

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.