotenindex 0.9) beide Allele zu unterscheiden.

In einer Region 80 bp stromaufwärts des Markers liegt eine CpG-Insel, die mehrere Schnittstellen des methylierungssensitiven Enzyms *HpaII* enthält. Diese Stellen sind auf dem inaktiven X-Chromosom methyliert und werden von *HpaII* nicht geschnitten. Die Primer sind so gelegt, dass nach dem Verdau nur das Allel auf dem inaktivem X-Chromosom amplifiziert werden kann. Liegt eine zufällige X-Inaktivierung vor, sind beide X-Chromosomen der Mutter im ungefähr gleichen Verhältnis nachweisbar. Ist die X-Inaktivierung dagegen nicht zufällig, wird nach dem Verdau mit *HpaII* nur ein Allel amplifiziert. In diesem Fall kann nachgewiesen werden, ob das Allel auf dem inaktiven X-Chromosom von der Mutter auf den Sohn vererbt wurde.

In Abbildung 18 ist der Zustand der DNA von Mutter und Sohn vor und nach dem Restriktionsverdau zu sehen. Nach dem Verdau ist ersichtlich, dass bei der Mutter eine nicht zufällige Inaktivierung des

Allels 1 vorliegt. Die Längenangabe der Allele macht deutlich, dass das Allel 1, das die Duplikation trägt, von der Mutter auf den Sohn vererbt wurde.

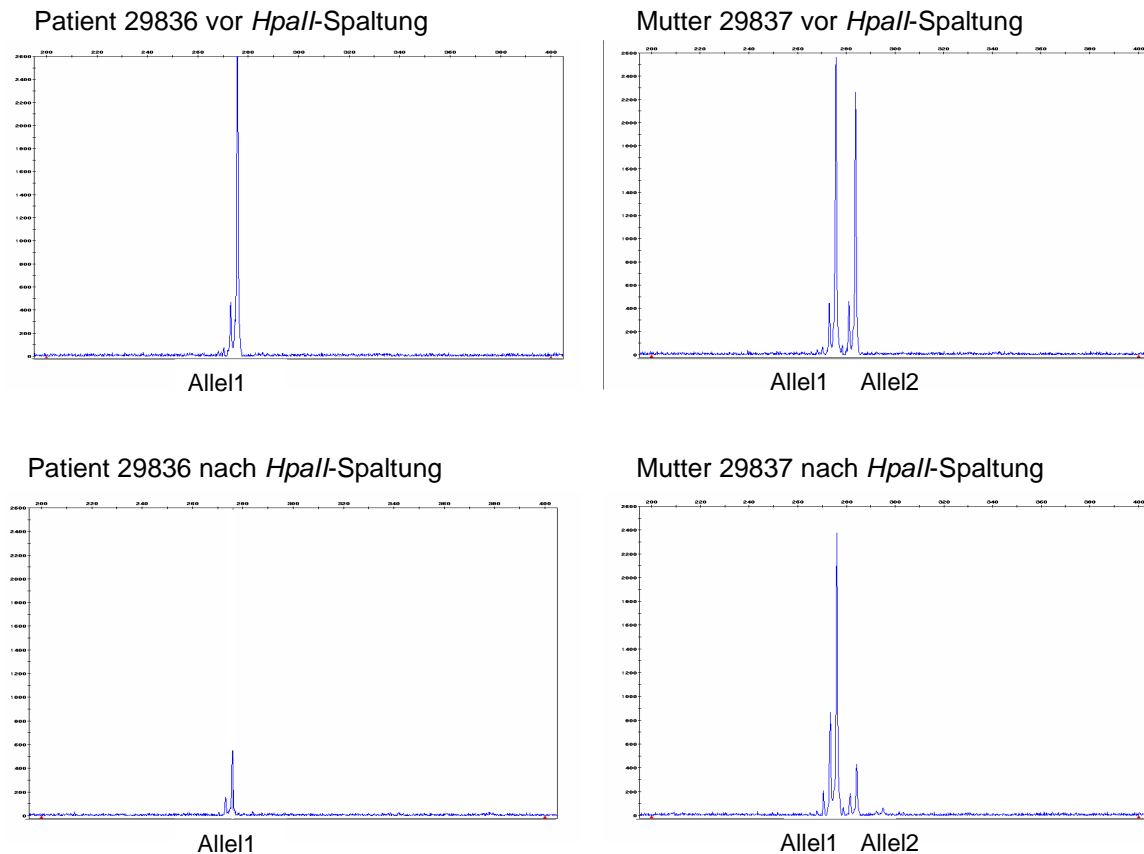


Abb. 18 Darstellung der X-Inaktivierung des Patienten 29836 und seiner Mutter 29837. Die Abbildung zeigt die Ergebnisse vor und nach der Spaltung mit dem Enzym *HpaII* in dem Patienten und seiner Mutter. Es ist deutlich zu erkennen, dass die Duplikation bei beiden auf dem Allel 1 auftritt und dass dieses Allel bei der Mutter inaktiviert ist.

Es wurde vermutet, dass die Duplikation bei dem Jungen die Ursache für die mentale Retardierung ist, weil das Chromosom, das die Duplikation trägt, bei der Mutter nicht zufällig inaktiviert ist und weil Deletionen in dieser Region eine mentale Retardierung hervorrufen können.

Auch die sprachliche Entwicklung ist bei dem Patienten verzögert. Im Alter von 9 Jahren konnte der Junge nur 2-Wort-Sätze anwenden. Er entwickelte postnatal eine Mikrozephalie, mit einem Kopfumfang von 50 cm (3. Perzentile) im Alter von 9 Jahren. Es wurden vor kurzem noch von 2 weiteren Fällen mit einer Duplikation Xp22.31 berichtet, die die Vermutung nahe legen, dass Duplikationen in dieser Region häufiger vorkommen (Horn et al. 2007).

2.4 CNVs, die nicht von low-copy repeats flankiert sind

Bei 7 Patienten wurden *de novo* CNVs detektiert, die nicht von LCRs umgeben waren. Die Größe dieser CNVs variierte von 200 kb bis zu 7.5 Mb.

- 1) Das Kind (ID 27384) mit der größten CNV, einer Deletion auf Chromosom 1q31.1-31.3, die 7.5 Mb umspannt, hat neben einer Entwicklungsverzögerung keine weiteren Symptome. In der deletierten Region, die durch 464 SNPs angezeigt wurde, liegen 17 Gene. Auch die DNA der Eltern wurde mit den 100K Arrays untersucht, um zu zeigen welches parentale Allel deletiert ist. 48 SNPs waren für diese Fragestellung informativ und zeigten, dass eines der väterlichen Allele deletiert ist. Um die Bruchpunkte der Deletion exakt bestimmen zu können, wurde versucht ein Junction Fragment zu isolieren. Dafür wurden die möglichen Bruchstellen mit quantitativer PCR näher eingegrenzt und der Versuch unternommen, ein PCR-Produkt über die Deletion zu amplifizieren. Mit Hilfe der quantitativen PCR konnten die Bereiche, in dem der Bruchpunkt liegt am proximalen Ende auf ungefähr 1 kb eingegrenzt werden und am distalen Ende auf ca. 1.5 kb. Trotzdem war es nicht möglich ein Junction Fragment zu amplifiziert, wahrscheinlich weil auf beiden Seiten repetitive Sequenzen in den Bruchpunktbereichen liegen.

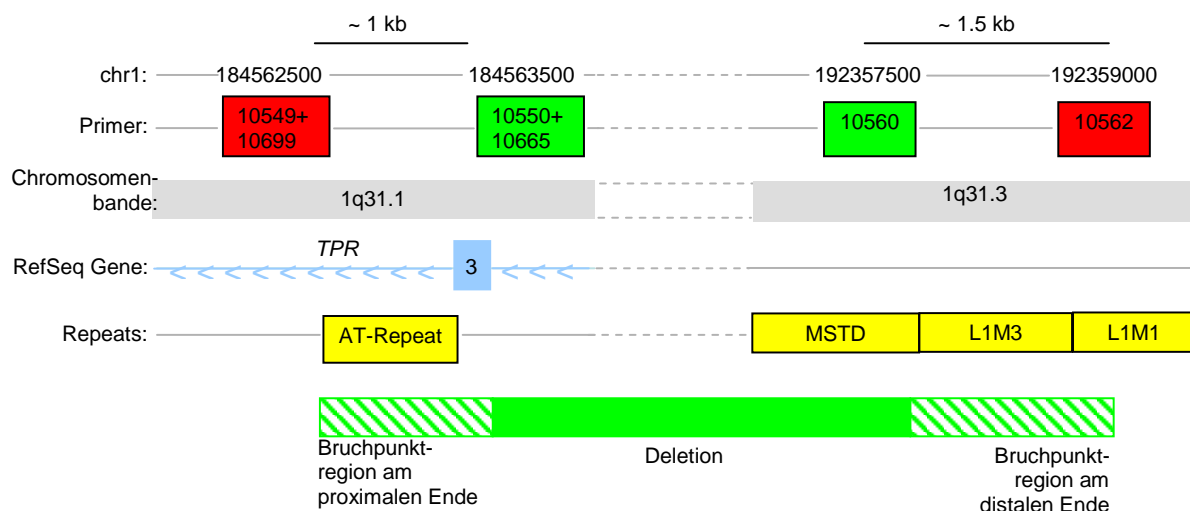


Abb. 19 Bestimmung der Bruchpunktregion bei Patient 27384
 In der Abbildung ist die Deletion auf Chromosom 1q31.1-q31.3 schematisch dargestellt. Oben in der Abbildung sind die genomische Position und die Primernummern der Produkte für die quantitative PCR dargestellt. Produkte, die in der quantitativen PCR eine zweifache Dosis zeigten, sind mit roten Kästchen markiert, Produkte, die eine einfache Dosis zeigten, mit grünen Kästchen. Am proximalen Ende

wurde der Bereich, in dem der Bruchpunkt liegt, auf 1kb zwischen den Produkten 10549+10699 und 10550+10665 eingegrenzt. In diesem Bereich liegt ein 372 bp langer AT-Repeat und das Exon 3 des Gens *TPR* (blaues Kästchen). Das distale Ende der Deletion wurde innerhalb von 1.5 kb im Bereich der Produkte 10560 und 10562 durch quantitativer PCR festgelegt. Diese Region enthält einen LTR- Repeat, MSTD (MaLR) und zwei LINE-Repeats (L1M3 und L1M1).

2) Auf Chromosom 2p25.3-25.1 wurde bei dem Patienten 30437 eine Deletion mit 3.8 Mb identifiziert.

In der Deletion liegen 193 SNPs und 4 Gene, darunter das *SOX11*-Gen. *SOX11* wird während der Entwicklung des Embryos in vielen Geweben transient exprimiert. Im Mausmodell wurde gezeigt, Mäuse mit einer homozygoten *Sox11*-Deletion kurz nach der Geburt aufgrund von Herzfehlern sterben. (Sock et al, 2004).

Der Junge hat folgende Symptome: Mikrozephalus (3. Perzentile), Adipositas, Hypospadie I. Grades und Klinodaktylie D5 an beiden Händen und Füßen. Die Untersuchung der Genotypen der Eltern zeigte, dass bei 18 informativen SNPs das maternale Allel deletiert war. Die Bruchpunkte der Deletion wurden mit quantitativer PCR näher bestimmt werden. Am distalen Ende wurde der mögliche Bereich auf 7.3 kb reduziert, allerdings liegen in dieser Region zahlreiche repetitive Elemente. Das proximale Ende ist eine 300 bp große Region ohne Sequenzauffälligkeiten. Auch bei dieser Deletion war es nicht möglich ein Junction Fragment zu amplifizieren.

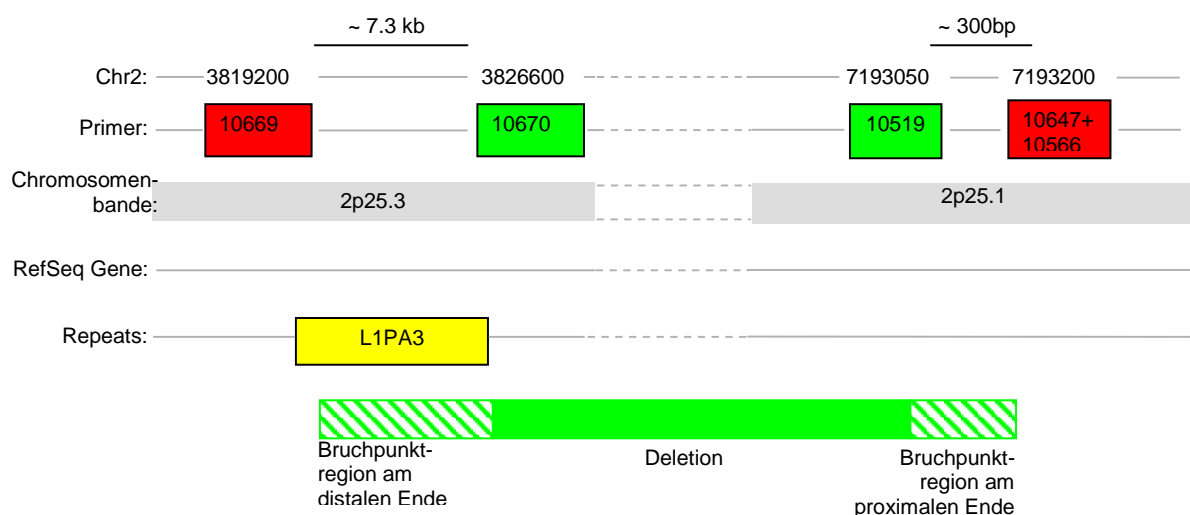


Abb. 20

Bestimmung der Bruchpunktregion bei Patient 30437

In der Abbildung ist die Deletion auf Chromosom 1q31.1-q31.3 schematisch dargestellt. Oben in der Abbildung sind die genomische Position und die Primernummern der Produkte für die quantitative PCR dargestellt. Produkte, die in der quantitativen PCR eine zweifache Dosis zeigten, sind mit roten Kästchen markiert, Produkte, die eine einfache Dosis zeigten, mit grünen Kästchen. Am distalen Ende

wurde der Bereich, in dem der Bruchpunkt liegt, auf 7.3 kb zwischen den Produkten 10669 und 10670 eingegrenzt. In diesem Bereich liegt ein 6 kb großer L1PA3-LINE Repeat. Das proximale Ende der Deletion wurde 300 bp zwischen den Produkten 10519 und 10647+10566 durch quantitativer PCR festgelegt.

- 3) Bei dem Patienten 30375 wurde eine 2.9 Mb große Deletion auf dem Chromosom 3p25.3-25.1 detektiert. Die Deletion liegt im proximalen Teil des 3p-Syndroms. Das 3p-Syndrom ist charakterisiert durch ein geringes Geburtsgewicht, mentale Retardierung und bei 30 % der Patienten durch kongenitale Herzfehler, vor allem Defekte des atrioventrikularem Septums. Bei dem Patienten wurden neben einer Entwicklungsverzögerung, ein Minderwuchs (<3. Perzentile) und ein Mikrozephalus (<3. Perzentile) diagnostiziert. Ein Schlaf-EEG war pathologisch und es wurde ein Vorhofseptumdefekt Typ II festgestellt. In der Deletion liegen 98 SNPs und 39 Gene. Unter den deletierten Genen befindet sich das Gen *CRELD1*. Mutationen in diesem Gen können der Grund für den atrioventricularen Septumdefekt 2 (AVSD2) sein (Zatyka et al. 2005). Auch bei diesem Patienten wurde untersucht, welches der parentalen Allele die Deletion betrifft. 8 informative SNPs zeigten einen Verlust des paternalen Allels in dieser Region.

Die Bruchpunktregionen wurden mit quantitativer PCR auf 700 bp am distalen Ende (Primer 10672 und 10650) und auf 400 bp am proximalen Ende (Primer 10660 und 10700) eingegrenzt. Mit dem Vorwärtsprimer 10672F und dem Rückwärtsprimer 10700R gelang es ein Junction Fragment zu amplifizieren. Um die genauen Bruchpunkte zu bestimmen, wurde das 750 bp große Junction Fragment sequenziert. Die Deletion des Patienten 30375 umspannt eine 2.9 Mb große Region. Beide Bruchpunkte liegen in Introns von Genen. Der distale Bruchpunkt im Intron 1 des Gens *MTMR14*, das bisher nicht weiter charakterisiert wurde, und der proximale Bruchpunkt im Intron 3 des Gens *MKRN2*, das für ein Ring-Finger Protein kodiert. Keiner der Bruchpunkte liegt direkt in einem Repeat, aber auf beiden Seiten gibt es 0.5-1 kb entfernt *Alu*-Repeats aus der *AluSx*-Familie. Am distalen Ende liegen zwei *AluSx*-Repeats je 1 kb distal und proximal des Bruchpunktes, am proximalen Ende liegt ein *AluSx*-Repeat 0.5 kb distal des Bruchpunktes.

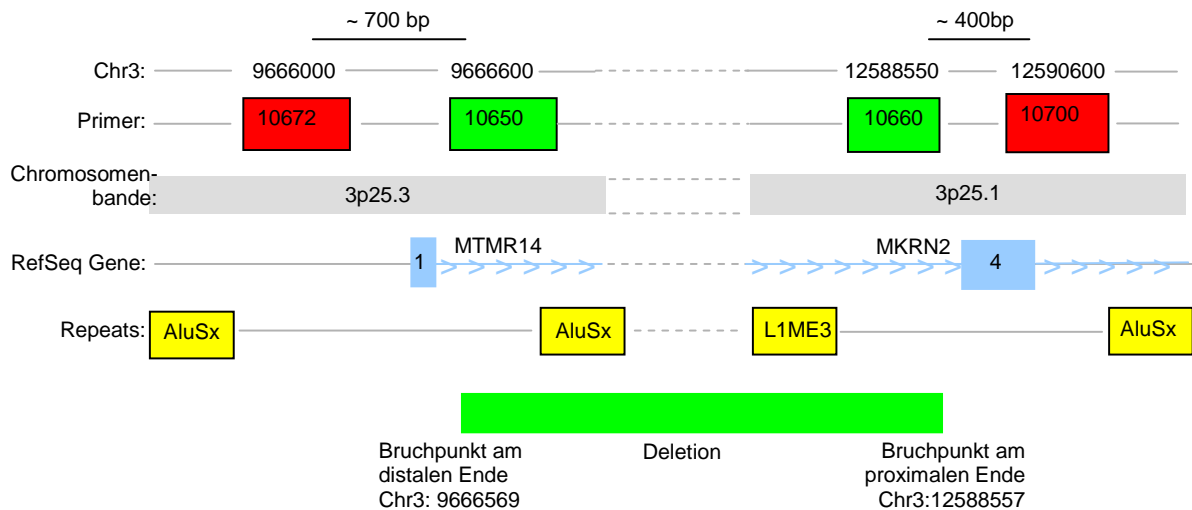


Abb. 21 Bestimmung der Bruchpunktregion bei Patient 30375

In der Abbildung ist die Deletion auf Chromosom 1q31.1-q31.3 schematisch dargestellt. Oben in der Abbildung sind die genomische Position und die Primernummern der Produkte für die quantitative PCR dargestellt. Produkte, die in der quantitativen PCR eine zweifache Dosis zeigten, sind mit roten Kästchen markiert, Produkte, die eine einfache Dosis zeigten, mit grünen Kästchen. Am distalen Ende wurde der Bereich, in dem der Bruchpunkt liegt, auf 700 bp zwischen den Produkten 10672 und 10650 eingegrenzt. Der exakte Bruchpunkt wurde durch die Amplifikation eines Junction Fragments bestimmt und liegt am distalen Ende auf Position 9666569 (UCSC Genome Browser hg18). Je 1 kb von diesem Bruchpunkt liegt ein *AluSx*-Repeat. Das proximale Ende der Deletion wurde innerhalb von 400 bp im Bereich der Produkte 10660 und 10700 durch quantitative PCR festgelegt. Der genaue Bruchpunkt liegt auf Position 12588557 (UCSC Genome Browser hg18) und ist im Abstand von je 500 bp von Repeats (*L1ME3* und *AluSx*) flankiert. Beide Bruchpunkte liegen in Introns von Genen. Der distale Bruchpunkt liegt im Intron 1 des Gens *MTMR14* und der proximale Bruchpunkt im Intron 3 des Gens *MKRN2*. Die Exons sind durch blaue Kästchen dargestellt, die Introns durch blaue Linien mit Pfeilen in Transkriptionsrichtung.

- 4) Auf Chromosom 12p13.33 wurde bei dem Patienten 28735 eine Deletion von 2,0 Mb durch 62 SNPs angezeigt. Es sind 18 Gene deletiert. Die Genotypen der Eltern zeigen, dass das paternale Allel deletiert ist. Der Junge hat neben einer Entwicklungsverzögerung eine sprachliche Verzögerung. Bei der Bestätigung der Deletion und der Eingrenzung der Bruchpunkte mit quantitativer PCR zeigte sich, dass die Deletion das terminale Ende des Chromosoms 12p betrifft und somit über 2 Mb groß ist. Am distalen Ende, ca. 17 kb nach Chromosomenbeginn, wurde die terminalste Probe 10479 für die quantitative PCR gelegt. Die Probe 10479 war nach Auswertung der quantitativen PCR deletiert. Die Bruchpunktregion am proximalen Ende wurde auf 3.6 kb im Intron 3 des Gens *CACNA1* (Primer 10483+10484) eingegrenzt. Ein Junction Fragment konnte nicht amplifiziert werden.

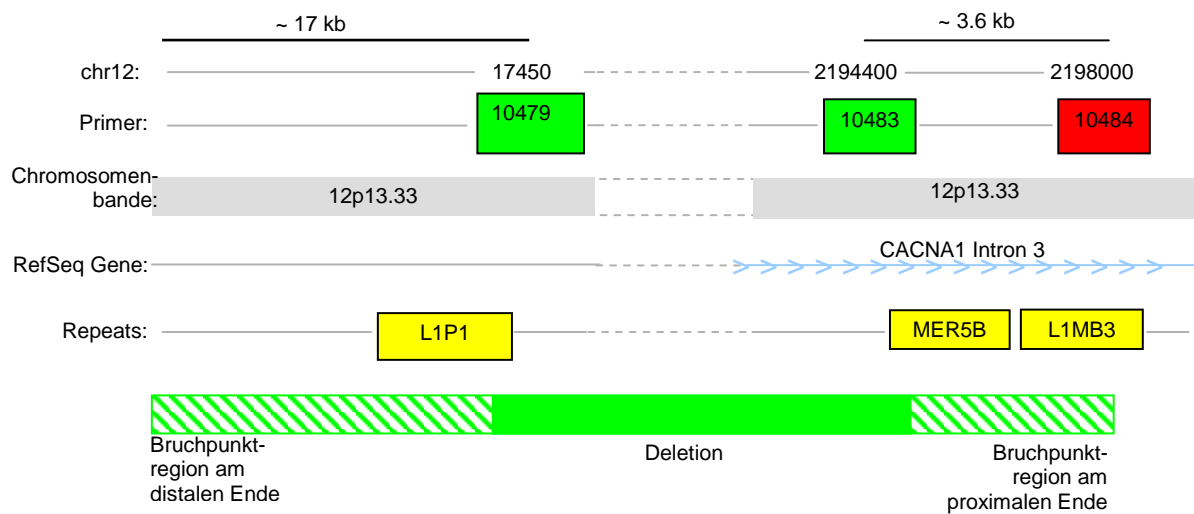


Abb. 22 Bestimmung der Bruchpunktregion bei Patient 28735

In der Abbildung ist die Deletion auf Chromosom 1q31.1-q31.3 schematisch dargestellt. Oben in der Abbildung sind die genomische Position und die Primernummern der Produkte für die quantitative PCR dargestellt. Produkte, die in der quantitativen PCR eine zweifache Dosis zeigten, sind mit roten Kästchen markiert, Produkte, die eine einfache Dosis zeigten, mit grünen Kästchen. Das distale Ende scheint bis zum Chromosomenanfang deletiert zu sein. Die letzte Probe (Primer 10479) die gelegt werden konnte, war in der quantitativen PCR deletiert. Vor der Probe 10479 liegt ein LINE-Repeat, an den sich die Telomerregion anschließt. Das proximale Ende der Deletion wurde im Intron 3 des Gens *CACNA1* innerhalb von 3.6 kb im Bereich zwischen den Produkten 10483 und 10484 durch quantitativer PCR festgelegt. In diesem Bereich liegt ein MER_1-Repeat (MER5B) und ein LINE-Repeat (L1MB3).

- 5) Bei dem Patienten 28701 wurde eine 525 kb große Duplikation auf dem Chromosom 13q12.11 festgestellt. Es sind insgesamt 6 Gene dupliziert. Neben einer Entwicklungsverzögerung hat der Patient einen Hydrocephalus, eine mediane Gaumenspalte, eine Trikuspidaldysplasie, Klumpfüße und dysplastische Ohren. In der duplizierten Region sind 3 Gap-junction Proteine enthalten, GJA3, GJB2 und GJB6. Alle 3 Gap-junction Proteine sind, wenn sie mutiert sind, mit einer Reihe von autosomal-dominanten und autosomal-rezessiven Erkrankungen assoziiert, die alle im Zusammenhang mit sensorineuronaler Taubheit stehen. Ein Befund über das Hörvermögen des Patienten lag nicht vor.

2.5 CNV mit einem deletierten Gen

Bei dem Patienten 29195 wurde eine nur 200 kb große Deletion auf Chromosom 2q23.1 von 13 SNPs angezeigt. In dieser Deletion liegen die ersten 7 von 10 kodierenden Exons des Methyl-CpG-Bindedomänenprotein 5 (*MBD5*). Das Gen *MBD5* wurde ursprünglich durch Sequenzhomologien zu den Genen *MBD1-MBD4* und *MECP2* entdeckt. Mutationen im *MECP2*-Gen führen zum Rett-Syndrom, einer neurologischen Erkrankung mit mentaler Retardierung, Sprachverlust, Autismus, Ataxie und stereotype Handbewegungen (Roloff et al, 2003).

Das Gen *MBD5* enthält eine PWWP-Domäne, eine Bindestelle für DNA, die häufig in Proteinen vorkommt, die in DNA-Methylierung, DNA Reparatur oder in die Regulation der Transkription involviert sind. Es wurden die Genotypen der Eltern bestimmt. Sie waren jedoch für die Fragestellung, welches parentale Allel deletiert ist, nicht informativ. Um auszuschließen, dass der Phänotyp autosomal-rezessiv verursacht ist, wurde der kodierende Teil des *MBD5*-Gens auf dem nicht deletierten Allel sequenziert. Die Sequenz des Patienten zeigte keine Unterschiede zur Referenzsequenz (NM_018328). Um die Bruchpunkte genau zu definieren, wurden zuerst mit quantitativer PCR die Bruchpunkte eingegrenzt und danach ein Junction-Fragment amplifiziert (GenBank Accession No EF504248) (Abb 23). Die Deletion erstreckt sich auf dem Chromosom 2 über 211870 bp von der Position 148752797 bis 148964667 (UCSC Genome Browser hg18). Der proximale Bruchpunkt liegt in einem LTR/ERVL-Repeat (MLT2B5) und der distale Bruchpunkt im Intron des Exons 7 im Gen *MBD5*. Im UCSC Genome Browser weisen ESTs (Expressed Sequence Tags) darauf hin, dass der Eintrag in der GenBank für *MBD5* im 5'-Bereich nicht vollständig ist. Mit 5'-RACE und RT-PCR wurde die RNA um 5 nicht-kodierende Exons erweitert (GenBank Accession No EF542797) (Abb 23).

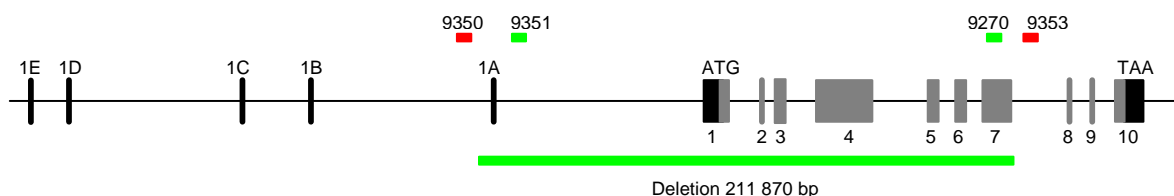


Abb. 23 Schematische Darstellung des Gens *MBD5*. Kodierende Exons sind durch nummerierte graue Kästchen dargestellt, nicht-kodierende Exons als schwarze Kästchen. Die Exons 1A bis 1E bezeichnen die nicht-kodierenden Exons am 5'-Ende, die in dieser Arbeit etabliert wurden. Die Primer für die quantitative PCR sind über der genomischen Darstellung als nummerierte rote

(normale Dosis in der quantitativen PCR) und grüne (deletierte Proben) Balken abgebildet. Die deletierte Region ist unter dem Schema in grün markiert.

Der Junge hat leichte faziale Dysmorphien und eine Sandalenfurchen zwischen dem ersten und zweiten Zeh. Neben der schweren mentalen Retardierung ist auch die motorische und sprachliche Entwicklung leicht verzögert. Der Junge hatte im Alter von 8 Monaten Krampfanfälle mit Fieber. Die ersten Anfälle ohne Fieber begannen mit 16 Monaten und erwiesen sich als therapieresistent. Im Verhalten ist der Junge hypoaktiv. Die soziale Interaktion ist gering.

Ein Hinweis, dass diese Deletion pathologisch sein könnte, zeigte sich in 2 Studien, in denen von Deletionen berichtet wurde, die mit der hier beschriebenen Deletion überlappen (Koolen et al. 2004, de Vries et al. 2005). In beiden Studien waren die Deletionen größer und es waren neben *MBD5* noch zusätzlich benachbarte Gene deletiert.

Für die weitere Bestätigung wurde untersucht, ob Mutationen in diesem Gen bei weiteren Kindern mit mentaler Retardierung vorhanden sind. Dazu wurden die DNAs von 415 Kindern untersucht. In 16 DNAs wurden 4 stille Varianten gefunden und 7 Varianten, die zu einem Aminosäureaustausch führten (Tab. 6). Darunter sind 4 Missense-Varianten, die in 660 Kontrollen (KORA) nicht gefunden wurden. Es konnte leider nicht überprüft werden, ob diese Varianten *de novo* entstanden sind, da in den meisten Fällen die DNA der Eltern nicht zur Verfügung standen.

Tabelle 6 Variationen im Gen *MBD5*

Tabelle Variationen in <i>MBD5</i>						
Variationen und Sequenzänderungen	ID	Exon/Intron	Kontrollindividuen			
			Genotyp 11	Genotyp 12	Genotyp 22	
Variationen mit Aminosäureaustausch						
c.431C>T	p.T144I	A12	4	CC: 660	CT: 0	TT: 0
c.1382G>A	p.R461H	A47, B217	4	GG: 649	GA: 0	AA: 0
c.1962C>A	p.D654E	C231	4	CC: 655	CA: 0	AA: 0
c.1963G>A	p.A655T	30224*	4	GG: 653	GA: 0	AA: 0
c.2030G>A	p.S677N	B134, C264, C281	4	GG: 662	GA: 3	AA: 0
c.2569G>A	p.A857T	31833*	5	GG: 670	GA: 0	AA: 0
c.3143C>T	p.T1048I	C260	7	CC: 640	CT: 0	TT: 0
stille Variationen						
c.1368G>T	p.S456S	A6, B135	4	GG: 663	GT: 1	TT: 0
c.1638C>T	p.A546A	A109, B139	4	CC: 658	CT: 3	TT: 0
c.2286C>T	p.H762H	B225	4	CC: 338	CT: 0	TT: 0
c.3279C>T	p.V1094V	A53	7	TT: 648	CT: 0	CC: 0

* in einem Elternteil wurde die gleiche Mutation gefunden

2.6 Potentiell pathologische CNV mit einem deletierten Gen

Eine zweite Deletion, die nur ein Gen, *FOXP1*, betrifft, ist eine 440 kb große Deletion auf Chromosom 3p13. Die Deletion wurde bei dem Patienten 29945 durch 22 SNPs angezeigt. Es stand nur die maternale DNA zur Verfügung, so dass nicht geklärt werden konnte, ob die Deletion *de novo* entstanden ist. In der DNA der Mutter wurden keine Veränderungen in dieser Region festgestellt.

Der Junge hat eine schwere Entwicklungsverzögerung und ein freundliches, hyperaktives Verhalten. Die aktive und passive Sprachentwicklung ist deutlich verzögert. Mit dem Alter von 7 Jahren sprach der Junge in Sätzen mit einem bis drei Worten. Im HAWIK-Test war der Intelligenzquotient kleiner als 50.

Die Deletion betrifft alle kodierenden Exons des *FOXP1*-Gens. Eine Sequenzierung der kodierenden Region des verbliebenen Allels zeigte keine Sequenzvariationen im Vergleich zur Referenzsequenz (NM_032682). Die Bruchpunkte der Deletion wurden, wie oben beschrieben, mittels quantitativer PCR und Amplifikation eines Junction Fragments (GenBank Accession No EF504249) bestimmt. Die Deletion umfasst mit 498299 bp alle kodierenden Exons (7 bis 21) des Gens *FOXP1* (Position Chr3:70807767-71306066; UCSC Genome Browser hg18). Der distale Bruchpunkt liegt in einem LINE Repeat (L1/L1PA5) vor *FOXP1* und der proximale in einem Repeat vom MER1-Typ im Intron 6. An den Bruchpunkten gibt es weder palindromische Sequenzen noch andere Ähnlichkeiten zwischen den beiden Bruchpunktregionen (Abb 24).

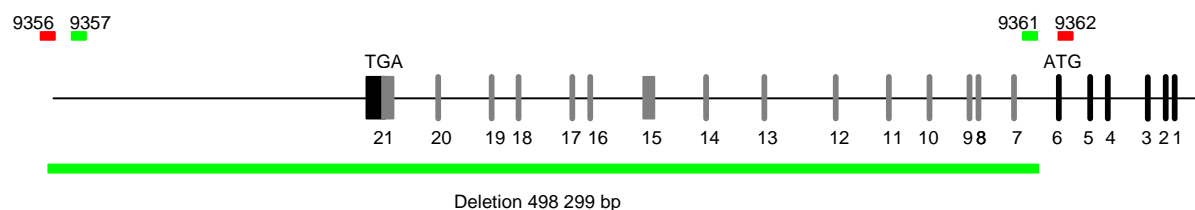


Abb. 24 Schematische Darstellung des Gens *FOXP1*. Kodierende Exons sind durch nummerierte graue Kästchen dargestellt, nicht-kodierende Exons als schwarze Kästchen. Die Primer für die quantitative PCR sind über der genomischen Darstellung als nummerierte rote (normale Dosis in der quantitativen PCR) und grüne (deletierte Proben) Balken abgebildet. Die deletierte Region ist unter dem Schema in grün markiert.

FOXP1 gehört zur Unterfamilie P der „Forkhead box transcriptional repressor“-Familie, die im menschlichen Genom aus vier Mitgliedern besteht. Es wurde vermutet, dass *FOXP1* als

Tumorsuppressor fungiert, aber *FOXP1* ist hochexprimiert in diffusen B-Zelllymphomen mit schlechter Prognose (Banham et al. 2001; Banham et al. 2005). Heterozygote Mutationen im Homolog *FOXP2* sind die Ursache einer verbalen Dyspraxie (Feuk et al. 2006; Lai et al. 2001). *FOXP1* und *FOXP2* sind in mehreren Hirnstrukturen gemeinsam exprimiert, was darauf hinweisen könnte, dass auch Mutationen in *FOXP1* zu Störungen in der Sprachentwicklung führen könnten (Teramitsu et al. 2004). Im Mausmodell entwickeln Mäuse mit homozygot deletiertem *Foxp1* schwere Herzfehler und sterben am Tag E14.5, während heterozygot deletierte Mäuse keinen pathologischen Phänotyp aufweisen (Wang et al. 2004).

Da bei dem Patienten 29945 nur die maternale DNA zur Verfügung stand, konnte nicht geklärt werden konnte, ob die Deletion *de novo* entstanden ist. Es kann also nur vermutet werden, dass die Deletion bei dem Patienten zu den vorliegenden Symptomen geführt hat.

Bei diesem Patienten wurde auch ein Mutationsscreening bei den oben genannten 415 Kindern mit mentaler Retardierung durchgeführt. Es wurden 1 nicht-kodierende Variante und 5 Varianten, die zu einem Austausch in der Aminosäuresequenz führten, in 6 DNAs gefunden (Tab. 7). Darunter sind 3 Missense-Varianten, die in 660 Kontrollen nicht gefunden wurden. Es konnte nicht überprüft werden, ob diese Varianten *de novo* entstanden sind, da in den meisten Fällen die DNA der Eltern nicht zur Verfügung standen.

Tabelle 7 Variationen im Gen *FOXP1*

Tabelle Variationen in <i>FOXP1</i>							
Variationen und Sequenzänderungen	ID	Exon/Intron	Kontrollindividuen				
			Genotyp 11		Genotyp 12	Genotyp 22	
Variationen mit Aminosäureaustausch							
c.13T>C	p.S5L	27734*	5	TT: 674	TC: 2	CC: 0	
c.643C>G	p.P215R	A115	10	CC: 668	CG: 2	GG: 0	
c.781T>C	p.S261P	28634	11	TT: 676	TC: 0	CC: 0	
c.1709A>G	p.N570S	A50	19	AA: 666	AG: 0	GG: 0	
c.1790A>C	p.N597T	29949	20	AA: 676	AG: 0	GG: 0	
stille Variationen							
c.664+6C>T	B146	Intron 10		CC: 666	CT: 0	TT: 0	

* in einem Elternteil wurde die gleiche Mutation gefunden

2.7 CNVs in Patienten mit Translokationen

Neben den Patienten mit mentaler Retardierung, die in der Routinediagnostik keine Auffälligkeiten zeigten, wurden noch 5 Patienten mit *de novo* Translokationen untersucht. Die Translokationen schienen, bis auf eine Ausnahme bei dem Patienten 27524, nach der GTG-Bänderung balanciert zu sein. Da diese Patienten trotz der balancierten Translokation eine Entwicklungsverzögerung zeigten, bestand der Verdacht, dass CNVs an den Bruchstellen vorhanden sein könnten. In 4 von 5 Patienten zeigten sich in der Arrayuntersuchung *de novo* CNVs an den Bruchpunkten (Tabelle 8). Die Größe der Deletionen variierte von 1.1 Mb bis 14.3 Mb, mit 3 bis 63 deletierten Genen. Die Angabe der Chromosomenbanden als Position der Translokationen wurde in fast allen Patienten nach der Auswertung der 100K-Arrays korrigiert.

Tabelle 8 Deletionen an Bruchpunkten in Patienten mit Translokationen

Tabelle Translokationen							
ID	GTG-Bänderung (500-550 Banden)	Chromosom	Start dbSNP rs ID	Ende dbSNP rs ID	Länge (Mb)	Anzahl der SNPs	Anzahl der Gene
31922	46,XY,t(12;13) (q22;q32)	13(q33.2-q33.3)	2149144	9301245	1.8	146	4
28526	46,XY,t(1;10) (p13.1;p13)	10(p14)	2259442	1243963	1.1	57	8
27524	46,XY,t(10;18),del(10q21.2-q11.2)	10(q21.1-q21.3)	6481257	10509188	6.3	320	41
28181	46,XY,t(4;6) (q28.3q31.1;q23.1q22.2)	4(q28.3-q31.1)	10519357	6841039	3.9	186	5
29200	46,XY,t(1;7) (q22;p13)	6(q16.1-q21)	9320518	6568702	14.3	625	85

Im Folgenden werden die identifizierten CNVs im Einzelnen beschrieben.

Patient 28181 hat eine *de novo* interstitielle Translokation zwischen den Chromosomen 4 und 6. Die Analyse der Daten zeigte Deletionen an zwei der vier Bruchpunkte der interstitiellen Translokation. Die Deletion auf Chromosom 4q28.3-q31.1 ist 3.9 Mb groß und beinhaltet fünf Gene. Die Deletion auf dem zweiten Chromosom, 6q16.1-q21, ist 14.3 Mb groß und beinhaltet 85 Gene.

Der Patient 28526 hat eine Entwicklungsverzögerung und eine sprachliche Retardierung. Die *de novo* aufgetretene terminale Translokation zwischen den Chromosomen 1 und 10 zeigte in der hochauflösenden GBG-Bänderung keine Auffälligkeiten an den Bruchpunkten. Mit den 100K-Arrays konnte an der Bruchstelle auf dem Chromosom 10 eine Deletion von 1.1 Mb identifiziert werden, die 8

Gene enthält. 3 der Gene sind Transkriptionsfaktoren (*SFMBT2*, *TAF3*, *GATA3*), 1 Gen (*KIN*) besitzt eine DNA-Bindestelle und 1 Gen (*ATP5C1*) ist an der ATP-Synthese beteiligt.

Der Patient 31922 hat neben einer Entwicklungsverzögerung und einer sprachlichen Retardierung einen Atriumseptumdefekt Typ 2. Es wurde eine *de novo* balancierte reziproke Translokation zwischen den Chromosomen 12 und 13 bei einer Amniozentese festgestellt. Die Analyse der Daten zeigte in diesem Patienten eine Deletion von 1.8 Mb auf dem Chromosom 13, die 4 Genen enthält. Von den vier Genen ist ein Gen, *EFNB2*, an der Entwicklung des Nervensystems beteiligt.

Beim Patienten 27524 wurde in der CT-Bänderung eine terminale *de novo* Translokation zwischen den Chromosomen 10 und 18 festgestellt. Dabei zeigte sich eine Deletion auf dem Chromosom 10 zwischen den Banden 10q21.2 bis 10q11.2 (Karyotypformel: 46,XY,t(10;18) (10pter→10q11.2::18q11.2→18qter; 18pter→18p11.2::10q21.2→10qter)). Die Größe der bereits bekannten Deletion auf Chromosom 10 konnte genauer, auf 6.3 Mb und 41 Genen eingegrenzt werden. Und auf dem Chromosom 18 wurde eine zweite Deletion gefunden, in der 8 Gene liegen und die 5.2 Mb groß ist.

Bei dem Patienten 29200 wurde zwischen den Chromosomen 1 und 7 eine *de novo* Translokation gefunden, die lichtmikroskopisch balanciert schien. Der Junge nicht symptomfrei und hat neben sprachlicher und psychomotorischer Entwicklungsverzögerung auch einen Herzfehler. Auch mit den 100K-Arrays wurden keine CNVs entdeckt.

D Diskussion

1 Kopplungsanalysen

Die Kopplungsanalyse, die bei vielen monogenen Erbkrankheiten erfolgreich angewendet wurde, wurde in der Vergangenheit meist mit Mikrosatellitenmarkern durchgeführt. Mikrosatelliten sind relativ gleichmäßig über das Genom verteilt und machen im menschlichen Genom etwa 60 Mb (ca. 2% des Genoms) aus (Strachan et al. 2004). Mikrosatellitenmarker haben den großen Vorteil, dass sie hoch polymorph und daher sehr informativ sind.

In der vorliegenden Arbeit wurden Kopplungsanalysen mit SNP-Arrays durchgeführt. Diese Methode war zu Beginn der Arbeit neu und wurde im Labor etabliert. SNPs kommen häufiger vor als Mikrosatelliten, sind aber, da sie nur biallelisch sind, weniger informativ. Durch den Informationsgehalt wird die genetische Information geschätzt, die mit den verwendeten Markern erhalten werden kann, im Vergleich zu einem hypothetisch unendlich dichten Set von Markern. Dabei ist der Informationsgehalt abhängig von der Dichte der Marker, dem Heterozygotiewert, der Familienstruktur und der Verfügbarkeit der DNAs. In einer Simulation von Kruglyak wurde der Informationsgehalt von Mikrosatellitenmarkern und SNPs unter verschiedenen Bedingungen verglichen und festgestellt, dass die Dichte eines Markersets wichtiger für den Informationsgehalt ist, als die Allelfrequenz der Marker (Kruglyak 1997). Um den gleichen Informationsgehalt wie mit Mikrosatelliten zu erhalten, muss daher eine größere Anzahl an SNPs untersucht werden. Inzwischen haben sich SNP-Arrays gegenüber den Mikrosatellitenmarkern durchgesetzt. Bei den üblichen Sets von 300-400 Mikrosatellitenmarker liegt der durchschnittliche Abstand zwischen zwei Markern bei ungefähr 10 cM, während beim 10K-SNP-Array der Abstand zwischen den Markern nur bei 0.34 cM liegt. Durch die höhere Dichte an Markern kann der niedrigere Informationsgehalt von SNPs kompensiert werden, wobei die Allelfrequenz nur einen geringen Effekt auf den Informationsgehalt hat (Kruglyak 1997). Ein weiterer Vorteil der SNP-Arrays liegt darin, dass damit wesentlich schneller Genotypen generiert werden, als es mit Mikrosatelliten möglich ist. Dabei ist die Fehlerrate, die beim Genotypisieren entsteht wesentlich geringer als mit Mikrosatelliten. Bereits eine Fehlerrate von 1% kann den beobachteten LOD Score um 50% reduzieren (Abecasis et al. 2001). Mit den SNP-Arrays sind auch ungefähr 10-mal niedrigere

Ausgangsmengen an DNA nötig (John et al. 2004). Zusammenfassend ist die Verwendung von SNP-Oligonukleotid-Mikroarrays mit einer Dichte von 6000 bis 10000 SNPs bei Kopplungsanalysen genauso effektiv, als ein Panel mit 300 bis 400 Mikrosatelliten (Evans et al. 2004, Kruglyak 1997, Matise et al. 2003, Middleton et al. 2004, John et al. 2004).

Die Programme zur Berechnung von Kopplungsanalysen setzen ein Kopplungsgleichgewicht zwischen den Markern voraus. Für die meisten Mikrosatellitenmarkersets trifft das zu. Da aber die Abstände zwischen den Markern bei SNP-Arrays geringer sind, kann zwischen eng benachbarten SNPs ein Kopplungsungleichgewicht (*Linkage Disequilibrium*, LD) vorliegen.

Die Definition des Kopplungsungleichgewichts besagt, dass sich zwei gekoppelte Genloci dann im Kopplungsungleichgewicht befinden, wenn ein Teil ihrer Allele, die die Mitglieder einer bestimmten Population produzieren, überzufällig häufig gemeinsam auftreten.

Marker, die in einem Kopplungsungleichgewicht liegen, können bei der Berechnung zu einer künstlichen Steigerung der LOD Score-Werte führen und die Ergebnisse verfälschen (Schaid et al. 2004, John et al. 2004). In 2 Studien wurden deshalb alle SNPs einer Region, die in einem Kopplungsungleichgewicht liegen, bis auf den SNP mit dem höchsten Informationsgehalt, aus der Analyse entfernt (John et al. 2004, Schaid et al. 2004). Die Ergebnisse unterscheiden sich in der Form, dass in der Studie von John et al. der Ausschluss von SNPs, die im Kopplungsungleichgewicht liegen, zu einem Informationsverlust führt, während Schaid et al. berichteten, dass nach dem Verwerfen dieser SNPs beinahe kein Unterschied im Informationsgehalt festgestellt werden konnte (John et al. 2004, Schaid et al. 2004).

Die in dieser Arbeit vorliegenden Kopplungsanalysen, wurden mit den 10K-Arrays von Affymetrix durchgeführt. Es sollte anhand von 3 verschiedenen Krankheitsbildern untersucht werden, inwieweit sich SNP-Arrays zur Kopplung von Krankheiten im Vergleich zu Mikrosatelliten eignen.

Die erste Kopplungsanalyse wurde mit 5 Familien mit insgesamt 7 betroffenen Familienmitgliedern durchgeführt, bei denen eine erbliche hypophosphatämische Rachitis mit Hyperkalziurie (HHRH) festgestellt wurde. Nach der Analyse der Genotypen, die mit den 10K-Arrays gewonnen wurden,

konnte keine Kandidatenregion identifiziert werden. Da eine zweite Plattform zur Verfügung stand, wurde in zwei der 5 Familien die Kopplungsanalyse mit dem Linkage IV Panel von Illumina wiederholt. Das Ergebnis zeigte einen signifikanten LOD Score am Ende des langen Armes von Chromosom 9. Um einen Hinweis auf die Ursache der unterschiedlichen Ergebnisse zu erhalten, wurden die zwei Array-Typen in dieser Region miteinander verglichen. Der Anfang der 2.7 Mb großen Kandidatenregion wurde durch die erreichten LOD-Score-Werte bei dem Marker *rs3132332* auf dem Illumina-Array festgelegt. Der Vergleich mit dem Affymetrix-Array zeigte, dass auf diesem Array lediglich ein SNP (*rs1105176*) in der Kandidatenregion liegt und deshalb nicht genügend Information für eine Kopplung zur Verfügung stand. Zur Bestätigung der Kopplung wurden noch 5 Mikrosatellitenmarker untersucht, die das Ergebnis verifizieren konnten. Die Sequenzierung des Kandidatengens *SLC34A3* zeigte Mutationen in allen betroffenen Individuen.

Die zweite Kopplungsanalyse wurde mit 3 Familien mit einer autosomal-rezessiven Hypophosphatämie (ARHP) durchgeführt, von denen 9 Familienmitglieder betroffen waren. Die Ergebnisse der 10K-Arrays konnten eine signifikante Kopplung am langen Arm des Chromosoms 4 identifizieren. Aus der Region, die 4.6 Mb groß war, wurden 5 Kandidatengene ausgewählt und sequenziert. In allen betroffenen Personen konnten Mutationen im Gen *DMP1* gefunden werden.

Die Identifizierung der Krankheitsgene in den Familien mit HHRH und ARHP und die anschließenden proteinbiochemischen Untersuchungen, die von Mitarbeitern der Arbeitsgruppe durchgeführt wurden ermöglichten weitere Einblicke in den Stoffwechselweg des Phosphats. Eine Schlüsselrolle in der Phosphathomöostase fällt dem Fibroblastenwachstumsfaktor FGF23 zu, ein sezerniertes Protein mit phosphaturischer Wirkung (Abb 25). Aktivierende Mutationen in FGF23 führen zur autosomal-dominanten Hypophosphatämie (ADHR) (The ADHR Consortium 2000, White et al. 2001, Benet-Pages et al. 2004). Die Familial Tumoral Calcinosis (FTC), bei der eine Hypophosphatämie vorliegt, wird durch inaktivierende Mutationen in FGF23 oder durch Mutationen in der UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 3 (GALNT3) verursacht (White et al. 2001, Shimada et al. 2001, Yamazaki et al. 2002, Jonsson et al. 2003). PHEX ist ein Membranprotein mit Endopeptidaseaktivität, das im Stoffwechselweg vor FGF23 liegt. Durch den Funktionsverlust von PHEX wird eine Inaktivierung von FGF23 verhindert (The HYP Consortium 1995). Bisher war es aber nicht möglich eine direkte Interaktion zwischen PHEX und FGF23 zu zeigen. Zirkulierendes FGF23

erniedrigt in den proximalen Tubuluszellen der Niere die Expression und Produktion von SLC34A1 und SLC34A3, zwei Natrium/Phosphat-Kotransportern. In dieser Arbeit konnte gezeigt werden, dass Mutationen in SLC34A3 zu der erblichen hypophosphatämischen Rachitis mit Hyperkalziurie (HHRH) führen. Weiterhin konnte die Ursache für die autosomal-rezessive Hypophosphatämie (ARHP) aufgeklärt werden, bei der erhöhte FGF23-Spiegel vorliegen. ARHP entsteht durch Mutationen in DMP1, ein Protein, das in Knockout-Mausmodellen zu verkürzten Knochen und Wirbeln und in älteren Tieren zu sklerotischen Veränderungen der Langknochen führt (Ye et al. 2005).

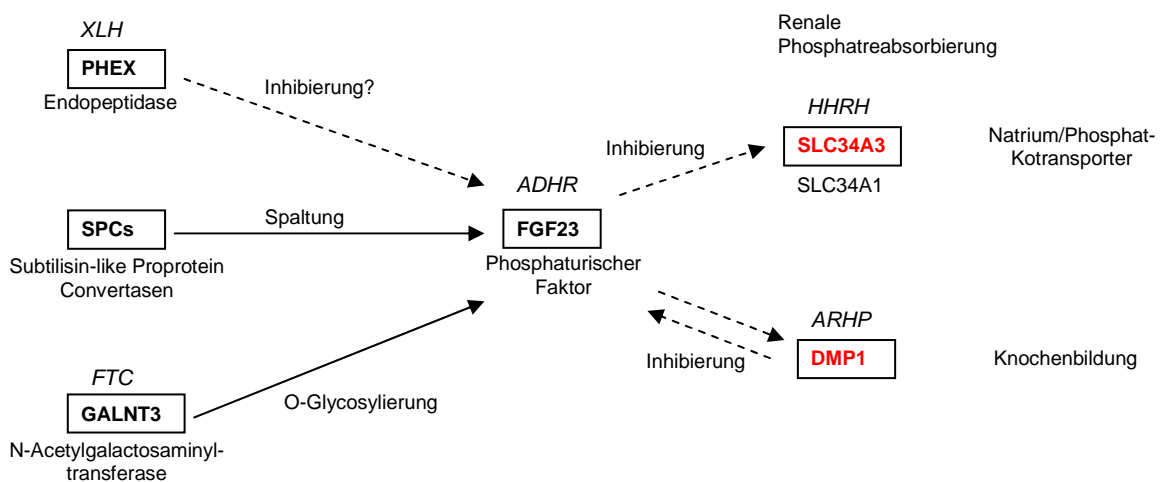


Abb. 25 Bisher bekannte Proteine im Phosphatstoffwechsel
 In der Abbildung sind die Proteine in Kästchen dargestellt, bereits bekannte Proteine in schwarz, die während der Arbeit neu identifizierte Proteine in rot. Die Abkürzungen der jeweiligen Krankheit sind kursiv geschrieben.

In der dritten Kopplungsanalyse wurde eine Familie mit einem spondylo-oculärem Syndrom (SOS) untersucht. Die Analyse der Daten, die aus den 10K-Arrays gewonnen wurden, zeigten eine signifikante Kopplung auf dem langen Arm des Chromosoms 17 an. Die 11 Mb große Region wird durch 7 SNPs angezeigt, die zum Teil weit auseinander liegen. Die Feinkartierung mit 11 Mikrosatellitenmarkern konnte die Region auf 5.3 Mb eingrenzen. Es wurden insgesamt 26 Kandidatengene ausgewählt und sequenziert. In keinem der Gene konnten Mutationen gefunden werden. Weitere Schritte zur Identifizierung der krankheitsverursachenden Mutation dieser Familie könnte die Sequenzierung weiterer Gene oder die Eingrenzung durch Kopplungsanalysen in weiteren Familien mit SOS sein. Im Rahmen dieser Arbeit konnte die Ursache für die Entstehung von SOS nicht geklärt werden.

Die Ergebnisse der Arbeit zeigen, dass SNP-Arrays zur Identifizierung von Kandidatenregionen bei Kopplungsanalysen geeignet sind, da sie schnell Genotypen erzeugen können, eine geringe Ausgangsmenge an DNA benötigen und die Möglichkeit des Hochdurchsatzes bieten. Es konnten aber auch einige Beeinträchtigungen festgestellt werden. In den Familien mit HHRH konnte mit den 10K-Arrays von Affymetrix keine signifikante Kopplung festgestellt werden, da in dieser Region die Dichte an SNPs nicht ausreichend war. In der Familie mit SOS konnte zwar eine Kandidatenregion identifiziert werden, aber die Anzahl an SNPs in dieser Region ist so gering, dass eine Eingrenzung der Region mit Mikrosatellitenmarkern möglich war. So müssen zum einen die Verteilung der SNPs auf dem jeweiligen Array und zum anderen die sich eventuell daraus ergebenden Lücken mit in Betracht gezogen werden. Eine Möglichkeit solche Lücken zu schließen, ist eine Erhöhung der Markerdichte auf dem Array. Dabei muss aber beachtet werden, dass sich die Marker nicht in einem Kopplungsungleichgewicht befinden und es dadurch zu einer Inflation der LOD Score-Werte kommt. Es besteht z.B. die Möglichkeit, statt der 10K-Arrays die 100K-Arrays zu verwenden, die eine 10-mal höhere Markerdichte haben, und dann gezielt SNPs auszuwählen, die für Kopplungsanalysen geeignet sind.

2 Kopienzahlvariationen

Die genetischen Ursachen von mentaler Retardierung sind heterogen und mit den heutzutage üblichen Diagnostikmethoden meist nicht zu klären. Eine Möglichkeit, um kleine Deletionen oder Duplikationen zu entdecken, die einen starken Einfluss auf seltene Phänotypen haben können, bietet die Hybridisierung von DNA-Proben auf SNP-Arrays. Diese Technologie, die ursprünglich entwickelt wurde, um SNPs für genomweite Assoziationsstudien zu identifizieren und zu untersuchen, eignet sich auch für die Detektierung von submikroskopischen CNVs. In dieser Studie wurden Arrays verwendet, die das gesamte Genom mit zweimal 50000 SNPs abdecken, um 71 Patienten mit mentaler Retardierung zu untersuchen.

2.1 Unterscheidung zwischen Polymorphismen und pathogenen CNVs

Die Suche nach pathogenen CNVs wird erschwert durch die Tatsache, dass jedes Individuum CNVs von einer Gesamtlänge von ca. 5 Mb hat (Redon et al. 2006; de Vries et al. 2005; lafrate et al. 2004; Sebat et al. 2004, Conrad et al. 2006; McCarroll et al. 2006; Hinds et al. 2006; Sharp et al. 2005; Tuzun et al. 2005). Je größer solche Polymorphismen sind, desto seltener kommen sie in der Population vor (Conrad et al. 2006). Es ist daher wichtig, Polymorphismen von pathogenen CNVs unterscheiden zu können.

In der vorliegenden Arbeit wurden folgende Kriterien für die Auswahl der pathogenen CNVs gesetzt:

- 1) Durch Einbeziehung der parental DNA untersucht, ob eine CNV *de novo* entstanden ist und damit möglicherweise pathogen sein könnte.
- 2) Anhand von Datenbanken und Literatur wurden die CNVs mit bekannten Syndromen verglichen.
- 3) Bereits bekannte Polymorphismen wurden durch einen Vergleich mit der DGV ausgeschlossen.
- 4) CNVs, in denen keine Gene lagen, wurden nicht weiter untersucht.

Ein wichtiges Kriterium zur Unterscheidung, ob eine CNV pathogen ist, ist der Nachweis, dass die CNV *de novo* entstanden ist. Da aber die Frequenz, mit der *de novo* CNVs in Neugeborenen vorkommen, auf 1:8 bzw. 1:50 geschätzt wird, ist es offensichtlich, dass nicht von allen angenommen werden kann, dass sie pathogen sind (van Ommen 2005). Im Gegensatz dazu ist es auch bekannt, dass eine vererbte CNVs bei einem Elternteil unauffällig sein kann und es bei dem Kind zur pathogenen Ausprägung kommt. Diese CNVs stellen ein Risiko dar, pathogen zu werden, durch Variationen in der Sequenz des zweiten Allels oder durch den genetischen Hintergrund. Datenbanken, die Informationen über CNVs von gesunden und kranken Individuen erfassen, werden immer unerlässlicher werden, um die Risiken richtig einschätzen zu können.

2.2 Pathogene CNVs

Mit dem 100K-Array von Affymetrix war es möglich, 11 Deletionen oder Duplikationen zu identifizieren, die als Ursache für die mentale Retardierung angesehen werden können. Diese CNVs sind entweder *de novo* entstanden oder sie entsprechen bereits bekannten Deletions-/Duplikations-Syndromen, die mit bestimmten Krankheitsbildern assoziiert sind. Die Detektionsrate von pathogenen CNVs in dieser Studie lag bei 16%. Eine von Friedman et al. durchgeführte Studie an mental retardierten Kindern, die mit den 100K-Arrays untersucht wurden, brachte bei 11 von 100 Kindern pathogene CNVs hervor (Friedmann et al., 2006). Eine weitere Studie auf diesem Gebiet wurde von de Vries und seinen Kollegen mit array-CGH durchgeführt (de Vries et al., 2005). Auch hier war der Prozentsatz von 10% der entdeckten pathogenen CNVs vergleichbar mit der vorliegenden Studie.

Pathologische CNVs verursachen die jeweilige Erkrankung durch Unterschiede in der Genexpression. Transkriptionsfaktoren sind die Gruppe von Genen, die am empfindlichsten auf Dosisunterschiede reagieren (Seidman et al. 2002). Alle CNVs in dieser Studie, mit Ausnahme der Deletionen in 2q23.1, 17q21.31 und der Duplikation in Xp22.31, enthalten mindestens ein Gen, das in den Transkriptionsprozess verwickelt ist. Insgesamt sind 189 Gene in den CNVs deletiert oder dupliziert, davon haben 129 einen Eintrag im *Gene Ontology Consortium* und 16 Gene sind direkt als Transkriptionsfaktoren bezeichnet. Dieser Anteil an Transkriptionsfaktoren ist aber nicht höher als im Genomdurchschnitt.

2.2.1 CNVs, die von *low-copy repeats* (LCRs) flankiert werden

CNVs entstehen durch nicht-allelische homologe Rekombination zwischen annähernd identischen, duplizierten Sequenzen, die die CNV flankieren (Stankiewicz et al. 2002; Lupski 2004; Lupski et al. 2006). Diese segmentalen Duplikationen, auch *low-copy repeats* (LCRs) genannt, sind Sequenzen im Genom mit über 90% Sequenzidentität und einer Größe von über 1 kb (Bailey et al. 2002; Cheung et al. 2003). Ungefähr 5% des menschlichen Genoms bestehen aus LCRs (Bailey et al. 2001; Bailey et al. 2002). In der vorliegenden Studie waren fünf CNVs (45%) von LCRs flankiert. Aber nicht nur LCRs markieren Regionen, in denen CNVs gehäuft vorkommen, auch andere hoch repetitive Sequenzen (*Alu*-Elemente, LINEs und SINEs) können eine Rekombination begünstigen (Deininger et al. 1999;

Shaw et al. 2004). Es wird auch vermutet, dass die Entstehung und Struktur von LCRs mit *Alu*-Sequenzen assoziiert ist. *Alu*-Sequenzen wurden an den Bruchpunkten von Gene oder Pseudogenen in LCRs auf dem Chromosom 22q11 gefunden (Babcock et al. 2003).

In zwei Fällen entsprachen die CNVs schon beschriebenen Deletions- oder Duplikationssyndromen. Von dem Patienten 29922, bei dem das 8p23.1-Deletionssyndrom vorliegt, wurden drei unabhängige Karyogramme angefertigt, die alle einen Verdacht auf Strukturänderungen im kurzen Arm des Chromosoms 8 zeigten. Es wurde daraufhin eine FISH-Analyse der entsprechenden Region durchgeführt, die aber keine strukturelle Auffälligkeiten in Chromosom 8p23 oder der Telomerregion finden konnte. Die Untersuchung des Patienten mit den SNP-Arrays konnte eine Deletion in Chromosom 8p23.1 nachweisen. Der zweite Patient (ID 27737) ist vom 17p11.2 Duplikationssyndrom betroffen, das klinisch nicht leicht zu diagnostizieren ist.

Bei beiden Patienten wurden in der Differentialdiagnose bereits Syndrome in Betracht gezogen, die aber mit Hilfe der vorliegenden Symptomatik und der diagnostischen Verfahren in der Routineuntersuchung nicht verifiziert werden konnten. Die Untersuchung des Genoms mit den 100K-Arrays hilft bei der Verifizierung oder Widerlegung von vermuteten Syndromen weiter, wenn in der Routinediagnostik keine klare Aussage getroffen werden kann.

Bei der Untersuchung des Patienten 30428 wurde eine Deletion auf dem Chromosom 17q21.3 festgestellt. Im Verlauf dieser Arbeit wurden von drei Arbeitsgruppen 10 Patienten mit der gleichen Deletion (Sharp et al., 2006, Koolen et al., 2006, Shaw-Smith et al., 2006) gefunden. Die Genotyp-Phänotyp-Beziehung aller Patienten führte zur Definierung des 17q21.3- Mikrodeletionssyndrom. Die Deletion liegt in einem Inversionspolymorphismus von ca. 900 kb, der vor ca. 1 Million Jahren entstanden ist. Die beiden Haplotypen (H1 und H2) lassen sich anhand von SNP-Markern unterscheiden. Liegt der H2-Haplotyp vor, kommt es bisher in allen Fällen zu einer nicht-allelischen homologen Rekombination, da der H2-Haplotyp einen direkt orientierten LCR hat. Auch der Patient 30428 hat die Deletion auf dem paternalen H2-Allel und bestätigt so den Mechanismus der CNV.

Auf den Chromosomen 15q25.2 und Xp22.31 wurden CNVs identifiziert, die möglicherweise neue Deletions- bzw. Duplikationssyndrome sein könnten. Da beide CNVs von direkt orientierten LCRs flankiert sind, ist die Wahrscheinlichkeit groß, dass diese CNVs häufiger vorkommen. Im Fall des

Patienten 29836 wurde die Duplikation von der phänotypisch gesunden Mutter auf den Sohn vererbt. Untersuchungen zur X-Inaktivierung zeigten, dass bei der Mutter das Allel mit der Duplikation nicht-zufällig inaktiviert ist. Daher fungiert die Mutter nur als Überträgerin der Duplikation, ohne selbst betroffen zu sein. Über weitere Fälle von Duplikationen auf Chromosom Xp22.31, assoziiert mit mentaler Retardierung und autistischem Verhalten, wurde inzwischen berichtet (Horn et al. 2007). Im Gegensatz dazu wurden bisher keine Patienten mit einer Deletion auf Chromosom 15q25.2 beschrieben.

2.2.2 CNVs, ohne flankierende LCRs

In insgesamt sieben Patienten traten CNVs auf, die nicht mit LCRs flankiert waren. Die Entstehung dieser CNVs lässt sich wahrscheinlich auf Doppelstrangbrüche oder Repeats zurückführen.

Es gibt die Möglichkeit, dass repetitive Sequenzen, wie LINEs (*long interspersed nuclear elements*), SINEs (*short interspersed nuclear elements*), LTRs (*long tandem repeats*) oder auch AT-Repeats die Bildung von CNVs begünstigen. Tandemrepeats sind Sequenzen, die häufig wiederholt und über das gesamte Genom verteilt sind. Zu der Familie der SINE-Sequenzen gehört die Familie der *Alu*-Repeats, die über 10% des menschlichen Genoms ausmachen (Batzer et al. 2002). Untersuchungen von Bruchpunkten des *Dystrophin*-Gens (*DMD*), das assoziiert ist mit der Duchenne'schen Muskeldystrophie, zeigten eine Verteilung der Bruchpunkte über das gesamte Gen. Bei 33% lag der Bruchpunkte in *Alu*-Sequenzen oder LTRs (*long tandem repeats*) (Nobile et al. 2002). AT-reiche Repeats können zur Bildung von Haarnadelstrukturen führen, die wiederum einen Doppelstrangbruch induzieren können (Shaw et al. 2004).

Der Mechanismus, der nach Doppelstrangbrüchen zur Bildung von Deletion oder Duplikationen führt, ist die nicht-homologe Endverknüpfung (NHEJ). Bei der NHEJ werden zwei DNA-Enden direkt miteinander verknüpft, die normalerweise keine oder nur sehr kurze Homologien zeigen. Für die Verknüpfung wird eine Reihe von Proteinen benötigt, die Doppelstrangbrüche in der DNA wieder zusammenfügen. Dabei kann es zu Umstrukturierung und damit zu Deletionen, Insertionen und Duplikationen der betroffenen Chromosomen kommen (Tsukamoto et al. 1998; Lupski et al. 2005; Sonoda et al. 2006; Lieber et al. 2006).

2.2.3 Bestätigung der Bruchpunkte

Die Größe der Deletionen, die aus den Ergebnissen der 100K-Arrays gewonnen wurden, stimmte nur ungefähr mit den tatsächlichen Größen überein. In 6 Patienten mit Deletionen wurde versucht, durch quantitative PCR die Bruchpunkte exakter zu bestimmen. In drei Patienten konnte ein Junction Fragment über das deletierte Allel amplifiziert werden und so die exakte Länge der Deletion mit den Bruchpunkten definiert werden. In den anderen drei Deletionen konnte die mögliche Bruchpunktregion nur eingegrenzt werden. In der Tabelle 9 wurden die Ergebnisse der 100K-Arrays mit den Ergebnissen der quantitativen PCR verglichen.

Tabelle 9 Vergleich der Größenangaben von Deletionen zwischen 100K-Arrays und quantitativer PCR

ID	Größe 100K (Mb)	Größe qPCR (Mb)	Position des proximalen Bruchpunkts	Position des distalen Bruchpunkts
27384	7.5	7.7	1845622500-184563500	192357500-192359000
30437	3.8	3.4	7193050-7193200	3819200-3826600
30375	2.7	2.9	12588557	9666569
28735	2.0	2.2	2194400-21980000	1-17450
29195	0.2	0.21	148752797	148964667
29945	0.44	0.5	70807767	71306066

2.2.4 CNVs mit nur einem deletierten Gen

In zwei Patienten konnten Deletionen identifiziert werden, die jeweils nur ein Gen enthalten. Bei einem Patienten ist es relativ wahrscheinlich, dass die Deletion die Ursache der Erkrankung ist. In dem zweiten Patienten kann nur vermutet werden, dass die Deletion pathogen sein könnte, da nicht nachgewiesen werden konnte, ob die Deletion *de novo* entstanden ist.

Der Patient 29195 hat eine Deletion auf Chromosom 2q23.1, die das Gen *MBD5* betrifft. Von dem Gen sind ein nicht-kodierendes und sieben kodierende Exons deletiert. In der Literatur finden sich zwei Berichte und ein Eintrag in der DECIPHER Datenbank, über Deletionen in dieser Region, die einige Gene, unter anderen auch *MBD5*, betreffen (de Vries et al., 2005, Koolen et al, 2004). Ein gemeinsames Symptom vom Patienten 29195 und von zwei Patienten aus den anderen Studien, sind epileptische Anfälle. Ein Mutationsscreening mit DNAs von 415 Kindern mit ungeklärter mentaler Retardierung brachte 4 Missense-Varianten, die in 660 Kontrollen nicht auftraten. Außer den

Missense-Varianten wurden keine weiteren Variationen, wie Deletionen oder Nonsense-Mutationen gefunden. Da die parentale DNA der 415 Kinder nicht zur Verfügung stand, konnte nicht überprüft werden, ob diese 4 Missense-Varianten *de novo* entstanden sind. Um bestätigen zu können, dass Mutationen im Gen *MBD5* pathogen sind, muss ein Screening mit mental retardierten Patienten durchgeführt werden, die an epileptischen Anfällen leiden und deren Eltern ebenfalls untersucht werden können.

Bei dem Patienten 29945 wurde ebenfalls eine Deletion gefunden, die nur ein Gen betrifft. Bei dieser CNV sind die kodierenden Exons 7 bis 21 des Gens *FOXP1* deletiert. *FOXP1* hat ein Homolog, *FOXP2*, mit dem es zusammen in mehreren Arealen des Gehirns koexprimiert wird. Es ist bekannt, dass heterozygote Mutationen in *FOXP2* zu verbaler Dyspraxie führen (Feuk et al. 2006). *FOXP1* und *FOXP2* werden im Gehirn koexprimiert (Teramitsu et al. 2004). Dies legt den Verdacht nahe, dass Mutationen in *FOXP1* auch zu Störungen der Sprachentwicklung führen könnten. Der Patient 29945 ist von einer Sprachretardierung betroffen, womit die Vermutung unterstützt wird, dass die Deletion von *FOXP1* pathogen sein könnte. Da aber der Vater nicht zur Untersuchung zur Verfügung stand, konnte nicht geklärt werden, ob die Deletion *de novo* entstanden ist. An den Bruchpunkten gibt es weder palindromische Sequenzen noch andere Ähnlichkeiten zwischen den Bruchpunktregionen. Die interstitielle Deletion ist vermutlich durch einen Doppelstrangbruch und nicht-homologe Endverknüpfung entstanden. Von der Deletion betroffen sind auch, mit Ausnahme einer Isoform, alle Isoformen des Gens *FOXP1*, die durch alternatives Splicen entstehen.

Ein Screening von Patienten mit mentaler Retardierung auf Veränderungen in *FOXP1* führte zwar zu der Identifizierung von 3 Missense-Varianten, brachte aber auch keinen endgültigen Beweis für die Pathogenität dieser Deletion, da wieder keine parentale DNA untersucht werden konnte. Ein weiteres Screening von Patienten, die zusätzlich zur mentalen Retardierung auch Probleme mit der Sprachentwicklung haben und ihren Eltern, kann neue Erkenntnisse bringen.

2.3 CNVs an Bruchpunkten von Translokationen

Es wurden 4 Patienten untersucht, die nach zytogenetischer Diagnostik scheinbar balancierte *de novo* Translokationen hatten, obwohl alle Patienten eine komplexe Symptomatik aufwiesen. Dazu wurde noch eine fünfte Translokation überprüft, die bereits in der Chromosomenanalyse eine Deletion an einem der Bruchpunkte zeigte. Nach der Auswertung der Daten der 100K-Arrays wurden in 4 Patienten Deletionen an einem oder mehreren Bruchpunkten gefunden. Nur im Fall des Patienten 29200 konnte keine CNV entdeckt werden. Die Ergebnisse zeigen, dass mit Hilfe der 100K-Arrays Bruchpunkte von Translokationen auf CNVs untersucht werden können, die in der Routinediagnostik der Zytogenetik balanciert erscheinen. Die Positionen der Deletionen, die mit den 100K-Arrays identifiziert wurden, weichen bei den meisten Patienten von den Positionen aus der Zytogenetik ab. Die Angabe der Lage von Translokationen ist mit Hilfe der FISH-Technik wesentlich ungenauer als es mit den SNP-Arrays möglich ist.

2.4 Eignung von SNP-Arrays zur Identifizierung von CNVs

In dieser Studie wurden CNVs mit einer Größe von 10 kb bis 7.5 Mb identifiziert und verifiziert. Mit mikroskopischen Techniken der Zytogenetik ist es nur möglich, CNVs mit einer Größe von mindestens 5-10 Mb zuverlässig nachzuweisen.

Die Qualität der Daten war in dieser Studie sehr variabel und abhängig von den experimentellen Umständen. Diese Einschränkung muss bei der Analyse der Daten berücksichtigt werden. Es wurde gezeigt, dass 20 aufeinander folgende SNPs eine CNV verlässlich anzeigen, während in dem Bereich von 5 bis 20 deletierten oder duplizierten SNPs, die CNVs eine Falsch-Positiv-Rate von 30% hatten.

Mit der Entwicklung von neuen Arrays und Protokollen, die experimentelle Fehler reduzieren und ein besseres Signal-Rausch-Verhältnis erreichen, wird sich die Auflösung weiter verbessern und kleinere CNVs bis zur Ebene einzelner Exons detektieren.

Literaturverzeichnis

- Abecasis, G. R., S. S. Cherny, et al. (2002). "Merlin--rapid analysis of dense genetic maps using sparse gene flow trees." Nat Genet **30**(1): 97-101.
- Allen, R. C., H. Y. Zoghbi, et al. (1992). "Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation." Am J Hum Genet **51**(6): 1229-39.
- Babcock, M., A. Pavlicek, et al. (2003). "Shuffling of genes within low-copy repeats on 22q11 (LCR22) by Alu-mediated recombination events during evolution." Genome Res **13**(12): 2519-32.
- Bailey, J. A., Z. Gu, et al. (2002). "Recent segmental duplications in the human genome." Science **297**(5583): 1003-7.
- Bailey, J. A., A. M. Yavor, et al. (2001). "Segmental duplications: organization and impact within the current human genome project assembly." Genome Res **11**(6): 1005-17.
- Banham, A. H., N. Beasley, et al. (2001). "The FOXP1 winged helix transcription factor is a novel candidate tumor suppressor gene on chromosome 3p." Cancer Res **61**(24): 8820-9.
- Banham, A. H., J. M. Connors, et al. (2005). "Expression of the FOXP1 transcription factor is strongly associated with inferior survival in patients with diffuse large B-cell lymphoma." Clin Cancer Res **11**(3): 1065-72.
- Batzler, M. A. and P. L. Deininger (2002). "Alu repeats and human genomic diversity." Nat Rev Genet **3**(5): 370-9.
- Benet-Pages, A., B. Lorenz-Depiereux, et al. (2004). "FGF23 is processed by proprotein convertases but not by PHEX." Bone **35**(2): 455-62.
- Bergwitz, C., N. M. Roslin, et al. (2006). "SLC34A3 mutations in patients with hereditary hypophosphatemic rickets with hypercalciuria predict a key role for the sodium-phosphate cotransporter NaPi-IIc in maintaining phosphate homeostasis." Am J Hum Genet **78**(2): 179-92.
- Bignell, G. R., J. Huang, et al. (2004). "High-resolution analysis of DNA copy number using oligonucleotide microarrays." Genome Res **14**(2): 287-95.
- Buckland, P. R. (2003). "Polymorphically duplicated genes: their relevance to phenotypic variation in humans." Ann Med **35**(5): 308-15.
- Cheung, J., M. D. Wilson, et al. (2003). "Recent segmental and gene duplications in the mouse genome." Genome Biol **4**(8): R47.
- Conrad, D. F., T. D. Andrews, et al. (2006). "A high-resolution survey of deletion polymorphism in the human genome." Nat Genet **38**(1): 75-81.
- de Vries, B. B., R. Pfundt, et al. (2005). "Diagnostic genome profiling in mental retardation." Am J Hum Genet **77**(4): 606-16.

- de Vries, B. B., L. A. Severijnen, et al. (2003). "FMRP expression studies in blood and hair roots in a fragile X family with methylation mosaics." J Med Genet **40**(7): 535-9.
- de Vries, B. B., A. M. van den Ouweland, et al. (1997). "Screening and diagnosis for the fragile X syndrome among the mentally retarded: an epidemiological and psychological survey. Collaborative Fragile X Study Group." Am J Hum Genet **61**(3): 660-7.
- Deininger, P. L. and M. A. Batzer (1999). "Alu repeats and human disease." Mol Genet Metab **67**(3): 183-93.
- Devriendt, K., G. Matthijs, et al. (1999). "Delineation of the critical deletion region for congenital heart defects, on chromosome 8p23.1." Am J Hum Genet **64**(4): 1119-26.
- Eichler, E. E. and K. A. Frazer (2004). "The nature, pattern and function of human sequence variation." Genome Biol **5**(4): 318.
- Evans, D. M. and L. R. Cardon (2004). "Guidelines for genotyping in genomewide linkage studies: single-nucleotide-polymorphism maps versus microsatellite maps." Am J Hum Genet **75**(4): 687-92.
- Evans, W., H. C. Fung, et al. (2004). "The tau H2 haplotype is almost exclusively Caucasian in origin." Neurosci Lett **369**(3): 183-5.
- Faivre, L., N. Morichon-Delvallez, et al. (1998). "Prenatal diagnosis of an 8p23.1 deletion in a fetus with a diaphragmatic hernia and review of the literature." Prenat Diagn **18**(10): 1055-60.
- Feuk, L., A. R. Carson, et al. (2006). "Structural variation in the human genome." Nat Rev Genet **7**(2): 85-97.
- Feuk, L., A. Kalervo, et al. (2006). "Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia." Am J Hum Genet **79**(5): 965-72.
- Feuk, L., C. R. Marshall, et al. (2006). "Structural variants: changing the landscape of chromosomes and design of disease studies." Hum Mol Genet **15 Spec No 1**: R57-66.
- Fisher, L. W., D. A. Torchia, et al. (2001). "Flexible structures of SIBLING proteins, bone sialoprotein, and osteopontin." Biochem Biophys Res Commun **280**(2): 460-5.
- Flint, J., A. O. Wilkie, et al. (1995). "The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation." Nat Genet **9**(2): 132-40.
- Freeman, J. L., G. H. Perry, et al. (2006). "Copy number variation: new insights in genome diversity." Genome Res **16**(8): 949-61.
- Friedman, J. M., A. Baross, et al. (2006). "Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation." Am J Hum Genet **79**(3): 500-13.
- Fukami, M., S. Kirsch, et al. (2000). "A member of a gene family on Xp22.3, VCX-A, is deleted in patients with X-linked nonspecific mental retardation." Am J Hum Genet **67**(3): 563-73.
- Gazit, D., M. Tieder, et al. (1991). "Osteomalacia in hereditary hypophosphatemic rickets with hypercalciuria: a correlative clinical-histomorphometric study." J Clin Endocrinol Metab **72**(1): 229-35.

- Greenberg, B. G., R. W. Winters et al. (1960) "The normal range of serum inorganic phosphorus and its utility as a discriminant in the diagnosis of congenital hypophosphatemia." J Clin Endocrinol Metab **20**: 364-379.
- Gudbjartsson, D. F., K. Jonasson, et al. (2000). "Allegro, a new computer program for multipoint linkage analysis." Nat Genet **25**(1): 12-3.
- Hefferin, M. L. and A. E. Tomkinson (2005). "Mechanism of DNA double-strand break repair by non-homologous end joining." DNA Repair (Amst) **4**(6): 639-48.
- Hinds, D. A., A. P. Kloek, et al. (2006). "Common deletions and SNPs are in linkage disequilibrium in the human genome." Nat Genet **38**(1): 82-5.
- Horn, D., S. Spranger et al. (2007) "Microdeletions and microduplications affecting the STS gene at Xp22.31 are associated with a distinct phenotypic spectrum." Medizinische Genetik **19**: 62.
- Huang, J., W. Wei, et al. (2004). "Whole genome DNA copy number changes identified by high density oligonucleotide arrays." Hum Genomics **1**(4): 287-99.
- Iafrate, A. J., L. Feuk, et al. (2004). "Detection of large-scale variation in the human genome." Nat Genet **36**(9): 949-51.
- Inoue, K. and J. R. Lupski (2002). "Molecular mechanisms for genomic disorders." Annu Rev Genomics Hum Genet **3**: 199-242.
- John, S., N. Shephard, et al. (2004). "Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites." Am J Hum Genet **75**(1): 54-64.
- Jonsson, K. B., R. Zahradnik, et al. (2003). "Fibroblast growth factor 23 in oncogenic osteomalacia and X-linked hypophosphatemia." N Engl J Med **348**(17): 1656-63.
- Knight, S. J., R. Regan, et al. (1999). "Subtle chromosomal rearrangements in children with unexplained mental retardation." Lancet **354**(9191): 1676-81.
- Koolen, D. A., W. M. Nillesen, et al. (2004). "Screening for subtelomeric rearrangements in 210 patients with unexplained mental retardation using multiplex ligation dependent probe amplification (MLPA)." J Med Genet **41**(12): 892-9.
- Koolen, D. A., L. E. Vissers, et al. (2004). "A novel microdeletion, del(2)(q22.3q23.3) in a mentally retarded patient, detected by array-based comparative genomic hybridization." Clin Genet **65**(5): 429-32.
- Koolen, D. A., L. E. Vissers, et al. (2006). "A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism." Nat Genet **38**(9): 999-1001.
- Kruglyak, L. (1997). "The use of a genetic map of biallelic markers in linkage studies." Nat Genet **17**(1): 21-4.
- Kruglyak, L., M. J. Daly, et al. (1996). "Parametric and nonparametric linkage analysis: a unified multipoint approach." Am J Hum Genet **58**(6): 1347-63.
- Lai, C. S., S. E. Fisher, et al. (2001). "A forkhead-domain gene is mutated in a severe speech and language disorder." Nature **413**(6855): 519-23.

-
- Lathrop, G. M. and J. M. Lalouel (1984). "Easy calculations of lod scores and genetic risks on small computers." Am J Hum Genet **36**(2): 460-5.
- Leonard, H. and X. Wen (2002). "The epidemiology of mental retardation: challenges and opportunities in the new millennium." Ment Retard Dev Disabil Res Rev **8**(3): 117-34.
- Lieber, M. R., K. Yu, et al. (2006). "Roles of nonhomologous DNA end joining, V(D)J recombination, and class switch recombination in chromosomal translocations." DNA Repair (Amst) **5**(9-10): 1234-45.
- Ling, Y., H. F. Rios, et al. (2005). "DMP1 depletion decreases bone mineralization in vivo: an FTIR imaging analysis." J Bone Miner Res **20**(12): 2169-77.
- Locke, D. P., A. J. Sharp, et al. (2006). "Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome." Am J Hum Genet **79**(2): 275-90.
- Lorenz-Depiereux, B., M. Bastepe, et al. (2006). "DMP1 mutations in autosomal recessive hypophosphatemia implicate a bone matrix protein in the regulation of phosphate homeostasis." Nat Genet **38**(11): 1248-50.
- Lorenz-Depiereux, B., A. Benet-Pages, et al. (2006). "Hereditary hypophosphatemic rickets with hypercalciuria is caused by mutations in the sodium-phosphate cotransporter gene SLC34A3." Am J Hum Genet **78**(2): 193-201.
- Lucito, R., J. Healy, et al. (2003). "Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation." Genome Res **13**(10): 2291-305.
- Lupski, J. R. (1998). "Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits." Trends Genet **14**(10): 417-22.
- Lupski, J. R. (2004). "Hotspots of homologous recombination in the human genome: not all homologous sequences are equal." Genome Biol **5**(10): 242.
- Lupski, J. R. (2006). "Genome structural variation and sporadic disease traits." Nat Genet **38**(9): 974-6.
- Matise, T. C., R. Sachidanandam, et al. (2003). "A 3.9-centimorgan-resolution human single-nucleotide polymorphism linkage map and screening set." Am J Hum Genet **73**(2): 271-84.
- McCarroll, S. A., T. N. Hadnott, et al. (2006). "Common deletion polymorphisms in the human genome." Nat Genet **38**(1): 86-92.
- Menten, B., N. Maas, et al. (2006). "Emerging patterns of cryptic chromosomal imbalance in patients with idiopathic mental retardation and multiple congenital anomalies: a new series of 140 patients and review of published reports." J Med Genet **43**(8): 625-33.
- Middleton, F. A., M. T. Pato, et al. (2004). "Genomewide linkage analysis of bipolar disorder by use of a high-density single-nucleotide-polymorphism (SNP) genotyping assay: a comparison with microsatellite marker assays and finding of significant linkage to chromosome 6q22." Am J Hum Genet **74**(5): 886-97.
- Miller, S. A., D. D. Dykes, et al. (1988). "A simple salting out procedure for extracting DNA from human nucleated cells." Nucleic Acids Res **16**(3): 1215.
- Murer, H., N. Hernando, et al. (2001). "Molecular aspects in the regulation of renal inorganic phosphate reabsorption: the type IIa sodium/inorganic phosphate co-transporter as the key player." Curr Opin Nephrol Hypertens **10**(5): 555-61.

- Nannya, Y., M. Sanada, et al. (2005). "A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays." Cancer Res **65**(14): 6071-9.
- Nguyen, D. Q., C. Webber, et al. (2006). "Bias of selection on human copy-number variants." PLoS Genet **2**(2): e20.
- Nobile, C., L. Toffolatti, et al. (2002). "Analysis of 22 deletion breakpoints in dystrophin intron 49." Hum Genet **110**(5): 418-21.
- O'Connell, J. R. and D. E. Weeks (1998). "PedCheck: a program for identification of genotype incompatibilities in linkage analysis." Am J Hum Genet **63**(1): 259-66.
- Ott, J. (1991). "Analysis of Human Genetic Linkage." 2nd Edition, . Baltimore, Johns Hopkins University Press.
- Potocki, L., W. Bi, et al. (2007). "Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype." Am J Hum Genet **80**(4): 633-49.
- Potocki, L., K. S. Chen, et al. (2000). "Molecular mechanism for duplication 17p11.2- the homologous recombination reciprocal of the Smith-Magenis microdeletion." Nat Genet **24**(1): 84-7.
- Qin, C., O. Baba, et al. (2004). "Post-translational modifications of sibling proteins and their roles in osteogenesis and dentinogenesis." Crit Rev Oral Biol Med **15**(3): 126-36.
- Redon, R., S. Ishikawa, et al. (2006). "Global variation in copy number in the human genome." Nature **444**(7118): 444-54.
- Reiter, L. T., P. J. Hastings, et al. (1998). "Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients." Am J Hum Genet **62**(5): 1023-33.
- Repping, S., S. K. van Daalen, et al. (2006). "High mutation rates have driven extensive structural polymorphism among human Y chromosomes." Nat Genet **38**(4): 463-7.
- Roeleveld, N., G. A. Zielhuis, et al. (1997). "The prevalence of mental retardation: a critical review of recent literature." Dev Med Child Neurol **39**(2): 125-32.
- Rudolph, G., P. Kalpadakis, et al. (2003). "Spondylo-ocular syndrome: a new entity with crystalline lens malformation, cataract, retinal detachment, osteoporosis, and platyspondyly." Am J Ophthalmol **135**(5): 681-7.
- Rueschendorf, F. and P. Nurnberg (2005). "ALOHOMORA: a tool for linkage analysis using 10K SNP array data." Bioinformatics **21**(9): 2123-5.
- Saiki, R. K., D. H. Gelfand et al. (1988) "Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase." Science **239**:487-491.
- Sanger, F., S. Nicklen, et al. (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A **74**(12): 5463-7.
- Schaid, D. J., J. C. Guenther, et al. (2004). "Comparison of microsatellites versus single-nucleotide polymorphisms in a genome linkage screen for prostate cancer-susceptibility Loci." Am J Hum Genet **75**(6): 948-65.

- Schmidt, H., G. Rudolph, et al. (2001). "Retinal detachment and cataract, facial dysmorphism, generalized osteoporosis, immobile spine and platyspondyly in a consanguinous kindred--a possible new syndrome." Clin Genet **59**(2): 99-105.
- Sebat, J., B. Lakshmi, et al. (2004). "Large-scale copy number polymorphism in the human genome." Science **305**(5683): 525-8.
- Seidman, J. G. and C. Seidman (2002). "Transcription factor haploinsufficiency: when half a loaf is not enough." J Clin Invest **109**(4): 451-5.
- Sewell, K. L. (1991). "Treatment of elderly-onset rheumatoid arthritis." J Am Geriatr Soc **39**(10): 1043-4.
- Sharp, A. J., S. Hansen, et al. (2006). "Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome." Nat Genet **38**(9): 1038-42.
- Sharp, A. J., D. P. Locke, et al. (2005). "Segmental duplications and copy-number variation in the human genome." Am J Hum Genet **77**(1): 78-88.
- Shaw, C. J. and J. R. Lupski (2004). "Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease." Hum Mol Genet **13 Spec No 1**: R57-64.
- Shaw, C. J. and J. R. Lupski (2005). "Non-recurrent 17p11.2 deletions are generated by homologous and non-homologous mechanisms." Hum Genet **116**(1-2): 1-7.
- Shaw-Smith, C., A. M. Pittman, et al. (2006). "Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability." Nat Genet **38**(9): 1032-7.
- Shaw-Smith, C., R. Redon, et al. (2004). "Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features." J Med Genet **41**(4): 241-8.
- Shimada, T., S. Mizutani, et al. (2001). "Cloning and characterization of FGF23 as a causative factor of tumor-induced osteomalacia." Proc Natl Acad Sci U S A **98**(11): 6500-5.
- Slater, H. R., D. K. Bailey, et al. (2005). "High-resolution identification of chromosomal abnormalities using oligonucleotide arrays containing 116,204 SNPs." Am J Hum Genet **77**(5): 709-26.
- Somer, H., A. Palotie, et al. (1988). "Osteoporosis-pseudoglioma syndrome: clinical, morphological, and biochemical studies." J Med Genet **25**(8): 543-9.
- Sonoda, E., H. Hohegger, et al. (2006). "Differential usage of non-homologous end-joining and homologous recombination in double strand break repair." DNA Repair (Amst) **5**(9-10): 1021-9.
- Speicher, M. R. and N. P. Carter (2005). "The new cytogenetics: blurring the boundaries with molecular biology." Nat Rev Genet **6**(10): 782-92.
- Spranger, J., A. Winterpacht, et al. (1997). "Kniest dysplasia: Dr. W. Kniest, his patient, the molecular defect." Am J Med Genet **69**(1): 79-84.
- Stankiewicz, P. and J. R. Lupski (2002). "Genome architecture, rearrangements and genomic disorders." Trends Genet **18**(2): 74-82.
- Stefansson, H., A. Helgason, et al. (2005). "A common inversion under selection in Europeans." Nat Genet **37**(2): 129-37.

- Stickler, G. B. and E. C. Burke (1965). "Studies in Nephrotic Syndrome. II. Color of Hair in Children with Nephrotic Syndrome." Mayo Clin Proc **40**: 391.
- Strachan, T., A. P. Read (2004). "Molekulare Humangenetik." 3. Auflage. Spektrum Akademischer Verlag, Heidelberg, Berlin, Oxford.
- Teramitsu, I., L. C. Kudo, et al. (2004). "Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction." J Neurosci **24**(13): 3152-63.
- Terwillinger, J. D., J. Ott (1994). "Handbook of Human Genetic Linkage." Baltimore, Johns Hopkins University Press.
- The ADHR Consortium (2000). "Autosomal dominant hypophosphataemic rickets is associated with mutations in FGF23." Nat Genet **26**(3): 345-8.
- The HYP Consortium (1995). "A gene (PEX) with homologies to endopeptidases is mutated in patients with X-linked hypophosphatemic rickets. The HYP Consortium." Nat Genet **11**(2): 130-6.
- Tieder, M., D. Modai, et al. (1985). "Hereditary hypophosphatemic rickets with hypercalciuria." N Engl J Med **312**(10): 611-7.
- Tieder, M., D. Modai, et al. (1987). "Idiopathic hypercalciuria and hereditary hypophosphatemic rickets. Two phenotypical expressions of a common genetic defect." N Engl J Med **316**(3): 125-9.
- Tieder, M., R. Samuel, et al. (1985). "Hypercalciuric rickets: metabolic studies and pathophysiological considerations." Nephron **39**(3): 194-200.
- Tsukamoto, K., N. Ito, et al. (1998). "Allelic loss on chromosome 1p is associated with progression and lymph node metastasis of primary breast carcinoma." Cancer **82**(2): 317-22.
- Tuzun, E., A. J. Sharp, et al. (2005). "Fine-scale structural variation of the human genome." Nat Genet **37**(7): 727-32.
- Tyson, C., C. Harvard, et al. (2005). "Submicroscopic deletions and duplications in individuals with intellectual disability detected by array-CGH." Am J Med Genet A **139**(3): 173-85.
- Van Esch, H., K. Hollanders, et al. (2005). "Deletion of VCX-A due to NAHR plays a major role in the occurrence of mental retardation in patients with X-linked ichthyosis." Hum Mol Genet **14**(13): 1795-803.
- van Karnebeek, C. D., M. C. Jansweijer, et al. (2005). "Diagnostic investigations in individuals with mental retardation: a systematic literature review of their usefulness." Eur J Hum Genet **13**(1): 6-25.
- van Ommen, G. J. (2005) "Frequency of new copy number variation in human." Nat Genet **37**: 333-334.
- Veltman, J. A., E. F. Schoenmakers, et al. (2002). "High-throughput analysis of subtelomeric chromosome rearrangements by use of array-based comparative genomic hybridization." Am J Hum Genet **70**(5): 1269-76.
- Vissers, L. E., B. B. de Vries, et al. (2003). "Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities." Am J Hum Genet **73**(6): 1261-70.
- Wang, B., J. Weidenfeld, et al. (2004). "Foxp1 regulates cardiac outflow tract, endocardial cushion morphogenesis and myocyte proliferation and maturation." Development **131**(18): 4477-87.

- White, K. E., G. Carn, et al. (2001). "Autosomal-dominant hypophosphatemic rickets (ADHR) mutations stabilize FGF-23." Kidney Int **60**(6): 2079-86.
- White, K. E., K. B. Jonsson, et al. (2001). "The autosomal dominant hypophosphatemic rickets (ADHR) gene is a secreted polypeptide overexpressed by tumors that cause phosphate wasting." J Clin Endocrinol Metab **86**(2): 497-500.
- Yamazaki, Y., R. Okazaki, et al. (2002). "Increased circulatory level of biologically active full-length FGF-23 in patients with hypophosphatemic rickets/osteomalacia." J Clin Endocrinol Metab **87**(11): 4957-60.
- Ye, L., Y. Mishina, et al. (2005). "Dmp1-deficient mice display severe defects in cartilage formation responsible for a chondrodysplasia-like phenotype." J Biol Chem **280**(7): 6197-203.
- Yeargin-Allsopp, M., C. C. Murphy, et al. (1997). "Reported biomedical causes and associated medical conditions for mental retardation among 10-year-old children, metropolitan Atlanta, 1985 to 1987." Dev Med Child Neurol **39**(3): 142-9.
- Zatyka, M., M. Priestley, et al. (2005). "Analysis of CRELD1 as a candidate 3p25 atrioventricular septal defect locus (AVSD2)." Clin Genet **67**(6): 526-8.
- Zhao, X., C. Li, et al. (2004). "An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays." Cancer Res **64**(9): 3060-71.

Tabelle I Primer		Rückwärtsprimer	Länge (bp)	Temperatur
Primer	Vorwärtsprimer			
Primer qPCR				
8920	5' - TGAAGGCTCTAAATCCCCAG - 3'	5' - AGCAACCGCTAAAACCCAG - 3'	126	60°C
8921	5' - GCGAGCCCGATTCTCTG - 3'	5' - ACCAATCCTCAGGTCCAGC - 3'	126	60°C
8922	5' - CTTCACTGGACCGAAACACC - 3'	5' - TTTATCCCGATTGCTTCTGC - 3'	135	60°C
8923	5' - AGCAGCCTCATGCTCTATGG - 3'	5' - TGAAGTGTCCAGTCCAAGGC - 3'	128	60°C
8924	5' - ATTTCCCTCTATGGAGCGTGG - 3'	5' - AATGGCTCCGATACACTTCC - 3'	127	60°C
8925	5' - AACTGGGAAGGAGGTATCCG - 3'	5' - AGGGAGCTCCAGCCAGC - 3'	133	60°C
8926	5' - GGACCACCCCTTCGGCTG - 3'	5' - GGCTGACGATGTTTCGAGG - 3'	124	60°C
8928	5' - TGGTGATTACTTTTGACAGTCTTTG - 3'	5' - GGAAAGAGTTGAGCTCCCG - 3'	129	60°C
8929	5' - CGAGTCGCTGACCAGTTACC - 3'	5' - ATCTGCTTGCCTGGCTGAC - 3'	126	60°C
8930	5' - GACATGTGCAGCCGTGTG - 3'	5' - ATGGGCTGTGTGTCAAGG - 3'	124	60°C
8932	5' - TTTTGGCCAGTTAGCC - 3'	5' - AATATGATGGTGGCTGGCTC - 3'	130	60°C
8933	5' - CTCATTGAGCCCTCATCTCACC - 3'	5' - CCTCCATGAAGGGCTAAAC - 3'	131	60°C
8934	5' - ATGCAAAATAATACCATCCAGAC - 3'	5' - GCCACTTGTGGGAAGTGC - 3'	137	60°C
8935	5' - CTTTTCATTAATGTGCAATTAACC - 3'	5' - TCTGCAAAAGCAACTGAAC TG - 3'	135	60°C
8936	5' - CCTCCTACTATATGAAAATTTTGGTCC - 3'	5' - GCTGCTTAAGACTGCAAAAGC - 3'	126	60°C
8937	5' - TCAACTATGACAGGGTGTCTGC - 3'	5' - CCCTTCAGAGGCAATAATAGG - 3'	121	60°C
8938	5' - AGATTTGGGCCCTCAAC - 3'	5' - AACAAACGAGGAAGAAATATTGG - 3'	133	60°C
8939	5' - GCAGCAGCAATATCCCTTTATG - 3'	5' - GCTGGTAGTCAAGTAAGGCTTTG - 3'	128	60°C
8940	5' - TGACTTAATATCAATTGAGGATCAC - 3'	5' - TTGAATTAAC TGAACCCCAATCTC - 3'	135	60°C
8941	5' - TCTGTCTGGATTCCTATTTGC - 3'	5' - TCAACACAAAGCTTGGTCTCC - 3'	130	60°C
8942	5' - GGAAGATGATTTGTCACTTTTCTTC - 3'	5' - GCACAGTCAACCAATTTTTCTC - 3'	132	60°C
8944	5' - AAAAGTGTGCAGAGCCAG - 3'	5' - GATGAAGCAAGCACATAGCG - 3'	130	60°C
8945	5' - ATTTGTCTTTGTCTCAATATGGG - 3'	5' - CCCTTAGGTTCTTCACTTCCC - 3'	130	60°C
8946	5' - TCACACCTTCTCTGTGGCAG - 3'	5' - GCAAAACAACCCCGAAAAC - 3'	127	60°C
8948	5' - AAGGAAGGAGGTCTTCCAGC - 3'	5' - CAAC TTTGCAGGTACCACAGG - 3'	118	60°C
9054	5' - CTTAACCTGATCACGGGGC - 3'	5' - AACAGTGGCCTCAACTCCTC - 3'	148	60°C
9055	5' - CTTCTAGGCTTGAAAATTGG - 3'	5' - TGATGAATTCGAGTTGCTG - 3'	130	60°C
9056	5' - GCCCACCTCATCTACCTG - 3'	5' - CGACGAGGGATTGTCTG - 3'	136	60°C
9057	5' - GTCTGCAACACACTGCAACC - 3'	5' - TGGTTTCGTGCCCTGTAGTAGG - 3'	154	60°C
9058	5' - TCTCTCAACTCACACGCC - 3'	5' - GAAACGGTGACCCGCTG - 3'	133	60°C
9059	5' - CTGGACATACACAGGCTCG - 3'	5' - CTGCAGAACTCTGTGGACTG - 3'	133	60°C
9060	5' - GAACATCAGCAGTTGCCCTC - 3'	5' - TCCCATGGAGATGCACATAG - 3'	152	60°C
9061	5' - CTCCCGAAGTGGGAGG - 3'	5' - AGAGGAGAGTTGCTGGAGCC - 3'	147	60°C
9062	5' - TAGATCACAGGGTCGGAAGG - 3'	5' - CAGGCAAGGAAGGCAG - 3'	131	60°C
9063	5' - CTACCTGCGGGTCTCTGG - 3'	5' - GCCCTGAAGCAGTCCCTC - 3'	141	60°C

9064	5' - TTGACTAACCTACGGCCACG - 3'	5' - TCCTAAGGAGAAGGGGAGGG - 3'	153	60°C
9065	5' - CCTGATGCTTGACTCTCCTC - 3'	5' - GACTTCCCTCTGCACCAAG - 3'	125	60°C
9066	5' - CTCGGATTGCTCAAGGACC - 3'	5' - ACTTTACGCTGGCAGGAGG - 3'	136	60°C
9067	5' - TCACAGAGCAGCTCCCATC - 3'	5' - GGCCACCAGGGAGATACAG - 3'	130	60°C
9068	5' - GATTGGCCAGAACCCCTGAG - 3'	5' - GCGCTGCATGGTCATATTTTC - 3'	125	60°C
9069	5' - TGTAGCAAAATGGTGGCAAG - 3'	5' - GAGACTATTGGCGATCGGAG - 3'	146	60°C
9070	5' - TCTCCTTGTTAATCTGAGCTCTTG - 3'	5' - GGTATGATGTCCATCCAGG - 3'	126	60°C
9073	5' - GGAACATCACCAACAGCC - 3'	5' - TTCAAACAAGGCAGAATACAGAC - 3'	133	60°C
9074	5' - TGGAGATAGATGCCATTTGC - 3'	5' - AGGGATCGATGTGCTAGGAG - 3'	133	60°C
9076	5' - TCAGTGGAGAAAGTGTGTGG - 3'	5' - ATCCAAGACCCAGAAAAGCC - 3'	141	60°C
9077	5' - TGGAGTTCACCTTTCTTGC - 3'	5' - ACTTGTGTAGCTGCCCAGG - 3'	120	60°C
9078	5' - TCAACCTGAACCAATTACGCC - 3'	5' - GGTGACCAGGGTGGTGTAG - 3'	140	60°C
9648	5' - AATAGCAGCTCAAATCCCAGTC - 3'	5' - CAGGAAGACGGGAACAAC - 3'	148	60°C
9649	5' - TGTGTTGTCACAGCCCTTGC - 3'	5' - CTAGAGATGGACCTCCACG - 3'	141	60°C
9650	5' - GTTGAAGTACGGCCGCTG - 3'	5' - CCTGAGAATGACTCCATGCTTC - 3'	149	60°C
9651	5' - GGACTCAGTGATCAAACCCC - 3'	5' - ATTGTTCTCAGAATGCACCC - 3'	130	60°C
9652	5' - CACGATAAATGTTTCTCTGTCC - 3'	5' - GCGTCCCTGGTATGGATG - 3'	109	60°C
9653	5' - ATTGCAACCAGAAAGCTGGAG - 3'	5' - TCCAGGTAAGCAGGTCCTC - 3'	137	60°C
9655	5' - GAAAGATCTCTGGGTGTCGG - 3'	5' - AATCAGTGATGGCCATGGAG - 3'	160	60°C
9656	5' - GTTCACTGCAGAGGAGTTG - 3'	5' - AATCTTGCCTGTTCTTTAGCG - 3'	143	60°C
9657	5' - CCCCTTTGGATTAAGTTGC - 3'	5' - ACTTCAGGGTCTGGGATGC - 3'	131	60°C
9820	5' - CAAGATGTGTCAGACATCCC - 3'	5' - CATGGTGGCGTATGAACCTGG - 3'	162	60°C
9821	5' - CAACCTTTTAAAACGCTTTGC - 3'	5' - TATAGCTGCACGAATCTCCG - 3'	119	60°C
9822	5' - GGTTCTTGAGTGATTAGTGTAGG - 3'	5' - CCTCAGTTGCGAATAATAAAAGGC - 3'	179	60°C

Primerwalking MBD5

9265	5' - TGCAAAACTGTGTTTATGATGTTAC - 3'	5' - AAGTCTGTGTTAAAAAGAAATTTGGTG - 3'	131	60°C
9266	5' - GCTATCCCTTAAAACTCTCACAGC - 3'	5' - TGCCTTTGGGAGCAATAGG - 3'	111	60°C
9267	5' - TGACTTTGCTTTCCAATAACCC - 3'	5' - CAAAACAATAACTGAAGAACTCATGG - 3'	113	60°C
9268	5' - TTTCACACGTGAACCAGGC - 3'	5' - GACCTGGGTAGAGCAGCATC - 3'	121	60°C
9269	5' - TTGCATATTTTATAGGCATAGATAGC - 3'	5' - TCCCTACTCAACACTAAAGAGCC - 3'	134	60°C
9270	5' - CAGGTTGTCATCACCACTGC - 3'	5' - GAATGACCGCACTGACAGC - 3'	133	60°C
9271	5' - AAAAGTAGGGGTTGATGCC - 3'	5' - AATATTTGTCAAAGGTGCCCC - 3'	140	60°C
9272	5' - ACAAGGGGATAACCACAGGG - 3'	5' - TTGTCAAAGAAAGCTGGGAGG - 3'	124	60°C
9273	5' - CAGCCAAGCCTTAGATAGCC - 3'	5' - TACCTATGCCACCCTTTCCC - 3'	124	60°C
9274	5' - TCCTTGCTTGCTTTGGTTG - 3'	5' - GACTCTACAGGTTATTTGGAGGC - 3'	127	60°C
9348	5' - TCAGGGAATGTAGGATTGC - 3'	5' - TCTCTTCGCTCCCTCCCTC - 3'	133	60°C
9349	5' - TCATGTGCTGATGCCTGTTTAC - 3'	5' - TGCTGAGAACAAGGATGCC - 3'	207	60°C
9350	5' - AATTTTCTTTCTTGACATAAATCTG - 3'	5' - GGAACAACATGCAGAACTGCTC - 3'	132	60°C
9351	5' - TGCTTAAGCCACAAATCTGAAG - 3'	5' - TTGTATGTGCACACTGGAACCTG - 3'	252	60°C

9352	5' - TTTCATACACAGAGGGAAGG - 3'	5' - ACATGGAGGAGATGCCGTC - 3'	136	60°C
9353	5' - CTAACATTTGGGAGGTGCGTC - 3'	5' - AGGATTCATTTCCAACAGCG - 3'	181	60°C
9354	5' - AGGTTAAGGCTCCCTCCC - 3'	5' - GTGTCTTCAACTTCTCGGGC - 3'	142	60°C
9355	5' - TTGCCAGTTGAAACATTTCTACC - 3'	5' - ACAGAAAGATTTGCTCCTCGC - 3'	264	60°C
MBD5 Idaho				
9275	5' - TCATCTATTGCTGATATCTTTGGAG - 3'	5' - TTGCAGGTACCACAGGTAATAAAG - 3'	193	60°C
9276	5' - AAAATGCTTTTCCCTAGTGGG - 3'	5' - GAAAAGGTAGAAAAGGTGTTTTAATG - 3'	482	60°C
9277	5' - TTTTACAGACATATCTAAACAAAGGC - 3'	5' - GGAGAGGAGAAATTCCTGAACC - 3'	261	60°C
9278	5' - TCCCTCCCACCACAAAAG - 3'	5' - GTGGTCAGTCCCTTGAGAG - 3'	395	60°C
9279	5' - GAAATCTCCATTCCTGGC - 3'	5' - AACATTTGCTCGTGGTATTTCC - 3'	354	60°C
9280	5' - TCTTTCCCAACCCTTGACTAC - 3'	5' - GCAATGGGAGATTACTTGGC - 3'	335	60°C
9281	5' - TTAATCCAACCAGTTTCCATTC - 3'	5' - TTGGGACCCTATTGTTGAC - 3'	375	60°C
9282	5' - ATGATGCCACCCTGTAGGACC - 3'	5' - GATTTGCTAGCTGCTTTGG - 3'	355	60°C
9283	5' - TGGGAATGCCCTTTAAATCAG - 3'	5' - TCCATTTGAGACTGCTGAGC - 3'	355	60°C
9284	5' - AGACGCATTGCGGAAAAG - 3'	5' - GTGTTGCATTTGTGGCAGTG - 3'	339	60°C
9285	5' - GCCTCAAATACTGCTTTGGC - 3'	5' - GAAGAGAATTTACAAATGGG - 3'	467	60°C
9286	5' - TCATGATTAATAACTGGTTTGTG - 3'	5' - AGTCTGTTGGATGGAATG - 3'	243	60°C
9287	5' - ACCCATTTGGGAGTGATTTTC - 3'	5' - AGGGTGGAGTTGATCTCAG - 3'	223	60°C
9288	5' - TCTCGGTACAAAGAGAGGC - 3'	5' - TTGTTTTCTTCTAAATGACACAGC - 3'	320	60°C
9289	5' - GAGCCCTCAAATTTATTTCCC - 3'	5' - AATGGGCTGATCTGAGTTG - 3'	351	60°C
9290	5' - TTACAAAGCAGTTGTCGATGC - 3'	5' - TGTGGCAACTTGTGACTTG - 3'	311	60°C
9291	5' - GAGGAATCAAGAGGGGCTC - 3'	5' - CATCTTTCCACAGTGTCTCTAGTC - 3'	387	60°C
9292	5' - CCACCAGAANAAGTCCAGG - 3'	5' - ATCTGAAGGGCTAGGCACAC - 3'	315	60°C
9293	5' - GGAGCAGTCTCCAAGTTCC - 3'	5' - GAACAAAAGTTAGTATTTGACAAATGG - 3'	361	60°C
9294	5' - TCTGGGTAATGTGGTTGGTC - 3'	5' - TCAGTGAGATTTTCATTGTCCC - 3'	197	60°C
9295	5' - TGGAAATGGTACTTTTGTCTG - 3'	5' - TTTGACAATACTCAAGAGACTTTACC - 3'	258	60°C
9296	5' - TGTGATTTTCATCCTCTGTTGTTG - 3'	5' - CACTGCGCATTAGTGGAGTAG - 3'	151	60°C
9658	5' - TCCCCAACCTTGACTACAAAAG - 3'	5' - AGAGCACAAAGAAAGGTGGAGG - 3'	140	60°C
9659	5' - TCCACTTGGCATTCTTGACC - 3'	5' - CTTGGCAAAGGAACAACAGC - 3'	147	60°C
RT und Race-Primer MBD5				
9432		5' - GACATTTCCAAGCCACACTTGC - 3'		64°C
9433		5' - CATGTTCCATCAGTAAGCAGG - 3'		62°C
9434		5' - ACACGACGCTGCCAACCCAC - 3'		66°C
9468 C1F	5' - AGAGGTACTCCCTTATAGGG - 3'			60°C
9468 C2F	5' - GGACTCGTAAAGACATAGAGC - 3'			60°C
9469 C1F	5' - TACTCCCTTATAGGGACTCG - 3'			60°C
9469 C2F	5' - GGGACTCGTAAAGACATAGAG - 3'			60°C
9470 C1F	5' - GAAATCAAGAAAGAGCACACAC - 3'			60°C

9470 C1F	5' - ACACATATTTCCCTTCATCAATCC - 3'		60°C
9725	5' - ATGTCAGTTTCTACATGTGG - 3'		60°C
9726	5' - ATGTCAGTTTCTACATGTGG - 3'		60°C
9727	5' - GGAAGTAGACATTTGAAGGC - 3'		60°C
9728	5' - GCCACCCGTCAGAGAGGGAC - 3'		60°C
9729	5' - GTCAGAGAGGGACATGCCG - 3'		60°C
10028		5' - GTTGACATCCTCTGTGGCCA - 3'	60°C
Primerwalking FOXP1			
9303	5' - AACCTAAGAATGTTCCAAAATATG - 3'	5' - GCTTCTTCATTTCTGATGCTGG - 3'	131
9305	5' - GGGCATAACTGTGGTCCCTAC - 3'	5' - TCTGGGTAAATTCAGGGTCC - 3'	124
9308	5' - GGGTGCAAGCTTAATCTCC - 3'	5' - CCAAAGGGTCTAAGTCTTATGTG - 3'	141
9309	5' - GAAACCCACTGCAACCAAC - 3'	5' - AGTGGAGCAAAAGATTTTGGG - 3'	183
9311	5' - AACTTTCTGGTGCATAGC - 3'	5' - AAATCATTCCTCTGTTCTGGC - 3'	140
9313	5' - TCAGTCCAATCCCAAATAAG - 3'	5' - GCAGTCTAAGTTGTAGGTAGCATCC - 3'	127
9314	5' - CACTTACTGTAATCTGTGCTCATTAG - 3'	5' - ACCAAGAGCCATTTTCATGTC - 3'	130
9315	5' - AATGAGACGTACCTTGAGTATAATTC - 3'	5' - AGAAGGACAGATTGTTAGGCAG - 3'	130
9316	5' - CAAGGTAGAGCTAGGGCAAATTC - 3'	5' - TTCAAATTTACAGGGATTTATGTTAG - 3'	129
9317	5' - GCTGCAAAGATGGATGAAGAG - 3'	5' - GGCTAGGTGGCTTCTCACC - 3'	115
9356	5' - ATGCTGAAGGTGGAATGGG - 3'	5' - GGGCCTTACTATAGCTTCTTTGTG - 3'	142
9357	5' - GCAATATTGCTCATCCACTG - 3'	5' - GACTATGCTTTGCTCCACTTTTC - 3'	110
9358	5' - GCTTCCAACCTCCCACTATC - 3'	5' - CATGTTATTCTCCATCCTGGAC - 3'	129
9359	5' - TCGGTATAACCAATGATTCAG - 3'	5' - AATCCTGAAGCCTGTAACGC - 3'	127
9360	5' - AATTCATAGGTGAGGCAGGC - 3'	5' - TCAGTTGCTCAAACGTTGGT - 3'	126
9361	5' - TGCAGAAAAGGCACTTTCTAATC - 3'	5' - AGGACTACCTCGAGCTCC - 3'	216
9362	5' - TCTGAATGACAAAACATTTTCAGG - 3'	5' - TCAGAGCTCTCAAGCATCTCTTAC - 3'	129
9363	5' - ACCTACTAGCAAAGGGAAGGTG - 3'	5' - TAGAAATGCCCTCAATTGGC - 3'	226
9364	5' - AATAAAATGGCTGAGAACATCAC - 3'	5' - TGCATTTAAGTTAGTGCAGATGG - 3'	132
9365	5' - AGTAGGCAAACGGTAGGGTG - 3'	5' - TCATCTGGGATTTCCATTGC - 3'	118
9366	5' - GGGCTGGAGCTGTTCTACTG - 3'	5' - TTCTGTTCTCAAACATTAGAGGG - 3'	125
9435	5' - GACTCCAAGTCAATGAGATGC - 3'		60°C
9436		5' - TGAAGTGGCCTATTGTCACC - 3'	60°C
9437		5' - GGCCACATACGTGTTGTCCAG - 3'	60°C
9438		5' - GCAGGCCCTTGTGTTTATCCAG - 3'	60°C
FOXP1 Idaho			
8797	5' - CTTGAAAATCCTTGTATCAGG - 3'	5' - GATCGCGATTAAAGTGAAAATTC - 3'	289
8798	5' - GTCAGGACCCAGTCCCAAAG - 3'	5' - TGCTCAAACACAATCCACTCC - 3'	287
8799	5' - GGCTGCGTGTCTCTGATTTTC - 3'	5' - GATTGCGAATGGAGTGATGG - 3'	257
8800	5' - CAAAAGGGACAGTGGATTGG - 3'	5' - AGCCACTAGATAGTCTCTCTG - 3'	210

8801	5' - CCATACTTTTAACTTCCTTGGG - 3'	5' - AAGAGACAACCCACCACCTC - 3'	283	62°C
8802	5' - GCTGTTAATTGATTAATTATCC - 3'	5' - CCACCCCTCTCATGTAAAG - 3'	299	58°C
8803	5' - CTCGTCGTGGTGAATGGTTC - 3'	5' - TGAGCAAAGCATGAACGGTG - 3'	239	60°C
8804	5' - GACCCGCTGCCTAGTTTATG - 3'	5' - CTACAGAAATCTGGAATTTGAG - 3'	245	60°C
8805	5' - TTTTGGTGAAGAGCCATGC - 3'	5' - GGAAGTAGAAAAGGAATACTG - 3'	209	58°C
8806	5' - CTGGTTTCTGAAGGTCTTC - 3'	5' - CTGGTCTACAAGAAATTCAG - 3'	305	58°C
8807	5' - TTGTAGACCAGCCCTTTTG - 3'	5' - AAAGGTTTCAGCATGCTTGC - 3'	252	58°C
8808	5' - GTGGAATGACGACTCTGCTG - 3'	5' - GCACAACTGCATTTTATCAAC - 3'	241	58°C
8809	5' - ACTGTTAGGCTGAATTTACTG - 3'	5' - CATCTAGCAGCCAAAGCCTC - 3'	285	58°C
8810	5' - AGCTAGAACTCAGCAGTCG - 3'	5' - TGAGTAGGGGAGACCTGTGC - 3'	186	62°C
8811	5' - GTAGTTCCCTGTCTTACATAC - 3'	5' - TGACCTAGGGGAGACTAGGG - 3'	283	60°C
8812	5' - TTTTCAGTGCACCTGAGCTG - 3'	5' - AACAATTTCACTGCTAACTTTTG - 3'	296	60°C
8825	5' - GAAGAGGACGGAATTAGCTG - 3'	5' - TGTTAAAGGCAGTTTGGACC - 3'	303	60°C
8826	5' - GTTTCTGAAGGCTTCTAAG - 3'	5' - GTCTACAAGAAATGCAGATTAC - 3'	299	58°C
RT and Race-Primer FOXPI				
8881	5' - GATAGAAGTACAGCCCAATG - 3'			58°C
8882		5' - GGGTTCTGTAGACTTCACAT - 3'		58°C
8883		5' - GTCGTTGGAGTATGAGGTAA - 3'		58°C
8884		5' - AGATGCATATGTAATGGTGG - 3'		58°C
Sonden FOXPI				
FOXP1_EX13	5' - GCAGTTAGACTACAGTAAAACCA - 3'	5' - CACAGGAACCTATTACAGAACACTACA - 3'	72	72°C
FOXP1_EX14	5' - GCCTGCAAGCCATGATGAC - 3'	5' - GCGGGCTTTGGGTTCTGTGA - 3'	55	60°C
X-Inaktivierung				
8774	5' - TCCAGAACTCTGTTCCAGAGCGTGC - 3'	5' - GCTGTGAAGGTTGCTGTTCCTCAT - 3'	287	58°C
SLC34A3				
7864	5' - TCITTTGACCTGCAACTGCTC - 3'	5' - CACCTGGACCCCTCCAC - 3'	265	60°C
7865	5' - TGTGAATCCAGCTTGTGAGG - 3'	5' - CTGTCTCCCCAGTCCCG - 3'	637	60°C
7866	5' - AGAGAATGAGGGCCTGG - 3'	5' - GACACCCGGACAGTCAGC - 3'	523	60°C
7867	5' - CAGCATGGTGGCTGCTAAG - 3'	5' - CAGACCCCTCCGCTGACC - 3'	595	60°C
7868	5' - CCCGGGTGAGTCTCTGAG - 3'	5' - GGCTGTGCTGTTCCTCTCC - 3'	448	60°C
7869	5' - GACAGGCTGCCCTGTGAG - 3'	5' - CACCACAGGACCCCCAGC - 3'	560	60°C
7870	5' - GACAGAGGCCTCGGGAAC - 3'	5' - CTGTCTGTGGGCTGAGAG - 3'	247	60°C
7871	5' - TAGGGTGGAGGAGGGCAG - 3'	5' - CTCCAGAGCCCTCCCAG - 3'	604	60°C
7901	5' - GTGGGTGCACACTCCCTC - 3'	5' - AGAGAGTGGGCAGGGTCTAG - 3'	271	60°C
7902	5' - GTGGGTGGAAGGGCTGG - 3'	5' - ACCCTCCATCAGGGCAG - 3'	366	60°C
7903	5' - GGTTACTGTCTCCTGCAGCGG - 3'	5' - GCTCCAGAGAATGGAGCCAG - 3'		60°C
7904		5' - GTGTCCACCAGGAGGCAG - 3'		60°C
7907	5' - CTCCCTCAGCTGAAGGACAC - 3'	5' - GACTGCGCCATTGAGACCAG - 3'		60°C

7908	5' - AAGCCAGCCCTGGAAAGAGC - 3'	5' - CAGGGTGTGGTGATGGATG - 3'	60°C
7909		5' - GACTGCGCCATTGAGACCAG - 3'	60°C
7910		5' - CAGGGTGTGGTGATGGATG - 3'	60°C
DMP1			
8237	5' - AAATTCCTGCACCATCAGAG - 3'	5' - CTCTACATGGTATAATGCCTC - 3'	233 60°C
8238	5' - GTAGGAAGAACTGTTAAAAAG - 3'	5' - AGCCATCTTGGTTCATAGCC - 3'	237 60°C
8239	5' - CTATCCCTACTCCTATGTATG - 3'	5' - AATGAAATCAATTTGGTTACTC - 3'	208 60°C
8240	5' - ATAATGAAATCCATCTGTGAGG - 3'	5' - TAACATCATTCCTCCATTCC - 3'	377 60°C
8241	5' - AGATAACTCCTTCTCTATCGG - 3'	5' - GTCCAGCAATCCCTTCAGG - 3'	1560 60°C
8243	5' - GAAACTCCAGACTGGGAAG - 3'	5' - AATTGGCTGCCACCTGAATC - 3'	389 60°C
8244	5' - GGAGTGAAGGGGAAACTCC - 3'	5' - GATGACAGGTTGGCCTCTTG - 3'	373 60°C
8245	5' - GGAGAGACAGCAAGGGTGAC - 3'	5' - CTGCTGAGCTGCTGTGAGAC - 3'	393 60°C
8246	5' - AGACAGTGACTCCAGCGAGG - 3'		60°C
8247	5' - CCAGTCTCACAGCAGCTCAG - 3'		60°C
MEPE			
8248	5' - GTTTGCTAGCGCTGAGTGAG - 3'	5' - ATACTTGGGTGGCAAGCATC - 3'	216 60°C
8249	5' - AGGATATGAAGTCCCAAGGG - 3'	5' - TAGATGCTTCTGTTTGTG - 3'	327 60°C
8250	5' - AATGAAACATTTGTTGAAGAAG - 3'	5' - GTGCCCAGACTATTTTGAC - 3'	227 60°C
8251	5' - TCCATGAAACCTGATTTGACC - 3'	5' - TCAACTCACAGGTGACGAGG - 3'	1700 60°C
8257	5' - GTAAGGGCAAAAGTCAGGGC - 3'	5' - CTTCTTGGTCATGTAGTTTGC - 3'	60°C
8258	5' - GAAGCCAGTGAGAATAAGGG - 3'	5' - AATACGATGGGTGCTTTTGC - 3'	60°C
8259	5' - ATCCCAGCAAGTATGAATTATG - 3'	5' - CAATGGTATTTCCACCATTTC - 3'	60°C
8260	5' - ACGGCCAACCTTTTAAGGAC - 3'	5' - GAGGGTGCAGGAGGGTAATG - 3'	60°C
8261	5' - TGCCATGTGTTATTATTCTC - 3'		60°C
DSPP			
8252	5' - GGATCCTAAACAGTGATTGGTTG - 3'	5' - CGCTTCTTAAAAATCAAACCTGGC - 3'	256 60°C
8253	5' - GGAAAAGGGCAAATGCTTAC - 3'	5' - GGAAAAGGGCAAATGCTTAC - 3'	377 60°C
8254	5' - TGTGCACGCTCACACAC - 3'	5' - TCAATGATCTTGAAGGAAAGC - 3'	251 60°C
8255	5' - CCTATGGCAACTTTTCCAG - 3'	5' - CCAACTGTTTGAATTTGATCCC - 3'	3029 60°C
8256	5' - GATTGCAATTTGCTTTCCTTC - 3'	5' - AGCATCAATAGATGGAGGG - 3'	60°C
8281	5' - TCCAGGGACAAGTAAGCATC - 3'	5' - CCTCCTGGCCCTTGCTG - 3'	60°C
8282	5' - TCCTAGTGGGAATGGAGCAG - 3'		60°C
8286	5' - GAGATGATCCCAATAGCAGTG - 3'		60°C
8287	5' - GACAGTGGCAAAGGTAATC - 3'		60°C
8292	5' - CGATAGCAGTGACGGCAGTG - 3'		60°C
8283		5' - TCCTCATTGTGACCTGCATC - 3'	60°C
8284		5' - CCTCGGCTACTGCTGTATTG - 3'	60°C
8293		5' - CACTGCCGCTCACTGCTATCG - 3'	60°C

8294	5' - CACTCTCATCATTGCTGTCAG - 3'		60°C
PNPO			
8359	5' - GTTCCGAACCTCAAAGGAACC - 3'	5' - CTCCCCGCTGTAGATGACTC - 3'	273
8360	5' - TGTAANAATGGGCACAGAATC - 3'	5' - CTTGCCCTAAGCCACAGAGC - 3'	417
8361	5' - ATCCTTTGGGAGAGGGAC - 3'	5' - TGACTGAGGAAGCAAGAGCC - 3'	240
8362	5' - TCCTGCCITTTCCCTGC - 3'	5' - ACAGCTTCTGGGGATAAGCC - 3'	571
8363	5' - CAGCACATATCCCCAGGAAG - 3'	5' - CCCACACCCCTCTTTGACAC - 3'	598
SNX11			
8364	5' - CCCTGGTAAAGAATGATAAGAACAG - 3'	5' - TCCTCCTAAATCCCCCTCTCC - 3'	181
8365	5' - GTGTTGTCAGAGCTGCCTTG - 3'	5' - CTCAGTACAAGCCCTCAGC - 3'	209
8366	5' - GCATGGCTCCTCAGTGAAG - 3'	5' - GCCTGGTATGTTGCAAGTTC - 3'	299
8367	5' - CAGAGGCCCGGAATGACTATC - 3'	5' - CTGGCACATTCCTCCACC - 3'	227
8368	5' - GGAGAGCAGTTCCTCAGGG - 3'	5' - ATGGACACCTCTCTGGTCC - 3'	336
8369	5' - ACATGCCCAACATCCCTGG - 3'	5' - CCCAGTCCCACCTGAGTAAG - 3'	446
AOC2			
11193	5' - TCAGAAATTTCTGCCCAAC - 3'	5' - CCTCTTTCAGATGCCCTCCAC - 3'	681
11194	5' - GTTTGCTGACCAGCCACAG - 3'	5' - AGAGGGAGGATACCACCAGG - 3'	919
11195	5' - CTGGACCCTGCCACTG - 3'	5' - AAGCAGCTGGACTGCC - 3'	516
11196	5' - TCGGTTCCAGGGTGAGC - 3'	5' - CAGCTTGACCTGGAGATTC - 3'	638
11197	5' - ACCTGGCCCCAGAGACCTTAC - 3'	5' - GGAAGGGAGTTTGTCTCCAAG - 3'	422
11198	5' - GATCTAAGGGACTCCTGGGC - 3'	5' - CAGCCAAAGATGCCATGC - 3'	259
11199	5' - GGATTGTGACTGAAGGGTGG - 3'	5' - AATCTGGGGAGCAGGTGTG - 3'	434
MEOX			
11200	5' - TTTAAGGAGAGACTTTAGTTTGGG - 3'	5' - AGTCTCCTGCATTTGGATGG - 3'	770
11201	5' - GGAATGGAGTCTCCCTCAG - 3'	5' - AAAGGAGGGAAGGGAGGG - 3'	289
11202	5' - GACATGGGGAGGAGGTAC - 3'	5' - AGAGGGTGAAGGTGGGATTG - 3'	263
TMUB2			
11203	5' - CCATCAGGGGTTAGTTGTGG - 3'	5' - ACCCCAGTCTTGGCTTAC - 3'	231
11204	5' - AAACAANACTAGAGACCTGGTTGC - 3'	5' - AGTGGAGTCTCCCCGACC - 3'	395
11205	5' - CAACCCCGAGCCAAC TG - 3'	5' - GTCTGCACCAGCCTCCTTAG - 3'	405
11206	5' - GTTGTCCAGGCCTGTTTG - 3'	5' - GCTCTGCCCAGGAAAAGTG - 3'	504
Junction Fragment 27384			
10549	5' - AACATATGGGGCTCCAAGAGG - 3'		60°C
10550	5' - AATCTCCCTGGGTAACCTCC - 3'		60°C
10560	5' - TTCCTGCTGAGTTAGTTAGCCC - 3'	5' - AGTGGGTTAGGTATGCCAGG - 3'	60°C
10562	5' - TTCAGTACAGTGTACACAAAAGTGG - 3'	5' - CCAAAGGAACTTACCTCAATACAG - 3'	60°C
10665		5' - GCTGCTAGTCCATAAGTAC - 3'	60°C

10699	5' - CTTTTAGGATTITGTTATCAATTAGACC - 3'	60°C
Junction Fragment 30437		
10519	5' - CCTGGAAATGCTTATTTGGG - 3'	120
10566	5' - CTTGCCAAGTCCAGATGATG - 3'	60°C
10647	5' - GAAAAGTGGTGGTATTTTATTTGC - 3'	60°C
10669	5' - GGGTGCAAGGCAGAGG - 3'	136
10670	5' - ACGCCACATTCATGCCAG - 3'	60°C
	5' - GTTGTCATGAGATGAGGATG - 3'	128
	5' - ATGACCTGGCCCTCTGCC - 3'	
	5' - GCAATACCTCTGCTTGAGGAACTCCTCC - 3'	
Junction Fragment 30375		
10650	5' - TTCTCCAGGGTCTGTATCCG - 3'	149
10660	5' - CTGCAGCTGCCTGGATTAC - 3'	132
10672	5' - AGCCCAATGGTTGAGCC - 3'	125
10700	5' - AAGTCTTGCACCCATTCCGAC - 3'	117
Junction Fragment 28735		
10479	5' - GAGCCACCACGTCAGCC - 3'	122
10483	5' - CAGTGTCTTCTCCCCACGTC - 3'	122
10484	5' - GAGCTCAGCGCATCTTTCAG - 3'	121
Junction Fragment 29195		
9350	5' - AATTTTCTTTCTTGACATAAATCTG - 3'	132
9351	5' - TGCTTAAGCCACAAAATCTGAAG - 3'	252
9270	5' - CAGGTTGTCATCACCCTGC - 3'	133
9353	5' - GGAAACATGCAGAACTGCTC - 3'	
	5' - TTGTATGTGCACTGGAACCTTG - 3'	
	5' - GAATGACCCGCACTGACAGC - 3'	
Junction Fragment 29945		
9356	5' - ATGCTGAAGGTGGAATGGG - 3'	142
9357	5' - GCAATATTGTCTCATTCCTG - 3'	110
9361	5' - TGCAGAAAAGGCACCTTCTAATC - 3'	216
9362	5' - TCTGAATGACAAAACATTTCCAGG - 3'	129
	5' - GGGCCTTACTATAGCTTCTTTGTG - 3'	
	5' - GACTATGCTTTGCTCCCATCTTC - 3'	
	5' - AGGACTACCCCTCGAGCTTCC - 3'	
	5' - TCAGAGCTCTCAAGCATCTCTTAC - 3'	

Tabelle Mikrosatelliten		
Mikrosatelliten	Vorwärtsprimer	Rückwärtsprimer
Mikrosatelliten für HHRH		
D9S164	5'-CTATGATCGTGCCACTGCAC-3'	5'-GTTTGCCTGGGGATTGATTT-3'
D9S1826	5'-ATGTTGATTGCCCGCAGTG-3'	5'-TGCAGAACCCAGGCTCTTATTCG-3'
D9S158	5'-TCTCAAGCGACAACAATCAC-3'	5'-GATTTGGCTAAAATAGGCTCA-3'
D9S905	5'-GTGGGAAAATTGGCCTAAGT-3'	5'-CTTCTGAGCCTCACACCTGT-3'
D9S1838	5'-ACCCAGCTACTGAGGAGGCTT-3'	5'-GCTTCTGCACTTTTGTAGAACCAAAT-3'
Mikrosatelliten für SOS		
D17S1802	5'-TGCTACAACATAAATCTCAGACAG-3'	5'-GGAATCACCTCTTGTGGATG-3'
D17S1861	5'-AGGGGCAGCAGTCCCTGTA-3'	5'-ACATCATCCTGAAATCTAATGGG-3'
D17S1804	5'-GCTTCCCTATTTAGACATTAGGTT-3'	5'-ATCGCACCCATTGCACTC-3'
D17S934	5'-TCTGAATGGCCCTTGG-3'	5'-TCCTATCTGAGGTGGGGT-3'
D17S920	5'-CTTCTGCATGGATCAGAAAA-3'	5'-ACTCACCCAGATGATTGCTT-3'
D17S931	5'-AACATTTGCTTGGGGCT-3'	5'-GGTGAGTAGTGCTCTGGGAA - 3'
D17S1868	5'-GCAAAATCCATCCAAGC-3'	5'-AGTCAAATTCAGCCTGGG-3'
D17S1795	5'-AGTGCCAGAGATATACCCGTG-3'	5'-GTCTGCAAGGCAAGTTGTC-3'
D17S1869	5'-TCATACAGATGCACCCGC-3'	5'-ATTGGAAGGTGGCCCCC-3'
D17S1820	5'-CATGAGGTCTCCAGAAGG-3'	5'-AACACACTTGCTGATGTGC-3'
D17S788	5'-CTAGGCAGCCACTACCAAAT-3'	5'-CAGCATCTTTGCTATAAGCATC-3'

Tabellenverzeichnis

Tabelle 1	Vergleich verschiedener Methoden zur Identifizierung von CNVs.....	12
Tabelle 2	Zusammenfassung der Standardabweichung und der mittleren absoluten Abweichung der \log_2 Intensitätswerte.....	72
Tabelle 3	Zusammenfassung des durchschnittlichen Signal-Rausch-Verhältnisses (SNR).....	74
Tabelle 4	Zusammenfassung der Kandidatenregionen.....	75
Tabelle 5	Deletionen und Duplikationen.....	77
Tabelle 6	Variationen im Gen <i>MBD5</i>	89
Tabelle 7	Variationen im Gen <i>FOXP1</i>	91
Tabelle 8	Deletionen an Bruchpunkten in Patienten mit Translokationen.....	92
Tabelle 9	Vergleich der Größenangaben von Deletionen zwischen 100K-Arrays und quantitativer PCR.....	103

Abbildungsverzeichnis

Abb. 1	Mechanismen der NAHR und ihre Auswirkungen auf die genomische Struktur.....	23
Abb. 2	Modell der Reparatur von DSBs durch NHEJ.....	25
Abb. 3	Der Einfluss von CNVs auf den Phänotyp.....	27
Abb. 4	Familienstammbäume der untersuchten Familien mit HHRH.....	60
Abb. 5	Genomweite Kopplungsanalyse der Familien 1 und 2.....	61
Abb. 6	Multipoint-Kopplungsanalyse am langen Arm des Chromosoms 9.....	62
Abb. 7	Mutationsanalyse in dem Gen <i>SLC34A3</i> in den 5 Familien mit HHRH.....	64
Abb. 8	Familienstammbäume der untersuchten Familien mit ARHP.....	65
Abb. 9	Genomweite Kopplungsanalyse der Familien 1 bis 3.....	66
Abb. 10	Mutationsanalyse in dem Gen <i>DMP1</i> in den 3 Familien mit ARHP.....	68
Abb. 11	Familienstammbaum der untersuchten Familie mit SOS.....	69
Abb. 12	Genomweite Kopplungsanalyse der Familie mit SOS.....	69
Abb. 13	Darstellung der Haplotypen nach der Analyse der SNP- und Mikrosatellitenmarker....	70
Abb. 14	Auswahl der Kandidatengene in der Region 17q21.32-q22.....	71
Abb. 15	Heatmap der median normalisierten Intensitätswerte von 169 <i>HindIII</i> -Arrays.....	73
Abb. 16	Überblick über die log ₂ Intensitätsprofile der 12 ausgewählten CNVs.....	76
Abb. 17	Schematische Darstellung der betroffenen Region Xq22.31 des Patienten 29836.....	81
Abb. 18	Darstellung der X-Inaktivierung des Patienten 29836 und seiner Mutter 29837.....	82
Abb. 19	Bestimmung der Bruchpunktregion bei Patient 27384.....	83
Abb. 20	Bestimmung der Bruchpunktregion bei Patient 30437.....	84
Abb. 21	Bestimmung der Bruchpunktregion bei Patient 30375.....	86
Abb. 22	Bestimmung der Bruchpunktregion bei Patient 28735.....	87
Abb. 23	Schematische Darstellung des Gens <i>MBD5</i>	88
Abb. 24	Schematische Darstellung des Gens <i>FOXP1</i>	90
Abb. 25	Bisher bekannte Proteine im Phosphatstoffwechsel.....	96

Diese Dissertation entstand in der Arbeitsgruppe von PD Dr. Tim-Matthias Strom, am Institut für Humangenetik von Prof. Dr. Thomas Meitinger im GSF-Forschungszentrum für Umwelt und Gesundheit in der Zeit von Januar 2004 bis Dezember 2007.

Danksagung

Meinem Projektbetreuer Dr. Tim-Matthias Strom danke ich herzlich für die Bereitstellung dieses Themas und sein beständiges Interesse an meiner Arbeit. Vielen Dank auch für die intensive Betreuung, besonders bei der Korrektur dieser Arbeit.

Prof. Dr. Thomas Meitinger danke ich für die Möglichkeit in seinem Institut zu arbeiten, seine wissenschaftlichen Anregungen zu dieser Arbeit, sowie für Hilfestellungen bei Vorträgen und Veröffentlichungen.

Prof. Dr. Martin Hrabé de Angelis danke ich für seine Bereitschaft, meine Dissertation vor der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zu vertreten. Für sein mir entgegengebrachtes Vertrauen möchte ich mich an dieser Stelle ausdrücklich bedanken.

Prof. Dr. Chris-Carolin Schön sei an dieser Stelle für die Übernahme des Zweitgutachtens und ihr dafür aufgebrachtes Engagement gedankt.

Ein großes Dankeschön gebührt meinen Laborkollegen aus dem Gebäude 35, Dr. Bettina Lorenz-Depiereux, Dr. Anna Benet-Pages, Sandy Lösecke und Nuria, sowie den Kolleginnen aus dem Gebäude 34, Dr. Gertrud Eckstein und allen Mädels aus dem TA-Zimmer, für die wissenschaftlichen Anregungen und die Unterstützung meiner Arbeit.

Ich möchte auch meiner Familie und meinen Freunden danken, die mich während meiner Arbeit unterstützt und auch immer wieder ermutigt haben.

Mein tiefster Dank gilt dabei meinen Eltern, die diese Arbeit durch ihre Unterstützung, Liebe und Geduld überhaupt ermöglicht haben.

Mein letzter Dank gehört meinem Freund Frank Graf, der mich durch alle Höhen und Tiefen während dieser Arbeit begleitet hat.

Lebenslauf

Persönliche Daten:

Name: Janine Wagenstaller
 Geburtsdatum: 30. Januar 1977
 Geburtsort: Ebersberg
 Familienstand: ledig

Institutsanschrift:

GSF-Forschungszentrum für Gesundheit und Umwelt
 Institut für Humangenetik
 Ingolstädter Landstr. 1
 85764 Neuherberg

Schulausbildung:

1984 – 88 Grundschole Hohenthann
 1988 - 89 Hauptschole Ostermünchen
 1989 – 97 Gymnasium Bad Aibling, Allgemeine Hochschulreife

Hochschulausbildung:

1997 – 2000 Grundstudium und Vordiplom in Biologie an der Universität München
 2000 – 2003 Hauptstudium Biologie (Diplom) an der LMU München,
 Fachrichtung: Anthropologie und Humangenetik
 2002 - 03 Diplomarbeit „Variabilität stabiler Strontium- und Bleisotope in archäologischen
 Funden von Hornzapfen als Indikatoren für frühneuzeitlichen Fernhandel mit
 Nutztvieh“ bei Prof. Dr. Gisela Grupe, Humangenetik der LMU München
 2003 Abschluß als Diplom Biologin mit Note 1,2

Berufstätigkeit:

04/2004 - 11/2007 wissenschaftliche Angestellte des Institut für Humangenetik der GSF Neuherberg
 (Ziel: Promotion)

Kurse und Kongresse:

06/2004 Europäischer Humangenetikkongress/ München
 08/2004 Fortgeschrittenenkurs Linkage Analysis an der GSF, München
 05/2005 Europäischer Humangenetikkongress/ Prag
 03/2006 Deutscher Humangenetikkongress/ Heidelberg
 05/2006 Europäischer Humangenetikkongress/ Amsterdam
 10/2006 Amerikanischer Humangenetikkongress/ New Orleans
 03/2007 Deutscher Humangenetikkongress/ Bonn
 06/2007 Europäischer Humangenetikkongress/ Nizza

Publikationsliste

Artikel:

Wagenstaller J, Spranger S, Lorenz-Depiereux B, Kazmierczak B, Nathrath M, Wahl D, Heye B, Glaser D, Liebscher V, Meitinger T, Strom TM. Copy-number variations measured by single-nucleotide-polymorphism oligonucleotide arrays in patients with mental retardation. *Am J Hum Genet.* 2007,81:768-79. Epub 2007 Aug 28.

Asmus F, Hjerminde LE, Dupont E, **Wagenstaller J**, Haberlandt E, Munz M, Strom TM, Gasser T. Genomic deletion size at the epsilon-sarcoglycan locus determines the clinical phenotype. *Brain* 2007,130:2736-2745.

Lorenz-Depiereux B, Benet-Pages A, Eckstein G, Tenenbaum-Rakover Y, **Wagenstaller J**, Tiosano D, Gershoni-Baruch R, Albers N, Lichtner P, Schnabel D, Hochberg Z, Strom TM. Hereditary Hypophosphatemic Rickets with Hypercalciuria Is Caused by Mutations in the Sodium-Phosphate Cotransporter Gene SLC34A3. *Am J Hum Genet* 2006,78:193-201. Epub 2005 Dec 9.

Lorenz-Depiereux B, Bastepe M, Benet-Pages A, Amyere M, **Wagenstaller J**, Muller-Barth U, Badenhoop K, Kaiser SM, Rittmaster RS, Shlossberg AH, Olivares JL, Loris C, Ramos FJ, Glorieux F, Vikkula M, Juppner H, Strom TM. DMP1 mutations in autosomal recessive hypophosphatemia implicate a bone matrix protein in the regulation of phosphate homeostasis. *Nat Genet.* 2006,38:1248-50. Epub 2006 Oct 8.

Kalay E, Li Y, Uzumcu A, Uyguner O, Collin RW, Caylan R, Ulubil-Emiroglu M, Kersten FF, Hafiz G, van Wijk E, Kayserili H, Rohmann E, **Wagenstaller J**, Hoefsloot LH, Strom TM, Nurnberg G, Baserer N, den Hollander AI, Cremers FP, Cremers CW, Becker C, Brunner HG, Nurnberg P, Karaguzel A, Basaran S, Kubisch C, Kremer H, Wollnik B. Mutations in the lipoma HMGIC fusion partner-like 5 (LHFPL5) gene cause autosomal recessive nonsyndromic hearing loss. *Hum Mutat.* 2006,27:633-639.

Langer S, Geigl JB, **Wagenstaller J**, Lederer G, Hempel M, Daumer-Haas C, Leifheit HJ, Speicher MR. Delineation of a 2q deletion in a girl with dysmorphic features and epilepsy. *Am J Med Genet A.* 2006, 140(A):764-768.

Kongressbeiträge

Vortrag mit publiziertem Abstract

Wagenstaller J, Spranger S, Heye B, Kazmierczak B, Cohen M, Freisinger P., Meitinger T, Speicher M. R., Strom TM. (2006) "Detection of copy number changes in patients with mental retardation using high density SNP microarrays." *Medizinische Genetik* **18**: 35.

Wagenstaller J, Spranger S, Heye B, Kazmierczak B, Cohen M, Freisinger P., Meitinger T, Speicher M. R., Strom TM. (2006) "Detection of copy number changes in patients with mental retardation using high density SNP microarrays." *European Journal of Human Genetics* **14** (Supplement 1): 99.

Wagenstaller J, Spranger S, Kazmierczak B, Nathrath M, Wahl D, Heye B, Gläser D, Liebscher V, Speicher M. R., Meitinger T, Strom TM.(2007) Genome-wide analysis of copy number variations in patients with mental retardation by single nucleotide polymorphism arrays. *Medizinische Genetik* **19**:55.

Poster mit publiziertem Abstract

Wagenstaller J, Spranger S., Heye B, Kazmierczak B, Cohen M., Freisinger P., Meitinger T, Speicher M. R., Strom TM. (2006) "Detection of copy number changes in patients with mental retardation using high density SNP microarrays." *The American Society of Human Genetics, 56th Annual Meeting*: 237.

Wagenstaller J, Spranger S, Lorenz-Depiereux B., Kazmierczak B, Nathrath M, Wahl D, Heye B, Gläser D, Liebscher V, Meitinger T, Strom TM. (2007) „Genome-wide analysis of copy number variations in patients with mental retardation by single nucleotide polymorphism arrays." *European Journal of Human Genetics* **15** (Supplement 1): 309.

Horn, D., Spranger S., Krüger G., **Wagenstaller J.**, Weschke B., Ropers H. H., Mundlos S., Ullmann R., Strom T. M., Klopocki E. (2007) "Microdeletions and microduplications affecting the STS gene at Xp22.31 are associated with a distinct phenotypic spectrum." *Medizinische Genetik* **19**: 62.

Erklärung

Ich erkläre an Eides statt, dass ich die dem Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Promotionsprüfung vorgelegte Arbeit mit dem Titel:

„Kopplungsanalysen und Bestimmung von Kopienzahlvariationen mit Oligonukleotid-Mikroarrays“

im Institut für Humangenetik des GSF-Forschungszentrums für Gesundheit und Umwelt

unter der Anleitung und Betreuung durch

PD Dr. Tim-Matthias Strom

ohne sonstige Hilfe erstellt und bei der Abfassung nur die gemäß § 6 Abs. 5 angegebenen Hilfsmittel benutzt habe.

- (x) Ich habe die Dissertation in keinem anderen Prüfungsverfahren als Prüfungsleistung vorgelegt.
- () Die vollständige Dissertation wurde in veröffentlicht. Die Fakultät für hat der Veröffentlichung zugestimmt.
- (x) Ich habe den angestrebten Doktorgrad noch nicht erworben und bin nicht in einem früheren Promotionsverfahren für den angestrebten Doktorgrad endgültig gescheitert.
- () Ich habe bereits am bei der Fakultät für der Hochschule unter Vorlage einer Dissertation mit dem Thema die Zulassung zur Promotion beantragt mit dem Ergebnis:.....

Die Promotionsordnung der Technischen Universität München ist mir bekannt.

München, den

.....

Unterschrift