

DHGP2-Forschungsvorhaben 01KW 9916

**Struktur- und Funktionsanalyse medizinisch relevanter
Regionen des humanen X-Chromosoms**

Schlussbericht

Teilprojekt Jena
Institut für Molekulare Biotechnologie
Genomanalyse

über den Projektverlauf vom 01.10.2000 bis 30.09.2004

Dr. Matthias Platzer & Dr. Gaiping Wen

Institut für Molekulare Biotechnologie, Genomanalyse
Beutenbergstr. 11
D-07745 Jena
Tel.: 03641-656241
Fax: 03641-656255

Koordinator:

Prof. Alfons Meindl

Abteilung für Medizinische Genetik der LMU
Goethestr. 29
D-80336 München
Tel.: 089-5160-4467
Fax: 089-5160-4780

Inhalt

1. Einleitung

- 1.1 Das Human-Genom-Projekt
- 1.2 Das Maus-Genom-Projekt
- 1.3 Funktionelle Genomanalyse
- 1.4 Hintergrund des Projektes
- 1.5 Ziel, inhaltliche und zeitliche Planung

2. Methoden

- 2.1 Kartierung eines kompletten BAC/PAC Contigs
- 2.2 Genomische Sequenzierung
- 2.3 Gen-Annotation
- 2.4 Analyse der Genexpression
- 2.5 Mutationsscreening in Patienten-Material
- 2.6 Komparative Genomanalyse

3. Ergebnisse

- 3.1 Konstruktion und Sequenzierung eines Contigs in der Region Xp11.4
- 3.2 Identifizierung und Charakterisierung der Gene in der Xp11 Region
- 3.3 Analyse der Genexpression und Mutationsscreening
- 3.4 Komparative Genomanalyse zwischen Mensch und Maus

4. Zusammenfassung

5. Gesamte Kosten

6. Erfolgte und geplante Publikationen

7. Literatur

8. Anhang

1. Einleitung

1.1 Das Human-Genom-Projekt

Das vorliegende Projekt war Teil des internationalen Human-Genom-Projektes (HGP), das die Entzifferung des gesamten menschlichen Genoms zum Ziel hatte und im April 2003, 50 Jahre nach der Entdeckung der Doppelhelix-Struktur der DNA, erfolgreich beendet wurde (IHGSC; 2004, 2001). Nach Angaben des „International Human Genome Sequencing Consortium“ (IHGSC) wurden von 20 Zentren in sechs Ländern Sequenzen generiert, die insgesamt 2,85 Mb repräsentieren und 20-25.000 protein-kodierende Gene beherbergen. Diese enormen Informationsmengen werden nicht nur zum Verständnis für erbliche Krankheiten beitragen, sondern auch die Entwicklung der molekularbiologischen Diagnostik, Therapie und neuer Medikamente in der Medizin unterstützen. Am Jenaer Institut für Molekulare Biotechnologie (IMB) wurden wesentliche Beiträge zur Sequenzierung und Analyse der Chromosomen 21 (Hattori *et al.*, 2000), X (Ross *et al.*, akzeptiert; Aradhya *et al.*, 2001; Strom *et al.*; 1998 ; Kioschis *et al.*, 1998) und 8 (Taudien *et al.*, 2004; Kalaydjieva *et al.*, 2000, Momeni *et al.*, 2000; Varon *et al.*, 1998) geleistet. Es wurden ca. 74 Mb in finaler Sequenzqualität („finished“: Genauigkeit >99,99%, weitestgehend lückenlos) generiert. Dies entspricht 2,1 % des gesamten HGP (IHGSC; 2004). Bei der Sequenzierung des X-Chromosoms haben sechs Sequenzierzentren kooperiert und ca. 154 Mb (5 % des Humangenoms) sequenziert. Insgesamt wurden 1.098 Gene identifiziert und charakterisiert (Ross *et al.*, akzeptiert).

1.2 Das Maus-Genom-Projekt

Eine erste „Rohfassung“ der Maus-Genomsequenz wurde vom internationalen Maus-Genom-Projekt 2002 veröffentlicht (Mouse Genome Sequencing Consortium, 2002). Aufgrund der Ähnlichkeit in ihrer biologischen und genomischen Organisation zum Mensch sowie der umfangreichen genetischen Ressourcen und Methoden ist die Maus derzeit der wichtigste Säugetier-Modellorganismus in der biomedizinischen Forschung. Durch den Vergleich des Humangenoms mit dem Mausgenom lassen sich Strukturen und Funktionen menschlicher Gene besser verstehen und mögliche regulatorische Elemente identifizieren. Dies unterstützt die Formulierung experimenteller Forschungsansätze und damit die Entwicklung neuer Strategien im Kampf gegen menschliche Krankheiten.

1.3 Funktionelle Genomanalyse

Da der physiologische Status einer Zelle oder eines Organismus' zu einem großen Teil durch die Genaktivitäten bestimmt ist, wird spätestens seit der kompletten Entzifferung des menschlichen Genoms die Analyse des globalen Genexpressionsmusters („expression profiling“) als wichtiger Bereich in der Genomforschung angesehen. Die Kenntnisse über die Änderung der

Genexpression ermöglichen die Untersuchung der genetischen Ursachen. Mit der raschen Entwicklung der Biotechnologie ist es möglich geworden, vollständige Genexpressionsprofile im experimentellen Modell oder in klinischen Proben durch Microarray-Analysen zu erhalten und das Wechselspiel von Krankheitsgenen aufzuklären. Daher ist die systematische Expressionsanalyse eine effektive Methode zur Identifikation potentieller Ziele für die Entwicklung diagnostischer und therapeutischer Methoden.

1.4 Hintergrund des Projektes

Die Arbeitsgruppen aus Jena und München haben in Zusammenarbeit in der ersten Phase des Deutschen Human-Genom-Projektes (DHGP; KW 01-9916) insgesamt 3 Mb in der Xp11.4 und der Xp11.23 Region sequenziert und analysiert. In der Region Xp11.23 wurden Mutationen im *CACNA1F*-Gens (calcium-channel, voltagedependent alpha 1F subunit) identifiziert. Diese Mutationen sind für die erbliche Form der stationären Nachtblindheit (*CSNB*, incomplete X-linked congenital stationary night blindness) verantwortlich (Strom *et al.*, 1998). Darüber hinaus gibt es drei erbliche Krankheiten, für die in Xp11 relevante Gene vermutet wurden: XLMR (X-linked mental retardation; Hedera *et al.*, 2002; Watty *et al.*, 1991), X-gekoppelte SMA (Spinale Muskelatrophie; Kobayashi *et al.*, 1995) und COD1 (X-linked cone dystrophy 1; Seymour *et al.*, 1998).

XLMR äussert sich als geistige Behinderung mit niedrigem IQ und ist oft von körperlichen Defekten begleitet. Die Krankheit kann monogenischer und polygenischer Natur sein und wird durch Mutationen in verschiedenen Genen verursacht. Bisher wurden 137 syndromische und nichtsyndromische XLMR Gene in der Literatur beschrieben. Neuere Untersuchungen (Ramser *et al.* 2004, in Vorbereitung) zeigen am Beispiel von *ATP6AP2*, daß auch Mutationen in einem regulatorischen Element (putative splice enhancer ESE) die Ursache für eine mit Epilepsie verbundene Form von XLMR sein können.

Als SMA wird eine Gruppe menschlicher motorischer Nervenzell-Krankheiten zusammengefasst. Dabei sind die Nervenzellen des Rückenmarks beeinträchtigt und führt zu Muskelschwund (Muskelatrophie), Lähmungen (Paresen) und verminderter Muskelspannung (Muskelhypotonie). Genetisch ist SMA meistens autosomal rezessiv, in einigen klinischen Fällen aber auch autosomal dominant und X-gekoppelt (Figlewicz & Orrell, 2003).

COD1 ist eine Erkrankung der Retina, bei der die primären Zapfen-Photorezeptoren angegriffen und dadurch die Sehschärfe und das Farbsehen gestört werden. Das *COD1*-Gen wurde ursprünglich in der Region Xp11.4 kartiert, durch erneute Untersuchungen unter Verwendung neuer Marker und klinischer Bewertung in drei betroffenen Familien unlängst als Kandidatenregion aber wieder ausgeschlossen (Demirci *et al.*, 2003).

1.5 Ziel, zeitliche und inhaltliche Planung

Um die Struktur und Funktion der oben beschriebenen Krankheitsgene zu identifizieren und zu

charakterisieren, wurde für das Projekt in Zusammenarbeit mit der Münchener Arbeitsgruppe ein Zeitraum von 3-4 Jahren veranschlagt. Innerhalb dieses Zeitrahmens sollten folgende Arbeitsschritte durchgeführt werden:

Ziel des Projektes

- 1) Konstruktion eines kompletten Klon-Contigs in der Region Xp11.4 zwischen Marker *OTC* und *DXS993* in Zusammenarbeit mit der Münchener Arbeitsgruppe,
- 2) Sequenzierung der kartierten Klone (ca. 2 Mb),
- 3) Konstruktion einer kompletten Transkriptionskarte,
- 4) Analyse von Expressionsprofilen,
- 5) Mutationsscreening sowie Analyse von Polymorphismen und alternatives Splicing mit Patienten-DNA,
- 6) Komparative Genomanalyse zwischen Mensch und Maus.

Zeitliche und inhaltliche Planung

Die Projektplanung erfolgte in enger Anlehnung an die Meilensteine des HGP. Insbesondere die Vorverlegung des Abschlusses der Sequenzierarbeiten von 2005 auf April 2003 und die Synchronisierung von DHGP (2. Förderphase) und Nationalem Genomforschungsnetz (NGFN; 1. Förderphase) hatten Einfluss auf die zeitliche und inhaltliche Planung.

Entsprechend wurden folgende Arbeitspakete umgesetzt:

- **01.10.2000-30.09.2001:** Konstruktion des kompletten Kloncontigs und Sequenzierung von 1.6 Mb zuerst in "draft" (4-fache Abdeckung), dann in "full top" (8-fache Abdeckung) Qualität,
- **01.10.2001-30.09.2002:** Fertigstellung von 2.6 Mb basierend auf 16 Klonen in finaler Qualität,
- **01.10.2002-30.04.2003:** automatische und manuelle Annotation aller Datenbank-Einträge der Regionen Xp11.4, Xp11.3 und Xp11.23 und Bestätigung aller annotierten Transkripte durch RT-PCR,
- **01.05.2003-30.09.2004:** Analyse von Expressionsprofilen und Mutationsscreening in Patientenmaterial mittels RT-PCR und Microarray-Technologie.

2. Methoden

2.1 Kartierung eines kompletten BAC/PAC Contigs

Es wurden drei Klonbanken von RPCI-11 (BAC), RPCI-1/3-5 (PAC), CTD (BAC) und X-spezifische STS Marker für das Klon-Screening durch PCR verwendet. Ein Teil der Klone wurde durch „Fluoreszenz *in situ* Hybridisierung“ (FISH) identifiziert. Anschlussklone bzw. lückenschließende Klone wurden durch Abgleich mit BAC-Endsequenzen (<http://tigrblast.tigr.org/erblast/?project=hmrbe>) sowie radioaktive BAC-Filter-Hybridisierung ermittelt.

2.2 Genomische Sequenzierung

Die kartierten BAC/PAC-Klone wurden unter Verwendung von „Nebulizern“ fragmentiert, die Fragmente zwischen 1.5 und 3.0 kb isoliert, in pUC18 Vektor subkloniert (Yanisch-Perron et al., 1985) und ca. 1.000 Subklone pro 100 kb BAC/PAC-Sequenz mit Universalprimern nach Sanger sequenziert (BigDye Terminator Chemie, ABI 377 bzw. 3700 Sequencer). Die Sequenzdaten wurden mit Hilfe des Programms GAP4 assembliert (Bonfield et al., 1995). Die angestrebte Sequenziergenauigkeit betrug mindestens 99,99%. Vier BAC Klone des Contigs (AC108879, AC108880, AC108881 AC108882), wurden am Whitehead Institute for Biomedical Research/MIT (WIBR; Cambridge, USA) bzw. dem Washington University Genome Sequencing Center (WUGSC; St. Louis, USA) in "working draft"-Qualität (HTGS1) generiert und anschließend in Jena entsprechende der internationalen HGP-Koordination in "finished sequence" (Sequenziergenauigkeit >99,99%; weitestgehend lückenlos) überführt.

2.3 Genannotation

Zur Analyse der Sequenzen wurden die Software-Pakete RUMMAGE (IMB, Taudien et al., 2000) und HAVANA (Sanger Institute, Hinxton, UK; <http://www.sanger.ac.uk/HGP/havana/>) verwendet.

Die identifizierten Gene, Pseudogene und Transkripte unterteilen sich in folgende Kategorien:

- (a) bekanntes Gen (**Known gene**): identifiziert durch bekannte humane cDNA- oder Protein-Sequenz(en) mit einem Eintrag in LocusLink oder GDB,
- (b) neuartiges protein-kodierendes Gen (**Novel CDS**): enthält einen ORF (open reading frame; >100 Aminosäuren); basiert auf gesplissenen ESTs und/oder zeigt Homologien zu bekannten Genen/Proteinen anderer Spezies,
- (c) neuartiges Transkript (**Novel transcript**): kein ORF>100 Aminosäuren; basiert auf gesplissenen humanen ESTs und/oder zeigt Homologie zu bekannten Genen/Proteinen anderer Spezies,
- (d) mutmaßliches Transkript (**Putative transcript**): basiert auf gesplissenen humanen ESTs, enthält aber keinen ORF>100 Aminosäuren und zeigt keine Homologie zu bekannten Genen/Proteinen anderer Spezies,
- (e) **Pseudogen**: entsprechen den Kriterien eines bekannten Gens (a), die CDS (protein-coding sequence) enthält aber Stop-Codons und/oder Verschiebungen des Leserasters.

2.4 Analyse der Genexpression

Zur Analyse der Genexpression aller annotierten Transkripte in der Xp11.4 Region wurden insgesamt 16 unterschiedliche humane Gewebe der CLONTECH-MTC Panel I (Hirn, Niere, Herz, Leber, Lunge, Pankreas, Plazenta, Muskel) und II (Milz, Thymus, Prostata, Testis, Ovarien, Dünndarm, Leukozyten, Dickdarm) mit RT-PCR untersucht. Für die Analyse des Genexpressionsprofils im größeren Maßstab mittels DNA-Microarray-Technik wurden I.M.A.G.E

Klone sowie cDNA im Bereich von 12 annotierten Transkripten mittels PCR amplifiziert, die PCR-Produkte gereinigt und danach jeweils zweimal auf Glaträgern immobilisiert. RNA wurde mit Cy3-dUTP bzw. Cy5-dUTP markiert und mit den o.g. Proben hybridisiert.

2.5 Mutationsscreening in Patienten-Material

Drei Patienten- und eine Kontroll-cDNA wurden von der Münchner Gruppe präpariert. Die für die PCR-Amplifikation verwendeten Primer wurden von den cDNA-Sequenzen der 11 ausgewählten Kandidatengene abgeleitet. Anschliessend wurden die PCR-Produkte sequenziert. Die Sequenzen wurden auf Mutationen, Polymorphismen und alternative Transkripte analysiert.

2.6 Komparative Genanalyse

Im Rahmen der komparativen Mensch-Maus-Sequenzanalyse wurden die nachfolgenden, jeweils repeatmaskierten Sequenzen (RepeatMasker; Huang *et al.*, 1990) des UCSC-Browsers (University of California, Santa Cruz; <http://genome.ucsc.edu/>) miteinander verglichen:

- Human: hg17; Mai 2004; chrX:37995708-40655592; 2.659.885 bp
- Maus: mm5; Mai 2004; chrX:8607125-11367363; 2.760.238 bp

Die graphische Auswertung erfolgte mit Dot-Plot und PipMaker (Schwartz *et al.* 2000)(<http://PipMaker.bx.psu.edu/PipMaker/>)

3. Ergebnisse

3.1 Konstruktion und Sequenzierung eines kompletten BAC/PAC Contigs

a. Konstruktion eines kompletten BAC/PAC Contigs

In Zusammenarbeit mit der Münchener Arbeitsgruppe wurde in der Region Xp11.4 ein kompletter BAC/PAC Contig konstruiert. Er überspannt ca. 2 Mb zwischen den Markern DXS8014 und DSX993 und besteht aus 14 Klonen, die zum Teil an anderen Einrichtungen (Whitehead Institute und WU Genome) bereits als "working draft" sequenziert waren und von uns in finale Sequenz überführt wurden. Außerdem wurde mit zwei weiteren Klonen in Xp11.3 eine bestehende Klon-Lücke geschlossen (**Abb. 1**; **Tabelle 1**; IHGSC, 2004, 2001; International Human Genome Mapping Consortium, 2001).

b. Genomische Sequenzierung

Insgesamt wurden 16 Klone mit einer Gesamtlänge von 2,669,435 bp (nicht redundant) in finaler Qualität (lückenlos, Genauigkeit >99,99%) sequenziert und liegen als Genbank- Einträge vor (**Tabelle 1**). Die sequenzierten Klone sind Teil des DHGP-Beitrages zum HGP (IHGSC, 2004, 2001).

3.2 Annotation der Xp11 Region

Die untersuchte Region in Xp11 überspannt 4.26 Mb (nicht redundant) und wird von insgesamt 83 in Jena sequenzierten BACs, PACs und Cosmid-Klonen abgedeckt. 39 Datenbank-Nummern mit der geringsten Überlappung ("minimum tiling path") wurden annotiert. Der Annotationsprozeß erfolgte in zwei Schritten. Im ersten wurden die Sequenzen automatisch durch Abgleich mit EST-, mRNA- und Proteindatenbanken auf ihre Homologien zu bereits bekannten humanen Genen sowie denen anderer Spezies untersucht (Software-Programmpakete Rummage und HAVANA). Auf Grundlage der dabei erhaltenen Resultate wurden anschließend die identifizierten Gene, Transkripte, Pseudogene und CpG-Inseln nach den in **2.3** genannten Kategorien manuell annotiert. Es konnten 55 bekannte Gene, 3 neuartige proteinkodierende Gene, 7 neuartige Transkripte, 4 mutmaßliche Transkripte sowie 25 Pseudogene und 71 CpG Inseln identifiziert werden (Ross *et al.*, akzeptiert). Besonders hohe Zahlen alternativer Transkripte wurden für *BCOR* (10 Varianten) und *WDR45* (17 Varianten) in Xp11.4 und Xp11.23 identifiziert (**Abb. 2**). Für 14 der gefundenen Gene finden sich in der OMIM-Datenbank Assoziationen zu Krankheiten. Die annotierten Datenbank-Nummern bzw. Klone und gefundenen Gene, Transkripte, Pseudogene und CpG Inseln wurden in den **Tabellen 2-5** gelistet.

3.3 Analyse der Genexpression und Mutationsscreening

a. RT-PCR

Für die neun bekannten Gene sowie die drei bisher unbekanntes Transkripte novel-1, -2 und -7 in Xp11.4 wurden RT-PCRs durchgeführt und das Genexpressionsprofil in 16 menschlichen Geweben analysiert (**Abb. 3a-b; Tabelle 6**). Für vier bisher unbekanntes Transkripte (novel-3, -4, -5 und -6) konnte kein PCR-Produkt erhalten werden. Somit ist die Existenz dieser Transkripte nicht bewiesen, kann aber auch nicht ausgeschlossen werden, falls ihre Expression zu schwach ist oder in anderen als den untersuchten Geweben erfolgt (Wen *et al.*, in Vorbereitung).

b. DNA Microarray Technik

135 cDNA Proben (12 aus annotierten Transkripten in Xp11.4 und 123 aus I.M.A.G.E Klonen in der Region Xp21.1-Xp11.21 zwischen den Markern *DMD* und *DXS1204*) wurden mittels DNA-Microarray-Technik auf ihr Genexpressionsprofil untersucht. Die Hybridisierung mit Cy3-dUTP-markierter cDNA (Bestätigung durch zweite Hybridisierung mit Cy5-dUTP-markierten cDNA) zeigt die starke Expression von *SYP* (Synaptophysin), *RBM3*-RNA binding motif protein und *FL11567* (**Abb. 4**).

c. Mutationsscreening in Patienten-Material

Die Arbeiten zur genomischen Kartierung, Sequenzierung und Annotation haben zur Entdeckung von krankheitsverursachenden Gen-Veränderungen bei XLMR (Lenski *et al.*, 2004; Ramser *et al.*,

2004; Ramser *et al.*, in Vorbereitung) bzw. zur Einengung der Kandidatenregion für COD1 (Demirci *et al.*, 2003, 2002) beigetragen.

Darüber hinaus erfolgte in Jena das Mutationsscreening und die Analyse der alternativen Transkripte von elf ausgewählten Kandidatengenen in der Region von Xp11.3 bis Xq12 mit cDNAs von drei Patienten (zwei SMA- und ein XLMR-Patient) und einer Kontroll-cDNA (**Abb. 5, Tabelle 7**). Auffällig ist der synonyme Nukleotidaustausch C228T (GCC>GCT) im *UREB1*-Gen (upstream regulatory element binding protein 1) eines SMA-Patienten. Ob dieser Austausch eine Mutation oder einen Polymorphismus repräsentiert und ob eine Assoziation zu SMA existiert, muss in nachfolgenden Untersuchungen geklärt werden.

3.4 Komparative Genomanalyse zwischen Mensch und Maus

Die Dot Plot- und PipMaker-Analyse von 2,6 Mb nicht redundanter humaner Sequenz aus Xp11.4 im Vergleich zu 2,7 Mb nicht redundanter Sequenz aus der orthologen Maus-Region XqA1.1 zeigt die hohe Ähnlichkeit zwischen beiden Spezies. (**Abb. 6**) Sieben bekannte Gene (*OTC*, *TM4SF2*, *BCOR*, *ATP6AP2*, *CRSP2*, *MIG12*, *MGC39350*) sind in beiden Organismen vorhanden, besonders hohe Konservierung findet sich erwartungsgemäß in den codierenden Regionen (**Abb.7**). Die Ergebnisse konnten durch BLAST-Analysen mit mRNA Sequenzen und Protein- Sequenzen bestätigt werden (Nukleotid-Identität 81-91%, Aminosäure-Identität 84-96%). Zusätzlich wurde mittels RT-PCR für sieben bekannte Gene eine Expressions-Analyse in jeweils zehn Geweben aus Mensch und Maus durchgeführt (**Abb. 8, Tabelle 8**). Für die meisten Gene ergeben sich keine signifikanten Unterschiede zwischen den Expressionsprofilen beider Spezies. Zwei bekannte Gene (*MKRN4* und *LOC286442*) sowie sieben der von uns annotierten, bisher unbekanntes Transkripte konnten in der Maus mittels PipMaker nicht identifiziert werden (**Abb.7**). Unter den sieben Transkripten sind allerdings vier, die auch mittels RT-PCR im Menschen nicht nachgewiesen werden konnten, so daß deren Existenz überhaupt fraglich ist. Das Fehlen der entsprechenden Maus-Sequenzen wurde auch durch entsprechende BLAST-Analysen bestätigt (Wen *et al.*, in Vorbereitung).

4. Zusammenfassung

Das DHGP-Projekt wurde in Zusammenarbeit mit der Münchener Arbeitsgruppe erfolgreich abgeschlossen und hat einen wichtigen Beitrag zum internationalen HGP geleistet. Die geplanten Aufgaben - Sequenzierung, Genanalyse, Genexpression und Mutationsscreening der humanen Xp11 Region wurden im Zeitraum vom 01.10.2000 bis 30.09.2004 erfüllt und sind in acht, überwiegend hochklassigen Journalen (Nature, Am. J. Hum. Genet., J. Med. Genet.) erschienen Publikationen dokumentiert. Zwei weitere Manuskripte sind in Arbeit.

Insgesamt wurden ca. 4,2 Mb Sequenz in finaler Qualität ("finished": Genauigkeit >99,99%, weitestgehend lückenlos) generiert (IHGSC, 2004). Das entspricht etwa 3% des menschlichen X

Chromosoms (Ross *et al.*, akzeptiert). Alle finalen Sequenzen in den Regionen Xp11.4, Xp11.3 und Xp11.23 (insgesamt 39 Datenbankeinträge, denen die Daten von 83 BACs, PACs und Cosmid-Klonen zugrunde liegen) wurden automatisch und manuell annotiert. Die Ergebnisse sind über die vom Sanger-Institut (Hinxton, GB) unterhaltene VegaDB (Vertebrate Genome Annotation Database) öffentlich zugänglich (<http://vega.sanger.ac.uk>). Über die 69 annotierten Transkripte der Region hinaus konnten elf bisher unbekannte Transkripte identifiziert werden. 16 der annotierten Gene in Xp11.4 wurden mittels RT-PCR bestätigt und ihr Genexpressionsprofil in 16 menschlichen Geweben bestimmt (Wen *et al.*, in Vorbereitung). Über die Bestimmung der Genexpressionsprofile mit DNA-Microarray-Techniken konnten drei besonders stark exprimierte Gene identifiziert werden. Die Arbeiten haben zur Entdeckung von krankheitsverursachenden Gen-Veränderungen beigetragen (Lenski *et al.*, 2004; Ramser *et al.*, 2004; Demirci *et al.*, 2003, 2002; Ramser *et al.*, in Vorbereitung) und bilden die Grundlage für nachfolgende krankheitsorientierte Forschungsansätze.

Die komparative Sequenzanalyse zwischen Mensch und Maus bestätigt die erwartete hohe Konservierung der beiden orthologen Regionen zwischen beiden Spezies. Einige Orthologe von im Menschen identifizierten Genen und Transkripten konnten im Maus-Genom nicht gefunden werden. Ob sie in der Maus tatsächlich nicht vorhanden sind oder in anderen Regionen lokalisiert sind, die vom noch lückenhaften "working draft" der Maus bisher nicht abgedeckt werden, kann momentan nicht geklärt werden. Sollte sich das Fehlen einzelner oder aller dieser Gene/Transkripte bestätigen, ergeben sich auch hieraus interessante Ansätze für weitere Untersuchungen (Wen *et al.*, in Vorbereitung).

5. Gesamtkosten

5.1 Die Pauschalkosten für das Projekt

Verbrauchsmittel	165.169,27	Euro
Geschäftsbedarf	672,50	Euro
Dienstreisen	3.031,10	Euro
Personalkosten	351.389,56	Euro
Summe	520.262,43	Euro

5.2 Gesamtkosten für den Projektzeitraum

Jahr 1 (01.10.2000 - 31.12.2000)	25660,04	Euro
Jahr 2 (01.01.2001 – 31.12.2001)	149.068,31	Euro
Jahr 3 (01.01.2002 – 31.12.2002)	136.279,69	Euro
Jahr 4 (01.01.2003 – 31.12.2003)	139.050,38	Euro
Jahr 5 (01.01.2004 – 30.09.2004)	70.204,01	Euro
Summe	520262,43	Euro

6. Erfolgte und geplante Publikationen

Gaiping Wen, Juliane Ramser, Karin Blechschmidt, Adam Frankish, Jennifer Ashurst, Alfons Meindl, Matthias Platzer. Comprehensive manual transcript annotation and experimental verification of a 2.6 Mb medically important Region of the human X chromosome between markers OTC and DXS993. (in progress).

Ramser J, Abidi FE, Burckle CA., Lenski C, Toriello H, Wen G, Lubs HA, Engert S, Stevenson RE, Meindl A, Schwartz CE, Nguyen G. Impairment of the renin receptor causes X-linked mental retardation-epilepsy syndrome. (in progress).

Ross MT et al. The Sequence of the Human X Chromosome. Nature (accepted).

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. Nature 431:931-945.

Lenski C., Abidi F., Meindl A., Gibson A., Platzer M., Frank Kooy R., Lubs H.A., Stevenson R.E., Ramser J., Schwartz C.E.(2004). Novel truncating mutations in the polyglutamine tract binding protein 1 gene (PQBP1) cause reppenning syndrome and X-linked mental retardation in another family with microcephaly. Am. J. Hum. Genet. 74:777-80.

Ramser J., Winnepenninckx B., Lenski C., Errijgers V., Platzer M., Schwartz C.E., Meindl A., Kooy RF. (2004). A splice site mutation in the methyltransferase gene FTSJ1 in Xp11.23 is associated with non-syndromic mental retardation in a large Belgian family (MRX9). J. Med. Genet. 41:679-683.

Demirci FY, Ramser J, White NJ, Rigatti BW, Meindl A, Lewis KF, Wen G, Gorin MB (2003). Refinement of the physical location and the genomic characterization of the CRSP2 (EXLM1) gene on Xp11.4. DNA Seq 14:123-127.

Demirci FY, Rigatti BW, Wen G, Radak AL, Mah TS, Baic CL, Traboulsi EI, Alitalo T, Ramser J, Gorin MB. X-linked cone-rod dystrophy (locus COD1): identification of mutations in RPGR exon ORF15. Am J Hum Genet. 70 (2002)1049-1053.

International Human Genome Mapping Consortium (2001). A physical map of the human genome. Nature 409: 934-41.

International Human Genome Sequencing Consortium (2001). Initial Sequencing and Analysis of the Human Genome. Nature 409: 860-921.

7 .Literatur

Bonfield J. K., Smith K., and Staden R. (1995). A new DNA sequence assembly program. *Nucleic Acids Res* 23: 4992-9.

Demirci F. Y., Ramser J., White N. J., Rigatti B. W., Meindl A., Lewis K. F., Wen G., and Gorin M. B. (2003). Refinement of the physical location and the genomic characterization of the CRSP2 (EXLM1) gene on Xp11.4. *DNA Seq* 14: 123-7.

Figlewicz D. A., and Orrell R. W. (2003). The genetics of motor neuron diseases. *Amyotroph Lateral Scler Other Motor Neuron Disord* 4: 225-31.

Aradhya S., Bardaro T., Galgoczy P., Yamagata T., Esposito T., Patlan H., Ciccodicola A., Munnich A., Kenwick S., Platzer M., D'Urso M., Nelson D.L. (2001). Multiple pathogenic and benign genomic rearrangements occur at a 35 kb duplication involving the NEMO and LAGE2 genes. *Hum Mol Genet* 10: 2557-2567.

Hattori M., Fujiyama A., Taylor T. D., Watanabe H., Yada T., Park H. S., Toyoda A., Ishii K., Totoki Y., Choi D. K., Groner Y., Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D., Reichwald K., Rump A., Schilhabel M., Schudy A., Zimmermann W., Rosenthal A., Kudoh J., Schibuya K., Kawasaki K., Asakawa S., Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S. E., Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brant P., Scharfe M., Schon O., Desario A., Reichelt J., Kauer G., Blocker H., Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E., Haaf T., Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F., Lehrach H., Reinhardt R., and Yaspo M. L. (2000). The DNA sequence of human chromosome 21. *Nature* 405: 311-9.

Hedera P., Alvarado D., Beydoun A., and Fink J. K. (2002). Novel mental retardation-epilepsy syndrome linked to Xp21.1-p11.4. *Ann Neurol* 51: 45-50.

Huang XQ, Hardison RC, Miller W. (1990). A space-efficient algorithm for local similarities. *Comput Appl Biosci.* 6(4):373-81.

Kalaydjieva L., Gresham D., Gooding R., Heather L., Baas F., de Jonge R., Blechschmidt K., Angelicheva D., Chandler D., Worsley P., Rosenthal A., King R.H., Thomas P.K. (2000). N-myc downstream-regulated gene 1 is mutated in hereditary motor and sensory neuropathy-Lom. *Am J Hum Genet* 67: 47-58.

Kioschis P, Wiemann S, Heiss NS, Francis F, Gotz C, Poustka A, Taudien S, Platzer M, Wiehe T, Beckmann G, Weber J, Nordsiek G, Rosenthal A. (1998). Genomic organization of a 225-kb region in Xq28 containing the gene for X-linked myotubular myopathy (MTM1) and a related gene (MTMR1). *Genomics* 1;54(2):256-66.

Kobayashi H., Baumbach L., Matise T. C., Schiavi A., Greenberg F., and Hoffman E. P. (1995). A gene for a severe lethal form of X-linked arthrogryposis (X-linked infantile spinal muscular atrophy) maps to human chromosome Xp11.3-q11.2. *Hum Mol Genet* 4: 1213-6.

Momeni P., Glöckner G., Schmidt O., v. Holtum D., Albrecht B., Gillissen-Kaesbach G., Hennekam R., Meinecke P., Zabel B., Rosenthal A., Hosthemke B., Lüdecke H.J. (2000). Mutations an a novel zinc finger gene cause tricho-rhino-phalangeal syndrome type I. *Nature Genetics* 24: 71-74.

Mouse Genome Sequencing Consortium (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520-562.

Impairment of the renin receptor causes X-linked mental retardation-epilepsy syndrome. Ramser J, Abidi FE, Burckle CA., Lenski C, Toriello H, Wen G, Lubs HA, Engert S, Stevenson RE, Meindl A,

Schwartz CE, Nguyen G. (in progress)

Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W. (2000) PipMaker--a web server for aligning two genomic DNA sequences. *Genome Res.* 10(4):577-86.

Seymour et al. (1998) Linkage analysis of X-linked cone-rod dystrophy: Localization to Xp11.4 and definition of a locus distinct from RP2 and RP3. *Am. J. Hum. Genet.* 62, 122-129.

Strom T.M., Nyakatura G., Apfelstedt-Sylla E., Hellebrand H., Lorenz B., Weber B.H., Wutz K., Gutwillinger N., Ruther K., Drescher B., Sauer C., Zenner E., Meitinger T., Rosenthal A., Meindl(1998). An L-type calcium-channel gene mutated in incomplete X-linked congenital stationary nicht blindness. *Nat Genet.* 19(3):260-3.

Taudien S., Petra Galgoczy, Klaus Huse, Kathrin Reichwald, Markus Schilhabel, Karol Szafranski, Atsushi Shimizu, Shuichi Asakawa, Adam Frankish, Ivan F. Loncarevic, Nobuyoshi Shimizu, Roman Siddiqui and Matthias Platzer (2004). Polymorphic segmental duplications at the defensin locus in human 8p23.1 challenge the assembly of a contiguous reference sequence. *BMC genomics* (in press).

Varon R., Vissinga C., Platzer M., Cerosaletti K.M., Chrzanowska K. H., Saar K., Beckmann G., Seemanova E., Copper P.R., Nowk N.J., Stumm M., Weemäs C.M.R., Digweed M., Rosenthal A., Reis (1998). Nibrin, a novel DNA double-strand break repair protein is mutated in Nijmegen Breakage Syndrome. *Cell* 1;93(3):467-76.

Watty et al. (1991) gene localisation in a family with X-linked syndromal mental retardation (Prieto syndrome). *Am. J. Med. Genet.* 38, 234-239.

Yanisch-Perron C., Vieira J., and Messing J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33: 103-19.

Der Abschlußbericht wurde erstellt durch die Projektverantwortliche Dr. Gaiping Wen.

Die sachliche Richtigkeit und Vollständigkeit der Angaben bestätigt:

Jena, den 21.12.2004

Dr. Matthias Platzer
Projektleiter

8. Anhang

Tabellen

- Tabelle 1 Klone und Sequenzen in der Region Xp11.4 zwischen den Markern DXS8113-DXS993
- Tabelle 2 Annotierte Accession-Nummern bzw. Klone und gefundene Gene, Transkripte, Pseudogene und CpG Inseln
- Tabelle 3 Annotierte Transkripte in der Region Xp11.4
- Tabelle 4 Annotierte Transkripte in der Region Xp11.3
- Tabelle 5 Annotierte Transkripte in der Region Xp11.23
- Tabelle 6 Expressionsprofile von sechzehn annotierten Transkripten in sechzehn humanen Geweben in der Region Xp11.4 mittels RT-PCR
- Tabelle 7 Mutationsscreening von elf ausgewählten Kandidatengenen
- Tabelle 8 Vergleich der Nukleotid- und Aminosäuresequenzen zwischen human und Maus

Abbildungen

- Abb. 1 Physikalische Klon- und Genkarte und GC Verteilung in der Region Xp11.4 zwischen OTC und DXS993. Klone werden als schwarze Linien mit dem Klonnenamen, Gene als schwarzer Balken mit der Orientierung gezeigt. Die G+C Verteilung ist als Kurve mit einer Fensterlänge von 100 kb (Schrittweite 10 kb) dargestellt.
- Abb. 2. Annotierte alternative Transkripte in der Region Xp11.4. Accession Nummern zeigen alternativ gesplissene Transkripte. 5'/3' gibt die Orientierung der Transkripte an. Exons werden durch schwarze Balken dargestellt, die verlängerte und verkürzte Exonbereiche rote bzw. weiße Balken.
- Abb. 3 RT-PCR Expressionsprofile annotierter Transkripte in der Region Xp11.4 in CLONTECH-MTC-Panels (a) I und (b) II von humanen Geweben.
- Abb. 4 Expressionsprofile durch Microarray-Technologie. Die Genexpression von FL11567, SYP und RBM3 zeigt nach der Hybridisierung eindeutig positive Signale.
- Abb. 5 Mutation T>C im Gen UREB1 zwischen Patienten cDNA (P) und Kontroll-cDNA (K).
- Abb. 6 Dot-Plot-Vergleich zwischen 2.6 Mb der menschlichen Region Xp11.4 und 2,7 Mb der Mausregion XqA1.1. Die hohe Ähnlichkeit zwischen Mensch- und Maussequenz wird durch die Diagonale repräsentiert, die G+C Verteilung durch die schwarze (Human) bzw. graue (Maus) Kurven (Fenster 100 kb, Schrittlänge 10 kb).
- Abb. 7 Pip-Maker-Analyse. Die Länge der Transkripte wird durch schwarze Linie dargestellt, die Orientierung durch den Richtungspfeil bzw. gelb (-) oder rot (+) gekennzeichnete Bereiche.
- Abb. 8 Expressionsprofile von sieben bekannten Genen in zehn Mensch- und Maus-Geweben mittels RT-PCR.