

Die *Fused toes* Mausmutation –
Identifizierung und Charakterisierung der Kandidatengene

Inaugural - Dissertation

zur

Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von
Thomas Peters
aus Peine

Düsseldorf
2002

Gedruckt mit der Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent : Prof. Dr. Ulrich Rüter
Korreferent : Prof. Dr. Uwe A.O. Heinlein

Tag der mündlichen Prüfung : 07. Mai 2002

Inhaltsverzeichnis

1	Einleitung	1
1.1	Die strukturelle Analyse von komplexen Genomen	1
1.2	Die Isolierung von phänotyp-assoziierten Genen	5
1.3	Die funktionelle Genomanalyse	7
1.4	Die Mausmutation <i>Fused toes</i>	9
1.4.1	Der Phänotyp heterozygoter <i>Fused toes</i> -Mäuse	9
1.4.2	Der Phänotyp homozygoter <i>Fused toes</i> -Embryonen	11
1.4.3	Die bisherige molekulare Analyse der <i>Fused toes</i> Mutation	14
1.5	Das Ziel dieser Arbeit	15
2	Material und Methoden	17
2.1	Chemikalien, Enzyme und sonstige Materialien	17
2.3	Hefestämmen und Nährmedien	18
2.5	Mäuse und Tierhaltung	19
2.6	Vektoren und Oligonukleotide	19
2.6.1	Plasmide	20
2.6.2	BAC-, PAC- und YAC-Klone	21
2.6.3	Oligonukleotide ('Primer')	22
2.7	DNA-Bibliotheken	25
2.8	Antikörper	26
2.9	Datenbanken	26
2.10	Isolierung von Nukleinsäuren	28
2.10.1	Präparation von Plasmid-DNA aus <i>E. coli</i>	28
2.10.2	Präparation von BAC- und PAC-DNA aus <i>E. coli</i>	29
2.10.3	Präparation von Gesamt-DNA aus Hefe	30
2.10.4	Präparation von genomischer DNA aus Mausgeweben	31
2.10.5	Präparation der Gesamt-RNA aus Mausgeweben	32
2.10.6	Präparation der Gesamt-RNA aus eukaryotischen Zelllinien	33
2.10.7	Konzentrationsbestimmung von Nukleinsäurelösungen	33
2.11	Klonierungstechniken	34
2.11.1	Sequenzspezifische Restriktionsspaltung von DNA	34
2.11.2	Gelelektrophoretische Auftrennung von DNA	34
2.11.3	Isolierung einzelner DNA-Fragmente aus Agarosegelen	35
2.11.4	Reinigung und Fällung von DNA	36
2.11.5	Ligation von DNA-Fragmenten	36
2.11.6	Herstellung kompetenter <i>E. coli</i> -Zellen für die Transformation	37
2.11.7	Transformation von kompetenten <i>E. coli</i> -Bakterien	38
2.12	Sequenzierung von DNA	38
2.13	Nachweis spezifischer DNA- und RNA-Sequenzen	39
2.13.1	Southern-Blot: Transfer von DNA	39
2.13.2	Northern-Blot: Transfer von RNA	40
2.13.3	Radioaktive Markierung von DNA-Sonden	41
2.13.4	Sequenzhomologie-Hybridisierung	42

2.14	Polymerase-Kettenreaktion	42
2.14.1	Standard-PCR.....	43
2.14.2	RT-PCR.....	43
2.14.3	Semiquantitative RT-PCR.....	45
2.14.4	RACE (' <i>Rapid Amplification of cDNA Ends</i> ').....	45
2.14.5	Reinigung von PCR-Reaktionsprodukten.....	47
2.15	Isolierung von Exon-Sequenzen aus genomischer DNA – 'Exon trapping'	47
2.16	<i>in vitro</i>-Transkriptions- und Translationsreaktion	48
2.17	SDS-Polyacrylamid-Gelelektrophorese	49
2.18	Zellkulturtechnik	51
2.18.1	Kultivierung und Passagieren von Zellen.....	51
2.18.2	Kryokonservierung von Zellen.....	51
2.18.3	Bestimmung der Gesamtzellzahl.....	52
2.18.4	Transfektion.....	52
2.19	Fluoreszenzmikroskopie	53
2.19.1	Indirekte Immunfluoreszenz.....	53
2.19.2	Direkte Immunfluoreszenz.....	54
2.20	<i>in situ</i> Hybridisierung von Ganzkörperpräparaten	55
2.20.1	Synthese der RNA-Sonde.....	55
2.20.2	Präparation und Vorbehandlung der Maus-Embryonen.....	56
2.20.3	Vorbehandlung und Hybridisierung der Embryonen.....	56
2.20.4	Antikörperinkubation und Nachweis der hybridisierten RNA-Sonden.....	57
3	Ergebnisse	58
3.1	Die 'Exon-trapping' Analyse des BAC-Klons MB3	58
3.2	Die Positionsklonierung von <i>Fantom (Ftm)</i>, ein zuvor unbekanntes Gen	59
3.2.1	Die Isolierung der vollständigen <i>Ftm</i> cDNA.....	60
3.2.2	Das <i>Ftm</i> Protein.....	63
3.2.3	Die Expression von <i>Ftm</i> im Embryo und im ausgewachsenen Tier.....	65
3.3	Die Positionsklonierung von <i>Fatso (Fto)</i>, ein zuvor unbekanntes Gen	67
3.3.1	Die Isolierung der vollständigen <i>Fto</i> cDNA.....	68
3.3.2	Die Analyse der <i>Fto</i> cDNA-Sequenz.....	70
3.3.3	Das <i>Fto</i> Protein.....	70
3.3.4	Die Bestimmung der subzellulären Lokalisation des <i>Fto</i> Proteins.....	72
3.3.5	Die Klonierung der menschlichen <i>FTO</i> cDNA.....	73
3.3.6	Die Sequenz der überlappenden Promotorregionen von <i>Fto</i> und <i>Ftm</i>	76
3.3.7	Die Expression von <i>Fto</i> im Embryo und im ausgewachsenen Tier.....	76
3.4	Die Identifizierung von <i>Ft</i>-Kandidatengenen durch die vergleichende Genomanalyse	80
3.5	Das Homöobox-Gen <i>Irx3</i> ist ein weiteres Kandidatengen für den <i>Ft</i>-Phänotyp	83
3.5.1	Die Mitglieder der <i>Iroquois</i> Genfamilie des Menschen und der Maus.....	84
3.5.2	Die Identifizierung von <i>Irx5</i> , ein zuvor unbekanntes Mitglied der <i>Irx</i> -Genfamilie der Maus.....	86
3.5.3	Die Identifizierung von <i>Irx6</i> , ein zuvor unbekanntes Mitglied der <i>Irx</i> -Genfamilie der Maus.....	88
3.5.4	Die Anordnung der <i>Irx</i> Gene der Maus in zwei Genkomplexen.....	88
3.5.5	<i>Irx5</i> und <i>Irx6</i> sind innerhalb des deletierten Genomabschnitts lokalisiert.....	91
3.6	Die Identifizierung von Kandidatengenen in den flankierenden Genomabschnitten	92
3.6.1	Die Kartierung des <i>Mmp2</i> Gens im distalen flankierenden Genomabschnitt.....	95
3.6.2	Die Kartierung des <i>KIAA0308/2FLJ</i> Gens im proximalen flankierenden Genomabschnitt.....	97

3.7	Menschliche Erbkrankheiten im Kandidatenbereich HSA 16q12-q13	98
3.7.1	Die Klonierung des <i>Sall1</i> Gens der Maus	101
3.7.2	Die Kartierung des <i>Sall1</i> Gens der Maus	101
3.7.3	Die Expression von <i>Sall1</i> zum Zeitpunkt der Ausprägung des <i>Ft</i> -Gliedmaßenphänotyps	103
3.8	Die Auswertung der Konsensussequenz des relevanten menschlichen Genomsegments	105
3.9	Die Betrachtung des <i>Fused toes</i> Locus unter evolutionären Aspekten	107
3.9.1	Die phylogenetische Analyse der <i>Iroquois</i> Genfamilie	107
3.9.2	Die Existenz der beiden <i>Irx</i> Genkomplexe beruht auf der Duplikation eines ausgedehnten Genomsegments	109
4	Diskussion	112
4.1	Die Kandidatengene im Bereich des deletierten Genomabschnitts	113
4.1.1	<i>Fantom (Ftm)</i>	114
4.1.2	<i>Fatso (Fto)</i>	118
4.1.3	Die <i>Iroquois</i> Gene <i>Irx3</i> , <i>Irx5</i> und <i>Irx6</i>	121
4.2	Die Kandidatengene in den flankierenden Genomabschnitten	127
4.2.1	<i>KIAA0308/2FLJ</i>	127
4.2.2	<i>Mmp2</i>	129
4.2.3	<i>Sall1</i>	132
4.3	Die Evolution der <i>Irx</i> Genfamilie der Wirbeltiere	133
4.4	Die Konsequenzen der <i>en masse</i> Duplikation für den <i>Ft</i>-Locus	137
4.5	Die konservierte Syntänie zwischen Mensch und Maus im Bereich des <i>Ft</i>-Locus	139
4.6	Ein Ausblick auf die weitere Vorgehensweise	140
5	Zusammenfassung	145
6	Literaturverzeichnis	147
7	Anhang	163
7.1	Abkürzungen	163
7.2	Sequenzen	165
7.2.1	<i>Ftm: Mus musculus</i> Fantom	165
7.2.2	<i>Fto: Mus musculus</i> Fatso	166
7.2.3	<i>FTO: Homo sapiens</i> Fatso	167
7.2.4	<i>Irx3: Mus musculus Iroquois</i> Homöobox Gen 3	168
7.2.5	<i>Irx5: Mus musculus Iroquois</i> Homöobox Gen 5	169
7.2.6	<i>Irx6: Mus musculus Iroquois</i> Homöobox Gen 6	170
7.2.7	<i>Mmp2: Mus musculus</i> Matrix Metalloproteinase 2	170
7.2.8	<i>KIAA0308/2FLJ: Homo sapiens</i> KIAA0308/2FLJ	171
7.2.9	<i>Sall1: Mus musculus sal (Drosophila)-like 1</i>	173

1 Einleitung

Die Identifizierung und Kartierung der Gene im Erbgut (Genom) verschiedener Spezies ist ein wichtiger Meilenstein auf dem Weg zum Verständnis der komplexen genetischen Wirkungsgefüge, die aus einer einzigen undifferenzierten Zelle einen hoch komplexen Organismus erwachsen lassen. So lassen sich mit Hilfe der aus der Genomanalyse gewonnenen Erkenntnisse grundlegende Prinzipien der molekularen und zellulären Interaktionen in lebenden Organismen erkennen. Im besonderen Maße trägt die Charakterisierung des menschlichen Genoms und auch des Genoms einiger Modellorganismen zur Aufklärung der molekularen Ursachen von genetisch bedingten Erkrankungen des Menschen bei. Die Genomforschung ist daher essentiell für die Entwicklung von Medikamenten, Impfstoffen und Diagnostika für zum Teil bis heute unheilbare Krankheiten wie z.B. Diabetes, Atherosklerose, Asthma, Alzheimer oder Krebs. Aber auch in anderen Bereichen wie der Landwirtschaft oder der Lebensmittelindustrie nimmt die Genomanalyse eine immer stärkere Rolle ein.

1.1 Die strukturelle Analyse von komplexen Genomen

Die Voraussetzung für die Identifizierung und Kartierung von einzelnen Genen in einem komplexen Genom ist die Verfügbarkeit detaillierter genetischer und physikalischer Karten. Genetische Karten geben einen statistischen Abstand zweier Orte (Loci) im Genom wieder und ermöglichen durch umfangreiche Kopplungsanalysen die relative Anordnung von mehreren Loci entlang eines Chromosomenabschnitts. Je dichter zwei Loci auf einem Chromosom beieinander liegen desto seltener werden sie durch einen Austausch (Rekombination) getrennt. Die statistische Rekombinationshäufigkeit kann somit als Maß für die Distanz zwischen zwei Loci herangezogen werden. Dabei beträgt der genetische Abstand zweier Loci, die eine Rekombinationshäufigkeit von einem Prozent zeigen, genau ein centiMorgan (cM). Dies entspricht im menschlichen Genom durchschnittlich einer physikalischen Distanz von 1Mb (Stephens *et al.*, 1990), während im Mausgenom durch 1cM im Durchschnitt 2Mb überspannt werden (Dietrich *et al.*, 1996). Allerdings ist die exakte Beziehung zwischen Rekombinationshäufigkeit und tatsächlichem Abstand von mehreren Faktoren abhängig. Dazu zählen beispielsweise die Nähe zu sogenannten 'Hot Spots' bzw. zum Telomerbereich und auch das Geschlecht (Petes, 2001).

Die effektive genetische Kartierung eines Genoms erfolgt durch die Verankerung von polymorphen genetischen Markern, deren Neukombination aufgrund unterschiedlicher Allele leicht nachgewiesen werden kann. Dabei werden vor allem Restriktionsfragment-

längenpolymorphismen (RFLPs) und auch Mini- und Mikrosatellitenmarker verwendet. Während den RFLPs Sequenzunterschiede in Restriktionsschnittstellen zugrunde liegen, beruhen Satellitenmarker auf tandemartigen, sich wiederholenden Sequenzen aus Di-, Tri- oder Tetranukleotideinheiten (Mikrosatelliten) bzw. aus Einheiten mit einer Länge von 15 bis 100 Nukleotiden (Minisatelliten). Die Anzahl dieser repetitiven Einheiten variiert dabei von Allel zu Allel. Der daraus resultierende Größenunterschied zwischen den einzelnen Allelen ('*simple-sequence length polymorphism*' – SSLP) kann leicht mit Hilfe spezifischer Primer aus den flankierenden Sequenzen in einer Polymerasekettenreaktion (PCR) bestimmt werden. Da insbesondere Mikrosatellitenmarker sehr häufig und gleichmäßig verteilt in komplexen Genomen zu finden sind, konnten für den Menschen hochauflösende genetische Karten bestehend aus 5826 (Murray *et al.*, 1994) bzw. 5264 (Dib *et al.*, 1996) polymorphen Markern mit einem durchschnittlichen Abstand von 0,7cM bzw. 1,6cM etabliert werden. Desgleichen wurden dichte genetische Karten auch für die Maus (Dietrich *et al.*, 1996; Rhodes *et al.*, 1998) und den Zebrafisch *Danio rerio* etabliert (Shimoda *et al.*, 1999).

Durch den verstärkten Einsatz von Einzelnukleotidpolymorphismen ('*single nucleotide polymorphisms*' – SNPs), die beispielsweise im menschlichen Genom fast gleichmäßig etwa alle 300 bis 1000bp auftreten (Cargill *et al.*, 1999), können noch sehr viel dichtere genetische Karten erstellt werden. Erste auf SNPs basierende genetische Karten stehen mittlerweile nicht nur für den Menschen (Wang *et al.*, 1998; Cargill *et al.*, 1999) sondern auch für die Maus (Lindblad-Toh *et al.*, 2000) und andere Modellorganismen (Koch *et al.*, 2000; Hoskins *et al.*, 2001) zur Verfügung. SNP-Karten werden aufgrund ihrer extrem hohen Auflösung in Zukunft die Isolierung von krankheitsrelevanten Genen erheblich erleichtern (Riley *et al.*, 2000). Das Ziel ist es jedoch, polymorphe Marker zu identifizieren, deren Allele kausal mit einem Merkmal oder einer Krankheit verknüpft sind. Dadurch gewinnen SNPs auch für die Diagnose bzw. Prognose komplexer Krankheiten sowie für die Voraussage der individuellen Reaktion auf pharmakologische Substanzen oder aber auch auf ultraviolettes Licht, Zigarettenrauch und bestimmte Diäten zunehmend an Bedeutung (Schork *et al.*, 2000).

Im Gegensatz zu den statistischen Distanzen der genetischen Karten werden in physikalischen Karten die tatsächlichen räumlichen Abstände zwischen zwei Loci ermittelt. Auf zytogenetischer Ebene erfolgte dies lange Zeit durch die grobe Kartierung eines Locus in Hinsicht auf das spezifische Bandenmuster, die Größe oder andere zytologische Merkmale des jeweiligen Chromosoms. Der Einsatz der Fluoreszenz *in situ* Hybridisierung (FISH) von Metaphasechromosomen mit markierten DNA-Sonden erlaubte später auch die präzise Bestimmung der relativen Anordnung mehrerer Loci mit einer Auflösung von wenigen Megabasen. Mittlerweile hat die effiziente Weiterentwicklung dieser Technik ('*Fiber-FISH*', '*released chromatin-FISH*') nochmals zu einer deutlichen Verbesserung der Auflösung bis in den Kilobasen-Bereich geführt (Raap, 1998). Besonders geeignet ist diese Methode für die

Darstellung von Chromosomenumordnungen wie z.B. Insertionen, Deletionen, Duplikationen und Translokationen.

Die Strahlungshybridkartierung (*'Radiation Hybrid (RH) Mapping'*) eignet sich ebenfalls sehr gut für die Erstellung von physikalischen Karten. In diesem Fall lässt sich die Entfernung zwischen verschiedenen Loci durch die Analyse von Mensch/Hamster- bzw. Maus/Hamster-Hybridzelllinien ermitteln. Allerdings kann wegen der begrenzten Auflösung dieser Methode nicht die exakte Abfolge entlang eines Chromosoms festgestellt werden. Die fusionierten Zellhybride enthalten aufgrund der zuvor erfolgten radioaktiven Bestrahlung der Donorzellen einzelne definierte Bruchstücke des Genoms einer zu untersuchenden Spezies. Auf diese Weise entsteht eine Kollektion von großen genomischen Fragmenten, deren Gehalt an sogenannten sequenzmarkierten Stellen (*'sequence-tagged sites'* – STS), Sequenzen die nur einmal im Genom vorkommen und zu denen unter anderen auch die bereits erwähnten SSLPs gehören, leicht mittels PCR analysiert und statistisch ausgewertet werden kann. Je häufiger dabei ein Bruch zwischen zwei STSs festgestellt wird, desto größer ist ihr physikalischer Abstand. Als Maßeinheit für diese Bestrahlungskarten dient das centiRay (cR). Eine Entfernung von 1cR_x bezieht sich auf eine Bruchfrequenz von 1% nach einer Bestrahlung mit x rad radioaktiver Strahlung. Dies entspricht in der RH-Karte des Mausgenoms im Durchschnitt einer Distanz von 100kb (Van Etten *et al.*, 1999). Weitere detaillierte RH-Karten sind für das Genom des Menschen (Deloukas *et al.*, 1998) und des Zebrabärblings (Geisler *et al.*, 1999) verfügbar. Dabei wurden in der menschlichen RH-Karte insgesamt 41664 STSs kartiert. Da sich darunter auch Mikrosatelliten-Marker befanden, konnte auf diese Weise ebenfalls ein direkter Bezug zu der bereits erwähnten genetischen Karte (Dib *et al.*, 1996) hergestellt werden. Ebenso wurden in diese Karte auch STSs integriert, die von bekannten Genen oder kurzen exprimierten sequenzmarkierten Stellen (*'expressed sequence tag'* – EST; Adams *et al.*, 1991) abgeleitet worden waren. So entstand eine erste physikalische Karte von mehr als 30.000 vermeintlichen menschlichen Genen (Deloukas *et al.*, 1998).

Eine weitere deutliche Verbesserung der Auflösung wird durch die Zusammenstellung von einzelnen überlappenden genomischen Fragmenten in einem sogenannten Contig (*'contiguous set of clones'*) erreicht. Innerhalb eines Contigs liegen die klonierten DNA-Fragmente in derselben ursprünglichen Reihenfolge wie im Genom vor. Voraussetzung für diese Form der Kartierung ist jedoch die Verfügbarkeit geeigneter Bibliotheken, welche die gesamten genomischen Fragmente eines Genoms enthalten. Die gesuchten Fragmente können durch die Verwendung von spezifischen Markern wie z.B. die zuvor beschriebenen STS-Marker (*'STS content mapping'*) oder auch mit Hilfe von isolierten Randsequenzen eines genomischen Fragments identifiziert und somit bestimmten Regionen eines Chromosoms zugeordnet werden. Auf diese Weise können auch unbekannte Genombereiche durch die gezielte

Erweiterung eines Contigs in einer Chromosomenwanderung (*'chromosome walking'*) erschlossen und lückenlos abgedeckt werden.

Für die Klonierung von großen DNA-Fragmenten stehen verschiedene Vektoren zur Verfügung, die sich vor allem in ihrer Kapazität, Stabilität und Handhabung bei der DNA-Gewinnung deutlich voneinander unterscheiden. Plasmide, Cosmide, das P1-Bakteriophagen-System, künstliche Bakterienchromosomen (*'bacterial artificial chromosome'* - BAC) und vom P1 Phagen abgeleitete künstliche Chromosomen (*'P1 derived artificial chromosome'* - PAC) können für die Klonierung größerer Fragmente eingesetzt und sehr effizient aus bakteriellen Wirtsorganismen (*E. coli*) präpariert werden. Sehr häufig kommen BACs und PACs zum Einsatz, da diese Systeme DNA-Fragmente bis zu einer Größe von 300kb aufnehmen. Die größten Klonierungskapazität (bis zu 2000kb) besitzen jedoch künstliche Hefechromosomen (*'yeast artificial chromosome'* - YAC), die in eukaryotischen Hefezellen (*S. cerevisiae*) vermehrt werden. Allerdings treten in diesem System häufig Umlagerungen und Deletionen im klonierten DNA-Fragment auf. Ein hoher Anteil von YAC-Klonen beinhaltet auch chimäre, artifiziell zusammengesetzte DNA-Fragmente, welche die Erstellung von physikalischen Karten unter Umständen erheblich behindern.

Physikalische Karten, die auf einem aus Genomfragmenten zusammengesetzten Contig beruhen, wurden zunächst vor allem für zahlreiche kleinere Genomregionen erstellt. In einigen Fällen ermöglichte diese Vorgehensweise die Identifizierung von bestimmten krankheits-assoziierten Genen (The Huntington's Disease Collaborative Research Group, 1993; Miki *et al.*, 1994). Später wurden auch gesamte komplexe Genome durch ausgedehnte YAC-Contigs physikalisch kartiert. Auf diese Weise konnte bisher das Genom des Menschen und der Maus nahezu vollständig in einer auf YACs basierenden physikalischen Karte erfasst werden (Hudson *et al.*, 1995; Nusbaum *et al.*, 1999).

Die Bestimmung der exakten Nukleotidabfolge in einem komplexen Genom führt letztendlich zu einer vollständigen physikalischen Karte, in der die Position eines jeden Locus direkt abgelesen werden kann. Die nahezu komplette Sequenzierung des Genoms einiger Modellorganismen wie z.B. der Hefe *S. cerevisiae* (Goffeau *et al.*, 1996), des Fadenwurms *C. elegans* (The *C. elegans* Sequencing Consortium, 1998) und der Fruchtfliege *Drosophila melanogaster* (Adams *et al.*, 2000) führte bereits zur Identifizierung zahlreicher zuvor unbekannter Gene. Kurz vor dem Abschluss dieser Arbeit wurde durch die Arbeiten des Humangenomprojektes und des US-Unternehmens *Celera Genomics* schließlich auch das menschliche Genom bis auf einige wenige Bereiche nahezu vollständig entschlüsselt (The International Human Genome Sequencing Consortium (IHGSC), 2001; Venter *et al.*, 2001). Somit sind die Sequenzen und Positionen fast aller menschlichen Gene öffentlich zugänglich. Eine komplette Sequenz des Mausgenoms wurde bisher jedoch noch nicht veröffentlicht. Allerdings generiert ein weltweites öffentliches Mausgenomprojekt eine ständig wachsende

Anzahl von kurzen Teilsequenzen, die bereits sehr gut für die Identifizierung und Kartierung von Genen im Mausgenom genutzt werden können. Im Rahmen dieses Projekts wird die Veröffentlichung der vollständigen Sequenz eines Großteils des Mausgenoms für das Jahr 2002 erwartet (Graham *et al.*, 2001; Rogers & Bradley, 2001).

1.2 Die Isolierung von phänotyp-assoziierten Genen

Mit Ausnahme der computergestützten Sequenzanalyse von ganzen Genomen können Gene nur durch das Auftreten von Mutationen identifiziert werden. Im Verlauf der letzten Jahre wurden verschiedene Strategien für die Identifizierung von mutierten Genen entwickelt. Die jeweilige Vorgehensweise ist dabei abhängig vom Informationsstand über die jeweilige Mutation. Beispielsweise bedient man sich bei der Analyse von Krankheiten, die auf bestimmten biochemischen Defekten beruhen, oft der funktionellen Klonierung. Dabei werden alle bekannten Informationen über die Funktion eines Gens für die Genisolierung herangezogen. Die gesuchten Gene können somit aufgrund von bekannten Proteinsequenzen oder durch das Durchsuchen einer cDNA-Expressionsbibliothek mit spezifischen Antikörpern isoliert werden. Eine ähnliche Vorgehensweise führte zum Beispiel im Fall der Hämophilie B zur funktionellen Klonierung des Faktor IX-Gens (Giannelli *et al.*, 1983).

Kennt man die genaue chromosomale Lokalisation eines mutierten Locus, so lässt sich das betroffene Gen im Zuge einer sogenannten Positionsklonierung isolieren. Dazu muss die Mutation zunächst über die Identifizierung von genetischen Markern, die eng an den mutierten Locus gekoppelt sind, so gut wie möglich genetisch kartiert werden. Dieser sehr zeitaufwendige und arbeitsintensive Prozess wird aufgrund der Etablierung von immer dichter werdenden genetischen Karten zusehends beschleunigt. Nachdem der mutierte Locus einer chromosomalen Region zugeordnet werden konnte, erfolgt in der nächsten Phase die physikalische Kartierung des relevanten Bereichs. Dazu wird in einer Chromosomenwanderung ein Contig von Genomfragmenten in der Regel auf der Basis von BAC-, PAC- oder YAC-Klonen erstellt, der die identifizierten Marker auf beiden Seiten der Mutation verbindet. Auf diese Weise wird sichergestellt, dass sich der gesuchte Locus im klonierten Intervall befindet. Anschließend werden die isolierten Genomfragmente mit Hilfe verschiedener Verfahren auf die Anwesenheit von transkribierten Sequenzen untersucht. Dabei werden die traditionellen Techniken wie z.B. das Durchmustern von genomischen Fragmenten nach evolutionär konservierten Sequenzen in einem Zoo-Blot oder die Suche nach CpG-Inseln, kurze hypomethylierte GC-reiche Sequenzen, die oft im 5' Bereich von Wirbeltiergenen gefunden werden können, immer mehr durch raffinierte Methoden ersetzt. Dazu gehört beispielsweise die '*cDNA Selection*'-Technik, bei der die gesuchte cDNA

aufgrund einer mehrfach wiederholten Hybridisierung des isolierten genomischen Fragments mit einer cDNA-Bibliothek angereichert wird (Lovett *et al.*, 1991). Voraussetzung dafür ist allerdings die Kenntnis der räumlichen und zeitlichen Expression des gesuchten Gens, da dieses in der ausgewählten cDNA-Bibliothek enthalten sein muss. Mit dem 'Exon trapping'-System steht eine weitere Methode für die gezielte Isolierung von transkribierten Sequenzen aus komplexen Genomabschnitten zur Verfügung. Diese Methode beruht auf der Selektion von funktionellen Spleiß-Akzeptor- bzw. Spleiß-Donorstellen (Exon-Intron-Übergänge), so dass keine Vorkenntnisse über die spezifische Expression des gesuchten Gens benötigt werden (Buckler *et al.*, 1991). Aufgrund dieses Vorteils hat sich das System bereits bei der Positionsklonierung mehrerer krankheitsrelevanter Gene bewährt (The Huntington's Disease Collaborative Research Group, 1993; Miki *et al.*, 1994; Zhang *et al.*, 1994).

Natürlich kann auch der gesamte relevante Genomabschnitt sequenziert und die erhaltene Sequenz mit Hilfe einer computergestützten Suche nach signifikanten Homologien zu bereits bekannten Genen der untersuchten oder auch anderer Spezies durchsucht werden. Andere Programme wurden speziell für die Lokalisation von potentiellen Exon-Sequenzen innerhalb einer analysierten Genomsequenz entwickelt. Zum Beispiel führte die Verwendung derartiger Programme im Rahmen des Humangenomprojekts zu der Aussage, dass im menschlichen Genom annähernd 38.000 Gene enthalten sind (IHGSC, 2001; Venter *et al.*, 2001).

Eine Methode in der sowohl funktionelle, strukturelle als auch alle anderen verfügbaren Informationen für die Isolierung von phänotyp-assoziierten Genen kombiniert werden, steht mit dem Kandidatengenansatz zur Verfügung. Bei dieser Vorgehensweise werden Gene beispielsweise wegen ihres spezifischen Expressionsmusters oder aufgrund von Schlussfolgerungen über den vorliegenden biochemischen Defekt als Kandidaten für eine Mutation ausgewählt. Ebenso führt auch die Ausprägung eines analogen Phänotyps in einer anderen Spezies und der in diesem Zusammenhang beschriebenen Gene zur Auswahl eines vielversprechenden Kandidatengens. Ausführliche Kartierungsstudien ermöglichen eine effiziente Überprüfung der erwähnten Kandidatengene, da Gene, die nicht mit der Mutation gekoppelt im Genom vorliegen, sofort ausgeschlossen werden können. Da im Rahmen der verschiedenen Genomprojekte immer genauere, hochauflösende genetische und physikalische Karten erstellt werden, gewinnt dieser positionelle Kandidatengenansatz zunehmend an Bedeutung. Dabei werden alle bekannten Gene eines betroffenen Genomabschnitts zunächst näher charakterisiert und somit einzelne Kandidatengene mit dem beobachteten Phänotyp in Verbindung gebracht. Eine besondere Rolle spielt hier die vergleichende Genomanalyse. Der Vergleich von Genkarten verschiedener Säugetierspezies enthüllte eine erstaunliche gute Konservierung der relativen Anordnung von Genen über weite Bereiche der verglichenen Genome. So teilen z.B. Maus und Mensch mindestens 183 konservierte Genomabschnitte (Nadeau & Taylor, 1984; IHGSC, 2001). Innerhalb dieser Abschnitte ist die Abfolge der

enthaltenen Gene annähernd vollkommen erhalten geblieben (konservierte Syntanie). Besonders verdeutlicht wurde die konservierte Syntanie bei der Kartierung von Genen auf dem X-Chromosom höherer Säuger. Dabei zeigte sich, dass die X-Chromosomen von Maus, Katze und Mensch die gleichen Gene (orthologe Gene) enthalten. Die relative Anordnung der Gene auf diesem Chromosom scheint zwischen Katze und Mensch sogar vollständig erhalten geblieben zu sein (Murphy *et al.*, 1999). Jedoch wurde die relative Abfolge von Gengruppen zwischen Maus und Mensch durch Umordnungen (Inversionen) in mindestens neun homologe Segmente verändert (Carver & Stubbs, 1997). Entsprechende konservierte Syntaniegruppen lassen sich ebenfalls als kürzere aber auch längere Teilstücke auf Autosomen nachweisen. Aufgrund der konservierten Syntanie von Gengruppen zweier Spezies können die erstellten Genkarten somit auch über Speziesgrenzen hinweg genutzt werden. Die vergleichende Genomanalyse kann daher die positionelle Klonierung von Genen erheblich erleichtern.

Ist das gesuchte Kandidatengen erst einmal vollständig kloniert, so müssen in den weiteren Analysen Veränderungen im isolierten Gen nachgewiesen werden, die verantwortlich für die Ausprägung des beobachteten Phänotyps sind. Im einfachsten Fall zeigen strukturelle Analysen, dass ein Teilabschnitt oder sogar das gesamte Gen im Genom des untersuchten Organismus deletiert ist. Für die Aufklärung von subtilen Mutationen stehen beispielsweise die Analyse von Einzelstrang-Konformationspolymorphismen (*'single-strand conformational polymorphism'* - SSCP) oder auch die Sequenzanalyse des identifizierten Gens zur Verfügung. Arbeitet man mit Modellorganismen, so kann die Kausalität auch durch die gezielte Inaktivierung des Kandidatengens oder durch die Komplementierung (*'rescue'*) des beobachteten Phänotyps in transgenen Tieren belegt werden (Probst *et al.*, 1998; Giraldo & Montoliu, 2001).

1.3 Die funktionelle Genomanalyse

Lange Zeit konnten Gene nur aufgrund von Mutationen und der damit einhergehenden Beeinträchtigung ihrer Funktion entdeckt werden. Dabei war man vor allem auf das Auftreten von spontanen Mutationen angewiesen, deren auffälliger Phänotyp über mehrere Generationen hinweg in Kreuzungsexperimenten beobachtet werden konnte. Daher wurde in Modellorganismen schon bald mit der Herstellung von künstlichen Mutationen begonnen, die mit Hilfe von Röntgenstrahlen oder Chemikalien erzeugt werden können. Die anschließende Analyse der Phänotypen führte zu der Identifizierung einer Vielzahl von Genen. So hatte zum Beispiel die chemische Mutagenese des Genoms der Fruchtfliege *D. melanogaster* einen großen Anteil an der Aufklärung von entwicklungsbiologisch relevanten Mechanismen (Nüsslein-Vollhard & Wieschaus, 1980). Großformatige Mutagenesen mit der Chemikalie

Äthylnitrosoharnstoff (ENU) werden zur Zeit auch beim Zebrafisch *D. rerio* (Haffter *et al.*, 1996) und der Maus (Hrabé de Angelis *et al.*, 2000; Nolan *et al.*, 2000) durchgeführt.

Mittlerweile werden im Rahmen der verschiedenen Genomprojekte immer mehr Gene identifiziert, denen auf Anhieb keine Funktion zugeordnet werden kann. Natürlich wird versucht, die biologischen Funktionen aller Gene eines Organismus zu bestimmen. Beispielsweise werden systematisch alle bekannten Gene des Fadenwurms *C. elegans* mit Hilfe der RNA-vermittelten Geninterferenz (RNAi) gezielt inaktiviert (Fraser *et al.*, 2000; Göncy *et al.*, 2000). Diese Technik der Genaktivierung wurde inzwischen auch für die Fruchtfliege (Misquitta & Paterson, 1999), die Maus (Wianny & Zernicka-Goetz, 2000) und sogar für menschlichen Zellen etabliert (Elbashir *et al.*, 2001). Im Zebrafisch kann die Funktion einzelner Gene mit der verwandten Morpholino-Technik ebenfalls gezielt beeinträchtigt werden (Nasevicius & Ekker, 2000). In der Maus erreicht man den vollständigen Funktionsverlust eines Gens über einen längeren Zeitraum nur mit Hilfe der sogenannten 'Gene-Targeting'-Technik (Capecchi, 1989; Joyner, 1993). Dabei wird eine zielgerichtete Mutation durch homologe Rekombination in das gewünschte Gen von embryonalen Stammzellen eingebracht, die dann für die Herstellung von Mausmutanten genutzt werden. Die ständige Weiterentwicklung dieser Technik erlaubt inzwischen die gezielte Umordnung von ganzen Chromosomen oder auch die Induktion des Funktionsverlusts zu einer bestimmten Zeit in einem bestimmten Gewebe (Ramirez-Solis *et al.*, 1995; Rajewsky *et al.*, 1996).

Eine Methode bei der die Aufklärung der biologischen Funktion mit der Identifizierung von unbekanntem Genen kombiniert wird, steht mit der 'Gene-trap'-Technik zur Verfügung. Dabei führt die ungerichtete Integration eines Transgens in das Genom zur Unterbrechung des betroffenen Gens und somit zum Funktionsverlust. Da mit dieser insertionellen Mutagenese ein exogenes DNA-Stück im mutierten Gen platziert wird, ist es möglich die flankierenden genomischen Sequenzen zu klonieren und so das betroffene Gen zu isolieren. Dieses Verfahren wird inzwischen nicht nur in der Maus sondern auch im großem Maßstab für die Genidentifizierung in *C. elegans*, *D. melanogaster* und in *D. rerio* verwendet (Übersicht in Durick *et al.*, 1999).

Die Maus eignet sich im besonderen Maße als Modellorganismus für die Aufklärung der biologischen Funktionen von Genen. Das Mausgenom lässt sich auf molekularer Ebene sehr gut manipulieren, so dass mit der 'Gene-trap'- und 'Gene-targeting'-Technik sowie der ENU-Mutagenese einige genetische Verfahren für funktionelle Studien zur Verfügung stehen (Übersicht bei Denny & Justice, 2000). Ebenso ermöglicht die Analyse von spontanen oder künstlich hergestellten Mausmutanten die Identifizierung von zuvor unbekanntem Genen. Da die Maus dem Menschen evolutionär und damit auch genetisch sehr nahe steht und zwischen Mensch und Maus große Übereinstimmungen hinsichtlich der Embryonalentwicklung,

Stoffwechselwege, Physiologie, Immunologie und Pathologie bestehen, leisten die in der Maus identifizierten Gene bzw. die im Mausmodell gewonnen Erkenntnisse einen großen Beitrag bei der Aufklärung der genetischen Grundlage und der Pathologie von menschlichen Krankheiten.

1.4 Die Mausmutation *Fused toes*

Die *Fused toes* (*Ft*) Mausmutation entstand im Verlauf der Etablierung eines transgenen Tiermodells, mit dem ursprünglichen Ziel ein aktiviertes menschliches *Harvey-ras* (*Ha-ras*) Onkogen unter der Kontrolle des menschlichen β -Interferon Promotors in das Mausgenom einzubringen. Die Integration des verwendeten DNA-Konstrukts bewirkte in einer transgenen Mauslinie die Ausprägung eines komplexen Phänotyps (van der Hoeven *et al.*, 1994). Da das *Ha-ras* Transgen selbst jedoch nicht in den betroffenen Mäusen exprimiert wird, also nicht aktiv ist, können die im folgenden näher erläuterten phänotypischen Veränderungen einzig und allein auf die insertionelle Mutagenese des betroffenen *Ft*-Locus zurückgeführt werden.

1.4.1 Der Phänotyp heterozygoter *Fused toes*-Mäuse

Der markante Phänotyp von heterozygoten *Ft*-Mäusen äußert sich morphologisch zunächst durch die Verwachsung der ersten drei oder der ersten vier Zehen an den vorderen Extremitäten (Abb. 1.1; van der Hoeven *et al.*, 1994). Kennzeichnend für diese Syndaktylie ist sowohl der Erhalt des Zwischenzehengewebes (kutane Syndaktylie) als auch die Verschmelzung der distalen Phalangen der einzelnen Zehen (ossäre Syndaktylie). In keinem Fall konnte jedoch eine Verwachsung mit der posterioren fünften Zehe beobachtet werden. Im Zuge der Entwicklung der Extremitäten erfolgt die Ausbildung von Fingern und Zehen durch die kontrollierte Aktivierung des programmierten Zelltods im Gewebe zwischen den einzelnen Fingerstrahlen und einer damit verbundenen Gewebsauflösung (Saunders & Fallon, 1967). Die Verminderung des programmierten Zelltods in diesem Gewebe zeigt an, dass dieser Prozess offensichtlich in den vorderen Extremitäten heterozygoter *Ft*-Mäuse gestört ist (van der Hoeven *et al.*, 1994). Auf molekularer Ebene ließ die Aktivitätsbestimmung von Genen, welche eine wesentliche Rolle bei der Gliedmaßenentwicklung spielen, keinerlei Beeinträchtigungen der grundlegenden Musterbildung in den sich entwickelnden vorderen Extremitäten erkennen. Allerdings zeigten einige der untersuchten Markergene (z.B. *Bmp4*, *Msx2*, *Noggin* und *Fgf8*) eine räumlich und zeitlich abweichende Expression im antero-

distalen Bereich der vorderen Extremitäten. Dies führt offenbar zu einer Unausgewogenheit zwischen der Bmp- und der Fgf-vermittelten Signaltransduktion, so dass es zu einer vorzeitigen Rückbildung des distalen Signalzentrums der Gliedmaßenanlagen, der sogenannten epithelialen Randleiste ('*apical ectodermal ridge*' - AER), kommt (Heymer & Rüther, 1999).

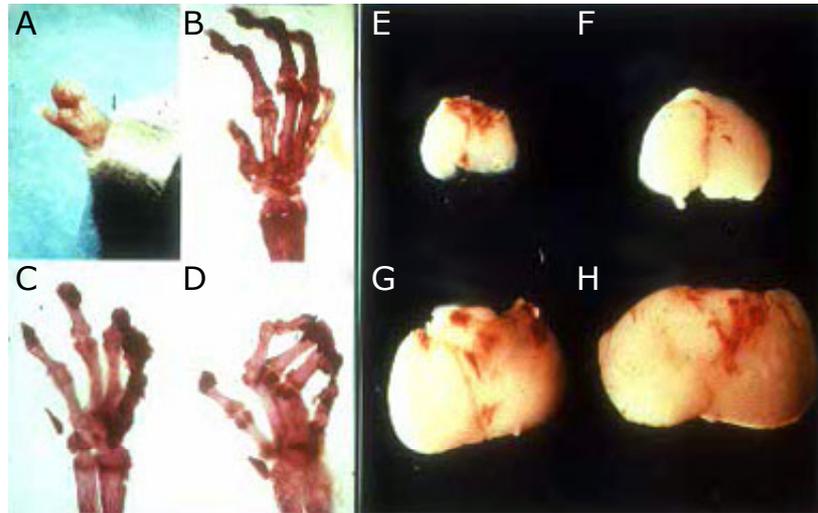


Abb. 1.1: Der Phänotyp heterozygoter *Ft*-Mäuse. **A:** Syndaktylie an der Vorderpfote einer heterozygoten *Ft*-Maus mit Verwachsungen zwischen den ersten vier Zehen. **B-D:** Knochenfärbungen der rechten Vorderpfote einer Wildtyp-Maus (B) und zweier heterozygoter *Ft*-Mäuse (C und D) offenbaren eine ossäre Syndaktylie zwischen den ersten drei (C) bzw. vier Zehen (D). **E:** Thymus einer sechs Wochen alten Wildtyp-Maus, der zu diesem Zeitpunkt seine maximale Ausdehnung erreicht hat. **F-H:** Hyperplasie des Thymus bei einer sechs Wochen (F), vier Monate (G) und sechs Monate (H) alten heterozygoten *Ft*-Maus. (Abb. aus van der Hoeven *et al.*, 1994)

Ausgewachsene heterozygote *Ft*-Mäuse fallen außerdem durch eine massive Vergrößerung (Hyperplasie) ihres Thymus auf (Abb. 1.1). Normalerweise erreicht der Thymus einer Wildtyp-Maus etwa sechs Wochen nach der Geburt seine maximale Ausdehnung. Danach wird das Wachstum eingestellt, und mit zunehmenden Alter nimmt die Größe des Thymus immer mehr ab. Der Thymus einer heterozygoten *Ft*-Maus unterliegt offenbar nicht dieser Wachstumsbeschränkung und erreicht durch sein ungebremstes, kontinuierliches Wachstum eine voluminöse Ausdehnung. Diese führt schließlich zu einer erheblichen Beeinträchtigung der Lungenfunktion und damit nach etwa zehn Monaten zum frühzeitigen Tod der Tiere. Wiederum kann dieser Phänotyp mit einer gestörten Regulation des programmierten Zelltods in Zusammenhang gebracht werden. So sind unreife Thymozyten von heterozygoten *Ft*-Mäusen weit weniger empfänglich für die *in vitro* Induktion des programmierten Zelltods als entsprechende Thymozyten, die aus Wildtyp-Mäusen isoliert wurden (van der Hoeven *et al.*, 1994). Transplantationsexperimente legen jedoch nahe, dass die beobachtete Hyperplasie nicht

auf Thymozyten, sondern auf Epithelzellen zurückzuführen ist (Volkman *et al.*, 1996). Einige heterozygote *Ft*-Mäuse zeigen neben der erwähnten Thymushyperplasie ebenfalls eine erhebliche Vergrößerung der Milz (Volkman *et al.*, 1996).

1.4.2 Der Phänotyp homozygoter *Fused toes*-Embryonen

Homozygote *Ft*-Embryonen sterben aufgrund schwerer Fehlbildungen bereits im Verlauf der Embryonalentwicklung zwischen Tag 10.0 p.c. und Tag 14.5 p.c., also in einem Zeitraum, in dem in etwa die Hälfte der Trächtigkeit bereits erreicht wurde. Diese Embryonen weisen im Vergleich zu gleichaltrigen Wildtyp-Embryonen ein deutlich verlangsamtes Wachstum auf (van der Hoeven *et al.*, 1994). Besonders drastischen Missbildungen unterliegt der Gesichts- und der vordere Gehirnbereich, so dass bei einem schweren Ausprägungsgrad sowohl das Gesicht als auch das Vorder- und Mittelhirn fast völlig unausgebildet vorliegen kann (Abb. 1.2-A). Ebenfalls stark betroffen sind die sich entwickelnden Gliedmaßen. Im Gegensatz zu heterozygoten *Ft*-Embryonen sind hier alle drei Achsen der Vorder- und Hinterbeinknospen von Fehlbildungen betroffen. So sind die Extremitäten in ihrer proximodistalen Achse stark verkürzt (Abb. 1.2-B,C). Am distalen Ende der Knospen erfolgt im anterioren Bereich der Aufbau einer massiven Gewebeschwulst. Dies führt zu einer enormen Verbreiterung der dorsoventralen sowie der anteroposterioren Achse. Eine Ausnahme bildet dabei die posteriore Region, welche den zukünftigen fünften Zeh umgibt. Dieser Bereich ist offenbar nicht von den eben erwähnten Fehlbildungen betroffen, so dass die Knochenelemente dieses Zehs ganz normal angelegt werden können. Alle anderen Zehenanlagen sind hingegen stark verkürzt. Beispielsweise fehlen die Anlagen für die distalen Phalangen der ersten vier Zehen völlig. In einigen Fällen kann auch eine dorsoventrale Verdopplung der Metacarpalia-Anlagen dieser verkürzten Zehen beobachtet werden (Abb. 1.2-D), was zu der Ausbildung von zwei übereinanderliegenden Kondensationsebenen mit jeweils vier Kondensationsstrahlen führt (Grotewold, 2001). Im Gegensatz dazu ist trotz einer ebenso deutlichen Verbreiterung der anteroposterioren Achse keinerlei präaxiale oder postaxiale Polydaktylie auf dieser Ebene erkennbar. Wie bereits erwähnt sind die Vorder- und Hinterbeinknospen gleichermaßen von den aufgeführten Missbildungen betroffen. Allerdings bilden die Hinterbeinknospen nicht die beschriebene dorsoventrale Verdopplung der Metacarpalia-Anlagen aus. Dafür fehlen zum Teil die Anlagen für einige anteriore Extremitätenelemente (Abb. 1.2-C), wie z.B. der erste und zweite Zeh oder die Tibia (Grotewold, 2001).

Ein weiteres markantes Merkmal der homozygoten *Ft*-Embryonen zeigt sich im Verlust der genetischen Kontrolle über die Ausbildung der Rechts-Links-Asymmetrie der Körperachse (Heymer *et al.*, 1997). Im Zuge der Herzentwicklung findet eine normalerweise zuerst nach

rechts gerichtete Bewegung des primitiven Herzschauchs statt. Dagegen lässt sich in homozygoten *Ft*-Embryonen auch eine vergleichbare Bewegung des Herzschauchs nach links erkennen (Abb. 1.2-E). Dies sollte zu einer spiegelbildlichen Anlage des Herzens auf der rechten Seite der Körperachse führen, was allerdings aufgrund des frühzeitigen Todes dieser Embryonen nicht beobachtet werden kann. Gleichermäßen erfolgt die embryonale Rotation ebenso zu einer der beiden Seiten. Dabei handelt es sich um eine für Nager spezifische Drehung, bei der sich der frühe Embryo in seine Membranen einwickelt, und die zu einer Platzierung des Schwanzes auf der rechten Seite des Rumpfes führt. Als Folge der zufällig festgelegten Rotationsrichtung befindet sich in einigen homozygoten *Ft*-Embryonen der Schwanz auch auf der linken Seite des Rumpfes (Abb. 1.2-E). Die Expressionsanalysen der asymmetrisch exprimierten Markergene *nodal* und *lefty2* untermauerten die morphologischen Beobachtungen. Beide Gene werden in Wildtyp-Embryonen im lateralen Plattenmesoderm der linken Körperhälfte exprimiert. Hingegen zeigten homozygote *Ft*-Embryonen auch eine beidseitige oder sogar eine fehlende Expression dieser beiden Gene (Heymer *et al.*, 1997).

Nicht zuletzt offenbarte die Expressionsanalyse von Markergenen im zentralen Nervensystem eine Desorganisation der ventralen Region des sich entwickelnden Neuralrohrs (Dr. J. Briscoe, persönliche Mitteilung). So spricht die fehlende Expression der beiden Gene *Shh* und *Hnf3 β* im ventralen Neuralrohr für einen Verlust der Bodenplatte (Abb. 1.2-F,G). In homozygoten *Ft*-Embryonen wird dieser Bereich von Zellen bevölkert, die normalerweise nur in einer weiter dorsal gelegenen Region zu finden sind, aus der später die Motorneuronen hervorgehen. Allerdings treten diese Veränderungen nur in der kaudalen Körperregion auf.

Insgesamt ist der Ausprägungsgrad der verschiedenen Missbildungen sehr variabel und zudem abhängig vom genetischen Hintergrund der Mäuse. Beispielsweise überleben homozygote *Ft*-Embryonen mit einem gemischten C57BL/6xC3H-Hintergrund wesentlich länger (bis Tag 14.5 p.c.) als vergleichbare Embryonen mit dem ursprünglichen C57BL/6xSJL-Hintergrund (bis Tag 10.5 p.c.). Desgleichen weisen heterozygote *Ft*-Mäuse mit einem angereicherten C57BL/6-Hintergrund auch eine kutane Syndaktylie zwischen der zweiten und vierten Zehe der hinteren Extremitäten auf (Prof. Dr. U. Rüter, persönliche Mitteilung).

Alles in allem bietet die *Ft*-Mutation eine ausgezeichnete Gelegenheit Gene zu identifizieren, die sowohl eine entscheidende Rolle bei der Entwicklung der Extremitäten, des Gesichtsbereichs und des zentralen Nervensystems als auch bei der Kontrolle des programmierten Zelltods und der Etablierung der Rechts-Links-Asymmetrie der Körperachse spielen.

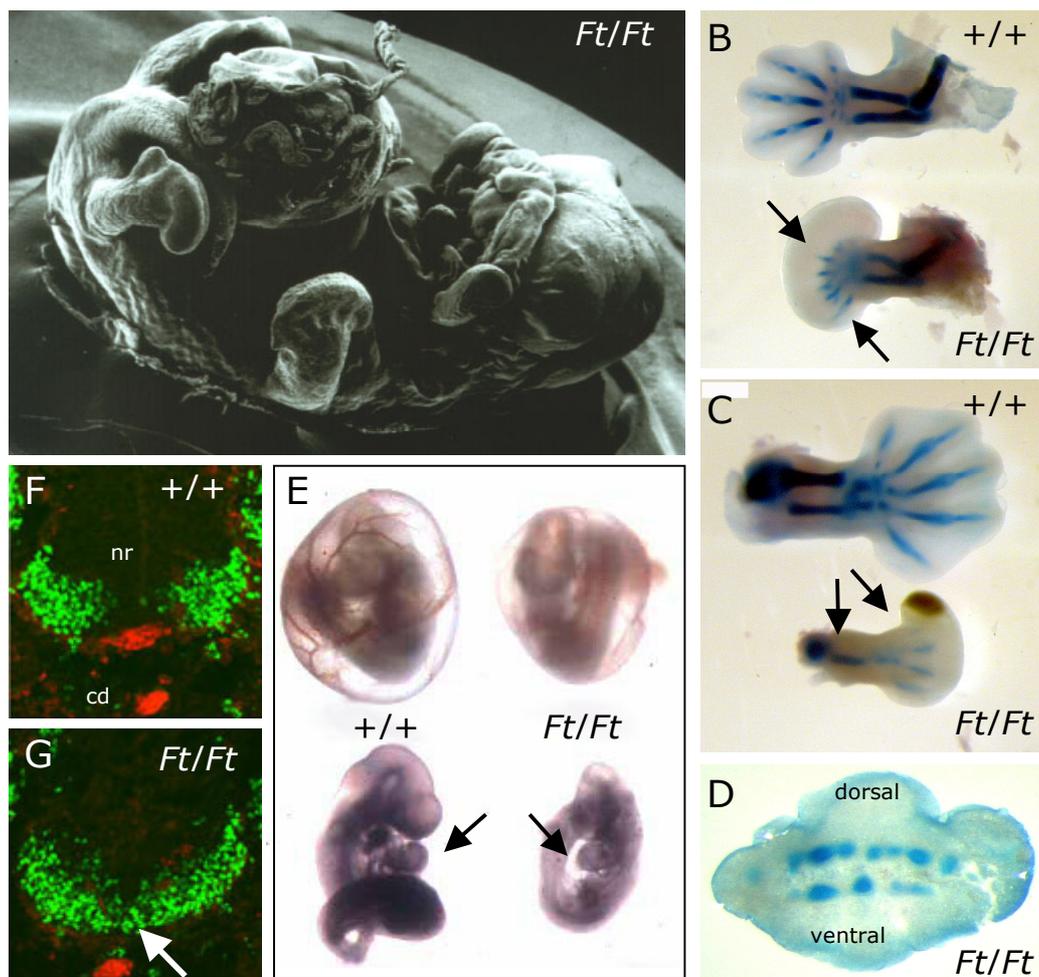
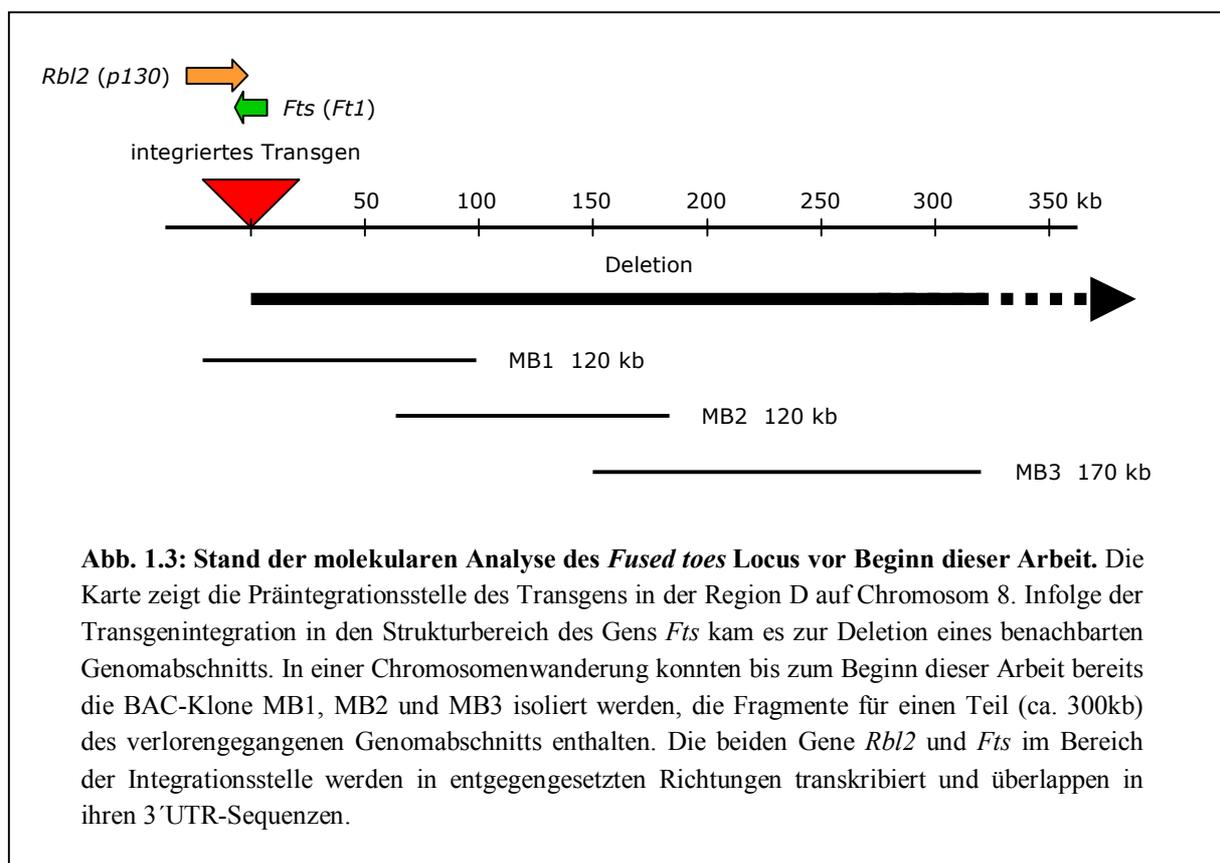


Abb. 1.2: Der Phänotyp homozygoter *Ft*-Embryonen. **A:** Elektronenmikroskopische Aufnahme eines 11.5 p.c. Tage alten homozygoten *Ft*-Embryos (*Ft/Ft*). Deutlich zu erkennen sind die Fehlbildungen im Gesichtsbereich und die massive Gewebeschwulst im antero-distalen Bereich der Gliedmaßen. **B und C:** Knochen/Knorpelfärbung der vorderen (B) und hinteren Gliedmaßen (C) von 14.5 p.c. Tage alten Wildtyp- (+/+) und homozygoten *Ft*-Embryonen (*Ft/Ft*). In den Gliedmaßen der *Ft*-Embryonen werden die distalen Phalange der ersten vier Zehenstrahlen nicht angelegt. Teilweise fehlen in den hinteren Gliedmaßen die Anlagen für die Tibia und die ersten beiden Zehen. **D:** Der Transverssschnitt durch die vordere Extremität eines 14.5 p.c. Tage alten homozygoten *Ft*-Embryos lässt die dorso-ventrale Verdopplung der Metacarpalia-Anlagen erkennen. **E:** Wildtyp- (links) und homozygote *Ft*-Embryonen (rechts) vor (oben) und nach (unten) der Entfernung der äußeren Membranen. Während der Schwanz von Wildtyp-Embryonen immer rechts vom Rumpf platziert wird, befindet sich der Schwanz bei einigen *Ft*-Embryonen auch auf der linken Seite. Desgleichen ist in einigen homozygoten *Ft*-Embryonen auch die Entwicklung des Herzens in der rechten Körperhälfte zu erkennen. **F und G:** Immunhistologische Färbung des ventralen Neuralrohrs (nr) und der *Chorda dorsalis* (cd) eines 11.5 p.c. Tage alten Wildtyp-Embryos (F) und eines gleichaltrigen homozygoten *Ft*-Embryos (G). *Shh*-exprimierende Zellen (rot) sind in der Bodenplatte des Neuralrohrs und in der *Chorda dorsalis* sichtbar. Dagegen wird der Bereich der Bodenplatte im Neuralrohr von nach ventral expandierenden *Isl1*-exprimierenden Zellen (grün) eingenommen.

(Abb. A,E: Dr. J. Heymer; Abb. B,C,D: Dr. L. Grotewold; Abb. F,G: Dr. J. Briscoe)

1.4.3 Die bisherige molekulare Analyse der *Fused toes* Mutation

Da die *Ft* Mutation auf der Integration eines Transgens in das Mausgenom beruht, konnten mit Hilfe einer transgen-spezifischen Sonde flankierende genomische Sequenzen von einer Seite der Integrationsstelle aus dem Genom von heterozygoten *Ft*-Mäusen isoliert werden. Die *in situ* Hybridisierung von Wildtyp-Metaphasechromosomen mit einer nahe der Integrationsstelle gelegenen Sonde (HPP500) zeigte, dass der betroffene Locus in der Region D des Maus Chromosoms 8 lokalisiert ist (van der Hoeven *et al.*, 1994). Allerdings ließ die weitere molekulare Charakterisierung der Mutation sehr bald erkennen, dass infolge der Transgenintegration auch benachbarte Genomsequenzen verlorengegangen sind (Abb. 1.3). Das genaue Ausmaß dieser Deletion konnte bis zum Beginn dieser Arbeit nicht eindeutig bestimmt werden. Eine erste Chromosomenwanderung zeigte aber, dass der deletierte Bereich eine Ausdehnung von mehr als 200kb aufweist (Ausmeier, 1997). Im Zuge einer parallel zu dieser Arbeit durchgeführten physikalischen Kartierung des *Ft* Locus wurden insgesamt 52 Genomfragmente (BACs und PACs) in einem Contig zusammengefasst, der die gesamten verlorengegangenen Wildtyp-Sequenzen enthält. Vom Gesamtumfang dieses Contigs lässt sich somit auf eine Größe der Deletion von annähernd 1650kb schließen (Ausmeier, 2000).



Die Suche nach betroffenen Genen führte zunächst zu der Identifizierung des bis dahin unbekanntes Gens *Fts/Ft1* (Abb. 1.3; Lesche *et al.*, 1997). Dabei zeigten strukturelle Analysen, dass in *Ft*-Mäusen durch die Integration des Transgens in den Strukturbereich von *Fts* große Teile dieses Gens (Exon 1-10) verlorengegangen sind. Die Expression von *Fts* in den Zellen heterozygoter *Ft*-Mäuse ist daher im Vergleich zum Wildtyp um 50% reduziert. In homozygoten *Ft*-Zellen ist *Fts* dementsprechend vollkommen inaktiv (Lesche *et al.*, 1997). Dennoch ließen funktionelle Studien bisher keinerlei Korrelation von *Fts* mit der Ausprägung des beschriebenen *Ft*-Phänotyps erkennen. Beispielsweise unterscheiden sich Embryonen, die aufgrund der gezielten Inaktivierung von *Fts* nur ein nichtfunktionelles Allel dieses Gens enthalten, nicht von ihren Wildtyp-Geschwistern (Lesche, 1997). Da *Fts* jedoch während der Embryogenese zum Zeitpunkt der Ausprägung des *Ft*-Phänotyps in den beeinträchtigten Strukturen exprimiert wird (z.B. in den Gliedmaßen und im Neuralrohr; Lesche, 1997), ist es durchaus möglich, dass der Verlust von *Fts* in Kombination mit dem Funktionsverlust von anderen betroffenen Genen ebenfalls zur Ausprägung des *Ft*-Phänotyps beiträgt.

Weiterhin konnte mit *Rbl2/p130* ein Mitglied der Retinoblastoma (*Rb*)-Genfamilie von Zellzyklusregulatoren in unmittelbarer Nähe der Transgenintegrationsstelle lokalisiert werden (Lesche *et al.*, 1997). Interessanterweise zeigen *Rbl2* und das entgegengesetzt orientierte Gen *Fts* eine evolutionär konservierte Überlappung in ihren 3' untranslatierten Regionen (Lesche & Rüther, 1998). Allerdings ist dieser Genabschnitt mehr als 1kb von der Integrationsstelle entfernt. Der transkribierte Strukturbereich von *Rbl2* ist also nicht von der *Ft*-Deletion betroffen. Da in umfangreichen Studien keine Abweichung der *Rbl2* Expression vom Wildtyp-Muster festgestellt werden konnte, scheint auch *Rbl2* nicht an der Ausprägung des *Ft*-Phänotyps beteiligt zu sein (Peetz, 1997).

1.5 Das Ziel dieser Arbeit

Der komplexe Phänotyp der *Ft*-Mäuse zeigt, dass die von der Mutation betroffenen Gene eine Schlüsselposition in grundlegenden molekularen, zellulären und physiologischen Prozessen vor allem während der Embryonalentwicklung einnehmen. Dabei lässt sowohl das Auftreten von dominanten und rezessiven Phänotypen als auch die Größe des deletierten Genomabschnitts von 1650kb den Schluss zu, dass die Funktion mehrerer Gene beeinträchtigt sein muss. Tatsächlich konnten mit *Rbl2* und *Fts* vor Beginn dieser Arbeit bereits zwei Kandidatengene am *Ft*-Locus identifiziert werden. Allerdings bewiesen die durchgeführten Studien, dass beide Gene bei der Ausprägung der komplexen Fehlbildungen in heterozygoten und homozygoten *Ft*-Mäusen wenn überhaupt nur eine untergeordnete Rolle spielen. Der vorliegende Phänotyp muss daher auf der Mutation von anderen Genen beruhen.

In dieser Arbeit sollten möglichst alle Gene im Bereich des *Fused toes* Locus sowohl innerhalb des deletierten Genomabschnitts als auch in den flankierenden Bereichen identifiziert werden. Dies sollte durch die Positionsklonierung, insbesondere durch die 'Exon-trapping'-Analyse der zu Beginn dieser Arbeit verfügbaren Genomfragmente aus dem betroffenen Bereich erfolgen. Gleichzeitig sollte dabei auch der positionelle Kandidatengenansatz vor allem durch die Auswertung von genetischen und physikalischen Karten des menschlichen Genoms verfolgt werden. Im Falle der Identifizierung von bisher völlig unbekannt Genen, sollten diese durch die Isolierung der vollständigen cDNA und die anschließende Auswertung der abgeleiteten Aminosäuresequenz näher charakterisiert werden. Zudem sollte eine erste Expressionsanalyse zur Bewertung der Kandidatengene beitragen. Schließlich sollten die identifizierten Gene in einer Transkriptkarte des *Ft*-Locus zusammengefasst und in die bestehenden genetischen und physikalischen Karten des Mausgenoms integriert werden.

2 Material und Methoden

2.1 Chemikalien, Enzyme und sonstige Materialien

Chemikalien in der Qualität reinst oder *pro analysis*, Enzyme, Kit-Systeme und sonstige Materialien wurden soweit nicht anders vermerkt von folgenden Firmen bezogen:

Amersham Pharmacia (Buckinghamshire, Großbritannien), Applichem (Darmstadt), J.T. Baker (Deventer, Niederlande), Biorad (München), Clontech (Heidelberg), Dianova (Hamburg), Difco (Detroit, USA), Fluka (Buchs, Schweiz), GibcoBRL Life Technologies (Karlsruhe), Invitrogen (Groningen, Niederlande), Kodak (New Haven, USA), MBI Fermentas (St. Leon-Rot), Merck (Darmstadt), New England Biolabs (Schwalbach), Promega (Heidelberg), Qiagen (Hilden), Riedel-de Hâen (Seelze), Roche (Mannheim), Roth (Karlsruhe), Schleicher & Schüll (Dassel), Serva (Heidelberg), Sigma-Aldrich (Steinheim), Stratagene (Heidelberg).

Radiochemikalien wurden von den Firmen Amersham Buchler (Braunschweig, [$\alpha^{32}\text{P}$] dCTP) und ICN Biomedicals (Cleveland, USA, Tran ^{35}S -label™) bezogen.

Sämtliche Lösungen für molekularbiologische Arbeiten wurden nach Sambrook *et al.* (1989) in bidestilliertem Wasser angesetzt und nach Bedarf autoklaviert oder sterilfiltriert.

2.2 Bakterienstämme und Nährmedien

Es wurden ausschließlich Derivate des *E. coli* Stammes *K12* verwendet (Tab. 2.1).

<i>E. coli</i> K12 Derivat	Verwendung
<i>JM109</i>	Standardwirt für diverse konstruierte Plasmide
<i>XL1 blue</i>	Wirt der verwendeten cDNA-Bibliotheken
<i>DH10B</i>	Wirt der verwendeten cDNA-, BAC-, und PAC-Bibliotheken
Tab. 2.1: eingesetzte Bakterienstämme	

Folgende Medien wurden durch den Zusatz der entsprechenden Antibiotika als Selektionsmedium eingesetzt:

- LB-Medium: 1% (w/v) NaCl
1% (w/v) Trypton
0,5% (w/v) Hefeextrakt
- LB-Agar: LB-Medium
1,5% (w/v) Agar
- Antibiotika: Ampicillin (50µg/ml), Kanamycin (25µg/ml), Chloramphenicol (12,5µg/ml)

2.3 Hefestämme und Nährmedien

YAC-Vektoren wurden aus dem Wirt *Saccharomyces cerevisiae* J57p der verwendeten YAC-Bibliothek isoliert. Dabei wurde der Hefestamm in YPD-Medium kultiviert.

- | | | | |
|---------------|----------------------|-------------|---------------|
| ▪ YPD-Medium: | 1% (w/v) Hefeextrakt | ▪ YPD-Agar: | YPD-Medium |
| | 2% (w/v) Pepton | | 2% (w/v) Agar |
| | 2% (w/v) D-Glucose | | |

2.4 Zelllinien und Nährmedien

Folgende eukaryotische Zelllinien wurden verwendet:

***AtT20* Zellen** (ATCC-Nr.: CCL 89; Furth & Upton, 1953; Buonassisi *et al.*, 1962)

Die neuroendokrine Zelllinie *AtT20* stammt ursprünglich aus der wiederholten Passage von Hypophysentumorzellen im Tier und in der Zellkultur. Die aus *AtT20* Zellen isolierte RNA wurden für den Nachweis der Expression von *Fto* und *Sst2* verwendet.

***COS-7* Zellen** (DSMZ-Nr.: ACC 60; Gluzman, 1981)

COS-7 Zellen wurden im Verlauf der *Exon-trapping*-Methode sowie für die Expression von Fusionsproteinen durch die transiente Transfektion mit eukaryotischen Expressionsvektoren eingesetzt. Diese Zellen stellen Derivate der fibroblastenartigen Affennierenzelllinie *CV-1* dar, die mit einer Mutante des *SV-40* Virus infiziert wurden. Daher tragen *COS-7* Zellen einen Teil des Genoms dieses Virus, der unter anderem für das große T-Antigen kodiert. Dieses virale Protein reguliert als DNA-bindendes Protein die Replikation des Virusgenoms. Das große T-Antigen ermöglicht somit die Replikation von Vektoren, die den Replikationsursprung des *SV-40* Virus tragen. Die daraus resultierende hohe Kopienzahl dieser Vektoren gewährleistet eine hohe Expression der kodierten Proteine.

***NIH-3T3* Zellen** (DSMZ-Nr.: ACC 59; Copeland *et al.*, 1979)

NIH-3T3 Zellen wurden für die Expression von Fusionsproteinen durch die transiente Transfektion mit eukaryotischen Expressionsvektoren eingesetzt. Diese Zelllinie stammt von kontakt-inhibierten NIH Swiss Mausembryo Fibroblasten ab.

MEF Zellen

Primäre Mausembryo-Fibroblastenzellen (*MEF*) wurden vor allem für die Isolierung von RNA verwendet. Aufgrund der Verpaarung von heterozygoten Tg7N Mäusen konnten MEF mit den verschiedenen Genotypen Wildtyp (Wt), heterozygot *Ft* (*Ft/+*) und homozygot *Ft* (*Ft/Ft*) gewonnen werden. Dazu wurden die Embryonen am Tag 11.5 p.c. präpariert. Nach der Entfernung des Kopfes und der Innereien wurden die MEF durch mechanisches Zerkleinern der Embryonen und wiederholte Trypsin-Behandlungen der Gewebestücken isoliert und kultiviert.

HeLa Zellen (DSZM-Nr.: ACC 57; Gey *et al.*, 1952; Scherer *et al.*, 1953)

Die Isolierung menschlicher RNA erfolgte aus kultivierten *HeLa*-Zellen. Diese Zelllinie wurde 1951 aus einem Zervixkarzinom einer 31 jährigen Frau isoliert.

Die Kultivierung der Zellen erfolgte wie unter 2.18 beschrieben in folgenden Medien:

- 90% DMEM (*COS-7*, *NIH-3T3*, *MEF*, *AtT20*; Gibco BRL) bzw. 90% RPMI 1640 (*HeLa*; GibcoBRL)
10% FCS (hitzeinaktiviert)
20mM L-Glutamin
100U/ml Penicillin
100mg/ml Streptomycin

2.5 Mäuse und Tierhaltung

Es wurden ausschließlich Mäuse der *Fused toes* Linie Tg7N (gemischter C57BL/6xNMRI-Hintergrund) verwendet. Als Wildtypkontrollen dienten NMRI Mäuse. Die Tierhaltung erfolgte unter konstanten Bedingungen (22-24°C, 60-80% Luftfeuchtigkeit, Tag-Nacht-Rhythmus von 12 Stunden) in der Tierversuchsanlage der Heinrich-Heine-Universität Düsseldorf.

2.6 Vektoren und Oligonukleotide

Die folgenden Tabellen (Tab. 2.2 - 2.8) liefern einen Überblick über die in dieser Arbeit verwendeten Vektoren und Oligonukleotide.

2.6.1 Plasmide

Die in Tab. 2.2 aufgeführten Vektoren wurden für die Klonierung von DNA verwendet.

Bezeichnung	Verwendungszweck	Referenz
pAMP10	UDG-Klonierung von PCR-Produkten (<i>Exon-trapping</i>)	GibcoBRL
pBluescriptIISK-	Klonierungsvektor	Stratagene
pCR2.1	Klonierung von PCR-Produkten	Invitrogen
pCRII	Klonierung von PCR-Produkten	Invitrogen
pEGFP-C1	eukaryotischer Expressionsvektor für GFP-Fusionsproteine	Clontech
pEVRF2	eukaryotischer Expressionsvektor	Matthias <i>et al.</i>
pEVRF2-HA	eukaryotischer Expressionsvektor für HA-Fusionsproteine	Dr. R. Janknecht
pGEM-T	Klonierung von PCR-Produkten	Promega
pSPL3	Spleißvektor für <i>Exon-trapping</i>	GibcoBRL
pSV-SPORT1	eukaryotischer Expressionsvektor, cDNA-Bibliotheken	GibcoBRL
Tabelle 2.2: verwendete Basisvektoren.		

Die konstruierten bzw. verwendeten cDNA-Klone sind in Tab. 2.3 zusammengefasst.

Gen	Bezeichnung	Grundvektor	cDNA-Fragment bzw. Klon-Identitätsnummer	Referenz
<i>Ftm</i>	pF1	pAMP10	<i>Exon trapping</i> Fragment F1 (Exons 5, 6)	diese Arbeit
	pF4	pAMP10	<i>Exon trapping</i> Fragment F4 (Exon 8)	diese Arbeit
	p24-5	pCR2.1	RT-PCR: F1-F1 und F4-B1	diese Arbeit
	p86-4	pGEM-T	3'RACE	diese Arbeit
	pFt2-L13	pGEM-T	RT-PCR : F4-F2 + Ftc-B1	diese Arbeit
	pFt2-P7	pGEM-T	RT-PCR : Ftc-F1 + Ftc-B2	diese Arbeit
	pFt2-bridge	pGEM-T	RT-PCR : FT2-F1 + Ft2U-B510	diese Arbeit
	pFt2RACE-R1.2.4	pGEM-T	5'RACE	diese Arbeit
<i>Fto</i>	p28-10	pAMP10	<i>Exon trapping</i> Fragment 28/10 (Exon 5)	diese Arbeit
	p29-10	pAMP10	<i>Exon trapping</i> Fragment 29/10 (Exon 4)	diese Arbeit
	p42-6	pCRII	RT-PCR: 29/10-F1 + 28/10-B1	diese Arbeit
	pB1598	pSV-SPORT1	cDNA-Klon MPMGc559B1598	RZPD
	pJ0125	pSV-SPORT1	cDNA-Klon MPMGc559J0125	RZPD
	pN2210	pSV-SPORT1	cDNA-Klon MPMGc559N2210	RZPD
	pΔJ	pSV-SPORT1	<i>EcoRV</i> -Derivat von J0125	diese Arbeit
	p75-3	pGEM-T	5'RACE	diese Arbeit
<i>FTO</i>	p627359	pBluescriptIISK-	cDNA-Klon IMAGE:627359	IMAGE
	p131-4	pGEM-T	RT-PCR: Ft3-F20 + hum3-B1	diese Arbeit
<i>Irx5</i>	p367171	pT7T3D-Pac	cDNA-Klon IMAGE:367171	IMAGE
	p177-4	pGEM-T	RT-PCR: Irx5-F30 + P05951-B2	diese Arbeit

<i>Irx6</i>	pIrx6	pBluescriptIISK-	cDNA-Klon IMAGE:1224322	IMAGE
<i>Sall1</i>	p537158	pCMV-SPORT	cDNA-Klon IMAGE:537158	IMAGE
	p199-2	pGEM-T	RT-PCR: rs1-F1 + ms1-B3	diese Arbeit
	p200-2	pGEM-T	RT-PCR: rs-F2 + rs-B1	diese Arbeit
Tab. 2.3: konstruierte bzw. verwendete cDNA-Klone.				

Tab. 2.4 zeigt die in dieser Arbeit klonierten Genomfragmente

Bezeichnung	Grundvektor	kloniertes Genomfragment
p94-5	pBluescriptIISK-	<i>Fto</i> : Exon 1 und Exon 2: ~15kb <i>SpeI</i> -Fragment aus MB3
p94-13	pBluescriptIISK-	<i>Fto</i> : Promotorregion und Exon 1: ~22kb <i>SpeI</i> -Fragment aus MB3
p172-1	pBluescriptIISK-	<i>Irx5</i> : <i>NotI/NheI</i> -Fragment aus PAC73
Tab. 2.4: klonierte Genomfragmente		

Für die Expression von *Fto*-Fusionsproteinen wurden die in Tab. 2.5 aufgeführten Plasmide konstruiert.

Bez.	Grundvektor	Protein	Konstruktion
p92-1	pEVRF-HA2	HA- <i>Fto</i>	<i>SmaI</i> -Fragment aus J0125
p152-4	pSV-SPORT1	<i>Fto</i> -GFP	800bp <i>NheI/EcoRI</i> -Fragment aus pEGFP-C1 in J0125 <i>PfIMI</i>
pLF529	pBluescriptIISK-	<i>Fto</i> -GFP	<i>Fto</i> -Promotor und Exon1: p94-13, ~15kb <i>SpeI</i> Fragment Exon2: 1.5kb <i>SpeI/Eco47III</i> -Fragment aus p94-5 Exon3: 2.5kb <i>EcoRV/XhoI</i> -Fragment aus p94-5 Exon3-9 cDNA und GFP : <i>AflII/XbaI</i> -Fragment aus p152-4 SV40 t-Intronregion und Polyadenylierungssignal: 580bp <i>BamHI</i> -Fragment aus pSV-SPORT1
Tab. 2.5: konstruierte Plasmide für die Expression von <i>Fto</i>-Fusionsproteinen			

2.6.2 BAC-, PAC- und YAC-Klone

Tab. 2.6 zeigt die in dieser Arbeit identifizierten bzw. verwendeten BAC-, PAC- und YAC-Klone.

Art	Bez.	Basisvektor	Identitätsnummer	bezogen von	Referenz
BAC	MB3	pBeloBAC11	FBA-4432/24I23	Genome Systems	Ausmeier, 2000
PAC	PAC73	pPAC4	RPCIP711G23488Q2	RZPD	Ausmeier, 2000
YAC	87-C-5	pRML1	WIBRY917C0587	RZPD	diese Arbeit
	90-F-12	pRML1	WIBRY917F1290D2	RZPD	diese Arbeit

YAC	91-F-3	pRML1	WIBRy917F0391D2	RZPD	diese Arbeit
	116-F-4	pRML1	WIBRy917F04116D2	RZPD	diese Arbeit
	125-A-1	pRML1	WIBRy917A01125D2	RZPD	diese Arbeit
	125-C-2	pRML1	WIBRy917C02125D2	RZPD	diese Arbeit
	152-C-1	pRML1	WIBRy917C01152D2	RZPD	diese Arbeit
	156-E-1	pRML1	WIBRy917E01156D2	RZPD	diese Arbeit
	223-A-9	pRML1	WIBRy917A09223D1	RZPD	diese Arbeit
	228-G-7	pRML1	WIBRy917G07228	RZPD	diese Arbeit
	228-H-8	pRML1	WIBRy917H08228D1	RZPD	diese Arbeit
	270-C-10	pRML1	WIBRy917C10270D2	RZPD	diese Arbeit
	352-H-11	pRML1	WIBRy917H11352D2	RZPD	diese Arbeit
	357-B-3	pRML1	WIBRy917B03357	RZPD	diese Arbeit
	388-A-10	pRML1	WIBRy917A10388D2	RZPD	diese Arbeit
	390-A-11	pRML1	WIBRy917A11390D2	RZPD	diese Arbeit
	390-C-11	pRML1	WIBRy917C11390D2	RZPD	diese Arbeit
	396-A-2	pRML1	WIBRy917A02396	RZPD	diese Arbeit
	401-G-10	pRML1	WIBRy917G10401D2	RZPD	diese Arbeit
	419-B-6	pRML1	WIBRy917B06419D2	RZPD	diese Arbeit
	421-H-10	pRML1	WIBRy917H10421D2	RZPD	diese Arbeit
Tab. 2.6: identifizierte bzw. verwendete BAC-, PAC- und YAC-Klone					

2.6.3 Oligonukleotide ('Primer')

Die in den Tabellen 2.7 und 2.8 aufgeführten synthetischen Oligonukleotide ('Primer') wurden im Rahmen dieser Arbeit verwendet.

Gen / Methode	Bezeichnung	Sequenz (5' → 3')
'Exon trapping'	SA2	ATCTCAGTGGTATTTGTGAGC
	SD6	TCTGAGTCACCTGGACAACC
	dUSD2	CUACUACUACUAGTGAAC'TGCAC'TGTGACAAGCTGC
	dUSA4	CUACUACUACUACACCTGAGGAGTGAATTGGTGC
RACE	OligodT-anchor	GACCACGCGTATCGATGTGCGACTTTTTTTTTTTTTTTTTT (AGC)
	PCR-anchor	GACCACGCGTATCGATGTGCGAC
	AP1	CCATCCTAATACGACTCACTATAGGGC
	AP2	ACTCACTATAGGGCTCGAGCGGC
<i>Ftm</i>	F1-F1	GGATGAAGCAGAAACTGTACAACCC
	F1-B1	TAGCCTGCTGTTCTCGAAGCTG
	F4-F1	TCAGGATCAGCCATGATGCC

<i>Ftm</i>	F4-F2	GCCAATGGAGATGAGCTGAACAAG
	F4-B1	CCTTCTGAAAACCTCACCAGAGTG
	F14-F1	CGGGACAATGTAGAAAACGATTAAG
	FT2-F1	GAAAGATACAGCTCTAAACCTCTTTGG
	Ft2U-B510	CAGCAGTT (CT) CCG (CT) TCCTTTTC (CT) AAATC
	Ftc-F1	CCAAACAGCCATCAGAAGAAATTCGG
	Ftc-B1	TGAGGCTAATCAGAGGCGTTCC
	Ftc-B2	AATCACATGACCACGTCTCAGCCCACAC
	Ft2-GSPB2(70)	GCCGCCACCACCCCGCTCATATCTTTTC
	B63.2	ATTCCGATCATCTGGGATGTTTGG
	Ft2-GSPB1	CGTCTGCCTGTGTCCCTGGACTTGGAG
<i>Fto</i>	28/10-F1	ACCCACCAGCACTGTGTTTTTGG
	28/10-B1	CCACACGGTGAGTGGAACATAAC
	29/10-F1	TCTGAGGATGAAAGTGAGGACGAG
	29/10-B1	AAATAGCAGTCTCCCTGGTGAAGAG
	PAF-B1	TAGGAAGGTCTGACATGCAGCG
	Ft3-B750	CAGGTTCTCATCGTGATGCCAG
	BIC-F1	TGGAAGCAAGATCCTAGAGG
	Ft3-B1700	TCAACACTTCAGAATCTGTGGGC
	Ft3-5'F1	GCCTCCTGTGTGTAGTTCTCTCAAC
	Ft3-5'B1	AAACAGTGAGTCCCGTTACCCAC
	Ft3-F3000	GGTACATTCTCACGGGTAAAATGG
	Ft3-B3400	TAAGGATGCGAGCAAGCAGCTC
	IR-F	CAAAGATGTGCTCACCCAGTG
	IR-B	TAGGAAGGTCTGACATGCAGCG
	Ft3-B1350	CCAGGATGGCAGACAGAATCTC
	Ft3-F20	AGGCGGCTTTAGTAGCAGCATG
	hum3-B1	CGTTGGAGGAGAAAAGCCGATCTC
	<i>Ft bzw. Fts</i>	Ft
Wt		GTGGAACCCTTCTGTACATG
Ft/Wt		CTGAAAGGTTGTACTGAGCC
<i>HPRT</i>	HPRT-5	CACAGGACTAGAACACCTGC
	HPRT-3	GCTGGTGAAAAGGACCTCT
<i>Irx1</i>	Irx1-F750	CATAGGCAAGTTTTCCAACCTGGAC
	Irx1-B1000	GCGACTTTAACTGTTGTGGGGG
<i>Irx2</i>	Irx2-F2050	GTTTGGGGCTCATTTTTACTGGAG
	Irx2-B2150	GGACGTTTATTTCACTGGCTCTTC
<i>Irx3</i>	Irx3-F1	CCCTATCCAATGTGCTTTTCATCAG
	Irx3-B1	GGCTGTCCTTCAGCTCATACTGAG
<i>Irx4</i>	Irx4-F1350	CATGCTGAAACGCCAAGGTC
	Irx4-B2050	CGACACAAAACAAGCCTCAGTACC

<i>Irx5</i>	Irx5-F30	ATGTCCTACCCGCAGCTACTTG
	P05951-F1	TCACCCTTATGCAGCACCTCTG
	P05951-B1	CATGATCTTCTCTCCCTTGGTGG
<i>Irx6</i>	Irx6-F1530	TCAGACTTGGAGGAGAGAAGGTGG
	Irx6-B1700	GACAGTCCTTGGAGACCATTTCAG
<i>KIAA0308/2FLJ</i>	KIAA0308-F1	TGAGCACGAGCCAAGATGAAAC
	KIAA0308-B1	ACAAATACTGTCCAGGCGGTTC
	KIAA0308-F2	CAGACAGTTCTGAAAGGGAAGTGG
	KIAA0308-B2	TCGTGCTCTTCTCTGACAGCAAC
	KIAA0308-F3	AACACGTACTGGGATAGCCTTCAG
	KIAA0308-B3	GAAGTTGGTCCAGGAATCTGCTAAG
<i>Mmp2</i>	Mmp2-F1	GACAGGTGACCTTGACCAGAACAC
	Mmp2-B1	TCAGGTGTGTAACCAATGATCCTG
	Mmp2-F2	GGTTGTCTGACATCGCTGCATG
	Mmp2-B2	TTGGTTCTCCTCCATCCAGTGG
<i>Sall1</i>	rs-F1	TCAACATTTCCAATCCGACCC
	rs-F2	TCCGAGCTGATGTTTTGAGCCAG
	rs-B1	CCGAGTTGAGGTAGAGAGGTTGTG
	msal1-F2	ATCCCTCCAATTCCTGGAAGCC
	msal1-B1/2	GCGACATTCTGAGAGGAAGCACTC
	ms1-B3	CACTCTGGCAGCTTTAGCTTGTG
<i>Sst2</i>	Sstr2-F	AACAGCTTCTGGGTGCTAGAGAAC
	Sstr2-B	GCAAAATGACGAGCAGAGACACAC

Tab. 2.7: Auf Genen basierende bzw. in PCR-Techniken verwendete Oligonukleotide.

genetischer Marker	Bezeichnung	Sequenz (5' → 3')
<i>D8Mit107</i>	D8Mit107-L	CCATAAAGGTCTTAGGATGGAA
	D8Mit107-R	CTAAAATGGGAGACCTCAATTG
<i>D8Mit266</i>	D8Mit266-L	AGGTGGGCCCTTCTACAATT
	D8Mit266-R	CTGATCATGGACAAGTGAGTAAGG
<i>19.MHAa58b7.seq</i>	19.MHAa58b7.seq-L	CTGTAGCTTCCGAAAGTGG
	19.MHAa58b7.seq-R	GGACTGTGGTCAGAGGGTGT
<i>D8Mit250</i>	D8Mit250-L	GGGAAGAAAGACAGCAGCAG
	D8Mit250-R	ACAGTAGTCTGGGTTGATCCTAGG
<i>D8Mit251</i>	D8Mit251-L	GGAAGCCCATGCAGTGAC
	D8Mit251-R	AAAGAATCCTTTCTCATTGAGC
<i>D8Mit252</i>	D8Mit252-L	GGGAAGAAAGACAGCAGCAG
	D8Mit252-R	ACAGTAGTCTGGGTTGATCCTAGG
<i>43.MMHAP10FRG10.seq</i>	43.MMHAP10FRG10.seq-L	AAATGGACCCAGCGTAGACA
	43.MMHAP10FRG10.seq-R	GTGTCTGGCTTAATTGCCT

<i>D8Mit196</i>	D8Mit196-L	CACAGGCTAGACACTAAGTATCAGTAC
	D8Mit196-R	TATGTGAGCATGAAGCAGCC
<i>D8Mit309</i>	D8Mit309-F	GAATTTGGGGAAATAGTGTCTCC
	D8Mit309-B	CTCAGCCCTGAGTTCCACTC
<i>D8Mit147</i>	D8Mit147-F	GAGGCTGTTGCCACTACAGA
	D8Mit147-B	GACCTGTGCCTGTGGGAC
<i>D8Mit267</i>	D8Mit267-L	CTTGACCATAAAGAGAAAAAATGC
	D8Mit267-R	GAGGTACATGGTATATTTGCCTACG
<i>D8Mit248</i>	D8Mit248-L	ATCCCTCAAGCAGTACCCCT
	D8Mit248-R	AGCAGAGGACCACACCTTACA
<i>D8Mit284</i>	D8Mit284-L	AATGCCCATCACTGTCTCTCC
	D8Mit284-R	TGACACTGAGTGCCAAGAGG
<i>D8Mit240</i>	D8Mit240-L	ACAGGGCCTAGGTCACACAC
	D8Mit240-R	TGTTGGATTTCAATTTTGATGC
<i>20.MMHAP88FRG8.seq</i>	20.MMHAP88FRG8.seq-F	CCCTCCACAATGAAGGCTAC
	20.MMHAP88FRG8.seq-B	GGGCAATCTGTGAAGGTTTT
<i>23.MMHAP27FRH8.seq</i>	23.MMHAP27FRH8.seq-F	TGGAGGGAGACATTGAGAGG
	23.MMHAP27FRH8.seq-B	GGTCCTTGGGAGGTGCTTAT
<i>36.MMHAP66FLD6.seq</i>	36.MMHAP66FLD6.seq-F	TAAAAGGCCCATGTCTTCA
	36.MMHAP66FLD6.seq-B	CAGGCATCCTCAGAGCAAAT
<i>D8Mit207</i>	D8Mit207-F	TACCCAGTGAGTCTCTGCTGG
	D8Mit207-B	TCCACTTCTGTAAGGGTGTGG
<i>D8Mit81</i>	D8Mit81-F	CCAGGCGCCAAAAGTAACTT
	D8Mit81-B	ACATGGCTGTAATATGTGCTGC
<i>D8Mit50</i>	D8Mit50-F	TGTGACTACAGCTCGACATGG
	D8Mit50-B	AACCACTTCAAGTTCTGGAAGC
Tab. 2.8: Für den Nachweis von genetischen Markern verwendete Oligonukleotide.		

2.7 DNA-Bibliotheken

Die verwendeten DNA-Bibliotheken sind in Tab. 2.9 zusammengefasst.

Art	Bezeichnung	hergestellt aus	bezogen von
cDNA	RZPD Nr. 559	Mausembryo Tag 9 p.c.	RZPD
	RZPD Nr. 522	Mausembryo Tag 9 p.c.	RZPD
	RZPD Nr. 523	Mausembryo Tag 12 p.c.	RZPD
	RZPD Nr. 400	Thymus einer ausgewachsenen Maus	RZPD

cDNA	RZPD Nr. 411	Testes einer ausgewachsenen Maus	RZPD
	RZPD Nr. 998	viele verschiedene Mausgewebe (IMAGE cDNA-Kollektion)	RZPD
BAC	FBAC4432	Genom Mausstamm 129/SvJ	Genome Systems
PAC	RPC121, RZPD Nr. 711	Genom Mausstamm 129/SvevTACfBr	RZPD
YAC	WIIMIT 820, RZPD Nr. 917	Genom Mausstamm C57BL/6J	RZPD
Tab. 2.9: verwendete DNA-Bibliotheken.			

2.8 Antikörper

Die Analyse von Proteinen bzw. der Nachweis von digoxigenin-markierten RNA-Sonden erfolgt mit Hilfe der in Tabelle 2.10 angegebenen Antikörper:

Art	Bezeichnung	Beschreibung	bezogen von
Primärantikörper	anti-HA	monoklonaler Antikörper (Klon 12CA5) aus der Maus gegen das Hämagglutinin-Epitop (HA) des humanen Influenza Virus	Roche
	anti-DIG	mit Alkalischer Phosphatase gekoppelter polyklonaler Antikörper (F(ab) ₂ -Fragmente) aus dem Schaf gegen Digoxigenin (DIG)	Roche
Sekundärantikörper	Cy3-Esel-anti-Maus	mit Indodicarbocyanin 3 (Cy3) markierter polyklonaler Antikörper (F(ab) ₂ -Fragmente) aus dem Esel gegen Maus ImmunglobulinG	Dianova
Tabelle 2.10: verwendete Antikörper			

2.9 Datenbanken

Die verwendeten Datenbanken sind in Tab. 2.11 zusammengefasst.

Datenbank	Beschreibung und Adresse
NCBI GenBank	Sequenzdatenbank des <i>National Center for Biotechnology Information</i> am <i>National Institutes of Health</i> , Bethesda, USA Datensammlung von allen öffentlich verfügbaren DNA- und Protein-Sequenzen (Benson <i>et al.</i> , 2000)
	http://www.ncbi.nlm.nih.gov/

NCBI BLAST	BLAST - ' <i>Basic Local Alignment Search Tool</i> ' Sammlung von Programmen, die Ähnlichkeiten zwischen DNA/DNA- und DNA/Protein- oder Protein/Protein Sequenzen identifizieren (Altschul <i>et al.</i> , 1994)
	http://www.ncbi.nlm.nih.gov/BLAST/
NCBI CD-Search	Suchprogramm der ' <i>Conserved Domain</i> '-Datenbank (CDD), das konservierte Domänen in einer Proteinsequenz identifiziert (Altschul <i>et al.</i> , 1997)
	http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml
NCBI GeneMap'99	Strahlungshybrid-Transkriptkarte des menschlichen Genoms (Schuler <i>et al.</i> , 1996; Deloukas <i>et al.</i> , 1998)
	http://www.ncbi.nlm.nih.gov/genemap/
OMIM	OMIM - ' <i>Online Mendelian Inheritance in Man</i> ' Katalog menschlicher Gene und Erbkrankheiten, herausgegeben von Dr. Victor A. McKusick, <i>Johns Hopkins University</i> , Baltimore, USA
	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
Ensembl Human Genome Server	durchsuchbare Annotierung des öffentlich verfügbaren Arbeitsentwurfs der Konsensussequenz des menschlichen Genoms
	http://www.ensembl.org
UCSC Human Genome Browser	durchsuchbare Annotierung des öffentlich verfügbaren Arbeitsentwurfs der Konsensussequenz des menschlichen Genoms an der <i>University of California</i> , Santa Cruz, USA
	http://genome.cse.ucsc.edu/index.html
Human-Mouse Homology Maps	Karten der konservierten Syntänie zwischen Mensch und Maus
	http://www.ncbi.nlm.nih.gov/Homology/
Celera	zusammengestellte Sequenzabschnitte der Konsensussequenz des menschlichen Genoms des US-Unternehmens <i>Celera Genomics</i> (Venter <i>et al.</i> , 2001)
	http://www.celera.com
MGI	MGI - ' <i>Mouse Genome Informatics</i> ' Datenbank des <i>Jackson Laboratory</i> , Bar Harbor, USA, die u.a. Informationen über Gene, Allele, genetische Marker, Phänotypen, konservierte Syntänien, Genexpression, Mausstämme, Polymorphismen sowie genetische und physikalische Karten enthält
	http://www.informatics.jax.org/
ENSEMBL Trace Server	durchsuchbares Archiv der primären Sequenzen (einzelne Sequenzierungen) aus ' <i>whole genome shotgun</i> '-Projekten oder anderen im großen Maßstab angelegten Sequenzierungsprojekten für verschiedene Modellorganismen
	http://trace.ensembl.org/
Ensembl Mouse Genome Server	durchsuchbare Annotierung der öffentlich verfügbaren Sequenz der abgeschlossenen bzw. bearbeiteten Sequenzierung von Fragmenten (BACs) des Mausgenoms
	http://mouse.ensembl.org/
WI/MIT-CGR	genetische und physikalische Karten des Mausgenoms des <i>Whitehead Institute for Biomedical Research</i> / <i>MIT Center for Genome Research</i> , Cambridge, USA
	http://www-genome.wi.mit.edu/cgi-bin/mouse/index
TRANSFAC	Transkriptionsfaktor-Datenbank der <i>GBF</i> , Braunschweig, welche die Identifizierung von vermeintlichen Transkriptionsfaktor-Bindungsstellen ermöglicht (Wingender <i>et al.</i> , 2000)
	http://transfac.gbf.de/TRANSFAC/

HUGE	HUGE - 'Database of Human Unidentified Gene-Encoded Large Proteins' Datenbank von bisher unbeschriebenen menschlichen Genen bzw. Proteinen am Kazusa DNA Research Institute, Japan (Ohara <i>et al.</i> , 1997)
	http://www.kazusa.or.jp/huge/
ExpASy Molecular Biology Server	ExpASy - 'Expert Protein Analysis System' Proteomics Server des Schweizerischen Instituts für Bioinformatik (SIB), der eine Programmsammlung für die Analyse von Proteinsequenzen bereitstellt
	http://www.expasy.ch/
Tab. 2.11: verwendete Datenbanken	

2.10 Isolierung von Nukleinsäuren

Je nach Ausgangsmaterial und Qualitäts- bzw. Mengenanspruch an die präparierte DNA bzw. RNA wurden verschiedene Protokolle für die Isolierung von Nukleinsäuren angewendet.

2.10.1 Präparation von Plasmid-DNA aus *E. coli*

Zur schnellen Gewinnung kleinerer Mengen Plasmid-DNA mit ausreichender Qualität für eine Restriktionsanalyse bzw. Sequenzierung wurde die Methode der Minipräparation von Plasmid-DNA angewendet (Sambrook *et al.*, 1989). Dazu wurden zunächst 2ml Selektionsmedium mit einer Einzelkolonie beimpft und über Nacht bei 37°C unter Schütteln inkubiert. Danach wurden 1,5ml dieser Übernachtskultur in ein Reaktionsgefäß überführt und zentrifugiert (10 min, 1000xg). Anschließend wurde das Bakteriensediment in 100µl Puffer P1 resuspendiert. Durch die Zugabe von 200µl frisch angesetzten Puffer P2 wurden die Zellen während einer Inkubation von 5 min auf Eis lysiert. Proteine und chromosomale DNA wurden durch die Zugabe von 150µl Puffer P3 gefällt und nach einer Inkubation von 10 min auf Eis durch eine Zentrifugation (10 min, 12000xg) vom plasmidhaltigen Überstand getrennt. Dieser wurde in ein neues Reaktionsgefäß überführt. Die Fällung der Plasmide erfolgte durch die Zugabe von 200µl Isopropanol. Nach einer Inkubation von 2 min bei Raumtemperatur wurde die Plasmid-DNA durch eine Zentrifugation (10 min, 12000xg) pelletiert. Die DNA wurde mit 500µl 70% Äthanol gewaschen, wie zuvor zentrifugiert und schließlich in 50µl TE-Puffer aufgenommen.

Die Gewinnung größerer Mengen hochreiner Plasmid-DNA erfolgte unter Verwendung des 'QIAGEN Plasmid Midi Kitsystems' (Qiagen), dessen Aufreinigungsprinzip ebenfalls auf der alkalischen Lyse der Zellen und einer zusätzlich nachfolgenden Anionenaustauschchromatographie beruht. Dafür wurden 50ml Selektionsmedium mit einer Bakteriensuspension beimpft und über Nacht bei 37°C unter Schütteln inkubiert. Diese Übernachtskultur wurde in ein 50ml

Reaktionsgefäß überführt und zentrifugiert (10 min, 6000xg, 4°C). Die Lyse der Zellen erfolgte wie bereits oben beschrieben. Allerdings wurden dabei jeweils 4ml der Puffer P1, P2 und P3 verwendet. Nach der Abtrennung der Zelltrümmer durch eine Zentrifugation (30 min, 6000xg, 4°C) wurde der plasmidhaltige Überstand auf eine Qiagen-Säule 100 gegeben, die zuvor mit 4ml QBT-Puffer äquilibriert worden war. Anschließend wurde die Säule zweimal mit 10ml QC-Puffer gewaschen, bevor die Plasmid-DNA mit 5ml QF-Puffer von der Säule eluiert werden konnte. Das plasmidhaltige Eluat wurde in einem 15ml Reaktionsgefäß aufgefangen, mit 3,5ml Isopropanol versetzt, gemischt und für 10 min bei Raumtemperatur inkubiert. Nach einer Zentrifugation (30 min, 6000xg, 4°C) wurde die gefällte Plasmid-DNA mit 2ml 70% Äthanol gewaschen, erneut zentrifugiert (10 min, 6000xg, 4°C) und schließlich in 100µl TE-Puffer aufgenommen.

Lösungen:

- Puffer P1: 50mM Tris/HCl
10mM EDTA, pH 8.0
100µg/ml RNaseA
- Puffer P2: 200mM NaOH
1% (w/v) SDS
- Puffer P3: 3M Kaliumacetat, pH5.5
- Puffer QBT: 750mM NaCl
50mM MOPS, pH 7.0
15% (v/v) Isopropanol
- Puffer QC: 1M NaCl
50mM MOPS, pH 7.0
15% (v/v) Isopropanol
- Puffer QF: 1,25M NaCl
50mM MOPS, pH 8.5
15% (v/v) Isopropanol
- TE-Puffer: 10mM Tris/HCl
1mM EDTA, pH 8.0

2.10.2 Präparation von BAC- und PAC-DNA aus *E. coli*

Zur Gewinnung von großen Mengen hochreiner BAC- bzw. PAC-DNA wurde das '*QIAGEN Large-Construct Kitsystem*' (Qiagen) verwendet. Dieses System basiert auf der alkalischen Lyse der Zellen, gefolgt von einem ATP abhängigen Exonuklease-Verdau der unerwünschten genomischen DNA und einer sich anschließenden Aufreinigung der BAC- bzw. PAC-DNA über eine Säule. Für die Isolierung von BAC-DNA wurden 500ml Selektionsmedium mit einer Bakteriensuspension beimpft und über Nacht bei 37°C unter Schütteln inkubiert. Dagegen wurden 25ml einer Übernachtskultur von PAC-DNA enthaltenden Bakterien zunächst in 500ml Selektionsmedium überführt und bei 37°C inkubiert, bis die Kultur eine optische Dichte von $OD_{600} = 0,1$ erreicht hatte. Anschließend wurde die Replikation der PAC-DNA durch Zugabe von 2,5ml 0,2M IPTG induziert. Danach wurden die Bakterien für

weitere vier Stunden bei 37°C kultiviert. Die auf diese Weise angesetzten Bakterienkulturen wurden zentrifugiert (15 min, 6000xg, 4°C). Die Lyse der Zellen erfolgte wie bereits unter 2.10.1 beschrieben. Allerdings wurden dabei jeweils 20ml der Puffer P1, P2 und P3 verwendet. Nach der Abtrennung der Zelltrümmer durch eine Zentrifugation (30 min, 15000xg, 4°C) wurde der DNA-haltige Überstand gefiltert und mit 36ml Isopropanol versetzt. Nach einer Zentrifugation (30 min, 15000xg, 4°C) wurde die gefällte DNA mit 5ml 70% Äthanol gewaschen, erneut zentrifugiert (10 min, 6000xg, 4°C) und schließlich in 9,5ml EX-Puffer aufgenommen. Nach der Zugabe von 200µl einer ATP-abhängigen Exonuklease (Qiagen) sowie 300µl ATP-Lösung wurde der Ansatz für 60 min bei 37°C inkubiert. Danach wurde der Ansatz mit 10ml QS-Puffer gemischt und auf eine Qiagen-Säule 500 gegeben, die zuvor mit 10ml QBT-Puffer äquilibriert worden war. Anschließend wurde die Säule zweimal mit 30ml QC-Puffer gewaschen, bevor die DNA mit 15ml auf 65°C erwärmten QF-Puffer von der Säule eluiert werden konnte. Das Eluat wurde in einem 50ml Reaktionsgefäß aufgefangen, mit 10,5ml Isopropanol versetzt, gemischt und für 10 min bei Raumtemperatur inkubiert. Nach einer Zentrifugation (30 min, 6000xg, 4°C) wurde die gefällte BAC- bzw. PAC-DNA mit 5ml 70% Äthanol gewaschen, erneut zentrifugiert (10 min, 6000xg, 4°C) und schließlich in 100µl TE-Puffer aufgenommen.

Lösungen:

- QS-Puffer: 1,5M NaCl
100mM MOPS, pH 7.0
15% (v/v) Isopropanol
- EX-Puffer: 50mM Tris/HCl, pH 8.5
10mM MgCl₂
- ATP-Lösung: 100mM ATP, pH 7.5
- alle weiteren Puffer wie unter 2.10.1

2.10.3 Präparation von Gesamt-DNA aus Hefe

Die Isolierung von YAC-DNA aus der Hefe *S. cerevisiae* erfolgte durch die Präparation der Gesamt-DNA nach einer Modifikation des Protokolls von Philippsen *et al.* (1991). Dazu wurden 10ml YPD-Medium mit einer Hefekolonie beimpft und über Nacht bei 30°C unter Schütteln inkubiert. Am folgenden Tag wurde die Kultur in 30ml YPD-Medium überführt und erneut bei 30°C über Nacht unter Schütteln inkubiert. Anschließend wurden die Hefezellen abzentrifugiert (15 min, 3000xg), in 10ml H₂O gewaschen und erneut zentrifugiert. Danach wurden die Zellen in 3ml Resuspensionspuffer I aufgenommen und für den Abbau der Zellwand 200µl Zymolyase-Lösung zugesetzt. Die Bildung von Sphäroblasten erfolgte durch eine Inkubation des Ansatzes für 60-120 min bei 37°C. Die Sphäroblasten wurden abzentrifugiert (5 min, 3000xg) und in 3ml Resuspensionspuffer II aufgenommen. Danach

wurden die Sphäroblasten durch die Zugabe von 150µl 20% SDS während einer Inkubation für 30 min bei 65°C lysiert. Daraufhin wurden die Zelltrümmer in einer Zentrifugation (3 min, 6000xg) abgetrennt und der DNA-haltige Überstand in ein neues Reaktionsgefäß überführt. Durch die Zugabe von 1ml 5M Kaliumacetat wurde während einer Inkubation für 60 min auf Eis der Großteil der verbliebenen Proteine aus dem Ansatz gefällt. Erneut wurde der Ansatz zentrifugiert (30 min, 6000xg) und der DNA-haltige Überstand in ein neues Reaktionsgefäß überführt. Zur Fällung der Nukleinsäuren wurden dem Überstand 4ml Äthanol abs. (eiskalt) zugesetzt. Nach einer Zentrifugation (15 min, 6000xg) wurde das Pellet mit 4ml 70% Äthanol gewaschen und schließlich in 300µl TE-Puffer aufgenommen. Bei Bedarf wurde die RNA durch den Zusatz von 15µl DNase-freie RNase A (10mg/ml) und einer Inkubation von 30 min bei 37°C aus dem Ansatz entfernt. Anschließend erfolgte eine Phenol/Chloroform-Extraktion. Die DNA wurde bei 4°C gelagert.

Zur schnellen Gewinnung der Gesamt-DNA wurde eine Hefekolonie in ein Reaktionsgefäß mit 200µl H₂O gegeben und für 30 min bei 99°C inkubiert. Anschließend wurde der Ansatz kurz auf Eis abgekühlt und zentrifugiert (15 min, 12000xg). Der Überstand wurde in ein neues Reaktionsgefäß überführt und in einer PCR-Reaktion analysiert.

Lösungen:

- Resuspensionspuffer I: 0,9M Sorbitol
0,1M EDTA
50mM DTT, pH 7.5
- Zymolyase-Lösung: 0,25mg Zymolyase
(100000U/g)
in 0,9 M Sorbitol
- Resuspensionspuffer II: 50mM Tris/HCl
50mM EDTA, pH 8.0
- 5M Kaliumacetat
- TE-Puffer: 10mM Tris/HCl
1mM EDTA, pH 8.0

2.10.4 Präparation von genomischer DNA aus Mausgeweben

Die Isolierung genomischer DNA aus der Schwanzspitze von Mäusen beruhte auf Modifizierungen eines von Laird *et al.* entwickelten Protokolls (Laird *et al.*, 1991). Durch den Einsatz von Proteinase K wurde dabei zunächst das eingesetzte Gewebe lysiert. Dazu wurde von drei Wochen alten Mäusen eine etwa 1cm lange Schwanzspitze präpariert und in 750µl Lysispuffer über Nacht bei 55°C unter Schütteln inkubiert. Am folgenden Tag wurden 250µl einer gesättigten NaCl-Lösung dazugegeben. Der gesamte Ansatz wurde für 5 min unter Schütteln inkubiert und danach zentrifugiert (10 min, 12000xg). Anschließend wurden 750µl des Gemisches entnommen, ohne dabei die Gewebereste am Boden des Reaktionsgefäßes zu berühren. Die enthaltene DNA wurde durch den Zusatz von 500µl Isopropanol in

einer anschließenden Zentrifugation (10 min, 12000xg) gefällt. Nach einem Waschschrift mit 1ml 70% Äthanol wurde die isolierte DNA in 100µl TE-Puffer aufgenommen.

Lösungen:

- Lysispuffer: 50mM Tris/HCl, pH 8.0
- gesättigte NaCl-Lösung (>6M)
- 100mM EDTA
- 100mM NaCl
- TE-Puffer: 10mM Tris/HCl
- 1% (w/v) SDS
- 1mM EDTA, pH 8.0
- 0,5 mg/ml Proteinase K

2.10.5 Präparation der Gesamt-RNA aus Mausgeweben

Die Isolierung von RNA erfolgte nach der Methode der sauren Guanidinthiocyanat-Phenol-Chloroform Extraktion (Chomczynski & Sacchi, 1987). Tiefgefrorenes Gewebe wurde dazu in 6ml Lösung D bei 4°C homogenisiert (Polytron). Anschließend wurden 0,8ml 2M Natriumacetat (pH 4.0), 8ml H₂O-gesättigtes Phenol und 1,6ml Chloroform zugesetzt. Der Ansatz wurde kurz durchmischt und für 15 min auf Eis inkubiert. Danach wurde das Gemisch in RNase-freie Zentrifugationsröhrchen (Corex) überführt und für 15 min bei 10000xg und 4°C zentrifugiert. Die obere RNA-haltige, wässrige Phase wurde abgenommen, in ein neues Reaktionsgefäß überführt und mit dem gleichen Volumen Isopropanol versetzt. Zur Fällung der RNA wurde der gesamte Ansatz für 60 min bei -20°C inkubiert und anschließend zentrifugiert (15 min, 6000xg, 4°C). Die RNA wurde in 2ml DEPC-H₂O aufgenommen und durch eine erneute Extraktion weiter aufgereinigt. Dazu wurden 2ml einer Phenol (H₂O-gesättigt) / Chloroform Lösung (1:1) zugegeben. Der Ansatz wurde kurz durchmischt und zur erneuten Phasentrennung zentrifugiert (15 min, 6000xg, 4°C). Wiederum wurde der wässrige Überstand abgenommen. Zur Fällung der RNA wurde der Überstand mit einem halben Volumen 7,5M Ammoniumacetat und 2,5 Volumen Äthanol versetzt und für 10 min bei Raumtemperatur inkubiert. Abschließend wurde die gefällte RNA abzentrifugiert (15 min, 6000xg, 4°C), mit eiskaltem 70% Äthanol gewaschen und nach kurzem Trocknen schließlich in 100µl DEPC-H₂O aufgenommen. Die isolierte RNA wurde bei -80°C gelagert.

Für die parallele Isolierung von DNA und RNA wurde zum Teil das TRIZOL-Reagenz der Firma GibcoBRL verwendet. Dabei wurde die entsprechende Vorschrift des Herstellers befolgt.

Lösungen:

- Lösung D: 4M Guanidiniumthiocyanat (GTC)
- 2M Natriumacetat
- 25mM Natriumcitrat, pH 7.0
- 0,5% (w/v) Lauroylsarcosin
- 7,5M Ammoniumacetat
- 0,1M β-Mercaptoethanol

2.10.6 Präparation der Gesamt-RNA aus eukaryotischen Zelllinien

Für die Isolierung von RNA aus Zelllinien wurden $3-10 \times 10^6$ Zellen mit 1ml Nährmedium in ein Reaktionsgefäß überführt und zentrifugiert (5 min, 1000xg). Der Überstand wurde sorgfältig abgenommen und die Zellen wurden in 1ml PBS gewaschen. Nach einer erneuten Zentrifugation (5 min, 1000xg) wurden die Zellen in 500µl Lösung D lysiert. Für die weitere Aufarbeitung wurden nacheinander 50µl Natriumacetat (2M, pH 4.0), 500µl Phenol (H₂O-gesättigt) und 100µl Chloroform/Isoamylalkohol (49:1) zugesetzt. Der Ansatz wurde gut durchmischt und für 15 min auf Eis inkubiert. Nach einer Zentrifugation (15 min, 12000xg, 4°C) wurde die obere RNA-haltige wässrige Phase abgenommen und in ein neues Reaktionsgefäß überführt, ohne dabei die Protein- bzw. DNA-haltige Inter- bzw. Phenolphase zu berühren. Der Ansatz wurde nach Zugabe von 400µl Chloroform/Isoamylalkohol (49:1) erneut zentrifugiert (15 min, 12000xg, 4°C). Wiederum wurde die wässrige Phase in ein neues Reaktionsgefäß überführt. Die Fällung der RNA erfolgte durch Zugabe des gleichen Volumens Isopropanol innerhalb von 90 min bei -20°C. Anschließend wurde die gefällte RNA abzentrifugiert (15 min, 6000xg, 4°C) und mit eiskaltem 100% Äthanol gewaschen. Nach einem weiteren Waschschrift mit eiskaltem 70% Äthanol wurde die RNA nach kurzem Trocknen schließlich in 20µl DEPC-H₂O aufgenommen und bei -80°C gelagert.

Lösungen: ▪ wie unter 2.10.5

2.10.7 Konzentrationsbestimmung von Nukleinsäurelösungen

Die Nukleinsäuren DNA und RNA weisen innerhalb ihres Absorptionsspektrums ein Absorptionsmaximum bei einer Wellenlänge von 260nm auf. Die Konzentration einer Nukleinsäurelösung kann daher durch die Messung der Absorption bei 260nm in einem Spektralphotometer bestimmt werden. Die Messung der Extinktion E_{260} von Nukleinsäurelösungen gegen die entsprechenden Pufferlösungen in Quarzküvetten mit einer Schichtdicke von 1cm entspricht dabei folgenden Konzentrationen:

$$1E_{260} \text{ doppelsträngige DNA} = 50 \mu\text{g/ml}$$

$$1E_{260} \text{ RNA oder einzelsträngige DNA} = 35\mu\text{g/ml.}$$

Der Grad der Verunreinigung der gemessenen Nukleinsäurelösung durch Proteine kann ebenfalls mit Hilfe dieser Methode festgestellt werden. Dabei sollte in gering verunreinigten Proben das Verhältnis E_{260}/E_{280} einem Wert von 1.8 bis 2.0 entsprechen.

2.11 Klonierungstechniken

2.11.1 Sequenzspezifische Restriktionsspaltung von DNA

Mit gereinigten Restriktionsendonukleasen bakteriellen Ursprungs lassen sich DNA-Moleküle an spezifischen Erkennungsstellen präzise spalten. In dieser Arbeit wurde die sequenzspezifische Spaltung von DNA mit Hilfe verschiedener Restriktionsenzyme unter den vom Hersteller (Roche, NE Biolabs) vorgegebenen Reaktionsbedingungen durchgeführt. Dabei wurden für einen Restriktionsverdau von 1µg DNA in einem Gesamtvolumen von 20µl üblicherweise Enzymaktivitäten von 1-10U eingesetzt. Die Hitzeinaktivierung von Restriktionsenzymen erfolgte, wenn möglich, durch die Inkubation des Ansatzes für 20 min bei 75°C bzw. gemäß den Angaben des Herstellers.

2.11.2 Gelelektrophoretische Auftrennung von DNA

Die gelelektrophoretische Auftrennung von DNA-Molekülen beruht auf der Wanderung der negativ geladenen Nukleinsäuren innerhalb eines Agarosegels unter der Wirkung eines elektrischen Feldes. Die Auftrennung der Nukleinsäuren erfolgt dabei in Abhängigkeit von der Molekülmasse, der angelegten Spannung, der Pufferzusammensetzung sowie der Konzentration des eingesetzten Agarosegels.

Je nach Größe der zu analysierenden DNA (restriktionsenzymatisch gespaltene DNA oder PCR-Fragmente) wurden 0,8 - 3,0% (w/v) Agarosegele verwendet. Als Gel- und Laufpuffer diente das TAE-Puffersystem. Für einen Gelträger der Größe von 11cm x 14cm wurde die gewünschte Menge Agarose in 100ml TAE-Puffer durch Aufkochen in der Mikrowelle gelöst. Der Gellösung wurden 4µl einer Ethidiumbromidlösung zugesetzt. Dabei handelt es sich um einen Fluoreszenzfarbstoff, der zwischen benachbarten Nukleinsäurebasenpaaren interkaliert und so einen Nachweis der DNA durch die Bestrahlung mit UV-Licht möglich macht. Nach kurzem Abkühlen konnte die Gellösung in den mit einem Kamm versehenen Gelträger gegossen werden. Danach wurde das erstarrte Gel in eine mit TAE-Puffer gefüllte Gelkammer überführt und der Kamm aus dem Gel entfernt. Die zu analysierende DNA-Probe wurde mit Auftragspuffer versetzt und in die Geltaschen eingefüllt. Zur Abschätzung der Größe sowie der Menge der aufgetragenen DNA-Fragmente wurden zusätzlich entsprechende DNA-Längenstandards aufgetragen. Die eigentliche Auftrennung der DNA erfolgte bei einer konstanten Spannung von 100V bei Raumtemperatur. Nach Beendigung des Gellaufs wurden die aufgetrennten DNA-Fragmente mit Hilfe eines UV-Tisches durch die Fluoreszenz des interkalierenden Farbstoffes Ethidiumbromid sichtbar gemacht und dokumentiert. Ein

Vergleich mit den parallel aufgetragenen DNA-Längenstandards ermöglichte die Bestimmung der ungefähren Größen sowie der Konzentration der aufgetrennten DNA-Moleküle. Als Längenstandard mit bekannten Fragmentgrößen dienten dabei die DNA-Molekulargewicht-Standards λ HindIII (Roche), ϕ X 174 HaeIII (Stratagene), die 1kb-DNA Leiter (GibcoBRL) und der DNA-Molekulargewichtsmarker VIII (Roche).

Lösungen:

- TAE-Puffer: 200mM Tris
200mM Essigsäure
5mM EDTA, pH 8.0
- DNA-Auftragspuffer: 50% (v/v) Glycerin
0,2% (w/v) Orange G
in TAE-Puffer
- Ethidiumbromidlösung (10mg/ml)

2.11.3 Isolierung einzelner DNA-Fragmente aus Agarosegelen

Einzelne DNA-Fragmente wurden mit Hilfe des 'QIAquick Gel Extraction Kitsystems' (Qiagen) aus Agarosegelen isoliert. Das Prinzip dieses Systems beruht auf der spezifischen Bindung von DNA an die Oberfläche einer Kieselerdemembran in Gegenwart einer hohen Salzkonzentration (Vogelstein & Gillespie, 1979). Das aufzureinigende DNA-Fragment wurde dazu möglichst exakt mit einer Skalpellklinge aus dem Agarosegel ausgeschnitten, gewogen und anschließend in der entsprechenden Menge QG-Puffer (Qiagen; 300 μ l Puffer je 100mg Gel) für 10 min bei 50°C unter Schütteln inkubiert. Danach wurde dem Gemisch eine entsprechende Menge Isopropanol zugesetzt (100 μ l Isopropanol je 100mg Gel) und gemischt. Anschließend wurde das Gemisch auf die im System enthaltene Mini-Säule aufgetragen. Die Säule wurde in ein Auffanggefäß platziert und zentrifugiert (1 min, 12000xg). Danach wurde der Durchfluss verworfen, die Säule mit 750 μ l PE-Puffer (Qiagen) gewaschen und erneut wie oben zentrifugiert. Wiederum wurde der Durchfluss verworfen und noch vorhandene Pufferreste durch eine weitere Zentrifugation entfernt. Nun wurde die Säule in ein Reaktionsgefäß platziert und die an die Säule gebundene DNA durch die Zugabe von 50 μ l TE-Puffer und einer anschließenden Zentrifugation (wie zuvor) eluiert. Abschließend wurde die erfolgreiche Aufreinigung des gewünschten DNA-Fragments durch eine Agarose-Gelelektrophorese überprüft.

Lösungen:

- TE-Puffer: 10mM Tris/HCl
1mM EDTA, pH 8.0

2.11.4 Reinigung und Fällung von DNA

Verunreinigungen (z.B. störende Enzyme und Salze) können durch eine Phenol/Chloroform-Extraktion und eine anschließende Äthanol-Fällung der DNA entfernt werden. Ebenso erreicht man auf diese Weise eine höhere Konzentration der Ausgangs-DNA.

Für die Reinigung wurde die zu reinigende DNA-Lösung zunächst mit TE-Puffer auf ein Gesamtvolumen von 100 µl aufgefüllt. Danach wurde die DNA-Lösung zusammen mit 100µl eines Gemischs aus Phenol (TE-gesättigt), Chloroform und Isoamylalkohol (im Verhältnis 25:24:1) maschinell emulgiert. Die anschließende Zentrifugation der Emulsion für 3 min bei 12000xg führte zur Abtrennung der proteinhaltigen Verunreinigungen, die sich aufgrund der denaturierenden Wirkung von Phenol und Chloroform in der Interphase zwischen der wässrigen Phase und der unteren organischen Phase ansammeln. Danach wurde die obere wässrige, DNA-haltige Phase vorsichtig abgenommen und in ein neues Reaktionsgefäß überführt, ohne dabei eine der anderen beiden Phasen zu berühren. Anschließend wurde die gereinigte DNA gefällt.

Die Präzipitation von DNA zur Konzentrierung und Entfernung von Salzen und Phenolspuren erfolgte durch Alkohol in Gegenwart großer Mengen monovalenter Kationen. Dazu wurde die zu präzipitierende wässrige DNA-Lösung mit 0,1 Volumen 3M Natriumacetat (pH 5.2) versetzt und gemischt. Nach Zugabe von 2-3 Volumen Äthanol abs. wurde die DNA durch eine Inkubation für mindestens 30 min (oder über Nacht) bei -80°C gefällt. Durch eine Zentrifugation (15 min, 12000xg, 4°C) wurde die DNA pelletiert und nachfolgend mit 70% Äthanol gewaschen. Nach einer weiteren Zentrifugation (wie zuvor) wurde das DNA-Pellet kurz bei Raumtemperatur getrocknet und abschließend in einem geeigneten Volumen TE-Puffer aufgenommen.

Lösungen:

- TE-Puffer: 10mM Tris/HCl
1mM EDTA, pH 8.0
- 3M Natriumacetat, pH 5.2

2.11.5 Ligation von DNA-Fragmenten

Die effiziente Verknüpfung von DNA-Fragmenten mit entsprechenden Plasmidvektoren war gewährleistet, wenn die zu verbindenden DNA-Fragmente komplementäre Enden aufwiesen. War dies nicht der Fall, ging der Ligation das Auffüllen der Fragmentenden mit zusätzlichen Nukleotiden voran. Dazu wurden die zu verknüpfenden DNA-Fragmente jeweils zusammen mit 5µl dNTP (330µmol/µl) und 1U Klenow-Fragment der *E.coli* DNA-Polymerase I (Roche) in einem Gesamtvolumen von 50µl eingesetzt. Der gesamte Ansatz wurde anschließend für

30 min bei 37°C inkubiert. Nach der Hitzeinaktivierung der DNA-Polymerase I für 20 min bei 75°C konnten die entstandenen glatten DNA-Fragmente direkt für die Ligation verwendet werden.

Das molare Verhältnis zwischen der Vektor-DNA und dem zu verknüpfenden DNA-Fragment betrug in einer üblichen Ligation 1:3. Die Menge der eingesetzten Gesamt-DNA sollte dabei 200ng nicht überschreiten. In einem typischen Ligationsansatz wurde die präparierte Vektor-DNA und das zu verknüpfende DNA-Fragment zusammen mit 4µl T4-DNA-Ligationspuffer (5x konz.; GibcoBRL) und 1µl T4-DNA-Ligase (5U/µl; GibcoBRL) in einem Gesamtansatz von 20µl eingesetzt. Die Inkubation des Ansatzes erfolgte über Nacht bei 16°C. Im Anschluss konnte der Ligationsansatz direkt für die Transformation von Bakterien verwendet werden.

2.11.6 Herstellung kompetenter *E. coli*-Zellen für die Transformation

Bakterienzellen sind unter normalen Umständen nur teilweise befähigt DNA in das Zellinnere aufzunehmen. Allerdings kann durch die Behandlung der Bakterienzellen mit einer eisgekühlten Salzlösung diese Aufnahmefähigkeit (Kompetenz) deutlich gesteigert werden. Dabei führt der Einsatz von zweifach geladenen Ca^{2+} -Ionen zusammen mit einer Temperatur von 0°C zu einer Abschwächung der Abstoßungskräfte zwischen den negativ geladenen DNA-Molekülen und den Lipopolysacchariden der äußeren Bakterienmembran sowie zu einer Verfestigung der ansonsten fluiden Membran. Beide Veränderungen begünstigen die Aufnahme von DNA durch die Bakterienzelle. Das verwendete Protokoll zur Steigerung der Kompetenz von Bakterienzellen stellt eine Modifikation der Methode nach Inoue (1990) dar. Dazu wurden zunächst Bakterienzellen des *E. coli* Stammes *JM109* in 20ml LB-Medium bei 37°C über Nacht auf einem Schüttler inkubiert. Am folgenden Tag wurde die gesamte Vorkultur in 250ml SOB-Medium überführt und bei 37°C auf einem Schüttler inkubiert, bis diese Hauptkultur eine optische Dichte von $\text{OD}_{600} = 0,93$ erreicht hatte. Danach wurde die Bakterienkultur für 10 min auf Eis inkubiert und anschließend bei 1300xg und 4°C abzentrifugiert. Das Bakteriensediment wurde in 80ml eiskaltem TB-Puffer resuspendiert und für weitere 10 min auf Eis inkubiert. Nach einer erneuten Zentrifugation (wie oben) erfolgte die Resuspendierung des Bakteriensediments in 20ml eiskaltem TB-Puffer sowie die Zugabe von DMSO (Endkonzentration: 7%). Danach wurden die Zellen erneut für 10 min auf Eis inkubiert und schließlich in 100µl Aliquots in flüssigem Stickstoff eingefroren. Die Lagerung der nun kompetenten *E. coli*-Zellen erfolgte bei -80°C.

Lösungen:

- | | |
|--|--|
| <ul style="list-style-type: none"> ▪ LB-Medium: 1% (w/v) NaCl
1% (w/v) Trypton
0,5% (w/v) Hefeextrakt ▪ TB-Puffer: 10 mM PIPES
55 mM MnCl₂
15 mM CaCl₂ | <ul style="list-style-type: none"> ▪ SOB-Medium: 2% (w/v) Bacto Trypton
0,5% (w/v) Hefeextrakt
10 mM NaCl
2,5 mM KCl
10 mM MgCl₂
10 mM MgSO₄
pH 6.7 – 7.0 |
|--|--|

2.11.7 Transformation von kompetenten *E. coli* - Bakterien

Die Aufnahme von DNA durch Bakterienzellen wird durch einen kurzzeitigen Hitzeschock induziert. Durch diesen Vorgang können beliebige Plasmide in die Bakterienzelle eingeschleust und mit Hilfe des internen Replikationsapparates der Wirtszelle vermehrt werden.

Für die Transformation wurden 100µl der aufgetauten, kompetenten Bakterienzellen (*E. coli JM109*) zusammen mit der jeweiligen Plasmid-DNA (20 - 100ng) für 10 min auf Eis inkubiert. Anschließend wurden die Bakterien kurzzeitig für 45 Sekunden bei 42°C sowie für weitere 2 min auf Eis inkubiert. Danach wurde den Bakterienzellen 900µl LB-Medium zugesetzt. Nach einer Inkubation der Bakterien für 45 min bei 37°C unter leichtem Schütteln wurde die Bakteriensuspension für 5 min bei 1000xg zentrifugiert. Danach wurde das Bakteriensediment in 100µl LB-Medium aufgenommen und auf einem Selektivnährboden ausgestrichen und über Nacht bei 37°C inkubiert. Am nächsten Tag wurden ausgehend von den über Nacht gewachsenen Bakterienkolonien transformierter Zellklone größere Bakterienkulturen angelegt, die schließlich für die Präparation von großen Mengen der eingesetzten Plasmid-DNA genutzt wurden.

2.12 Sequenzierung von DNA

DNA-Sequenzierungen wurden von Peter Sikorski mit dem 'Thermo Sequenase Fluoreszent Labelled Primer Cycle Sequencing Kit with 7-deaza-dGTP'-System (Amersham Pharmacia Biotech) auf einem LICOR 'DNA Sequenzer Modell 4000' durchgeführt.

2.13.2 Northern-Blot: Transfer von RNA

Die zu analysierende RNA wurde zunächst unter denaturierenden Bedingungen gelelektrophoretisch aufgetrennt. Die Denaturierung der RNA-Moleküle verhindert dabei die Ausprägung von Sekundärstrukturen, die das Laufverhalten der RNA-Moleküle verändern und falsche Molekülmassen vortäuschen könnten.

Die gelelektrophoretische Auftrennung der RNA-Moleküle erfolgte in einem 0,9%igen (w/v) Agarosegel. Für die Herstellung von 250ml Gellösung wurden 2,25g Agarose, in 5ml MOPS (50x konzentriert) und 203ml H₂O durch Erhitzen gelöst. Nach dem Abkühlen auf 65°C wurden dieser Lösung 42ml Formaldehyd zugesetzt, bevor die gesamte Gellösung auf einen präparierten Gelträger gegossen werden konnte. In der Regel wurden 20µg der zu analysierenden Gesamt-RNA zusammen mit 20µl Formamid (100%), 7µl Formaldehyd (37%), 1µl MOPS (50x konz.), 4µl Bromphenolblau (1%) und 1µl Ethidiumbromid (1mg/ml) in einem Gesamtansatz von 46µl vermischt. Der Ansatz wurde vor dem Auftragen für 5 min bei 65°C inkubiert, auf Eis abgekühlt und schließlich in die Geltaschen eingefüllt. Die Elektrophorese erfolgte bei 100 - 150V innerhalb von zwei bis drei Stunden. Dabei wurde der RNA-Laufpuffer ständig zwischen den Pufferkammern umgepumpt. Nach der Elektrophorese wurde das Gel zunächst in H₂O gewaschen, unter UV-Licht fotografiert und für 15 - 20 min in 20x konz. SSC-Puffer inkubiert. Für die Übertragung der RNA auf eine Nylonmembran (GeneScreen, DuPont) wurden insgesamt sechs Whatmanfilter (Stärke 3mm) zusammen mit der Nylonmembran in 20x konz. SSC-Puffer getränkt. Danach wurden nacheinander drei angefeuchtete Whatmanfilter, das angefeuchtete Gel (mit der dem Gelträger zugewandten Seite nach oben), die Nylonmembran, wiederum drei angefeuchtete Whatmanfilter sowie mehrere Lagen Zellstoff auf eine Glasplatte aufgelegt. Dabei wurden eventuelle Luftblasen zwischen den feuchten Whatmanpapieren, dem Gel und der Membran möglichst entfernt. Anschließend wurde der ganze Aufbau mit einer Glasplatte und einem Gewicht beschwert. Die Übertragung der RNA erfolgte durch die wirkenden Kapillarkräfte über Nacht. Die Position der Geltaschen wurde nach Beendigung des Transfers mit einem Bleistift markiert. Schließlich wurde die Membran für 20 min in 2x konz. SSC-Puffer gewaschen und für zwei Stunden bei 80°C getrocknet. Die endgültige Fixierung der DNA an die Nylonmembran erfolgte durch die Bestrahlung mit UV-Licht (UV Stratalinker 2400; Stratagene). Die Membran konnte bei Raumtemperatur gelagert werden.

Lösungen:

▪ 50x MOPS:	1M MOPS	▪ 1% (w/v) Bromphenolblau
	0,25M Natriumacetat	
	0,05M EDTA, pH 7.0	▪ Ethidiumbromid (1mg/ml)
▪ 20x SSC:	3M NaCl	▪ RNA-Laufpuffer: 2% (v/v) 50x MOPS
	0,3M Natriumcitrat, pH 7.0	11,5% (v/v) Formaldehyd

2.13.3 Radioaktive Markierung von DNA-Sonden

DNA-Sonden wurden nach dem Verfahren des *'random priming'* markiert. Bei diesem Verfahren dient eine Oligonukleotid-Mischung von chemisch synthetisierten Hexameren mit einer statistisch verteilten Basenabfolge als Primer für die Neusynthese von DNA durch das Klenow-Fragment der *E. coli* DNA-Polymerase I. Die eingesetzte DNA wird dabei als Matrize genutzt. Neusynthetisierte DNA wird durch den Einbau von radioaktiven dNTPs markiert und kann anschließend als spezifische Sonde eingesetzt werden.

Im Rahmen dieser Arbeit erfolgte die radioaktive Markierung von Nukleinsäuren mit Hilfe des *Prime-It* Kitsystems (Stratagene). Dazu wurden zunächst 25ng der als Matrize dienenden DNA mit H₂O in einem Gesamtansatz von 23µl verdünnt. Anschließend wurden 10µl der *Random Oligonucleotid* Primer (Stratagene) hinzugegeben. Zum Denaturieren der DNA wurde der gesamte Ansatz für 5 min bei 95°C inkubiert. Nach einer kurzen Zentrifugation wurden 10µl des 5x konzentrierten dCTP-Primer-Puffers (Stratagene), 5µl [α -³²P]dCTP (10mCi/ml) und 1µl Exo(-) Klenow-Enzym (5U/µl; Stratagene) hinzugegeben. Der gesamte Ansatz wurde für 20 min bei 37°C inkubiert. Anschließend wurde die Markierungsreaktion durch die Zugabe von 2µl Stop-Mix (Stratagene) beendet und das Gesamtvolumen des Ansatzes mit 18µl STE-Puffer auf 70µl aufgefüllt.

Nichteingebaute, freie Nukleotide wurden mit Hilfe einer *'Nuc Trap'* Aufreinigungssäule (Stratagene) von der markierten DNA-Sonde abgetrennt. Dazu wurde zunächst eine *Nuc Trap* Säule mit 70µl STE-Puffer äquilibriert. Danach wurde die vorbereitete Säule in eine Säulen-Beta-Schutz-Einrichtung (Stratagene) eingespannt und der gesamte Markierungsansatz auf die Säule gegeben. Das Laden der Säule erfolgte jeweils unter Zuhilfenahme einer aufgesetzten Spritze. Anschließend wurde die markierte DNA-Sonde durch das zweimalige Spülen mit 70µl STE-Puffer von der Säule eluiert, während die freien Nukleotide in den Poren der Säulenmatrix verblieben. Zum Abschluss der Aufreinigung wurde noch zweimal Luft durch die Säule gepresst. Die Cerenkov-Strahlung von 1µl der aufgereinigten Sonde wurde mit dem Szintillationszähler (Beckman) bestimmt. Vor dem Einsatz der Sonde in der Sequenzhomologie-Hybridisierung wurden jeweils 100µl der Sonde mit 30µl 2M Natronlauge denaturiert. Die denaturierte Sonde konnte nun in Abhängigkeit von der einzusetzenden Radioaktivität (üblicherweise 2x10⁶ cpm/ml Hybridisierungslösung) für die Hybridisierung eingesetzt werden.

Lösungen:

- STE-Puffer: 100mM NaCl
20mM Tris/HCl, pH 7.5
10mM EDTA
- 2M Natronlauge

2.13.4 Sequenzhomologie-Hybridisierung

Für die Sequenzhomologie-Hybridisierung wurde die zu analysierende Membran zunächst an drei Seiten in eine Plastikfolie eingeschweißt. Danach wurde je nach Größe der Membran eine entsprechende Menge CHURCH-Puffer sowie die denaturierte DNA-Sonde zugesetzt. Nach dem Versiegeln der Plastikfolie wurde die Membran über Nacht bei 65°C in einem Wasserbad unter leichtem Schwenken inkubiert. Am folgenden Tag wurden die überschüssigen bzw. unspezifisch gebundenen DNA-Sonden durch das dreimalige Waschen der Membran in 400ml Waschpuffer entfernt. Der erste Waschschrift erfolgte dabei innerhalb von 5-10 min bei Raumtemperatur, der zweite für 5 min bei 65°C und der dritte Waschschrift für eine Stunde ebenfalls bei 65°C. Danach wurde die noch leicht feuchte Membran in eine Zellophanfolie verpackt und konnte anschließend zusammen mit einer Verstärkerfolie auf einen Röntgenfilm (X-OMAT AR, Kodak) gelegt werden. Die Exposition erfolgte in einer Filmkassette bei -80°C über Nacht.

Lösungen:

- Waschpuffer: 50mM Na₂HPO₄, pH 7.2
0.5% (w/v) SDS
- CHURCH-Puffer: 250mM Na₂HPO₄, pH 7.2
1mM EDTA
7% (w/v) SDS

2.14 Polymerase-Kettenreaktion

Die Polymerase-Kettenreaktion (PCR) ermöglicht die *in-vitro*-Vermehrung (Amplifikation) eines ausgewählten DNA-Fragments aus heterogener DNA. Dabei werden die Enden dieses Fragments durch zwei bekannte Oligonukleotide (*Primer*) definiert, die an die entgegengesetzten Einzelstränge der vorher durch Hitze denaturierten Ziel-DNA hybridisieren und daher den zu vermehrenden DNA-Abschnitt flankieren. Die Verlängerung dieser Primer und somit die Synthese der komplementären Stränge des gewünschten DNA-Fragments erfolgt mit Hilfe einer hitzestabilen DNA-abhängigen DNA-Polymerase (z.B. *Taq*-Polymerase). Die sich mehrfach wiederholende Abfolge (Zyklus) aus Aufschmelzen der DNA, Hybridisierung der Primer und Strangsynthese führt schließlich zu einer exponentiellen Vermehrung eines spezifischen DNA-Fragments. Zahlreiche Modifikationen und Weiterentwicklungen dieses Grundprinzips erlauben eine vielfältige Anwendung der PCR in der Molekularbiologie (z.B. RT-PCR, RACE, *in vitro* Mutagenese, *Differential Display* und *in situ* PCR).

Die PCR-Reaktionen wurden mit der *Taq* DNA Polymerase oder dem *Expand High Fidelity PCR* System der Firma Roche in einem Thermozykler *MWG Primus 96 plus* (MWG Biotech) durchgeführt. Die Anzahl und Länge der Zyklen sowie die verwendeten Temperaturen

wurden jeweils der Länge des gewünschten DNA-Fragments und der Schmelztemperatur der Primer angepasst. Die verwendeten Primer sind im Abschnitt 2.6.3 aufgeführt. Die Aufreinigung der PCR-Produkte erfolgte mit dem 'High Pure PCR Product Purification Kitsystem' (Roche). Aufgereinigte PCR-Produkte wurden hauptsächlich in den pGEM-T Vektor (Promega) kloniert.

2.14.1 Standard-PCR

Die Standard-PCR auf genomischer DNA wurde mit der *Taq*-Polymerase der Firma Roche durchgeführt. Dabei wurden folgende Einzelkomponenten in einem 50µl Reaktionsansatz verwendet:

- 2µl Vorwärts-Primer (10µM)
- 2µl Rückwärts-Primer (10µM)
- 1µl dNTP-Mix (10mM dATP, 10mM dTTP, 10mM dGTP, 10mM dCTP)
- 5µl 10x-Reaktionspuffer (100mM Tris/HCL, 15mM MgCl₂, 500mM KCl, pH 8.3)
- 0,5µl Taq-Polymerase (5U/µl)
- 1µl DNA (z.B. ca. 50ng genomische DNA)
- 38,5µl ddH₂O

Die PCR-Reaktion im Thermozykler erfolgte mit folgendem Standard-Programm:

Temperatur	Dauer	Vorgang	
94°C	2 min	Hitzenaturierung der DNA	
94°C	30 sek	Hitzenaturierung der DNA	35x Wdh.
52-60°C (je nach Primer)	30 sek	Hybridisierung der Primer	
72°C	30 sek	DNA-Synthese	
72°C	7 min	finale DNA-Synthese	
4°C	∞	Reaktionsende	
Tab. 2.12: Programm für Standard-PCR			

2.14.2 RT-PCR

Die RT-PCR ermöglicht den Nachweis und die Analyse von mRNA-Transkripten in einer PCR-Reaktion (z.B. für Expressionsstudien, Sequenzanalysen, Diagnosen). Da RNA jedoch nicht als Ausgangsmaterial in einer PCR verwendet werden kann, wird vor der eigentlichen PCR-Reaktion eine einzelsträngige DNA-Kopie (cDNA) der eingesetzten mRNA angelegt. Diese sogenannte reverse Transkription (RT) geschieht mit Hilfe von RNA-abhängigen DNA-

Polymerasen (Reverse Transkriptasen), die aus verschiedenen Retroviren (z.B. *Moloney murine leukemia virus* (M-MuLV) oder *avian myeloblastosis virus* (AMV)) gewonnen werden konnten. Häufig besitzen die verwendeten Reversen Transkriptasen ebenfalls eine RNase H-Aktivität für den spezifischen Abbau der RNA in den entstehenden RNA:cDNA-Hybriden (z.B. M-MuLV, AMV). Der Start einer RT-Reaktion erfolgt durch die Hybridisierung eines universellen Oligo-dT-Primers an den polyA-Schwanz der im Ausgangsmaterial vorhandenen mRNAs. Gleichfalls kann auch ein genspezifischer Primer für die Synthese des komplementären cDNA-Strangs einer gewünschten mRNA-Subpopulation eingesetzt werden. Anschließend wird der synthetisierte cDNA-Strang durch eine Standard-PCR vermehrt.

Die RT-Reaktionen erfolgten mit der *SUPERSCRIPT II RNase H⁻ Reverse Transkriptase* (GibcoBRL). Dazu wurden für einen 20µl Reaktionsansatz zunächst

- 1µl Oligo-dT- oder genspezifischer Primer (10µM)
- 1µl RNA (1-3µg Gesamt-RNA)
- 10µl DEPC-H₂O

gemischt, für 5 min bei 70°C inkubiert (Denaturierung der RNA) und anschließend für 1 min auf Eis gekühlt (Hybridisierung des Primers). Danach wurden zu dem Ansatz

- 1µl dNTP-Mix (10mM dATP, 10mM dTTP, 10mM dGTP, 10mM dCTP)
- 4µl 5x-Reaktionspuffer (250mM Tris/HCl (pH 8.3), 375mM KCl, 15mM MgCl₂)
- 2µl DTT (0,1M)
- 1µl *SUPERSCRIPT II RNase H⁻ Reverse Transkriptase* (200U/µl)

gegeben. Der gesamte Reaktionsansatz wurde für 60 min bei 42°C inkubiert. Danach wurde die Temperatur für 5 min auf 55°C erhöht. Abschließend wurde dem Ansatz 1µl RNase H (GibcoBRL) zugegeben und für weitere 10 min bei 55°C inkubiert (Abbau der eingesetzten RNA).

Um falsch positive Ergebnisse aufgrund einer Kontaminationen des Ausgangsmaterials (Gesamt-RNA) mit genomischer DNA ausschließen zu können, wurden für die anschließende PCR-Reaktion Primer aus unterschiedlichen Exons eines Gens eingesetzt. Die Vermehrung der cDNA von schwach exprimierten mRNAs (z.B. *Ftm*) wurde mit Hilfe der *Marathon-Ready* cDNA (Clontech) durchgeführt. Für die Vervielfältigung von längeren cDNAs (2-5kb) wurde die *Advantage Taq*-Polymerase (Clontech) verwendet. Die PCR-Reaktion im Thermozykler erfolgte mit dem Programm für eine Standard PCR (siehe 2.14.1).

2.14.3 Semiquantitative RT-PCR

Neben qualitativen Analysen können PCR und RT-PCR auch für die quantitative Bewertung der Ausgangsmenge einer zu untersuchenden DNA herangezogen werden. Dazu wird die exponentielle Vermehrung des Ausgangsmaterials und das damit nahezu lineare Verhältnis zwischen der Anzahl der durchlaufenden Zyklen und dem Logarithmus der Anzahl der synthetisierten DNA-Moleküle in einer PCR ausgenutzt. Da jedoch die Effizienz der Reaktion aufgrund zunehmender störender Einflüsse (z.B. Primerverbrauch, Enzymermüdung, kompetitive Reaktionen) im Verlauf einer PCR immer mehr abnimmt, geht die exponentielle Vermehrungsphase nach einer bestimmten Zyklenzahl in einen Sättigungsbereich über. Eine Quantifizierung kann also nur in der exponentiellen Phase einer PCR durchgeführt werden. Der Übergang von der exponentiellen Phase in den Sättigungsbereich muss daher für jede PCR genau definiert werden. Dies kann beispielsweise durch die Probenentnahme nach verschiedenen Zyklenzahlen und die anschließende visuelle Auswertung der PCR-Produkte in einem Agarosegel erfolgen.

In der vorliegenden Arbeit wurde diese Methode für die Bestimmung der Expression eines Gens im Verhältnis zu der konstanten Expression eines Haushalts-Gens in verschiedenen Geweben herangezogen. Dieser Vergleich erlaubt eine semiquantitative Bewertung des Expressionsniveaus des untersuchten Gens in einem bestimmten Gewebe. Die Expressionsanalyse der Gene *Ftm* und *Fto* wurde in der exponentiellen Phase der jeweiligen PCR Reaktion nach 30 (*Ftm*) bzw. 32 Zyklen (*Fto*) durchgeführt und mit der Expression des Haushalts-Gens Hypoxanthin-Phosphoribosyl-Transferase (HPRT, 30 Zyklen) verglichen. Die Vermehrung der cDNA erfolgte unter Standard-PCR-Bedingungen.

2.14.4 RACE ('Rapid Amplification of cDNA Ends')

Die RACE-Methode ist eine modifizierte RT-PCR, in der die Sequenzinformation einer partiellen cDNA für die Erstellung einer vollständigen cDNA genutzt werden kann. Ausgehend von einer bereits bekannten Teilsequenz einer cDNA lassen sich durch die Kombination von genspezifischen Primern mit speziell entworfenen Ankerprimern unbekannte Sequenzen in Richtung des 5'-Endes (5'RACE) bzw. in Richtung des 3'-Endes (3'RACE) der vollständigen cDNA isolieren.

5'RACE

Für die Isolierung von unbekanntem 5'-Sequenzen wurde das *5'/3'RACE Kitsystem* (Roche; für das Gen *Fto*) oder die *Marathon-Ready cDNA* (Clontech; für das Gen *Ftm*) eingesetzt.

Generell wurde bei dieser Methode zunächst ein genspezifischer Primer für die Synthese der einzelsträngigen cDNA aus der eingesetzten mRNA (2µg) verwendet (Reverse Transkription, 2.14.2). Nach der Aufreinigung dieser Erststrang-cDNA durch das *'High Pure PCR Product Purification Kitsystem'* (siehe 2.14.5) wurde das 5'-Ende der cDNA durch die Anheftung eines homopolymeren dA-Nukleotidschwanzes mittels Terminaler Transferase verlängert. Dazu wurde die aufgereinigte cDNA für 3 min bei 94°C denaturiert, kurz auf Eis abgekühlt und zusammen mit 10U Terminaler Transferase für 20 min bei 37°C inkubiert. Danach wurde die eingesetzte Terminale Transferase für 10 min bei 70°C inaktiviert. In der nachfolgenden PCR-Reaktion wurden 5µl der nun mit einem polyA-Schwanz (*5'/3' RACE Kitsystem*) bzw. Adapter (*Marathon-Ready cDNA*) verlängerten Erststrang-cDNA zusammen mit einem genspezifischen Primer GSP1 und einem Ankerprimer AP1 eingesetzt. Mit einer weiteren PCR-Reaktion und dem Einsatz des genspezifischen Primers GSP2 und des Ankerprimers AP2 wurde die Sensitivität dieser Methode deutlich verbessert. Die PCR-Reaktionen erfolgten mit dem *Expand High Fidelity PCR System* (Roche) oder dem *Advantage 2 Polymerase Mix* (Clontech). Tabelle 2.13 zeigt die jeweils eingesetzten Primer.

3'RACE

Für die Isolierung von unbekanntem 3'-Sequenzen wurde das *5'/3' RACE Kitsystem* (Roche) eingesetzt. Dabei erfolgte die Synthese der einzelsträngigen cDNA durch einen Oligo-dT-Ankerprimer, der an den natürlichen polyA-Schwanz der eingesetzten mRNA (2µg) hybridisiert. Nach der cDNA Synthese konnte diese direkt in eine erste PCR-Reaktion mit dem genspezifischen Primer GSP1 und dem Ankerprimer AP1 eingesetzt werden. Zur Erhöhung der Sensitivität folgte eine zweite PCR-Reaktion mit dem genspezifischen Primer GSP2 zusammen mit dem Ankerprimer AP2. Die PCR-Reaktionen wurden mit dem *Expand High Fidelity PCR System* (Roche) durchgeführt. Tabelle 2.13 zeigt die jeweils eingesetzten Primer.

RACE	Gen	cDNA-Synthese: Primer, eingesetzte mRNA	Primer PCR I: GSP1, AP1	Primer PCR II: GSP2 , AP2	Plasmid
5'	<i>Ftm</i>	Marathon-Ready cDNA, Embryo 11.5 p.c.	Ft2-GSPB1, Adaptor P1	Ft2-GSPB2(70), Adaptor P2	pFt2RACE- R1.2.4
3'	<i>Ftm</i>	Oligo d(T)-anchor, Embryo 11.5 p.c.	F4-F1, Oligo-dT-anchor	F4-F2, PCR anchor	p86-4
5'	<i>Fto</i>	Ft3-B1350, Embryo 12.5 p.c.	Ft3-B750, Oligo-dT-anchor	PAF-B1, PCR anchor	p75-3

Tab. 2.13: Zusammenfassung der im RACE verwendeten Primerkombinationen.

2.14.5 Reinigung von PCR-Reaktionsprodukten

Die in einer PCR-Reaktion eingesetzten thermostabilen Polymerasen, Primer, freien Nukleotide oder Salze können nachfolgende Enzymreaktionen (Sequenzierung, Klonierung) erheblich stören. Daher wurden die erhaltenen PCR-Produkte mit Hilfe des 'High Pure PCR Product Purification Kitsystems' (Roche) aufgereinigt. Dieser beruht auf der Absorption von Nukleinsäuren an ein Glasfaservlies in Gegenwart eines chaotropen Salzes (Vogelstein & Gillespie, 1979). Dazu wurden 100µl eines PCR-Reaktionsansatzes mit 500µl Bindungspuffer gemischt und in ein High Pure Filtrationsgefäß überführt. Dieses wurde für 30 Sekunden bei 12000xg zentrifugiert. Der Durchlauf wurde verworfen und in das Filtrationsgefäß wurden 500µl Waschpuffer gegeben. Nach einer erneuten Zentrifugation (wie oben, Durchlauf verworfen) wurde das Filtrationsgefäß in ein 1,5ml Reaktionsgefäß gesetzt. Durch die Zugabe von 50µl Elutionspuffer oder bidestilliertem Wasser und einer weiteren Zentrifugation (wie oben) wurden die PCR-Produkte vom Glasfaservlies gelöst und aufgefangen. Die Qualität der aufgereinigten PCR-Produkte wurde in einer Agarose-Gelelektrophorese überprüft. Danach konnten die PCR-Produkte direkt für die Klonierung eingesetzt werden.

Lösungen:

- Bindungspuffer: 3M Guanidin-Thiocyanat
10mM Tris/HCl, pH 6.6
5% (v/v) Äthanol
- Waschpuffer: 20mM NaCl
10mM Tris/HCl, pH 7.5
- Elutionspuffer: 10mM Tris/HCl
1mM EDTA, pH 8.5

2.15 Isolierung von Exon-Sequenzen aus genomischer DNA – 'Exon trapping'

Das *Exon-trapping*-System ermöglicht die Isolierung von exprimierten Sequenzen aus einem genomischen DNA-Fragment. Das von Buckler *et al.* (1991) entwickelte System beruht auf der Selektion von funktionellen Spleiß-Akzeptor- bzw. Spleiß-Donorstellen (Exon-Intron-Übergänge), so dass die in einem genomischen Fragment enthaltenen Exonsequenzen mit Hilfe des natürlichen Spleiß-Apparates eukaryotischer Zellen isoliert werden können. Dazu werden die zu untersuchenden Genomfragmente in ein Intron zwischen zwei Exons eines Spleißvektors kloniert. Dieses Konstrukt wird durch einen starken Promotor in eukaryotischen Zellen exprimiert. Wenn das integrierte Genomfragment ein Exon in der richtigen Orientierung enthält, wird dieses in die resultierende mRNA hineingespleißt und somit die Größe der ursprünglichen mRNA verändert. Das zusätzlich enthaltene Exon lässt sich leicht aufgrund des Größenunterschieds der mRNA mittels PCR detektieren und klonieren. Die

Methode wurde kontinuierlich weiterentwickelt und erlaubt mittlerweile die Isolierung von internen und 3' terminalen Exons (Krizman & Berget, 1993).

Die *Exon-trapping*-Analyse des BAC-Klons MB3 erfolgte unter Verwendung des *Exon-trapping*-Systems der Firma GibcoBRL nach den Angaben des Herstellers. Für die Subklonierung von MB3 BAC-Fragmenten wurden jeweils 2,5µg der BAC-DNA mit entsprechenden Restriktionsenzymen (s. Tab. 3.1) verdaut und über Nacht bei 37°C inkubiert. Anschließend wurden die Enzyme für 10 min bei 80°C inaktiviert und die DNA durch eine Phenol-Chloroform-Extraktion gereinigt und gefällt. Danach wurden die erzeugten DNA-Fragmente eines Restriktionsansatzes willkürlich mit dem *Exon-trapping*-Vektor pSPL3 ligiert (*'shotgun'*), der zuvor mit dem entsprechenden Restriktionsenzym linearisiert und bei Bedarf dephosphoryliert worden war. Die Rekombinationsfrequenz wurde in einigen analytischen Plasmidisolierungen überprüft. Die Plasmid-DNA aller erhaltenen Bakterienkolonien wurden in einem Gesamtansatz präparativ isoliert. Anschließend wurden 20µg Plasmid-DNA für die Transfektion von *COS-7* Zellen verwendet. Die transient transfizierten Zellen wurden 24 Stunden nach der Transfektion geerntet. Im Anschluss wurde die Gesamt-RNA der Zellen isoliert und für die cDNA-Synthese mit dem Primer SA2 verwendet. Nach der cDNA-Synthese wurde die eingesetzte Gesamt-RNA durch die Behandlung mit RNase H entfernt. Danach konnten die rekombinanten cDNAs mittels zweier PCR-Reaktionen mit den aus den flankierenden Exons stammenden und ineinander verschachtelten Primerpaaren SA2 und SD6 bzw. dUSA4 und dUSD2 vermehrt und analysiert werden. Unerwünschte religierte Vektorprodukte wurden zum größten Teil nach der ersten PCR durch einen *Bst*XI-Verdau des PCR-Ansatzes eliminiert. Abschließend wurden die nach der zweiten PCR erhaltenen PCR-Produkte auf einem Agarosegel analysiert und gegebenenfalls durch die Behandlung mit Uracil-DNA-Glykosylase (UDG) direkt in den Vektor pAMP10 kloniert und sequenziert.

2.16 *in vitro*-Transkriptions- und Translationsreaktion

Die *in vitro*-Transkription- und Translation ermöglicht die Expression von plasmidkodierten Proteinen in einem zellfreien System. Dabei erfolgt zunächst die Transkription einer im eingesetzten Plasmid enthaltenen proteinkodierenden Sequenz mit Hilfe spezifischer RNA-Polymerasen. Je nach Art der vorgeschalteten Promotorsequenzen werden dafür vorrangig T3-, T7- oder SP6-RNA-Polymerasen eingesetzt. Die auf diese Weise synthetisierte mRNA wird anschließend in einem aus Weizenkeimen, Oozyten oder Retikulozyten präparierten Lysat, welches alle notwendigen Komponenten für eine Translation bereitstellt, und in Gegenwart eines Aminosäuregemisches translatiert. Die synthetisierten Proteine können

danach aufgrund des erfolgten Einbaus von ^{35}S -markiertem Methionin oder auch durch immunologische Techniken nachgewiesen werden.

Im Rahmen der vorliegenden Arbeit erfolgte die *in vitro*-Transkription- und Translation (TnT) mit Hilfe eines TnT gekoppelten Retikulozytenlysatsystems (Promega). In diesem System können die *in vitro*-Transkription und die *in vitro*-Translation simultan in einem Ansatz ablaufen. Da die Expression der zu untersuchenden DNA-Sequenzen (*Fto* cDNA-Sequenzen) unter der Kontrolle des SP6-Promotors stand, wurde in allen Ansätzen die SP6 RNA-Polymerase für die *in vitro*-Transkription verwendet. Die synthetisierten Proteine wurden durch den Einbau von ^{35}S -markiertem Methionin während der Translation radioaktiv markiert. Dabei bestand der jeweilige Reaktionsansatz aus folgenden Einzelkomponenten:

- 25µl TnT Kaninchen Retikulozytenlysat
- 2µl TnT Reaktionspuffer
- 1µl TnT SP6 RNA-Polymerase
- 1µl methioninfreies Aminosäuregemisch (1mM)
- 4µl ^{35}S -Methionin (10µCi/µl Trans ^{35}S -Label, ICN)
- 1µl RNase-Inhibitor (RNase-Block, Stratagene, 40U/µl)
- 1µl Plasmid-DNA (1µg/µl)
- 15µl DEPC-H₂O

Die Inkubation des Ansatzes erfolgte für 90 min bei 30°C. Anschließend wurden 5µl des Ansatzes mit 20µl SDS-Probenpuffer versetzt und wie unter 2.17 beschrieben in einer SDS-Gelelektrophorese analysiert. Die markierten Proteine wurden dabei in einer Autoradiografie detektiert.

2.17 SDS-Polyacrylamid-Gelelektrophorese

Die SDS-Polyacrylamid-Gelelektrophorese wird vor allem für die Auftrennung von Proteinen innerhalb eines Gels unter der Wirkung eines elektrischen Feldes eingesetzt (Laemmli, 1970). Die Proteine wandern in Abhängigkeit von ihrer Molekülmasse unterschiedlich schnell zur Anode. Durch den Zusatz von Natrium-Dodecylsulfat (SDS) werden die Proteine dabei denaturiert. Außerdem bindet SDS in einem konstanten Massenverhältnis an die gestreckten Proteine, deren Eigenladung durch die stark negative Ladung des SDS überdeckt wird. Das für die Auftrennung verwendete Polyacrylamidgel weist eine Zweiteilung in ein oberes Sammelgel und ein unteres Trenngel auf. Die aufgrund der Bindung des SDS negativ geladenen und gestreckten Polypeptidketten werden dabei zunächst an der Grenze des Sammelgels zum Trenngel fokussiert und anschließend im Trenngel ihrer Größe nach entsprechend aufgetrennt. Die Porengröße der Gele kann mit Hilfe der Konzentration bzw.

des Vernetzungsgrades der Gelbildner dem jeweiligen Trennproblem genau angepasst werden. Die aufgetrennten Proteine können später beispielsweise durch eine Coomassie-Blau-Färbung sichtbar gemacht werden (Sambrook *et al.*, 1989). Radioaktiv markierte Proteine werden mittels einer Autoradiografie detektiert.

Im Rahmen dieser Arbeit wurden die in einer *in vitro*-Transkriptions- und Translationsreaktion synthetisierten Fto Proteine in einem 15%igen Polyacrylamidgel aufgetrennt. Dabei wurde eine Gelapparatur verwendet, die aus einer vorderen Ausschnitts- und einer hinteren Rückenglasplatte (beide mit einer Größe von 15cm x 17,5cm) sowie insgesamt drei 1,5mm dicken Trennleisten bestand. Zunächst wurde das untere Trenngel über eine Länge von 10 bis 11cm in die Gelapparatur gegossen und für die gleichmäßige Polymerisierung unter Luftausschluss mit 1ml Butanol (H₂O gesättigt) überschichtet. Nach der vollständigen Polymerisierung des Trenngels (etwa nach 60 min) wurde das Butanol abgegossen und die Geloberfläche vorsichtig mit H₂O gespült. Anschließend konnte das Sammelgel über eine Länge von 4 bis 5cm auf das Trenngel gegossen werden. Danach wurde der Taschenkamm eingesetzt und das fertige Gel polymerisierte innerhalb von 60 min. Nach der Entfernung des Taschenkamms und der unteren Trennleiste wurde das Gel in die Gelkammer eingesetzt und die beiden Pufferreservoirs der Kammer mit EB-Laufpuffer gefüllt. Die mit 2x konz. SDS-Probenpuffer entsprechend versetzten Proteinproben wurden für 5 min bei 99°C erhitzt und nach einer kurzen Abkühlung (5 min auf Eis) in die Geltaschen eingefüllt. Zur Abschätzung der Größe der aufgetragenen Proteine wurde zusätzlich ein Protein-Längenstandard aufgetragen. Die Separation der Proteine erfolgte über Nacht bei 50V. Danach wurde das Gel über einen Zeitraum von 30 min wiederholt mit insgesamt 11 Fixierlösung befeuchtet. Im Anschluss wurde das Gel vorsichtig auf ein Whatman-Filterpapier überführt und auf einem Vakuumtrockner bei 80°C für mehrere Stunden getrocknet. Schließlich wurde das getrocknete Gel auf einen Röntgenfilm (BIOMAX MR, Kodak) in eine Filmkassette gelegt. Die Exposition erfolgte über Nacht bei Raumtemperatur. Ein Vergleich mit dem parallel aufgetragenen Protein-Längenstandards ermöglichte die Bestimmung der ungefähren Größe der detektierten Proteine. Als Längenstandard mit vorgefärbten bekannten Proteinen diente dabei der Protein-Molekulargewicht-Standard DA4101 (GibcoBRL).

Lösungen:

- | | |
|--|--|
| <ul style="list-style-type: none"> ▪ Trenngel (40ml): 14,1ml Acrylamid 40% 1,6ml Bisacrylamid 2% 10,05ml Tris/HCl 1,5M, pH 8.7 11,5ml H₂O 187,5µl SDS 20% (w/v) 13,13µl TEMED 252µl Ammoniumpersulfat 10% (w/v) | <ul style="list-style-type: none"> ▪ Sammelgel (15ml): 1,9ml Acrylamid/Bisacrylamid, 30%:0,8% 1,9ml Tris/HCl 1M, pH 6.8 11ml H₂O 75µl SDS 20% (w/v) 10,5µl TEMED 180µl Ammoniumpersulfat 10% (w/v) |
|--|--|

- | | |
|--|---|
| <ul style="list-style-type: none"> ▪ 2x SDS-Probenpuffer: <ul style="list-style-type: none"> 8mM Tris/HCl, pH 6.8 20% (v/v) Glycerin 10% (w/v) SDS 8% (v/v) β- Mercaptoethanol 0,004% (w/v) Bromphenolblau | <ul style="list-style-type: none"> ▪ EB-Laufpuffer: 384mM Glycin
50mM Tris/HCl, pH 8.0 ▪ Fixierlösung: 10% (v/v) Essigsäure
10% (v/v) Äthanol |
|--|---|

2.18 Zellkulturtechnik

2.18.1 Kultivierung und Passagieren von Zellen

Die Zellen wurden in Zellkulturschalen (Nunc) in einem Brutschrank bei 37°C und 7,5% CO₂ kultiviert. Dies geschah in den unter 2.4 angegebenen Medien. Unter diesen Kultivierungsbedingungen mussten die Zellen alle 3 bis 4 Tage passagiert werden. Dazu wurden die adhärennten Zellen nach dem Erreichen der Konfluenz zunächst mit PBS gewaschen und durch die Behandlung mit einer 0,05% Trypsin/EDTA-Lösung (GibcoBRL) von den Kulturschalen abgelöst. Anschließend wurden die Zellen für 2 bis 5 min bei 37°C inkubiert. Nach dem Ablösen der Zellen vom Gefäßboden wurden diese in Medium aufgenommen und durch wiederholtes Aufziehen in der Pipette vereinzelt. Schließlich wurde die Zellsuspension in einer 1:5 oder 1:10 Verdünnung mit frischem Medium erneut in Kulturschalen ausgesät.

Lösungen:

- PBS: 120mM NaCl
17mM Na₂HPO₄
3mM KH₂PO₄, pH 7.2

2.18.2 Kryokonservierung von Zellen

$5 \times 10^6 - 1 \times 10^7$ Zellen wurden wie bereits beschrieben von den Kulturschalen abgelöst und für 8 min bei 530xg zentrifugiert. Anschließend wurde das Zellpellet mit 5ml PBS gewaschen und erneut zentrifugiert. Schließlich wurden die Zellen in 1ml Einfriermedium (10% Dimethylsulfoxid (DMSO), 45% Medium, 45% FCS) aufgenommen und in ein Kryoröhrchen (Nunc) überführt. Das Einfrieren der Zellen erfolgte zunächst für 4 Stunden bei -20°C. Danach wurden die Zellen für 3 Tage bei -80°C gelagert, bevor sie endgültig in flüssigen Stickstoff überführt wurden.

Eingefrorene Zellen wurden bei 37°C im Wasserbad kurz aufgetaut. Die aufgetauten Zellen wurden möglichst schnell in Medium überführt, mit PBS gewaschen und anschließend

ausgesät. Die verbliebenen DMSO-Rückstände wurden am nächsten Tag durch den Wechsel des Kulturmediums entfernt.

2.18.3 Bestimmung der Gesamtzellzahl

Die Bestimmung der Gesamtzellzahl einer Zellsuspension erfolgte mit Hilfe einer Neubauer-Zählkammer. Dabei wurde die Zellzahl von acht Großquadraten dieser Zählkammer ausgezählt. Anschließend wurde die Gesamtzellzahl aus dem Mittelwert der gezählten Zellen pro Großquadrat nach folgender Formel berechnet:

Zellzahl pro ml = Mittelwert der gezählten Zellen pro Großquadrat $\times 10^4$

2.18.4 Transfektion

Das Einbringen von Plasmid-DNA in eukaryotische Zellen erfolgte durch die Copräzipitation der zu transfizierenden DNA mit Kalziumchlorid und Phosphat (Graham & van der Eb, 1973). Dabei bilden sich im neutralen pH-Bereich Kalzium-Phosphat-DNA-Komplexe, die von den Zellen mit einer erhöhten Effizienz vermutlich über Endocytose aufgenommen und in den Zellkern transportiert werden. Mit dieser Methode können bis zu 20% der eingesetzten Zellen transfiziert werden. Die aufgenommene Plasmid-DNA liegt schließlich zumindest in transient transfizierten Zellen extrachromosomal vor. Die Expression dieser DNA wird somit nicht durch den Ort der Integration ins Zellgenom beeinflusst.

Einen Tag vor der Transfektion wurden 5×10^5 *COS-7* oder *NIH3T3* Zellen in einer Kulturschale ($\varnothing 10$ cm, Nunc) ausgesät und über Nacht inkubiert. Vor der eigentlichen Transfektion wurden die Zellen zweimal mit PBS gewaschen, in 10ml frischem Kulturmedium aufgenommen und für mindestens 2 Stunden kultiviert. Anschließend wurden 20 μ g Plasmid-DNA in 500 μ l 2x konz. Hebs-Puffer aufgenommen. Danach wurden diesem Ansatz tropfenweise 500 μ l einer 0,25M CaCl_2 -Lösung zugesetzt, die kurz zuvor durch die 1:10 Verdünnung einer 2,5M Stammlösung hergestellt worden war. Innerhalb von 10 Sekunden wurde der vorher gemischte Ansatz langsam auf die Zellen gegeben. Schließlich wurden die Zellen sofort in den Brutschrank überführt und dort für 4 bis 6 Stunden kultiviert. Die Zellen befanden sich während der Prozedur nur kurzzeitig außerhalb des Brutschranks, da der korrekte CO_2 -Gehalt des Mediums die wichtigste Voraussetzung für die Ausbildung der Kalzium-Phosphat-DNA-Komplexe ist. Die erfolgreiche Copräzipitation konnte nach ca. einer Stunde visuell durch das Auftreten eines feinen Präzipitats im Kulturmedium überprüft werden. Zum Abschluss der Transfektion wurde das Präzipitat durch zweimaliges Waschen

der Zellen mit Hepes entfernt. Die transfizierten Zellen wurden für 36 bis 48 Stunden kultiviert und danach geerntet und analysiert.

Lösungen:

- PBS: 120mM NaCl
17mM Na₂HPO₄
3mM KH₂PO₄, pH 7.2
- Hepes: 6,7mM KCL
142mM NaCl
10mM HEPES, pH 7.3
- 2x Hebs-Puffer: 274mM NaCl
42mM HEPES,
9,6mM KCL,
1,5mM Na₂HPO₄, pH 7.1
- 2,5M CaCl₂-Lösung

2.19 Fluoreszenzmikroskopie

Die Fluoreszenzmikroskopie liefert Informationen über die Lokalisation und Funktion zellulärer Proteine. Durch dieses Verfahren können z.B. Zellkernproteine, Gerüstproteine, membrangebundene Proteine oder sogar individuelle Proteindomänen in der Zelle lokalisiert werden.

2.29.1 Indirekte Immunfluoreszenz

Die indirekte Immunfluoreszenz erlaubt den spezifischen Nachweis eines Antigens (Epitop) mit Hilfe eines unmodifizierten Primärantikörpers, der in einem zweiten Inkubationsschritt von fluoreszenzmarkierten Sekundärantikörpern visualisiert wird. Dazu wird die zu analysierende Zell- oder Gewebeprobe auf einem Deckglas fixiert und permeabilisiert. Die Fixierung und Permeabilisierung sichert den freien Eintritt des Antikörpers und immobilisiert zugleich die Antigene, während die Zellarchitektur erhalten bleibt. Die Wahl der Fixierungsmethode wird dabei von der Art des zu analysierenden Antigens bestimmt. So werden z.B. Proben für den Nachweis von Oberflächenantigenen anders fixiert als für den Nachweis von intrazellulären Antigenen. Die jeweiligen Eigenschaften der eingesetzten Antikörper spielen ebenfalls eine wichtige Rolle für die richtige Auswahl der Fixierungsbedingungen.

Die im Rahmen dieser Arbeit durchgeführten Immunfluoreszenzanalysen dienten ausschließlich der Bestimmung der subzellulären Lokalisation von HA-Fto-Fusionsproteinen. Dazu wurden *COS-7* oder *NIH3T3* Zellen mit dem Expressionsvektor p92-1 transient transfiziert. Einen Tag nach der Transfektion wurden die Zellen auf Deckgläser (Ø 12 mm) ausgesät und für einen weiteren Tag in 12-Loch-Platten kultiviert. Danach wurden die Deckgläser zunächst dreimal mit PBS gewaschen. Dieser Waschschrift wurde zwischen allen

nachfolgenden Schritten jeweils wiederholt. Die Zellen wurden anschließend mit frischem 4%igen Paraformaldehyd (PFA) in PBS für 25 min bei 4°C inkubiert. Dazu wurden die Deckgläser mit der von den Zellen bewachsenen Seite auf einen 40µl Tropfen PFA gelegt, der zuvor auf Parafilm im Deckel einer 12-Loch-Platte pipettiert worden war. Anschließend wurde der Boden der Platte auf den im Deckel befindlichen Parafilm aufgesetzt. Nach der Fixierung wurden die Zellen durch die 10 min Inkubation mit 0,2% Triton X-100 in PBS bei Raumtemperatur permeabilisiert. Unspezifische Antigen-Antikörper-Bindungen sollten durch die Absättigung der Antigene mit FCS vermieden werden. Dazu wurden die Zellen für 30 min bei 37°C mit 20% FCS in PBS inkubiert. Danach folgte die Zugabe von 40µl des verdünnten anti-HA-Antikörpers (10µg/ml; siehe Tab. 2.10). Die Deckgläser wurden für 60 min bei 37°C inkubiert, gewaschen und anschließend mit dem verdünnten Cy3-markierten Sekundärantikörper (1:100; siehe Tab. 2.10) erneut für 60 min bei 37°C inkubiert. Als Referenz wurden die Zellkerne durch den Hoechst Farbstoff 33258 (blau fluoreszierendes Bisbenzimid) innerhalb von 3 min bei Raumtemperatur angefärbt. Abschließend wurden die Deckgläser zweimal mit H₂O gewaschen, kurz getrocknet und in einem Tropfen Eindeckelmedium (Moviol) auf einem Objektträger fixiert. Nach dem Aushärten des Eindeckelmediums konnten die Zellen im Fluoreszenzmikroskop (Axioskop 2, Zeiss) analysiert werden.

Lösungen:

- PBS: 120mM NaCl
- 17mM Na₂HPO₄
- 3mM KH₂PO₄, pH 7.2
- 4% (w/v) Paraformaldehyd in PBS
- 0,2% (w/v) Triton X-100
- Hoechst Farbstoff 33258 (1µg/ml) in PBS

2.19.2 Direkte Immunfluoreszenz

Der direkte Nachweis eines Proteins im Fluoreszenzmikroskop gelingt durch die Expression als Fusion mit dem Grün-Fluoreszierenden-Protein (GFP). Das GFP wurde aus der Qualle *Aequoria victoria* isoliert und kann durch Anregung mit UV bzw. blauem Licht leicht aufgrund einer hellen grünen Eigenfluoreszenz nachgewiesen werden. Es ist daher ideal für die Darstellung der subzellulären Lokalisation eines Proteins verwendbar. Mit Hilfe des GFP lassen sich sogar in lebenden Zellen oder Organismen ablaufende Prozesse ohne Fixierung oder operative Eingriffe in Echtzeit beobachten. Die Fluoreszenzeigenschaften des Wildtyp-GFP wurden im Lauf der Zeit verändert und verbessert, so dass mittlerweile weitere fluoreszierende GFP-Varianten für verschiedene Anwendungen verfügbar sind.

Zur Bestimmung der subzellulären Lokalisation des Fto-GFP-Fusionsproteins wurden COS-7 Zellen mit dem Expressionsvektor p152-4 bzw. pLF529 transient transfiziert. Einen Tag nach

der Transfektion wurden die Zellen auf Deckgläser (\varnothing 12mm) ausgesät und für einen weiteren Tag in 12-Loch-Platten kultiviert. Danach wurden die Deckgläser dreimal mit PBS gewaschen und direkt im Fluoreszenzmikroskop auf einem Objektträger analysiert.

2.20 *in situ* Hybridisierung von Ganzkörperpräparaten

Die Methode der *in situ* Hybridisierung von Ganzkörperpräparaten ('*whole mount*') ermöglicht den Nachweis der gewebe- und stadienspezifischen Expression von Genen während der Embryogenese (Tag 6.5-12.5 p.c.). Die Methode basiert auf der Hybridisierung der Transkripte eines zu untersuchenden Gens mit einer markierten komplementären RNA-Sonde ('*Antisense*'-Sonde), deren Lokalisation sich im Embryo durch die Bindung spezifischer Antikörper gefolgt von einer Farbreaktion leicht nachweisen lässt. Die angewendete Prozedur beruht auf Modifizierungen des von Dr. D. Wilkinson entwickelten Protokolls (Wilkinson & Nieto, 1993). Alle verwendeten Lösungen sind in Abschnitt 2.20.4 aufgeführt.

2.20.1 Synthese der RNA-Sonde

Zunächst wurde das Plasmid, welches die Sequenzen für die RNA-Sonde kodiert, linearisiert. Dazu wurden in einem 50 μ l Reaktionsansatz 3 μ g des gewünschten Plasmids mit dem benötigten Restriktionsenzym linearisiert und aufgereinigt. Für die Herstellung einer *SallI* RNA-Sonde wurde dazu das Plasmid *IMAGE:537158* mit dem Restriktionsenzym *SalI* linearisiert. Für die anschließende Synthese und Digoxigenin (DIG)-Markierung der RNA-Sonde wurden in einem 20 μ l Reaktionsansatz 1 μ g des linearisierten Plasmids, 2 μ l 10x konz. Transkriptionspuffer (Roche), 2 μ l 10x DIG-Nukleotidmix (Roche), 1 μ l RNase Block (40U) sowie 2 μ l RNA-Polymerase (Roche; im Fall von *SallI* T7-RNA-Polymerase) in DEPC-H₂O vermischt und für 2 Stunden bei 37°C inkubiert. Danach wurde der Reaktionsansatz mit DEPC-H₂O auf 100 μ l aufgefüllt und die synthetisierte RNA-Sonde mit 25 μ l 7,5M Ammoniumacetat und 200 μ l Äthanol abs. für 30 min bei -80°C gefällt. Nach einer Zentrifugation (20 min, 12000xg, 4°C) wurde die RNA-Sonde mit 70% Äthanol zweimal gewaschen und schließlich in 60 μ l 50% Formamid (in DEPC-H₂O) aufgenommen. Die Integrität und Konzentration der synthetisierten RNA-Sonde wurde durch die gelelektrophoretische Auftrennung eines 2 μ l Aliquots überprüft. Die RNA-Sonde wurde bei -20°C gelagert.

2.20.2 Präparation und Vorbehandlung der Maus-Embryonen

NMRI oder Tg7N Mäuse wurden verpaart. Konnte bei den weiblichen Tieren am folgenden Tag ein Vaginalpfropfen festgestellt werden, wurde dieser Tag als Entwicklungstag 0.5 p.c. bezeichnet. Ausgehend von diesem Datum wurden die Embryonen an den gewünschten Tagen präpariert. Dazu wurden die trächtigen weiblichen Tiere durch zervikale Dislokation getötet. Nach der Entnahme der Uteri wurden die Embryonen in eiskaltem PBS aus dem Uterus präpariert, kurz in PBS gewaschen und über Nacht in 4% Paraformaldehyd fixiert. Am nächsten Tag wurden die Embryonen kurz in PBT gewaschen und danach in einer aufsteigenden Methanolreihe (25%, 50%, 75% (v/v) Methanol in PBT) für jeweils 10 min bei Raumtemperatur inkubiert. Die Dehydrierung der Embryonen wurde durch die Inkubation in Methanol abs. abgeschlossen. Danach konnten die Embryonen bis zur weiteren Verwendung bei -20°C gelagert werden.

2.20.3 Vorbehandlung und Hybridisierung der Embryonen

Die Rehydrierung der Embryonen erfolgte durch die Inkubation in einer absteigenden Methanolreihe (25%, 50%, 75% (v/v) Methanol in PBT) für jeweils 10 min bei Raumtemperatur. Danach wurden die Embryonen für 10 min in PBT gewaschen und je nach Bedarf durch die Behandlung mit 6% H₂O₂ für 10 min gebleicht. Wiederum wurden die Embryonen in PBT gewaschen und anschließend je nach Alter (Tag 10.5 p.c.: 10 min, Tag 11.5 p.c.: 15 min) für einige Minuten mit 10µg/ml Proteinase K behandelt. Danach wurden die Embryonen erneut in PBT gewaschen und für 20 min in eiskaltem 4% Paraformaldehyd / 0,1% Glutaraldehyd nochmals fixiert. Nach einem weiteren Waschschrift in PBT erfolgte über Nacht die Prähybridisierung der Embryonen in 1,5ml Hybridisierungsmix bei 65°C unter leichtem Schwenken. Am nächsten Tag wurde der Hybridisierungsmix ausgetauscht und 0,5µg/ml der markierten RNA-Sonde hinzugegeben. Die Hybridisierung erfolgte über Nacht bei 65°C unter leichtem Schwenken.

Die Entfernung von unspezifisch gebundenen RNA-Sonden erfolgte durch mehrere Waschschriffe. Dazu wurden die Embryonen zunächst mit vorgewärmten Lösungen jeweils für 30 min bei 65°C zweimal mit frischem Hybridisierungsmix, zweimal mit Lösung1 sowie einmal mit 50% (v/v) Lösung1 (verdünnt in MABT) gewaschen. Anschließend wurden die Embryonen zweimal für jeweils 10 min in MABT bei Raumtemperatur inkubiert.

2.20.4 Antikörperinkubation und Nachweis der hybridisierten RNA-Sonden

Durch die Inkubation der Embryonen in Blocklösung für 3 bis 4 Stunden bei Raumtemperatur wurden zunächst unspezifische Antikörperbindungsstellen abgesättigt. Danach wurde die Blocklösung durch frische Blocklösung ausgetauscht. Nach Zugabe des mit einer Alkalischen Phosphatase (ALP) gekoppelten α -Digoxigenin-Antikörpers (1:1000 Verdünnung; Tab. 2.10) erfolgte eine Inkubation unter leichtem Schwenken über Nacht bei 4°C.

Unspezifisch gebundene Antikörper wurden am Folgetag durch eine sich sechsmal wiederholende Inkubation der Embryonen in frischem MABT für jeweils 60 min bei Raumtemperatur und eine abschließende Inkubation in MABT bei 4°C über Nacht entfernt.

Am nächsten Tag wurden die Embryonen zweimal in NTMT für jeweils 20 min bei Raumtemperatur gewaschen, bevor der Nachweis der gebundenen Antikörper in einer ALP-vermittelten Farbreaktion erfolgte. Dazu wurden die Embryonen in Färbelösung unter Lichtausschluss (*Sall1*: bis zu 2 Stunden) inkubiert, bis die spezifische räumliche Verteilung der hybridisierten RNA-Sonde eindeutig erkennbar war. Danach wurden die Embryonen mehrmals in PBT gewaschen und schließlich in 4% Paraformaldehyd bei 4°C fixiert bzw. gelagert. Die Auswertung erfolgte mit einer Digitalkamera unter einem Lichtmikroskop.

Lösungen:

- PBS: 120mM NaCl
17mM Na₂HPO₄
3mM KH₂PO₄, pH 7.2
- 4% (w/v) Paraformaldehyd in PBS
- 7,5M Ammoniumacetat
- PBT: 0,1% Tween-20 in PBS
- H₂O₂ 6% (v/v) in PBT
- Proteinase K 10µg/ml in PBT
- 4% (w/v) Paraformaldehyd
0,1% (v/v) Glutaraldehyd
- Hybridisierungsmix: 50% Formamid
25% 20x SSC
0,1% (v/v) Triton X-100
50µg/ml Heparin
0,1mg/ml tRNA
5mM EDTA, pH 8.0
in DEPC-H₂O
- Lösung1: 50% Formamid
5% 20x SSC
0,1% (v/v) Triton X-100
- 20x SSC: 3M NaCl
0,3M Natriumcitrat, pH 7.0
- MABT: 100mM Maleinsäure, pH 7.5
150mM NaCl
0,1% (v/v) Triton X-100
- NTMT: 100mM NaCl
100mM Tris/HCl, pH 9.5
50mM MgCl₂
0,5% (v/v) Tween-20
0,2mg/ml Levamisol (Sigma)
- Blocklösung: 2% (w/v) Blockierreagenz (Roche)
20% hitzeinaktiviertes Schafserum
in MABT
- Färbelösung: 2% (v/v) NBT/BCIP (Roche)
in NTMT

3 Ergebnisse

Der *Ft*-Phänotyp beruht auf der Funktionsbeeinträchtigung von Genen, die eine entscheidende Rolle in verschiedenen entwicklungsbiologischen Prozessen spielen. Da zu Beginn dieser Arbeit die Lokalisation des *Fused toes* Locus im Mausgenom bereits annähernd bekannt war und auch schon erste Genomfragmente des betroffenen Genomabschnitts vorlagen, wurde die Identifizierung der beeinträchtigten Gene weitgehend über positionelle Ansätze verfolgt. Dies geschah zum einen über den positionellen Kandidatengenansatz und zum anderen über die Positionsklonierung mittels der 'Exon-trapping' Analyse der vorliegenden Genomfragmente.

3.1 Die 'Exon-trapping' Analyse des BAC-Klons MB3

'Exon-trapping' ist eine schnelle und effiziente Methode, um transkribierte Sequenzen innerhalb eines komplexen Genomabschnittes aufzuspüren. Da das System auf der Selektion von funktionellen Exon-Intron-Übergängen beruht, werden keine Vorkenntnisse über die gewebespezifische Expression des zu identifizierenden Gens benötigt. Aufgrund dieses immensen Vorteils gegenüber anderen Methoden sollte dieses Verfahren für die Identifizierung von weiteren transkribierten Sequenzen, die von der *Ft*-Mutation betroffen sind, genutzt werden.

Die *Ft*-Mutation basiert auf der Deletion eines Genomabschnitts mit einer Größe von insgesamt 1650kb (Ausmeier, 2000). Zu Beginn dieser Arbeit stand jedoch nur ein Teilabschnitt (ca. 300kb) des betroffenen Genomabschnitts für die Positionsklonierung von Kandidatengen zur Verfügung (Abb. 1.3). Größtenteils wurde dieser Bereich durch den BAC-Klon MB3 abgedeckt, der ein 170kb langes Genomfragment enthält, welches vollständig durch die *Ft*-Mutation deletiert worden ist. Die 'Exon-trapping' Analyse dieses Fragments führte zu der Isolierung mehrerer vermeintlicher Exons (genaue Vorgehensweise unter Abschnitt 2.15 und Tab. 3.1). Die Authentizität der jeweiligen Exons wurde durch die PCR-Analyse von genomischer DNA überprüft. Dabei wurde verlangt, dass authentische Exon-Sequenzen zumindest auf dem BAC-Klon MB3 (Ausgangsfragment dieser Analyse) lokalisiert sind. Ebenso wurde erwartet, dass die isolierten Sequenzen mittels RT-PCR als Bestandteil eines Transkripts in der RNA von Wildtyp-Mäusen nachgewiesen werden können. Dagegen sollten authentische Exon-Sequenzen nicht in der genomischen DNA von homozygoten *Ft*-Embryonen enthalten sein. Die in der Tabelle 3.1 aufgeführten 'Exon-trapping'-Fragmente erfüllten alle drei Kriterien. Weitere strukturelle Analysen zeigten, dass die Fragmente 28/10 und 29/10 nur auf dem BAC-Klon MB3 lokalisiert sind. Dagegen stammten

die beiden Fragmente *F1* und *F4* aus einem Bereich von MB3, der auch auf dem BAC-Klon MB2 zu finden ist (siehe Abb. 1.3). Weiterhin ließen diese Analysen erkennen, dass sich das Fragment *F1* im Gegensatz zu den anderen Sequenzen aus zwei aneinandergespleißten Exons zusammensetzt. Die isolierten Exonsequenzen zeigten zu diesem Zeitpunkt keine Homologien zu anderen bereits beschriebenen Sequenzen in der Datenbank des *National Center for Biotechnology Information (NCBI-GenBank)*.

'Exon trapping' Fragment	für die Fragmentierung von MB3 verwendetes Restriktionsenzym	Größe	Anzahl der enthaltenen Exons	Lokalisation
F1	<i>Bgl</i> II	247bp	2	MB2, MB3
F4	<i>Pst</i> I	147bp	1	MB2, MB3
28/10	<i>Bgl</i> II	78bp	1	MB3
29/10	<i>Bam</i> HI	143bp	1	MB3

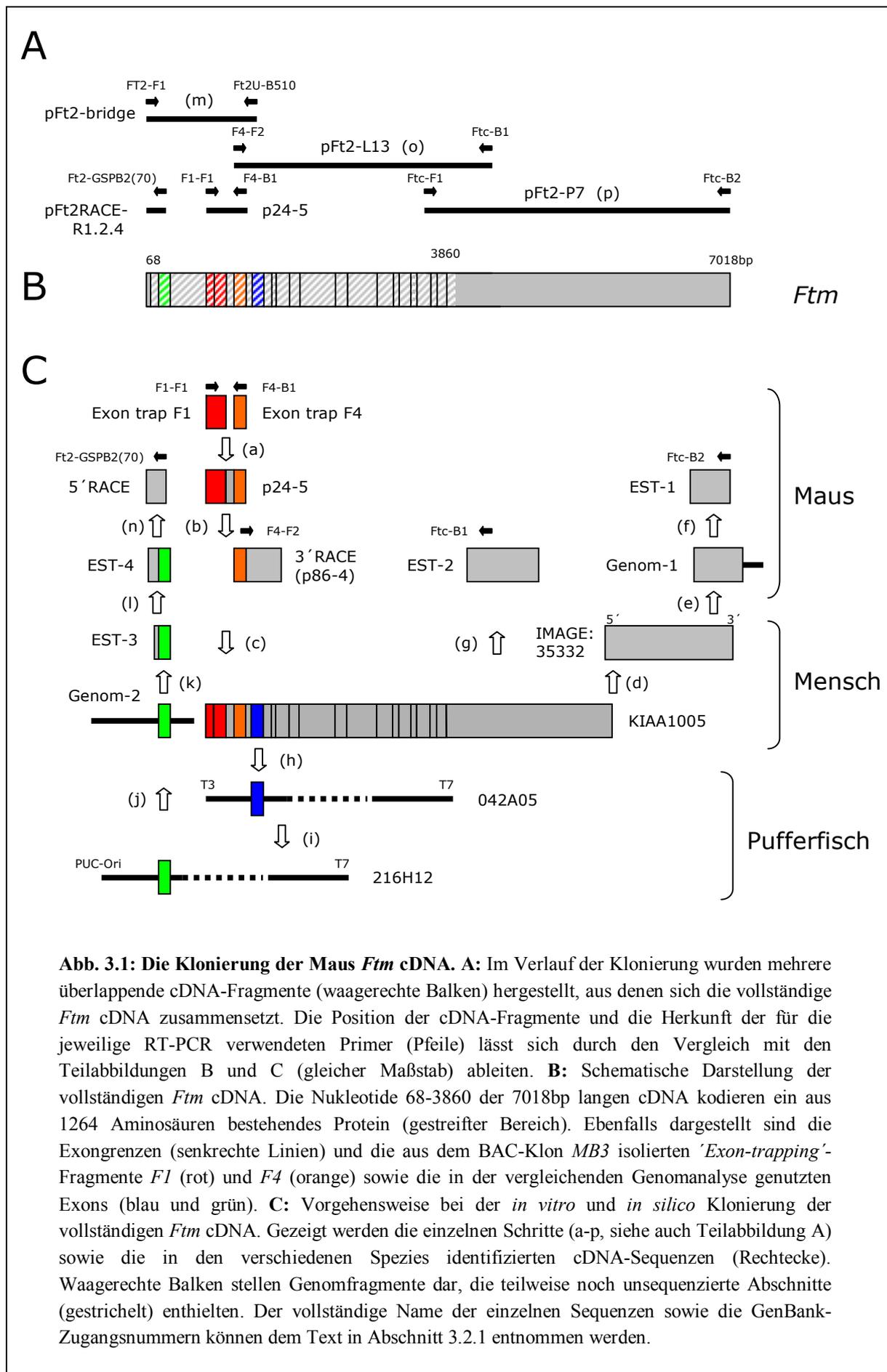
Tab. 3.1: Übersicht über die aus dem BAC-Klon MB3 isolierten vermeintlichen Exons. Informationen über die genaue Vorgehensweise bzw. die Sequenzen der isolierten 'Exon-trapping'-Fragmente können den Abschnitten 2.15 bzw. 7.2.1 und 7.2.2 entnommen werden.

3.2 Die Positionsklonierung von *Fantom (Ftm)*, ein zuvor unbekanntes Gen

Die Erstellung eines Expressionsprofils für jede einzelne 'Exon-trapping'-Sequenz verriet bereits einige auffällige Übereinstimmungen in der räumlichen und zeitlichen Expression. Insbesondere die Sequenzen der Fragmente *F1* und *F4* zeigten in der RT-PCR Analyse der RNA aus embryonalen Geweben bzw. Organen ausgewachsener Mäuse ein identisches Expressionsmuster (ohne Abb.). Um zu ermitteln, ob die isolierten Exon-Sequenzen tatsächlich einem Gen zugeordnet werden können, wurde zunächst versucht die Fragmente *F1* und *F4* mittels RT-PCR auf der Ebene eines Transkripts zu verbinden. In der Tat führte die Kombination der Primer F1-F1 und F4-B1 (Abb. 3.1-C/a) zum Nachweis eines 469bp langen PCR-Produkts (*p24-5*). Beide 'Exon-trapping'-Fragmente sind also Bestandteil ein und desselben Transkripts. Der Einsatz des erhaltenen PCR-Produkts als Sonde in einer Northern-Blot-Analyse der Gesamt-RNA von Embryonen und einigen Geweben ausgewachsener Mäuse ließ eine Gesamtgröße von ungefähr 7kb für das identifizierte Transkript erkennen (Abb. 3.2). Auf diese Weise führte die Kombination der 'Exon-trapping'-Fragmente *F1* und *F4* zur Identifizierung eines Transkripts, dessen kodierende Sequenzen im Bereich des verlorengegangenen Genomabschnitts lokalisiert sind.

3.2.1 Die Isolierung der vollständigen *Ftm* cDNA

Um die vollständige cDNA-Sequenz des identifizierten Transkripts zu isolieren, wurden zunächst alle verfügbaren cDNA-Bibliotheken (siehe Abschnitt 2.7) nach cDNA-Klonen durchsucht, die eventuell weitere Teilstücke der gesuchten cDNA enthielten. Allerdings konnte auf diese Weise kein einziger positiver cDNA-Klon identifiziert werden. Da die gesuchte cDNA in den verwendeten Bibliotheken offensichtlich nicht enthalten war, sollten nun weitere Sequenzen durch ein 5'- bzw. 3'-RACE (*'rapid amplification of cDNA ends'*) isoliert werden. Dabei gelang es jedoch nicht weiter 5' gelegene cDNA-Sequenzen zu bestimmen. Ein 3'-RACE (*p86-4*) führte aufgrund des Erreichens eines internen Polyadenylatbereichs nur zu einer Verlängerung des bisherigen 3'-Endes um 378bp (Abb. 3.1-C, Schritt b). Da sowohl *in vitro* durch weitere Experimente als auch *in silico* durch das Durchsuchen der *NCBI-GenBank* nach überlappenden Maus-Sequenzen keine weiteren cDNA-Sequenzen isoliert werden konnten, wurde die *NCBI-GenBank* nach verwandten orthologen Sequenzen (homologe Sequenzen zwischen verschiedenen Spezies) durchsucht. Dadurch konnte die 4557bp lange menschliche cDNA-Sequenz *KIAA1005* [AB023222] identifiziert werden, deren 5'-Ende zu 87% identisch mit der bisher bekannten Maus-cDNA war (Abb. 3.1-C/c). Ausgehend von dieser menschlichen Sequenz konnte im weiteren Verlauf die vollständige 3'-Region der gesuchten Maus cDNA identifiziert und kloniert werden. Dabei gelang es zunächst eine überlappende, weiter in Richtung 3' gelegene menschliche EST-Sequenz [R24721], zu identifizieren (Abb. 3.1-C/d). EST-Sequenzen (*'expressed sequence tag'*) werden im Rahmen mehrerer groß angelegter Projekte durch die Ansequenzierung der Enden von beliebigen Klonen aus verschiedenen cDNA-Bibliotheken gewonnen und in einer öffentlichen Datenbank zur Verfügung gestellt (Boguski *et al.*, 1993). Die Sequenz des anderen Endes (*IMAGE:35332-3'*) des dem EST R24721 zugrundeliegenden cDNA-Klons *IMAGE:35332* war ebenfalls in der *NCBI-GenBank* verfügbar [R45454]. Mit Hilfe dieser noch weiter 3' gelegenen cDNA-Sequenz konnte eine zu 86% identische orthologe genomische Maus-Sequenz (*Genom-1*) [AZ874705] *in silico* identifiziert werden (Abb. 3.1-C/e), deren Teilbereich wiederum identisch mit einer exprimierten Maus-EST-Sequenz (*EST-1*) [AI837847] war (Abb. 3.1-C/f). Dieses Fragment konnten unter Einbeziehung einer weiteren Maus EST-Sequenz (*EST-2*) [BF141615] (85% identisch mit *KIAA1005*; Abb. 3.1-C/g) in zwei RT-PCR Reaktionen basierend auf der RNA von 11.5 p.c. Tage alten Embryonen mit dem cDNA-Fragment *p24-5* verbunden werden (Abb. 3.1-A/o,p). Damit wurde die Identifizierung des 3'-Bereichs der gesuchten cDNA abgeschlossen. Das vollständige 5'-Ende der gesuchten cDNA wurde in Zusammenarbeit mit Dr. Renate Dildrop isoliert. Da zunächst keine weiteren überlappenden orthologen Sequenzen im



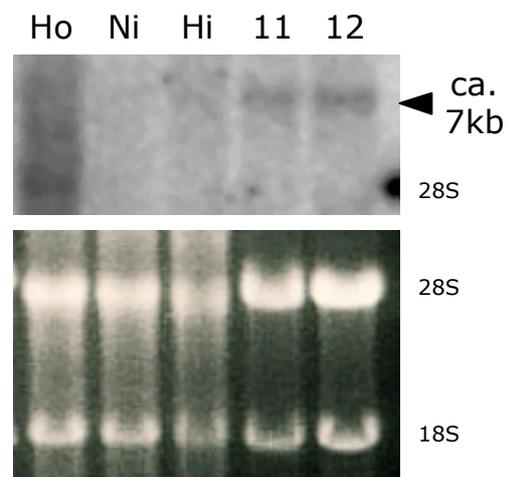
5'-Bereich erkennbar waren, wurde auf die Analyse des entsprechenden homologen Genomabschnitts des Pufferfisches *Tetraodon nigroviridis* zurückgegriffen. Der große Vorteil dieses Modellorganismus in der vergleichenden Genomanalyse ist in der geringen Größe des Genoms dieser Fischfamilie zu finden. Dies wurde durch die Analyse von Genomabschnitten des Pufferfisches *Takifugu rupiens*, der ebenfalls der Familie der *Tetraodontidae* angehört, besonders veranschaulicht. Obwohl das Genom des Pufferfisches etwa 7,5 mal kleiner ist als das Menschenom, enthalten doch beide Genome eine ähnliche Anzahl von kodierenden Sequenzen (Brenner *et al.*, 1993; Venkatesh *et al.*, 2000). Der Gehalt an nicht-kodierenden Sequenzen (z.B. Intronsequenzen) in *Tetraodontidae*-Genomen ist somit relativ gering. Diese Tatsache sollte für die Identifizierung von 5' gelegenen orthologen Exonsequenzen im *Tetraodontidae*-Genom genutzt werden. In einem Genomprojekt am *Genoscope - Centre National de Séquençage* in Evry/Frankreich werden derzeit die Enden von kurzen genomischen Fragmenten ansequenziert und über die *NCBI-GenBank* öffentlich zur Verfügung gestellt. Innerhalb der Sequenzen des genomischen Fragments *042A05* [AL308921, T3-Ende des Fragments] konnte eine Exon-Sequenz identifiziert werden, die ortholog (83% Ähnlichkeit) zu einem Teilbereich (späteres Exon 10) der aus der Maus cDNA *p86-4* und der Mensch cDNA *KIAA1005* abgeleiteten Aminosäuresequenz war (Abb. 3.1-C/h). Ausgehend von diesem Genomfragment wurde im weiteren Verlauf *in silico* eine kurze chromosomale Wanderung durchgeführt. Ein Abschnitt des identifizierten Fragments *042A05* überlappte mit der gegenläufigen Sequenz des genomischen Fragments *216H12* [AL230582, T7-Ende des Fragments]. Aufgrund dieser Überschneidung (Abb. 3.1-C/i) konnte die weiter 5' gelegene Sequenz des anderen Endes dieses Fragments [AL230583, *PUC-Ori*-Ende des Fragments] für die Identifizierung von orthologen 5'-Sequenzen im menschlichen Genom verwendet werden (Abb. 3.1-C/j). Tatsächlich war es somit möglich eine menschliche genomische Sequenz (*Genom-2*) [AC084795] auf Chromosom 16, einem Bereich der homolog zur Region D des Maus-Chromosoms 8 ist (s. Abb. 3.13), zu erkennen, welche die gesuchte orthologe 5'-Sequenz enthielt (82% Ähnlichkeit auf Ebene der abgeleiteten Aminosäuresequenz). Diese Exon-Sequenz führte zunächst zu der Identifizierung einer identischen menschlichen EST-Sequenz (*EST-3*) [AV754125] (Abb. 3.1-C/k), die wiederum für die Erkennung einer orthologen (92% Identität) Maus EST-Sequenz (*EST-4*) [AV045526] (Abb. 3.1-C/l) genutzt werden konnte. Diese konnte in einer RT-PCR basierend auf der RNA von 11.5 p.c. Tage alten Embryonen mit dem cDNA-Fragment *p24-5* verbunden werden (*pFt2-bridge*, Abb. 3.1-A/m). Abschließend wurde die cDNA durch ein 5'-RACE mit dem Primer *Ft2-GSPB2(70)* vervollständigt (*pFt2RACE-R1.2.4*, Abb. 3.1-C/n).

Die Gesamtlänge der isolierten cDNA beträgt 7kb und stimmt mit der im Northern-Blot beobachteten Transkriptlänge gut überein (Abb. 3.2). Außerdem sind am 3'-Ende der isolierten Sequenz drei klassische Polyadenylierungssignale ('AATAAA'; Colgan & Manley,

1997) enthalten. Es kann daher davon ausgegangen werden, dass die isolierte Sequenz der vollständigen cDNA entspricht. Die nun vorliegende Sequenz des identifizierten Transkripts wurde zuvor noch nicht beschrieben. Die weiteren Analysen (s. Abschnitt 3.2.2) zeigten, dass es sich um die cDNA eines neuen, zuvor noch unbekanntes Gens handelte. Aufgrund der nur schwer nachweisbaren Expression (Abschnitt 3.2.3) wurde dieses neue Gen *Fantom* (*Ftm*) genannt.

Die Orientierung von *Ftm* wurde durch die Southern-Blot- und PCR-Analyse genomischer DNA bestimmt. Das 5'-Ende der *Ftm* cDNA war auf dem genomischen BAC-Klon MB3 zu finden, während das 3'-Ende auf den BAC-Klonen MB2 und MB1 lokalisiert ist (Abb. 3.18). Weiterhin bestätigten diese Analysen, dass *Ftm* vollständig im Genom homozygoter *Ft*-Embryonen deletiert ist (Abb. 3.17).

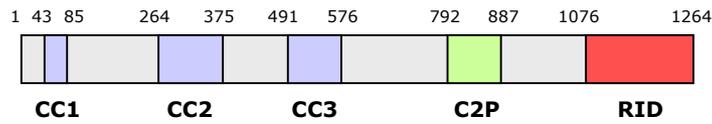
Abb. 3.2: Die Identifizierung eines Transkripts, welches die Sequenzen der beiden *Exon-trapping* Fragmente *F1* und *F4* enthält. Gezeigt ist die Northern-Blot-Analyse der Gesamt-RNA einiger ausgewählter Gewebe (Hoden (Ho), Niere (Ni), Hirn (Hi), Embryo Tag 11.5 p.c. (11) und Tag 12.5 p.c. (12)). Als Sonde wurde das cDNA-Fragment *p24-5* eingesetzt, welches die beiden *Exon-trapping* Fragmente *F1* und *F4* enthält. Der Pfeil markiert die Position des ungefähr 7kb großen Transkripts oberhalb der Position der 28S rRNA. Als Ladekontrolle ist im unteren Abschnitt die UV-Belichtung der eingesetzten Gesamt-RNA mit den charakteristischen Banden der 28S und 18S rRNA im verwendeten Agarose-Gel dargestellt.



3.2.2 Das *Ftm* Protein

Die Sequenzanalyse der *Ftm* cDNA enthüllte einen offenen Leserahmen, der für ein vermeintliches Protein bestehend aus 1264 Aminosäuren und mit einem theoretischen Molekulargewicht von 145kDa kodiert. Der abgeleitete Translationsstart beginnt an Position 68 und die das Startcodon umgebenden Sequenzen entsprechen gut den beschriebenen Konsensussequenzen für einen Translationsstart (Kozak, 1987). An den Translationsstopp (Position 3860) schließt sich eine 3'-untranslatierte Region (UTR) von 3158bp an. Der strukturelle Aufbau des sehr hydrophilen *Ftm* Proteins ließ eine hohe Ähnlichkeit zu dem vor

A



B

Ftm:	1	MSGPSDETAGDLPVKDTGLNLFVGGGLQETSTARTVTKTRQAVSRVSRREEEDRFLRLHDENILLKQHARKQEDKIKRMTAKLIRLVNDKK	90
RPGRIP:	1	DLPV+DT + L + + + V + + R + + + EL R L LH+E L+K+ + KQ DK +R T + R +	85
Ftm:	91	RYERVGGPKRLGRDVEEMIEQLQEKVHELERQNEVLKNRLISAKQQLQVQGHR-----QTSYSRVQARVNTGRRRASASAGSQE	140
RPGRIP:	86	+ LQE E R+ V + SA + V G + T+ Q RV+ GRR +	152
Ftm:	141	CPGKGLRFQNVDEAEVTPQTLTKY---SNSLLEEARGEIRNLENNVQSQRGQIEELEHLAEILKTKQRKNEIELSLLQLREQQATDQR	259
RPGRIP:	117	N + T P Y N + E R E L + + + +EE+ E K ++ E Q+A + R	228
Ftm:	260	SNIRDNVETIKLHKQLVEKSNALSVEIGKFIQLQE-----KQRTLRI SHDALMANGDELNKQLKEQRLKCCSLEKQLHVSFRSERRV	342
RPGRIP:	229	++I+NVE I+L K L E++ +L+ E + ++QE Q L +H+A ++ +EL +L E+ K SL QL V + +	318
Ftm:	343	EELQDRINDLEKERELLKENYDKLYNSAFSAHEE-----QW--KLKEQQMKVQIAQLETALKSDLTDKTEVLDKLTKTERDQNEKLVQ	422
RPGRIP:	319	+E Q R+ DLEKER+LL ++YD+L + ++H+ W +L +Q+ Q+ L + + L ++T+V + + Q+ KL	407
Ftm:	423	ENRDLQLQCLQKQRLHQLSRLKFNQESDINADDLSEALLIKAQKEXKNGDLSFLEKVDKINKDLDRSMKELQATHAETVQELEKT	512
RPGRIP:	408	++ D+ Q Q+++ LQS + ++ LL Q+E + ++ DL + + ELQ +HAET ELEKT	486
Ftm:	513	QDITDILYQHEQEEE---SLQSTATVASSPEELCELAQPTLLPQTDQRESS-----EPKAQDENDLSQVSELQVSHAETPLELEKT	591
RPGRIP:	487	RDMLLQRLKINMICYQEELEATLTKADRENDRHEEKLERLNHLLDFKNSRIKQLEGILRSHGLPTSEQLKDVAYGTLPPSLCLLEPLAAHRG	576
Ftm:	592	DEFDETIHLERGENLFEIHKVTFSSVLRASGDKELVTFCTYAFYDFELQTTPIVRGLYPEYNFTSQYLHVHNDLFLQYIQKNTVTL	681
RPGRIP:	577	D+ + L ENLFE+H+++ + L +GD+ TPCTY+FYDFE TP+ L G P Y+PYSQY+V + L L Y+Q +V L+	666
Ftm:	682	LHQAHSTDYETIAACQLRFHEILEKSGRIFCTTSLVGTGKGDIPNFGTVVEYFRLRVPMDQAIRLYRERAKALGYITSSFKKPEKMQLSQ	771
RPGRIP:	667	LHQAMASYHVLATGWI SLDKVLGTVERVHGLATLAGAGGE--DLGVLEYWMRLCLPLKPSLQACNKRKKAQYLSVSVLGARKVQSNES	754
Ftm:	772	QAATTAQISPAESTDGNLNLHVTVRCCITGLQSRASYLPHAYVYKFFDFPDHDTAIVPSSNDPQFDDHMCFPVPMNMDLDRYLKSESL	861
RPGRIP:	755	++ T A NEL V + C GL+SR QP YV+Y+FF FPDHDT I+P+S++P F D FPV + DLD+YL+ E+L	834
Ftm:	862	RSETWAP-----QNELRVEITRCCGLRSRRLGRQPSYMYRFFTFPDHDTIIIPASSNPFYKQALFPVLVTSDDLQYLRREAL	938
RPGRIP:	835	SVYVFDDEDPEPGSYLGRAQVPLPLAQNKSIKGFNLDSDGSEKSNKSVKQVLDWKSXYLAPEGFQMSAEKPEGEKEEVEEGVEEVEE	924
Ftm:	939	-----EEPEAVQRLPKKSDVTSVAPKPKRQRLTFVDKVKVSFADTISHSPSPETSPPKDKIDSS-----	999
RPGRIP:	925	EE E + + + + K + ++ D+K + + E KD+ ++S	1014
Ftm:	1000	-----PEVGPKPENGLSAVAYPS-----KXSGVAKVEENVGEMQOGKED---	1038
RPGRIP:	1015	P + K E +Y + K E+ + Q +E+	1104
Ftm:	1039	PIEAGQYPEKRPVIAEKKEREHQVASYSRKHKPKGVQDKNRMEYLSNINLNGNTQOMHYTEWKFSLGKKAEDGGGLKAQDKREPPS	1127
RPGRIP:	1105	DISFLSEQQLASGSVASSEDETEITEELEPEDEDRSADSDDCIIPSSVSTNTKQPSSEIRIEIIALNL-NDSQITREDTIIQRLFIECRF	1193
Ftm:	1128	S L + + A S + E + E + +DSDD I+ T K SE++ IEI+L ++ + +TIQ+++E +F	1215
RPGRIP:	1194	PRASALRQEHSPHRNFAFLADQESCEQAS-EVSETQTTSDSDIIVTPQAQTVPKADSEKMCIEIVSLAFCAPEADVMSDETIQVYVVEYKF	1282
Ftm:	1216	YSLPAE--ETPMSLKPQSQGWVYVYNSNVIYLDKKNPAVRDILKAILQRRELPHRSVRFTVSDPPEDEQDLECEDIGVAHVLDLADLF	1264
RPGRIP:	1283	LP ETMPSL KP+++ +++++S VI LD + + R L A+L ++ +FTVSD P DE+ EC+DIG A+++L +F	1331
		CDLPLSETETPMSLRKPRAGEEIHFFHSKVIDLDPVEHQSRRLQFLFAMLAHQDSDEGRFKFTVSD-PLDEEKKECQDIGAYLWQIF	
		QKGRDIEQDIDVLDARTDGGTIGKLVTVVEALHALRSVYQNRKDL	
		Q G+DI+EQ++++ R IG+LKV+++A AL +Y++ +DL	
		QSGKDILEQELEIVSPRNQAIQIGRLKVSQAALHGIYKEMTEDLFS	

Abb. 3.3: Das vermeintliche Ftm Protein. A: Das aus der isolierten cDNA abgeleitete Ftm Protein enthält neben den drei *coiled-coil* Domänen CC1 (AS 43-85), CC2 (AS 264-375) und CC3 (491-576) eine C2-artige Domäne (C2P, AS 792-887) sowie eine vermeintliche RPGR-Interaktionsdomäne (RID, AS 1076-1264). B: Ein Vergleich der Aminosäuresequenzen des Ftm und des RPGRIP Proteins der Maus [NP_076368]. Identische und ähnliche (+) Aminosäuren sind in der mittleren Zeile aufgeführt. Ebenfalls markiert sind die drei CC-Domänen (blau), die C2P-Domäne (grün) sowie die RID-Domäne (rot).

kurzem in der Maus (Hong *et al.*, 2001), im Menschen (Boylan & Wright, 2000) und im Rind (Roepman *et al.*, 2000) identifizierten *Retinitis Pigmentosa* GTPase Regulator-interagierenden Protein (RGPRIP) erkennen. Die computergestützte Analyse der abgeleiteten Aminosäuresequenz zeigte, dass die *N*-terminale Hälfte (AS 43-576) des vermeintlichen Ftm Proteins genauso wie RGPRIP insgesamt drei verwickelte Knäuel-Domänen ('*coiled-coil*'; Lupas, 1996) enthält, die dem gesamten Bereich eine α -helikale Sekundärstruktur geben. Diese Region zeigt eine signifikante Ähnlichkeit zu der '*coiled-coil*'-Struktur der beiden schweren Ketten eines Myosin-Schwanzes und könnte daher ebenfalls bei der Bildung von Proteindimeren eine Rolle spielen. Weiterhin ist in diesem Zusammenhang ebenfalls die hohe Ähnlichkeit dieses Bereichs zu den '*coiled-coil*'-Domänen einiger intermediärfilament-bindenden Proteine (Desmoplakin I und II), einiger Motor- bzw. Transportproteine (neben Myosin z.B. Hefe USO1) und eines mikrotubuli-bindenden Proteins (Restin) zu erwähnen. An die drei *coiled-coil* Domänen schließt sich eine C2-artige Domäne an (AS 792-887). Diese Domäne wurde ursprünglich in der Proteinkinase C Familie beschrieben (Azzi *et al.*, 1992). Homologe Regionen wurden jedoch auch in einigen anderen phospholipid-bindenden Proteinen identifiziert (z.B. Synaptotagmin; Sutton *et al.*, 1995). Es wird angenommen, dass die C2-Domäne bei der kalzium-abhängigen Bindung oder aber auch bei der kalzium-unabhängigen Bindung von Phospholipiden (z.B. von Inositol-1,4,5-triphosphat) eine Rolle spielt. Am *C*-terminalen Ende des Ftm Proteins (AS 1076-1264) befindet sich eine Domäne, die eine sehr hohe Ähnlichkeit zu der mit dem *Retinitis Pigmentosa* GTPase Regulator-Protein interagierenden Domäne des RGPRIP Proteins aufweist. Insgesamt beträgt die Ähnlichkeit zu RPGRIP 70%. Damit ist RPGRIP das mit Ftm am nächsten verwandte Protein.

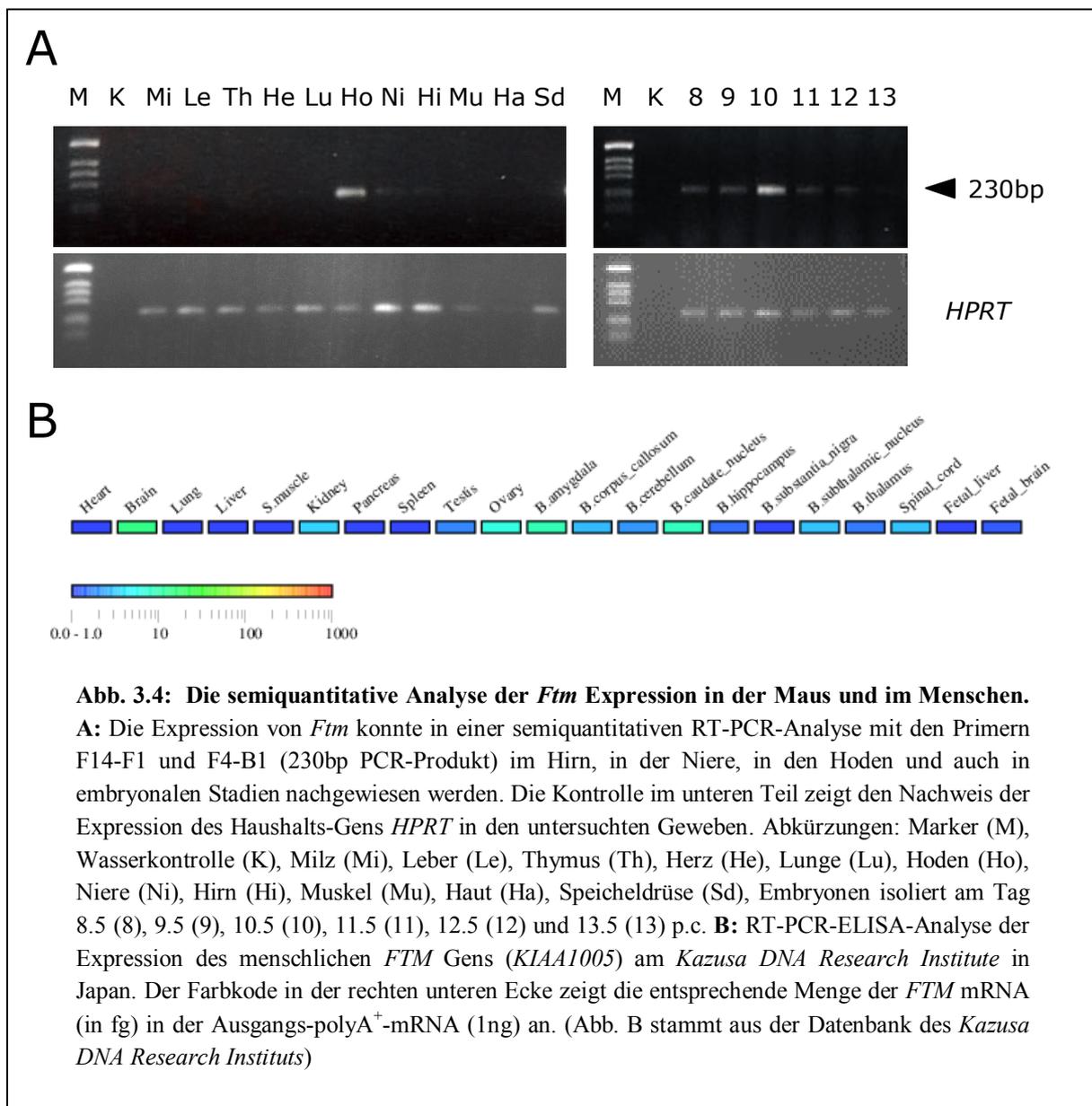
3.2.3 Die Expression von *Ftm* im Embryo und im ausgewachsenen Tier

Die Bestimmung des Expressionsmusters eines Kandidatengens erlaubt eine erste Bewertung, ob eine Mutation des identifizierten Gens als Ursache für den zugrundeliegenden Phänotyp in Betracht gezogen werden kann. So wird von einem Kandidatengen erwartet, dass dessen räumliche und zeitliche Expression mit der Ausbildung des entsprechenden Phänotyps korreliert.

Mittels einer semiquantitativen RT-PCR wurde das Expressionsmuster von *Ftm* während der Embryogenese und in der ausgewachsenen Maus bestimmt. Diese Analyse ließ eine sehr eingeschränkte Expression in den untersuchten Organen einer ausgewachsenen Maus erkennen. Nur im Gehirn, den Nieren und in den Hoden konnte eine signifikante Expression nachgewiesen werden. Während der Embryogenese wurde eine Expression in allen untersuchten Stadien von Tag 8.5 bis Tag 13.5 festgestellt. Weiterhin konnte durch eine RT-

PCR die Expression von *Ftm* in differenzierenden embryonalen Stammzellen nachgewiesen werden (ohne Abb.). Die Expression des *Ftm* Gens erfolgt daher schon in früheren Stadien der Embryogenese.

Die räumliche Expression von *Ftm* Transkripten während der Embryogenese sollte durch die *in situ* Hybridisierung von vollständigen Embryonen dargestellt werden. Dazu wurden unterschiedliche Bereiche der *Ftm* cDNA für die Herstellung entsprechender RNA-Sonden genutzt. Aufgrund der schwachen Expression von *Ftm* konnte jedoch keine der eingesetzten Sonden einen signifikanten Nachweis der spezifischen räumlichen Verteilung der *Ftm* Transkripte in 8.5 p.c. bis 12.5 p.c. Tage alten Embryonen erbringen.



Seit 1994 werden im Rahmen eines cDNA Projekts am *Kazusa DNA Research Institute* in Japan im großen Maßstab lange cDNA-Sequenzen aus der RNA des menschlichen Gehirns isoliert (Nomura *et al.*, 1994; Ohara *et al.*, 1997). Dabei wird mit Hilfe einer semi-quantitativen RT-PCR-ELISA-Analyse das Expressionsprofil der identifizierten Gene in ausgewählten Geweben bestimmt und in einer Datenbank öffentlich zur Verfügung gestellt (Kikuno *et al.*, 2000). Die dort durchgeführte Expressionsanalyse des menschlichen *FTM* Gens (Teilsequenz *KIAA1005*, Nagase *et al.*, 1999) zeigte ebenfalls eine sehr schwache Expression von *FTM* Transkripten in den untersuchten Geweben von erwachsenen Menschen und menschlichen Embryonen (Abb. 3.4-B). Eine signifikante Expression konnte nur im Gehirn, besonders in der Amygdala und im kaudalen Nukleus nachgewiesen werden. Das Expressionsprofil des menschlichen *FTM* Gens bestätigt somit die in der Maus gewonnenen Erkenntnisse (Abb. 3.4).

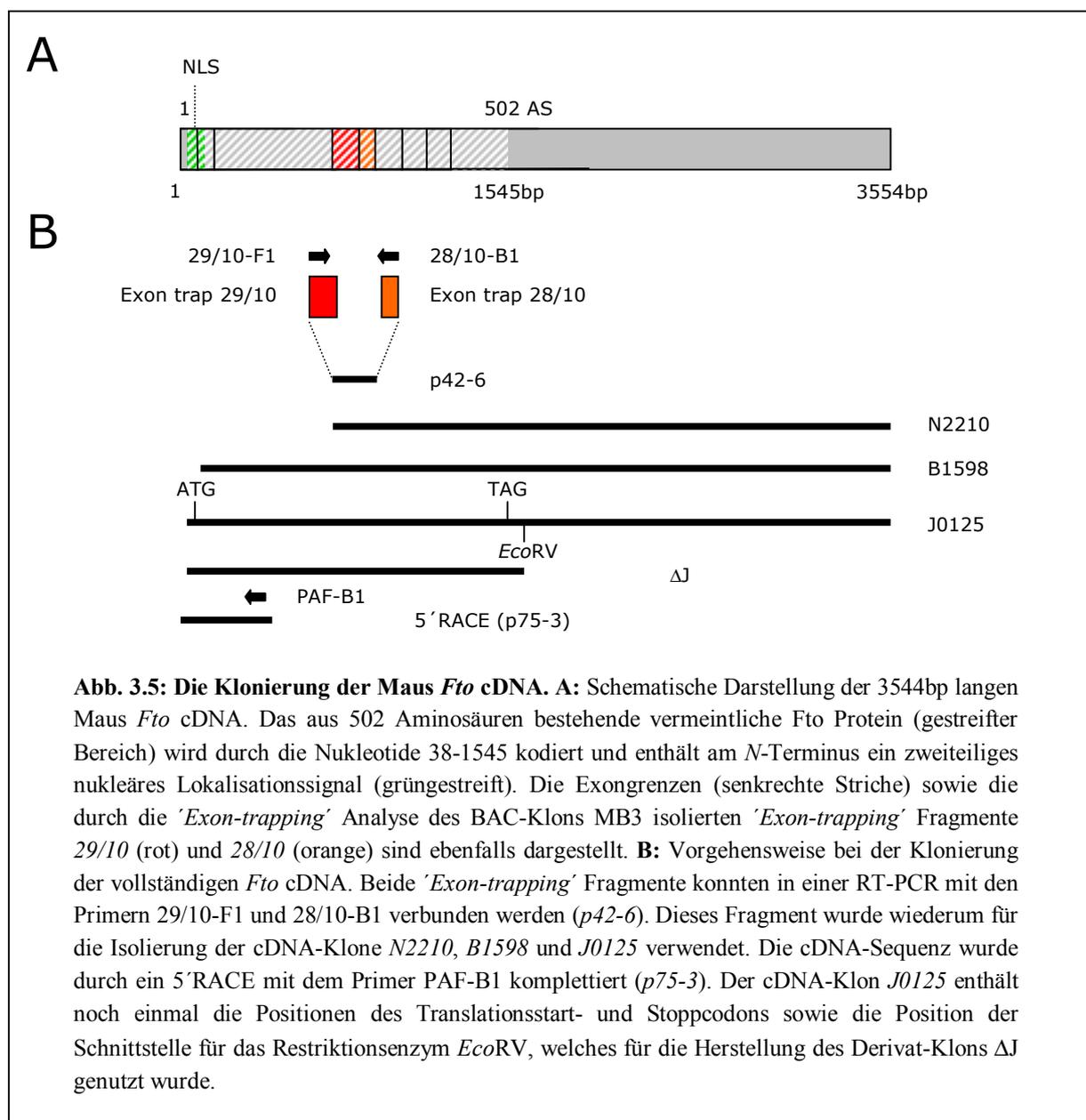
Zusammenfassend lässt sich sagen, dass *Ftm* in der Maus während der Embryogenese zum Zeitpunkt der Ausprägung des *Ft*-Phänotyps exprimiert wird. Allerdings konnte keine Expression im Thymus ausgewachsener Mäuse festgestellt werden. Trotzdem kommt *Ftm* als Kandidatengenen, dessen Mutation für die Ausprägung der beobachteten Missbildungen mit verantwortlich ist, in Betracht.

3.3 Die Positionsklonierung von *Fatso* (*Fto*), ein zuvor unbekanntes Gen

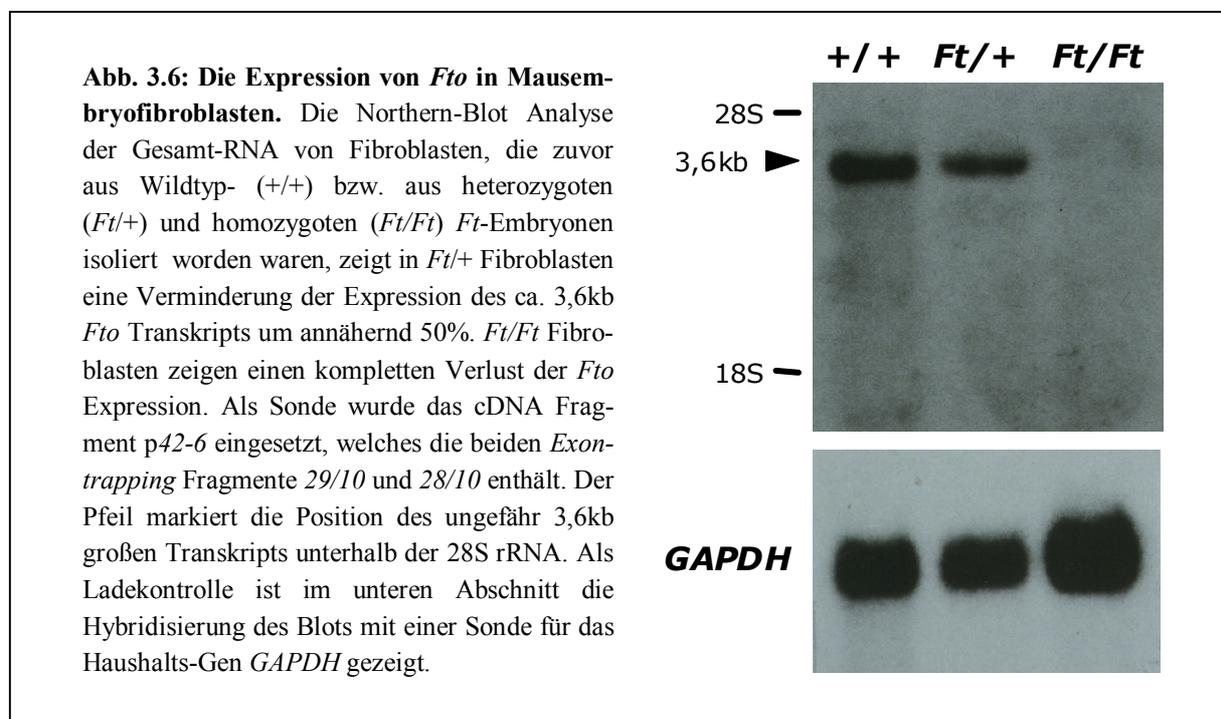
Auch die in einer RT-PCR Analyse erstellten Expressionsprofile der 'Exon-trapping'-Fragmente 28/10 und 29/10 (Tab. 3.1) zeigten, dass beide Sequenzen in der RNA aus embryonalen Geweben bzw. Organen ausgewachsener Mäuse in einem übereinstimmenden Muster exprimiert werden (ohne Abb.). Daher wurde auch hier versucht, beide Fragmente mit Hilfe einer RT-PCR auf der Ebene eines Transkripts zu verbinden. Die Kombination der Primer 29/10-F1 und 28/10-B1 führte zum Nachweis eines 221bp PCR-Produkts (Abb. 3.5 u. 3.12). Dabei zeigte die Größe der erhaltenen cDNA an, dass beide Sequenzen direkt nebeneinander auf ein und demselben Transkript lokalisiert sind. Zudem bewies die Northern-Blot-Analyse der Gesamt-RNA von Mausembryo-Fibroblasten (MEF), dass das identifizierte Transkript in homozygoten *Ft/Ft* MEF nicht exprimiert wird (Abb. 3.6). Ebenso zeigten heterozygote *Ft/+* MEF im Vergleich zum Wildtyp eine um annähernd 50% verminderte Expression des ungefähr 3,6kb großen Transkripts. Die 'Exon-trapping'-Sequenzen 29/10 und 28/10 sind daher Bestandteil eines weiteren Transkripts, dessen kodierende Sequenzen durch die *Ft* Mutation betroffen sind.

3.3.1 Die Isolierung der vollständigen *Fto* cDNA

Um die vollständige cDNA-Sequenz dieses Transkripts zu identifizieren, wurde eine cDNA-Bibliothek (Maus-Embryo, Tag 9 p.c.) mit Hilfe des zuvor isolierten 221bp PCR-Produkts nach Klonen durchsucht, die weitere Bestandteile der gesuchten cDNA enthalten. Dabei konnten drei verschiedene cDNA-Klone identifiziert werden: *N2210* (MPMGc559N2210Q3), *B1598* (MPMGc559B1598Q3) und *J0125* (MPMGcJ0125Q3). Der cDNA-Klon *J0125* enthielt dabei das längste cDNA-Fragment (3531bp), gefolgt von *B1598* (3499bp) und *N2210* (2742bp). Dieser Längenunterschied ist ausschließlich auf einen vorzeitigen Abbruch der cDNA-Synthese im 5'-Bereich zurückzuführen, da alle drei Klone identische 3'-Enden aufwiesen. Die Sequenz des längsten cDNA-Klons *J0125* konnte durch ein 5'-RACE um



weitere zehn Nukleotide verlängert werden. Insgesamt korreliert die Länge der isolierten cDNA von 3,54kb sehr gut mit der in einem Northern-Blot beobachteten Transkriptlänge (Abb. 3.6). Zudem sind im 3'-Ende der isolierten Sequenz zwei klassische Polyadenylierungssignale ('AATAAA') enthalten. Es kann daher davon ausgegangen werden, dass die erhaltene Sequenz der vollständigen cDNA des identifizierten Transkripts entspricht. Ein Sequenzvergleich zeigte keinerlei authentische Übereinstimmungen zu bisher bekannten cDNA-Sequenzen (siehe Abschnitt 3.3.2). Es handelte sich also um das Produkt eines neuen, zuvor unbekanntes Gens. Die Southern-Blot- und PCR-Analyse von genomischer DNA mit Proben, die vom 5'-Ende bzw. 3'-Ende der isolierten cDNA stammten, bewiesen, dass diese Sequenzen vollständig in homozygoten *Ft*-Mäusen verlorengegangen sind (Abb. 3.17, 3.18). Weiterhin konnte dadurch gezeigt werden, dass das 5'-Ende ausschließlich auf dem BAC-Klon MB3 zu finden ist. Das 3'-Ende hingegen lag in einem zum damaligen Zeitpunkt unbekanntem Abschnitt des in der *Ft*-Mutation deletierten Bereichs und konnte in einer parallel angefertigten Dissertation (Ausmeier, 2000) zur Isolierung neuer genomischer Fragmente genutzt werden (*'chromosome jumping'*). Ebenso offenbarte diese erste strukturelle Analyse, dass das neu identifizierte Gen trotz seiner Transkriptlänge von nur 3,54kb einen genomischen Bereich von mehr als 350kb überspannt. Aufgrund dieser Tatsache wurde das identifizierte und zuvor unbekanntes Gen *Fatso* (*Fto*) genannt.

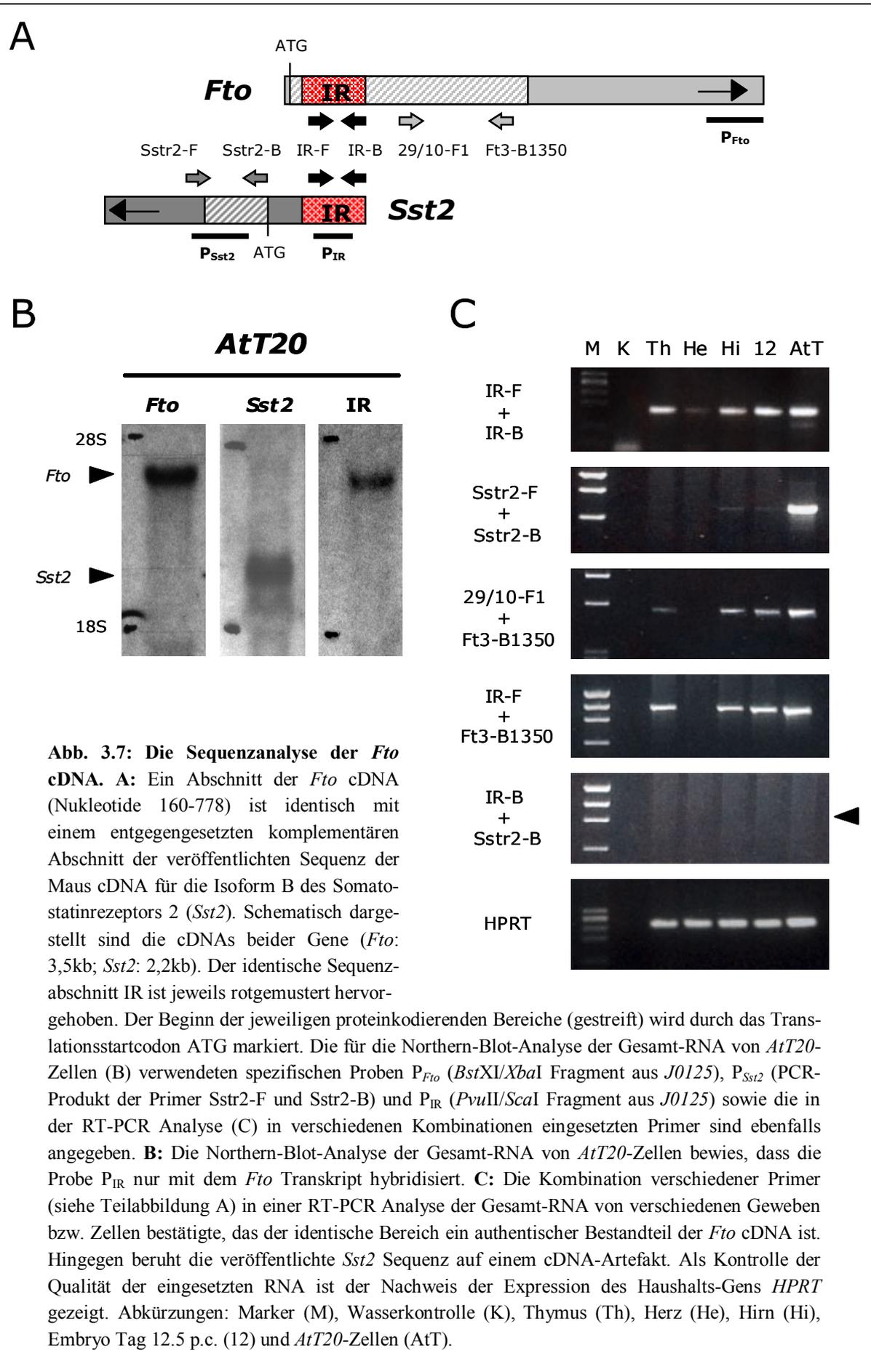


3.3.2 Die Analyse der *Fto* cDNA-Sequenz

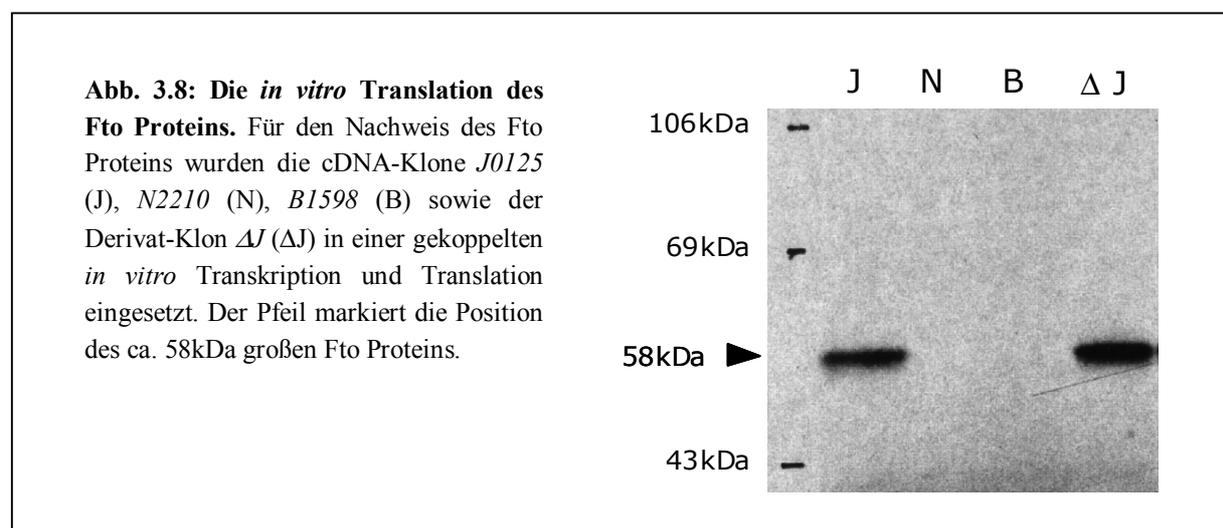
Interessanterweise zeigte ein Sequenzvergleich mit DNA-Sequenzen der *NCBI*-GenBank, dass ein Abschnitt der *Fto* cDNA (Nukleotide 160-778) fast identisch (98%) mit einem Teil der entgegengesetzten komplementären Sequenz der Maus cDNA für die Isoform B des Somatostatinrezeptors 2 (*Sst2*) ist. Diese Sequenzen wurden als Exon 1 und 5' flankierende Sequenzen eines alternativen Promotors des *Sst2* Gens beschrieben, die auf der Ebene des Genoms jedoch nicht mit anderen *Sst2* Sequenzen verbunden werden konnten (Kraus *et al.*, 1998). Da das *Sst2*-Gen im Gegensatz zu *Fto* (MMU8) auf dem Maus-Chromosom MMU11 lokalisiert ist (Brinkmeier & Camper, 1997), sollte durch Northern-Blot- und RT-PCR-Analysen überprüft werden, ob beide Gene tatsächlich diesen identischen Sequenzabschnitt enthalten. Die Northern-Blot-Analyse der Gesamt-RNA von *AtT20*-Zellen mit Proben, die spezifisch für *Fto* bzw. *Sst2* sind, zeigte, dass diese Zellen sowohl das 3,6kb lange *Fto* Transkript als auch das 2,2kb lange *Sst2* Transkript exprimieren. Eine für den identischen Sequenzabschnitt spezifische Probe hybridisierte jedoch nur mit dem *Fto* Transkript (Abb. 3.7-B). Dieses Northern-Blot Ergebnis wurde durch weitere RT-PCR-Analysen bestätigt. Dabei stimmte das Expressionsmuster des identischen cDNA-Abschnitts mit der Expression des *Fto* Gens exakt überein (Abb. 3.7-C). Hingegen unterschied sich die Expression des *Sst2* Gens deutlich von diesem Expressionsprofil. Ebenso führte die Kombination des Primers IR-F aus dem identischen Sequenzabschnitt mit dem *Fto*-Primer Ft3-B1350 zum Nachweis des erwarteten 1049bp langen PCR-Produkts. Im Gegensatz dazu konnte niemals ein PCR-Produkt mit dem Primer IR-B (ebenfalls aus dem identischen Sequenzabschnitt) und dem *Sst2*-Primer Sstr2-B erzeugt werden. Sowohl die Northern-Blot- als auch die RT-PCR-Analysen bewiesen somit, dass der identische Sequenzabschnitt ein exklusiver Bestandteil der *Fto* cDNA ist. Die veröffentlichte cDNA-Sequenz für die Isoform B des Somatostatinrezeptors 2 (Vanetti *et al.*, 1992) basiert daher sehr wahrscheinlich auf der Sequenzanalyse eines Hybrid-cDNA-Klons.

3.3.3 Das *Fto* Protein

Die Sequenzanalyse der vorliegenden *Fto* cDNA enthüllte einen offenen Leserahmen, der für ein vermeintliches Protein bestehend aus 502 Aminosäuren und mit einem theoretischen Molekulargewicht von 57,98kDa kodiert. Der abgeleitete Translationsstart beginnt an Position 38. Allerdings entsprechen die das Startcodon umgebenden Sequenzen nicht den publizierten Konsensussequenzen für einen Translationsstart (Kozak, 1987). An den Translationsstopp



(Position 1544) schließt sich eine langer 3'UTR von 1997bp an. Um den abgeleiteten offenen Leserahmen experimentell zu bestätigen, sollten die vorhandenen cDNA-Klone (Abb. 3.5) in einer gekoppelten *in vitro* Transkription und Translation für den Nachweis des vermeintlichen Fto Proteins genutzt werden. Wie erwartet konnten in diesem Experiment mit Hilfe des cDNA-Klons *J0125*, der im Gegensatz zu *B1598* und *N2210* den abgeleiteten Translationsstart enthielt, ein ungefähr 58kDa großes Protein synthetisiert werden (Abb. 3.8). Ebenso erlaubte die Verwendung des Derivat-Klons ΔJ , der durch die restriktions-enzymatische Abspaltung (*EcoRV*) der 3'UTR-Sequenzen (Position 1582 bis 3541) von *J0125* erzeugt worden war (Abb. 3.5), den Nachweis desselben Proteins. Die *in vitro* Translation des Fto Proteins bestätigte somit den abgeleiteten offenen Leserahmen. GleichermäÙen stimmt die Größe des *in vitro* synthetisierten Proteins mit der theoretisch errechneten Größe des Fto Proteins von 57,98kDa überein.



Die Aminosäuresequenz des Fto Proteins zeigte keine signifikanten Homologien zu bisher bekannten Proteinfamilien. Allerdings legt das Vorhandensein eines zweiteiligen nukleären Lokalisierungssignals (NLS) am *N*-terminalen Ende des Proteins eine Funktion im Zellkern nahe. Diese Signalsequenzen steuern den selektiven Transport von Zellkernproteinen nach ihrer Synthese im Zytoplasma über die Kernporen der Kernmembran in das Innere des Zellkerns.

3.3.4 Die Bestimmung der subzellulären Lokalisation des Fto Proteins

Um erste Hinweise auf die Funktion des Fto Proteins zu erhalten, sollte dessen subzelluläre Lokalisation in Fto-exprimierenden Zellen bestimmt werden. Aufgrund der Präsenz eines

vermeintlichen zweiteiligen nukleären Lokalisierungssignals wurde für das Fto Protein nach dessen Synthese im Zytoplasma eine Verlagerung in den Kernraum erwartet. Da keine spezifischen Antikörper gegen das Fto Protein zur Verfügung standen, sollte die Bestimmung der subzellulären Lokalisation mit Hilfe von Säugetier-Zellen (*COS-7* Affennierenzellen und *NIH3T3* Fibroblasten) erfolgen, die transient verschiedene Fto-Fusionsproteine exprimierten. Dazu wurde zunächst für den Nachweis durch die indirekte Immunfluoreszenz der Expressionsvektor *p92-1* konstruiert, der für ein Fto-Fusionsprotein kodiert, welches am N-terminalen Ende das 15 Aminosäure lange Hämagglutinin (HA)-Epitop des menschlichen *Influenza* Virus trägt (HA-Fto, Abb. 3.9). Weiterhin wurden für den direkten Fluoreszenznachweis die Expressionsvektoren *p152-4* und *pLF529* hergestellt, die beide jeweils für ein Fto-Fusionsprotein kodieren, welches am C-terminalen Ende das 265 Aminosäure lange Grün-Fluoreszierende-Protein trägt (Fto-GFP, Abb. 3.9-E). Die Expression dieser Fusionssequenzen wurde bei *p92-1* und *p152-4* unter die Kontrolle des *CMV* Promotors gestellt, um ausreichende Mengen der Fusionsproteine für diesen Nachweis zu produzieren. Im Fall des Expressionsvektors *pLF529*, der im Gegensatz zu den anderen beiden Vektoren ein Konstrukt aus genomischer DNA und cDNA enthielt, wurde die Expression des Fto-GFP Proteins durch den endogenen *Fto*-Promotor kontrolliert.

Die Immunfluoreszenzanalyse der mit dem Expressionsvektor *p92-1* transient transfizierten Zelllinien (*COS-7*, *NIH3T3*) zeigte, dass die Lokalisation des HA-Fto-Fusionsproteins immer auf den Zellkern beschränkt war (Abb. 3.9). Diese exklusive Präsenz im Zellkern konnte durch die direkte Fluoreszenzmikroskopie von *COS-7* Zellen, die aufgrund einer vorherigen Transfektion mit den Expressionsvektoren *p152-4* und *pLF529* das Fto-GFP-Fusionsprotein exprimierten, bestätigt werden (ohne Abb.). Aus den vorliegenden Ergebnissen lässt sich somit schließen, dass es sich bei dem endogenen Fto Protein um ein Zellkernprotein handelt.

3.3.5 Die Klonierung der menschlichen *FTO* cDNA

Die Charakterisierung von unbekanntem Genen kann durch die vergleichende Analyse orthologer cDNA-Sequenzen aus verschiedenen Spezies erheblich erleichtert werden. Daher sollte die vollständige cDNA eines orthologen menschlichen *FTO* Gens identifiziert und kloniert werden. Dazu wurde die *NCBI*-GenBank nach EST-Sequenzen durchsucht, die für Teile der orthologen *FTO* cDNA kodieren. Zum damaligen Zeitpunkt konnte in der Datenbank nur eine 422bp lange EST-Sequenz [AA190384] identifiziert werden, die einen Sequenzabschnitt (Nukleotide 13–296) enthielt, welcher zu 83 % der Maus cDNA im Bereich der Nukleotide 1285-1570 entsprach (Abb. 3.10). Der dieser EST-Sequenz zugrunde liegende

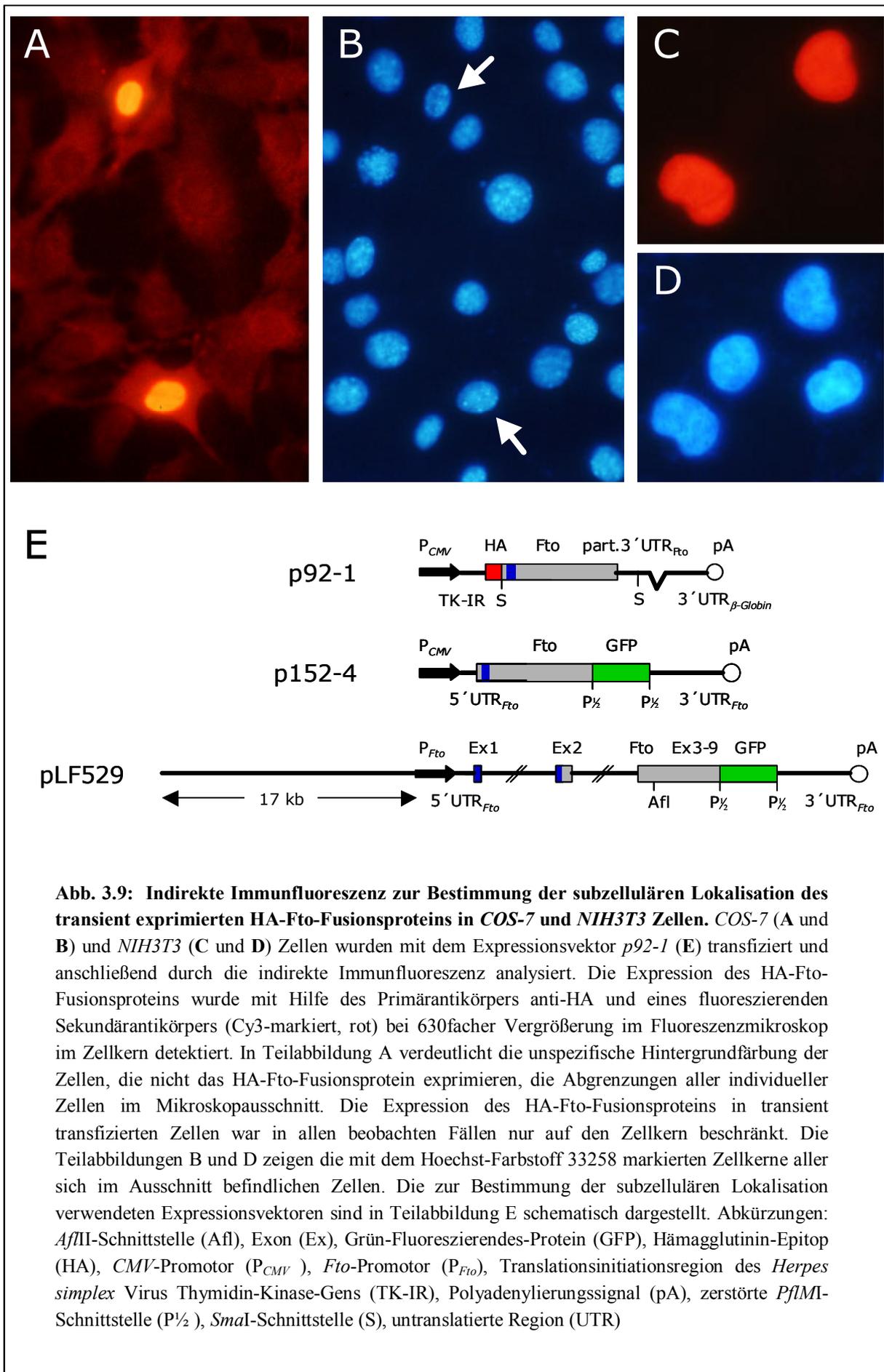
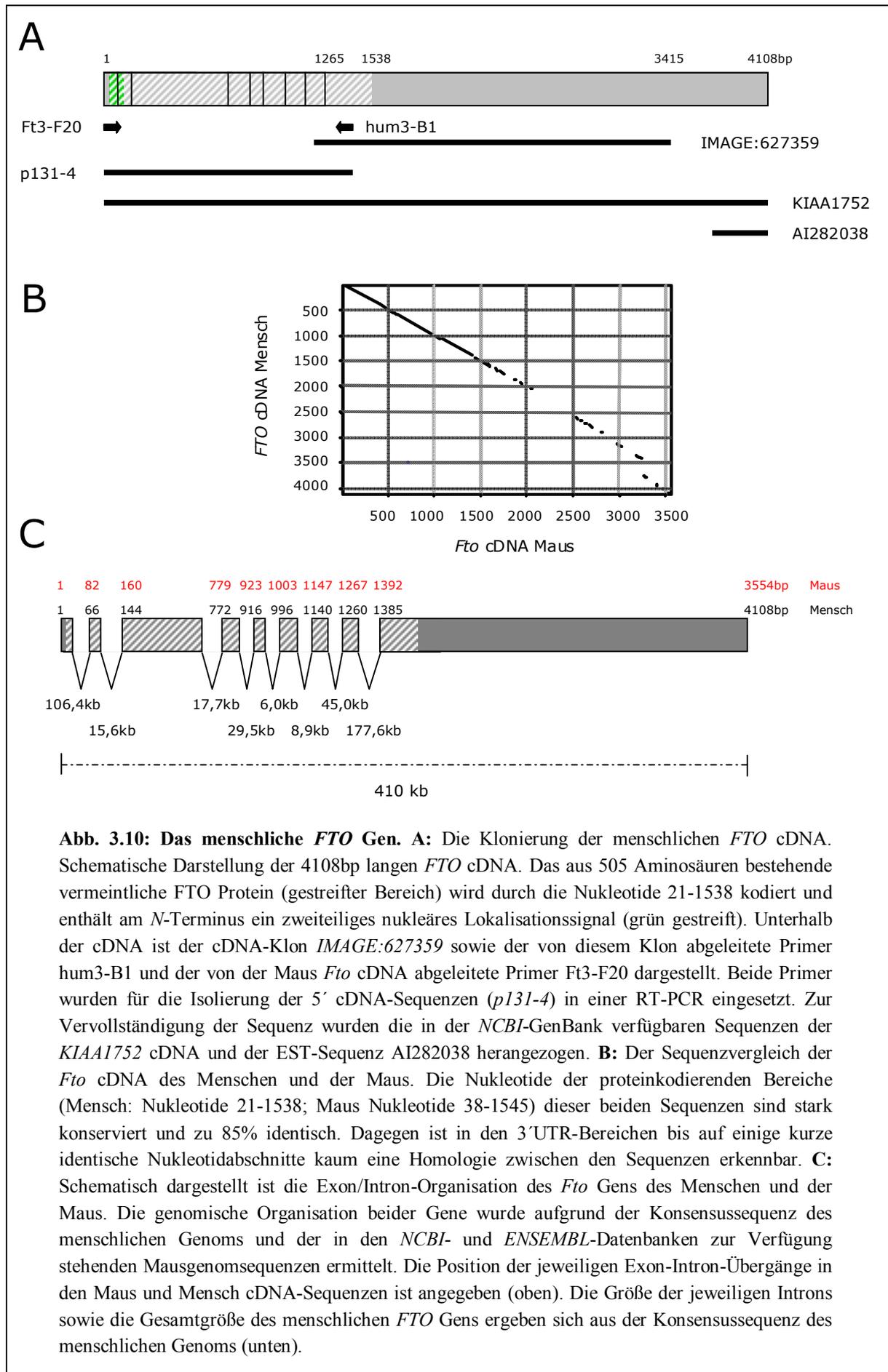


Abb. 3.9: Indirekte Immunfluoreszenz zur Bestimmung der subzellulären Lokalisation des transient exprimierten HA-Fto-Fusionsproteins in *COS-7* und *NIH3T3* Zellen. *COS-7* (A und B) und *NIH3T3* (C und D) Zellen wurden mit dem Expressionsvektor *p92-1* (E) transfiziert und anschließend durch die indirekte Immunfluoreszenz analysiert. Die Expression des HA-Fto-Fusionsproteins wurde mit Hilfe des Primärantikörpers anti-HA und eines fluoreszierenden Sekundärantikörpers (Cy3-markiert, rot) bei 630facher Vergrößerung im Fluoreszenzmikroskop im Zellkern detektiert. In Teilabbildung A verdeutlicht die unspezifische Hintergrundfärbung der Zellen, die nicht das HA-Fto-Fusionsprotein exprimieren, die Abgrenzungen aller individueller Zellen im Mikroskopausschnitt. Die Expression des HA-Fto-Fusionsproteins in transient transfizierten Zellen war in allen beobachteten Fällen nur auf den Zellkern beschränkt. Die Teilabbildungen B und D zeigen die mit dem Hoechst-Farbstoff 33258 markierten Zellkerne aller sich im Ausschnitt befindlichen Zellen. Die zur Bestimmung der subzellulären Lokalisation verwendeten Expressionsvektoren sind in Teilabbildung E schematisch dargestellt. Abkürzungen: AflII-Schnittstelle (Afl), Exon (Ex), Grün-Fluoreszierendes-Protein (GFP), Hämagglutinin-Epitop (HA), CMV-Promotor (P_{CMV}), Fto-Promotor (P_{Fto}), Translationsinitiationsregion des *Herpes simplex* Virus Thymidin-Kinase-Gens (TK-IR), Polyadenylierungssignal (pA), zerstörte *PfIMI*-Schnittstelle ($P_{1/2}$), *SmaI*-Schnittstelle (S), untranslatierte Region (UTR)

cDNA-Klon *IMAGE:627359* wurde vom RZPD zur Verfügung gestellt und vollständig sequenziert. Die so erhaltene cDNA kodierte jedoch nur den C-terminalen Bereich des vermeintlichen FTO Proteins und den größten Teil des 3'UTR. Die fehlenden cDNA-Sequenzen des proteinkodierenden Bereichs konnten durch eine RT-PCR mit Hilfe der Primer Ft3-F20 und hum3-B1 aus der RNA von *HeLa* Zellen (Zelltyp: menschliches Zervix-Tumorgewebe; Scherer *et al.*, 1953) isoliert werden (Abb. 3.10). Während der Anfertigung dieser Arbeit wurde der größte Teil der menschlichen *FTO* cDNA ebenfalls im Zuge des *Kazusa* cDNA Sequenzierungsprojekts isoliert (Nagase *et al.*, 2000). Die in diesem Projekt identifizierte cDNA *KIAA1752* konnte *in silico* am 3'-Ende durch eine EST-Sequenz [AI282038] um 15 Nukleotide verlängert werden. Die Gesamtlänge der identifizierten humanen *FTO* cDNA beträgt somit 4108bp. Der längste offene Leserahmen wird von den Nukleotiden 21–1538 der *FTO* cDNA kodiert. Das vermeintliche FTO Protein besteht aus 505 Aminosäuren und ist zu 86 % identisch bzw. zu 92% ähnlich mit dem Maus *Fto* Protein. Gleichermäßen enthält der *N*-Terminus des menschlichen Proteins ein zweiteiliges nukleäres Lokalisierungssignal (AS 2-18). Weitere bekannte Proteindomänen konnten in einer computergestützten Analyse der Aminosäuresequenz jedoch nicht identifiziert werden.

Ein Vergleich der Maus und Mensch *Fto* cDNA-Sequenzen zeigte, dass die Homologie zwischen beiden Genen nur im proteinkodierenden Bereich erhalten geblieben ist (85% Identität). Dagegen ist in den 3'UTR-Bereichen mit Ausnahme von einigen kurzen identischen Abschnitten von 27 bzw. 75 Nukleotiden die Homologie fast vollständig verloren gegangen (Abb. 3.10-B). Ein Vergleich der Exon/Intron-Organisation beider Gene, die aufgrund der bereits in verschiedenen Datenbanken verfügbaren Genomsequenzen des Menschen und der Maus aufgestellt werden konnte, zeigte, dass die Genstruktur während der Evolution zwischen beiden Spezies erhalten geblieben ist (Abb. 3.10-C). Beide Gene verfügen über neun Exons und auch im Genom des Menschen beansprucht die Organisation des *FTO* Gens mehrere hundert kb (ca. 410kb). Interessant ist dabei, dass insbesondere die Exons 1 und 2 bzw. 8 und 9 durch sehr lange Intronsequenzen (Intron 1: 106,4kb; Intron 8: 177,6kb) getrennt werden.

Die Sequenzuntersuchungen der menschlichen cDNA und des menschlichen Proteins bestätigten somit die aus den Mausequenzen gewonnenen Erkenntnisse, konnten aber nicht zu einer weiteren Charakterisierung des *Fto* Gens beitragen. Weitere homologe Sequenzen konnten in der *NCBI*-GenBank in der Ratte *Rattus norvegicus* [289555], im Krallenfrosch *Xenopus laevis* [BG410196] und im Rind *Bos taurus* [BF653272] sowie im Pufferfisch *Tetraodon nigroviridis* [AL192638] identifiziert werden. Die Analyse der nahezu vollständigen Genomsequenzen der Hefe *Saccharomyces cerevisiae*, des Fadenwurms *Caenorhabditis elegans* und der Fruchtfliege *Drosophila melanogaster* ließ jedoch keine



homologen Sequenzen in diesen Organismen erkennen. Insgesamt konnten orthologe *Fto* Sequenzen bisher nur in Vertebraten nachgewiesen werden.

3.3.6 Die Sequenz der überlappenden Promotorregionen von *Fto* und *Ftm*

Aufgrund des Mausgenomprojekts wächst die Zahl der öffentlich zugänglichen Genomsequenzen ständig an. In den entsprechenden Datenbanken (*NCBI*, *ENSEMBL*) konnten somit auch einige kurze Sequenzabschnitte identifiziert werden, die jeweils den Transkriptionsstart sowie einen Teil der Promotorregion der Gene *Fto* und *Ftm* enthielten. Interessanterweise deckte die Analyse der jeweiligen Promotorregionen eine unmittelbare Nachbarschaft der beiden Gene auf (Abb. 3.11). Obwohl die Kopf an Kopf orientierten Gene im Genom der Maus und des Menschen jeweils einen beträchtlichen Genomabschnitt überspannen (Maus: *Ftm* ca. 100kb, *Fto* ca. 350kb (Abb. 3.18); Mensch: *FTM* 120kb, *FTO* 410kb), werden die vermutlichen Transkriptionsinitiationsstellen im Mausgenom nur durch 284bp voneinander getrennt. Die enge Kopplung von *Fto* und *Ftm* wurde während der Evolution konserviert und kann daher auch im menschlichen Genom nachgewiesen werden, wo die Distanz zwischen den entsprechenden Initiationsstellen 322bp beträgt. Durch die vergleichende Sequenzanalyse der gemeinsamen Promotorregion zwischen Mensch und Maus konnten mit Hilfe des Match-Programms einige konservierte vermeintliche Promotorelemente identifiziert werden. Interessant ist, dass sich jeweils ca. 50bp vor den beiden Transkriptionsstarts ein konserviertes Element von 10bp befindet. Genau wie die eigentlichen Gene zeigen diese beiden Elemente eine Kopf-an-Kopf-Orientierung. Im Zentrum des jeweiligen Elements ist eine vermeintliche Engrailed1 (*En-1*) Bindungsstelle enthalten (Abb. 3.11).

3.3.7 Die Expression von *Fto* im Embryo und im ausgewachsenen Tier

Auch in diesem Fall sollte durch die Erstellung eines Expressionsprofils eine erste Bewertung von *Fto* als Kandidatengen für den *Ft*-Phänotyp erfolgen. Wiederum wurde das Expressionsmuster während der Embryogenese und im ausgewachsenen Tier mittels einer semiquantitativen RT-PCR bestimmt (Abb. 3.12-A). Dabei zeigte sich, dass *Fto* in fast allen untersuchten Organen einer ausgewachsenen Maus exprimiert wird. Nur im Herzen und in der Speicheldrüse konnte keine signifikante Expression nachgewiesen werden. Während der Embryogenese konnte ebenfalls eine starke Expression in allen untersuchten Stadien von Tag 8.5 bis Tag 13.5 p.c. festgestellt werden (Abb. 3.12-A). Desgleichen konnte auf diese Weise

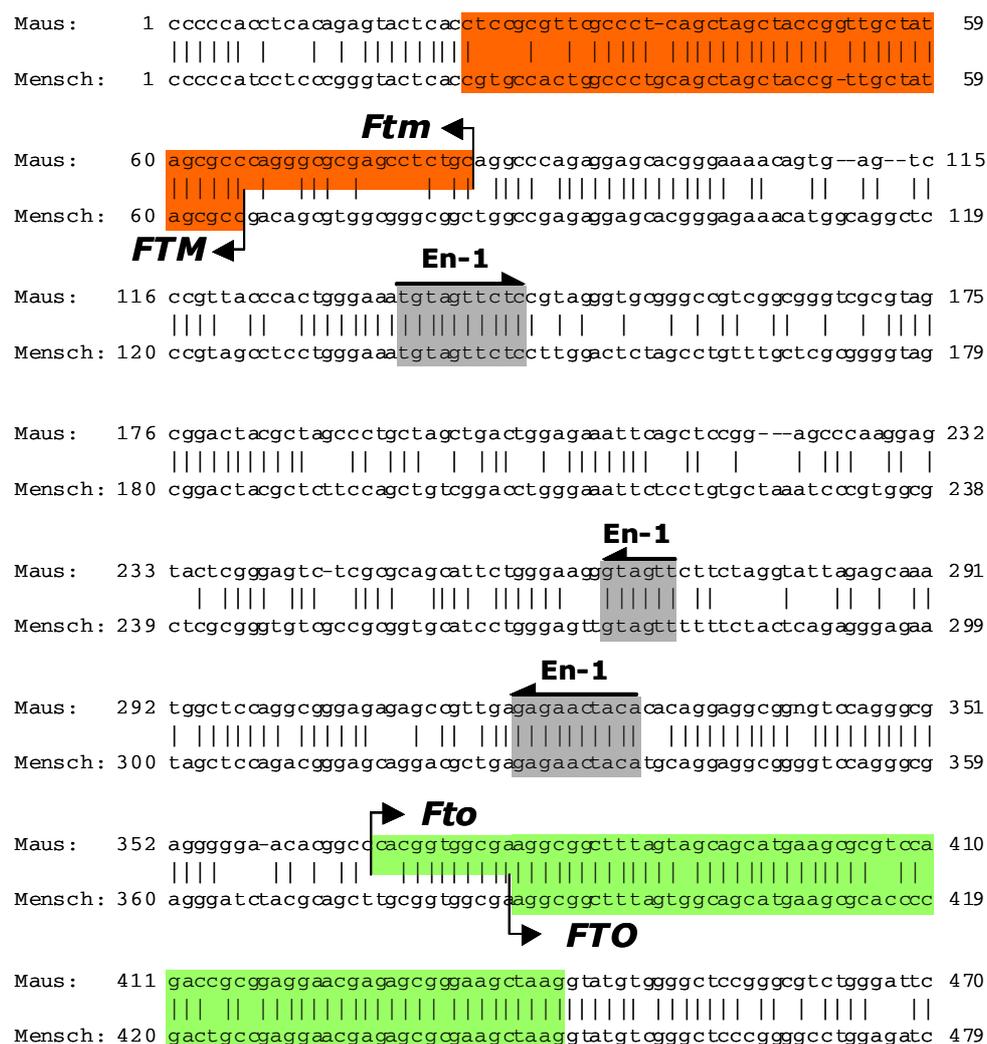


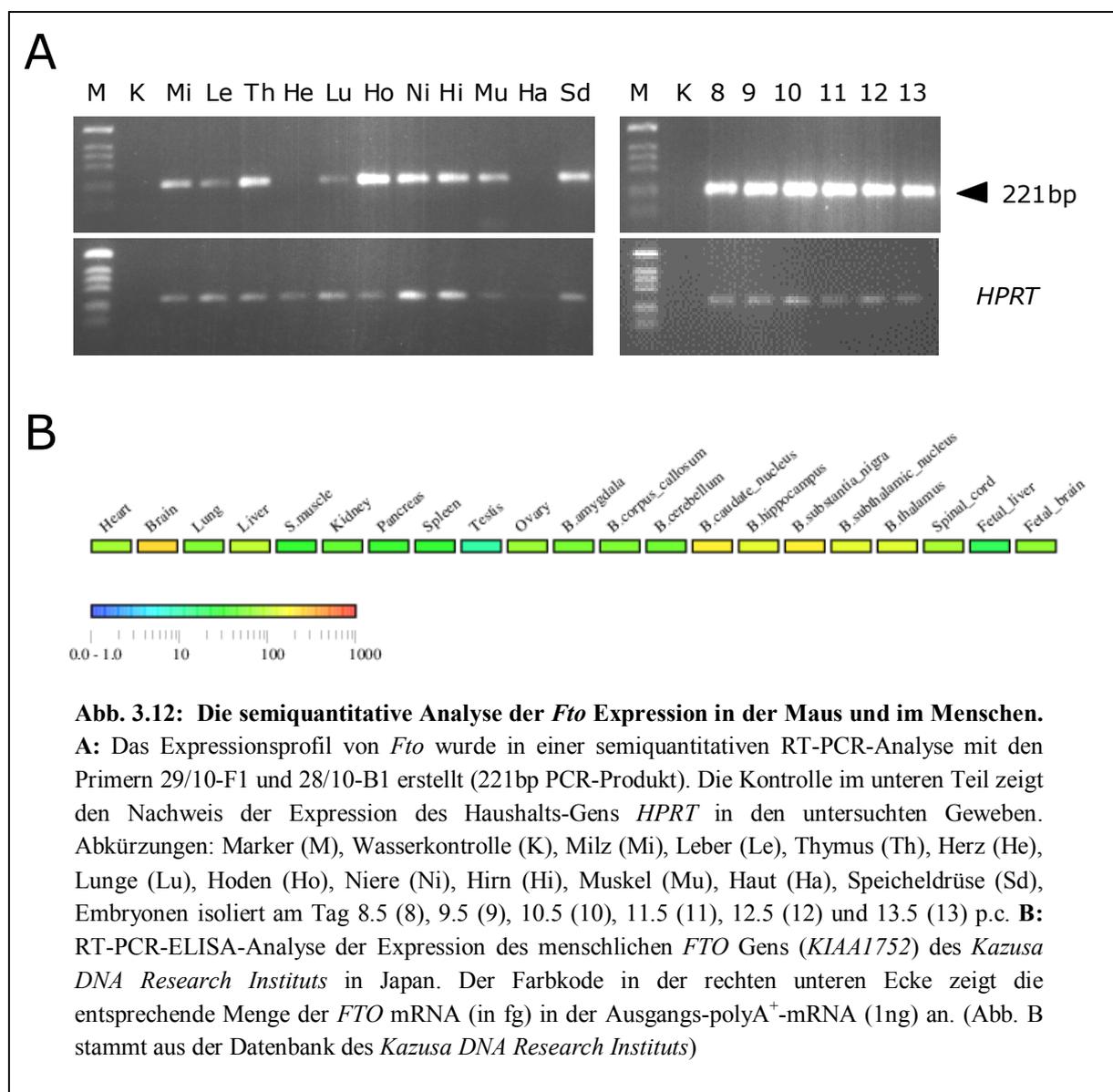
Abb. 3.11: Die Sequenzanalyse der Promotorregionen von *Fto* und *Ftm*. Gezeigt ist der Genomabschnitt zwischen den Transkriptionsstarts der Gene *Fto* und *Ftm*. Beide Gene werden in der Maus (obere Sequenz) nur durch 284bp voneinander getrennt. Im Menschen beträgt die entsprechende Distanz 332bp. Die Sequenzen des jeweils ersten Exons der beiden Gene sind durch Rechtecke gekennzeichnet (*Ftm*: rot; *Fto*: grün). Graue Rechtecke markieren vermeintliche Bindungsstellen für den Transkriptionsfaktor Engrailed-1 (En-1), die zwischen Mensch und Maus hoch konserviert sind. Die halbierten Pfeile geben die Orientierung dieser Elemente an.

die Expression von *Fto* in undifferenzierten embryonalen Stammzellen nachgewiesen werden (ohne Abb.). Die Expression von *Fto* erfolgt also schon in sehr frühen Stadien der Embryogenese.

Die räumliche Expression von *Fto* Transkripten während der Embryogenese sollte mit Hilfe der *in situ* Hybridisierung von Ganzkörperpräparaten hinterfragt werden. Obwohl dazu mehrere RNA-Sonden aus unterschiedlichen Bereichen der *Fto* cDNA eingesetzt wurden, konnte keine dieser Sonden eine spezifische räumliche Verteilung der *Fto* Transkripte in 8.5

bis 12.5 p.c. Tage alten Embryonen aufdecken. Alle eingesetzten Sonden enthüllten vielmehr eine weitreichende, nahezu ubiquitäre Expression im gesamten Embryo (ohne Abb.).

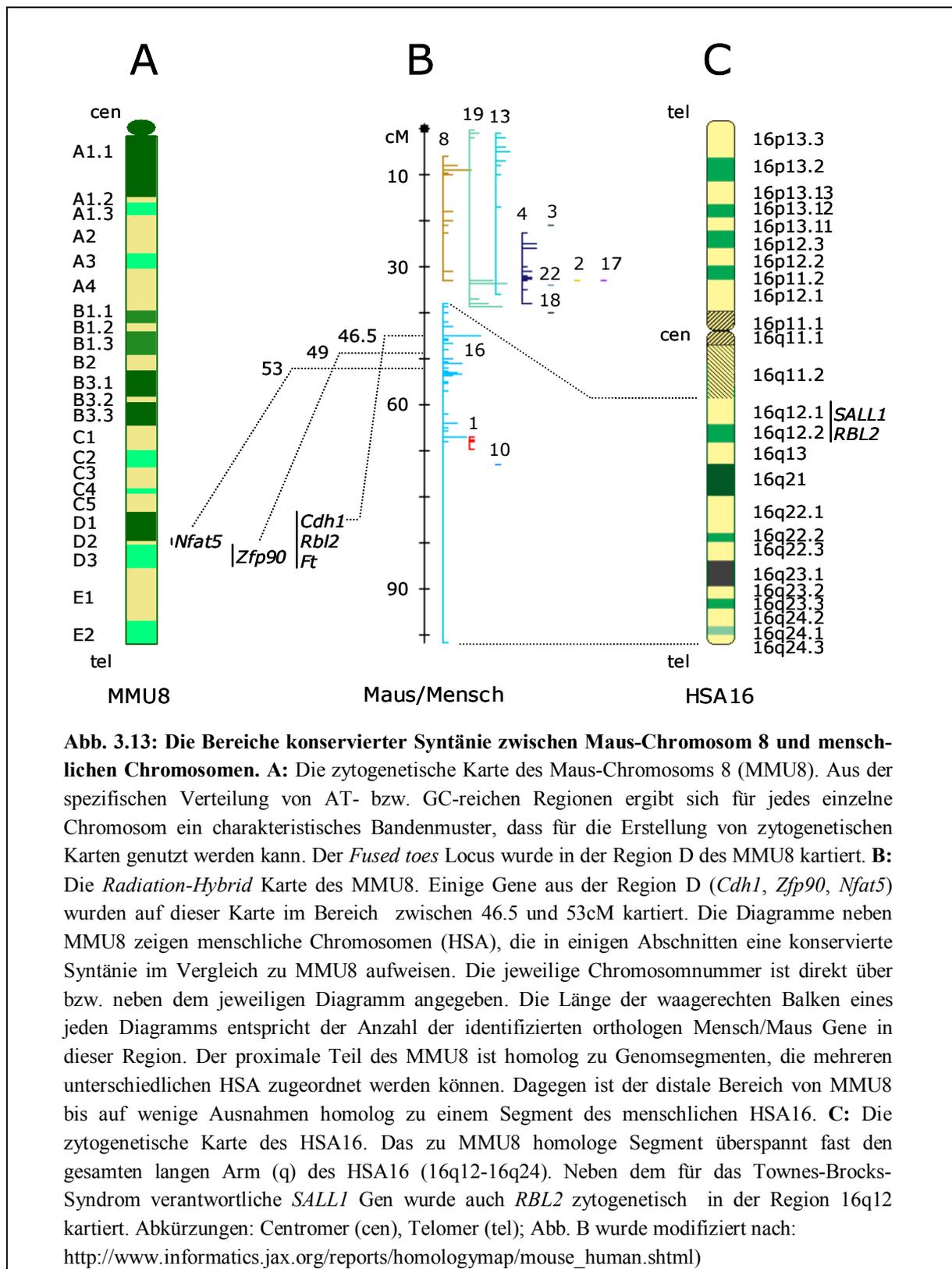
Zur Bestätigung der Expressionsdaten konnte auch hier die Expressionsanalyse des menschlichen *FTO* Gens (*KIAA1752*) am *Kazusa DNA Research Institute* (Japan) herangezogen werden (Abb. 3.12-B). Die über eine Datenbank verfügbaren RT-PCR-ELISA-Ergebnisse ließen ebenfalls eine sehr ausgedehnte Expression der *FTO* Transkripte in ausgewachsenen und embryonalen Geweben erkennen und bestätigten somit die in der Maus gewonnenen Erkenntnisse. Zusammengefasst wird *Fto* in der Maus in den von der *Ft*-Mutation betroffenen Geweben zum Zeitpunkt der Ausprägung der verschiedenen Phänotypen exprimiert und kommt daher als Kandidatengen in Betracht.



3.4 Die Identifizierung von *Ft*-Kandidatengenegen durch die vergleichende Genomanalyse

Sowohl im Menschen als auch in der Maus überspannt das gesamte *Fto* Gen einen bemerkenswert großen Genomabschnitt von ca. 400kb. Aufgrund dieser beträchtlichen Größe ragte das 3'-Ende von *Fto* weit über den zum damaligen Zeitpunkt klonierten Genomabschnitt (Abb. 1.3) hinaus und konnte später in einer parallel angefertigten Arbeit (Ausmeier, 2000) für die Isolierung neuer Genomfragmente im relevanten Genomabschnitt genutzt werden ('*chromosome jumping*'). Da jedoch zunächst keine weiteren Genomfragmente für die positionelle Klonierung von Kandidatengenegen zur Verfügung standen, wurde im weiteren Verlauf der Arbeit auf die vergleichende Genomanalyse und somit auf den positionellen Kandidatengenegenansatz zurückgegriffen. Dies geschah insbesondere durch die Auswertung der verfügbaren physikalischen und genetischen Karten des menschlichen Genoms.

Durch die *in situ* Hybridisierung von Metaphasechromosomen wurde die *Fused toes* Mutation und damit auch die bisher identifizierten Gene *Fts*, *Ftm* und *Fto* sowie das mit *Fts* überlappende Gen *Rbl2* in der Region D des Maus-Chromosoms 8 (MMU8) kartiert (van der Hoeven *et al.*, 1994). Die Kartierung von anderen Genen in dieser Region (Lange *et al.*, 1995; Sage *et al.*, 1997; Hebinck *et al.*, 2000) zeigte, dass dieser Bereich ungefähr dem Abschnitt von 45-55cM der genetischen Karte des MMU8 entspricht (Abb. 3.13). Vergleicht man die genetischen und physikalischen Karten von Maus und Mensch, wird deutlich, dass der distale Bereich des MMU8 (~38-104cM) bis auf wenige Ausnahmen homolog zu einem Segment ist, welches fast den gesamten langen Arm des menschlichen Chromosoms 16 (HSA16) überspannt (Ceci, 1994). Auf einer zytogenetischen Karte des HSA16 entspricht dies den Regionen 16q12 bis 16q24 (Abb. 3.13). Da bereits nachgewiesen werden konnte, dass die Genanordnung innerhalb der homologen Segmente beider Chromosomen anscheinend sehr gut konserviert ist (Becker-Follmann *et al.*, 1997), bestand die Möglichkeit, eine erste, jedoch noch ungenaue Transkriptkarte (*GeneBridge4* (GB4) Karte der GeneMap'99; Deloukas *et al.*, 1998) des HSA16 für die Identifizierung weiterer von der *Ft*-Mutation betroffener Gene heranzuziehen. Als Eintrittspunkt in diese Transkriptkarte diente dabei das menschliche *RBL2* Gen, welches bereits zytogenetisch in der chromosomalen Region 16q12.2 kartiert worden war (Yeung *et al.*, 1993). Innerhalb der GB4-Karte wurde *RBL2* von zwei verschiedenen Arbeitsgruppen in einem Segment zwischen den beiden Mikrosatellitenmarkern *DI6S419* und *DI6S3032* lokalisiert (Tab. 3.2). In unmittelbarer Nähe von *RBL2* wurden von drei verschiedenen Arbeitsgruppen Teilsequenzen des menschlichen *FTO* Gens kartiert. Die Kopplung beider Gene ist zwischen Mensch und Maus also erhalten geblieben (Tab. 3.2). Um weitere Gene in der näheren Umgebung von *Fto* und *Rbl2* zu identifizieren, wurde die NCBI-



GenBank nach Sequenzen durchsucht, die für homologe Mausgene der in diesem GB4-Kartenabschnitt lokalisierten menschlichen Sequenzen kodieren. Konnte ein entsprechendes orthologes Mausgen identifiziert werden, wurde das Vorhandensein dieses Gens in der genomischen DNA von homozygoten *Ft*-Embryonen mittels PCR überprüft. Dabei stellte sich

GB4 Koordinaten (in cR ₃₀₀₀)	STS Marker	Gen-Symbol (Arbeitsgruppe)	Marker- und Genbeschreibung und Ähnlichkeiten zu bekannten Proteinen	Auswertung in dieser Arbeit
347.05	AFM225zf2	D16S419	Microsatellite marker AFM225zf2	
349.17	stSG2036		Human mRNA for KIAA0308 gene	= KIAA0308/2FLJ
349.17	SGC38139		ESTs	
349.27	WI-16741	(WI/CGR)	Human clone 23745 mRNA, complete cds	= FTO
349.40	sts-AA005225		ESTs	
349.48	Cda11d08	(G)	Human clone 23745 mRNA, complete cds	= FTO
349.79	stSG16320		ESTs	
349.79	stSG52822		ESTs, Weakly similar to HPV16 E1 protein	
349.86	Bda63b12	RBL2 (G)	retinoblastoma-like 2 (p130)	
349.89	stSG52750		ESTs	
349.99	WIAF-1971		EST	
350.10	H02267	RBL2 (G)	retinoblastoma-like 2 (p130)	
350.17	WI-18486		ESTs	
350.52	stSG16339	RBL2 (SC)	retinoblastoma-like 2 (p130)	
351.01	stSG51954	RBL2 (SC)	retinoblastoma-like 2 (p130)	
351.05	sts-T03230		EST	
351.05	FB5C6		EST	
351.05	stSG31698	RBL2 (SC)	retinoblastoma-like 2 (p130)	
351.11	stSG31520	(SC)	Human clone 23745 mRNA, complete cds	= FTO
351.15	SGC31994		Human mRNA for KIAA0308 gene,	= KIAA0308/2FLJ
352.11	stSG10226	(SC)	ESTs [R55185]	= IRX3
352.96	sts-M55593	MMP2	matrix metalloproteinase 2 (gelatinase A)	
353.70	sts-U07807	<i>MT4</i>	unknown	
353.74	sts-X91220	<i>SLC12A3</i>	solute carrier family 12 (sodium/chloride...)	
354.26	R34127	<i>SLC6A2</i>	solute carrier family 6 (neurotransmitter...)	
354.96	stSG52056		ESTs	
355.58	stSG62364		ESTs	
355.69	R07176		ESTs	
357.91	AFMa183wd9	D16S3032	Microsatellite anchor marker AFMa183wd9	

Tab. 3.2: Auszug aus der Genebridge4 Radiation-Hybrid Transkript-Karte des menschlichen Chromosoms 16. In der Tabelle sind die in einem Segment des HSA16 zwischen den beiden Mikrosatellitenmarkern *D16S419* und *D16S3032* kartierten Transkripte aufgeführt. Für die RH-Kartierung wurden zum größten Teil STS-Marker verwendet, die zuvor aus EST-Sequenzen abgeleitet worden waren. Die erste Spalte enthält die für jeden einzelnen STS-Marker bestimmten Kartenkoordinaten (in cR₃₀₀₀). In einigen Fällen wurden verschiedene STS-Marker an derselben cR-Position bzw. STS-Marker von ein und demselben Gen an unterschiedlichen cR-Positionen kartiert. Diese Unregelmäßigkeiten ergeben sich aus der engen Nachbarschaft der untersuchten Sequenzen und zeigen, dass die exakte lineare Anordnung von Markern auf dem Chromosom nicht durch diese Methode aufgelöst werden kann. Die weiteren Spalten bezeichnen den Namen des kartierten STS-Markers sowie das dazugehörige etablierte Gensymbol. Da die verwendeten STS-Marker z.T. aus bisher nicht näher charakterisierten EST-Sequenzen stammten, wurden diese der Einfachheit halber als EST bezeichnet. In einigen Fällen wurden auch Ähnlichkeiten bzw. Übereinstimmungen zu bereits beschriebenen Genen oder Proteinen mit angegeben. Die GB4-Karte basiert auf den Arbeiten eines internationalen RH-Konsortiums (*Généthon* (G), *Sanger Center* (SC), *Whitehead Institute for Biomedical Research / MIT Center for Genome Research* (WI/CGR), *Stanford Human Genome Center*, *Wellcome Trust Centre for Human Genetics*). Verschiedene Arbeitsgruppen konnten dabei die Gene *RBL2* und *FTO* im gezeigten Chromosomenabschnitt kartieren. Die letzte Spalte zeigt die in dieser Arbeit durchgeführten Zuordnungen von EST-Sequenzen zu bekannten Genen. Die Auswertung der Konsensussequenz des menschlichen Genoms ergab, dass einige der aufgeführten STS-Marker tatsächlich außerhalb des relevanten Genomabschnitts lokalisiert sind. Diese STS-Marker sind durch eine graue Schrift gekennzeichnet. (Modifikation der GeneMap '99; <http://www.ncbi.nlm.nih.gov/genemap>; Deloukas *et al.*, 1998)

heraus, dass *Irx3*, ein Homöobox-Gen, ebenfalls durch die *Ft*-Mutation vollständig verlorengegangen ist. Außerdem konnte im weiteren Verlauf dieser Arbeit gezeigt werden, dass die zu *MMP2* und *KIAA0308* orthologen Mausgene den in der *Ft*-Mutation deletierten Bereich flankieren und somit in unmittelbarer Nachbarschaft zum betroffenen Genomabschnitt liegen.

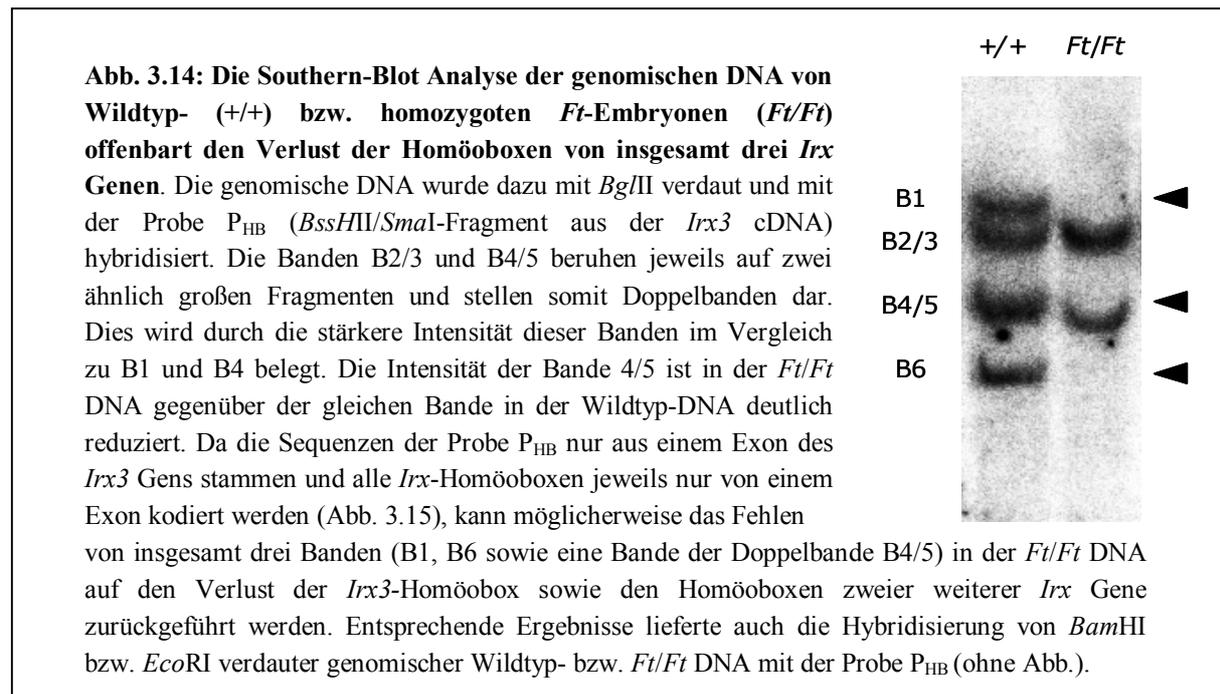
3.5 Das Homöobox-Gen *Irx3* ist ein weiteres Kandidatengen für den *Ft*-Phänotyp

Die *GeneBridge4* Karte der GeneMap'99 zeigte, dass EST-Sequenzen [R55185] des menschlichen *IRX3* Gens in der unmittelbaren Umgebung von *RBL2* und *FTO* lokalisiert sind (Tab. 3.2). In der Tat bewies die PCR-Analyse der genomischen DNA von homozygoten *Ft*-Embryonen, dass das orthologe Mausgen *Irx3* (Bosse *et al.*, 1997) durch die *Ft*-Mutation vollständig entfernt wurde (Abb. 3.17 und 3.18).

Iroquois Proteine bilden eine Familie von Transkriptionsfaktoren, die sich vor allem durch eine 63 Aminosäuren lange DNA-bindende Homöodomäne des TALE Typus sowie eine für diese Familie spezifische sogenannte *Iro*-Domäne auszeichnen (Abb. 3.15; Bürglin, 1997). Erste funktionelle Analysen zeigten, dass *Iroquois* Proteine insbesondere bei der Differenzierung und Regionalisierung von embryonalen Geweben eine bedeutende Rolle spielen (Übersicht bei Cavodeassi *et al.*, 2001). Ursprünglich wurden die ersten Mitglieder dieser Familie aufgrund der Analyse der *Iroquois*-Mutation in *Drosophila melanogaster* beschrieben (Gómez-Skarmeta *et al.*, 1996; McNeill *et al.*, 1997). Dabei stellte sich heraus, dass die drei untereinander sehr eng verwandten *Iroquois* Gene *araucan* (*ara*), *caupolican* (*caup*) und *mirror* (*mirr*) aneinandergereiht in einem Genkomplex, dem *Iroquois*-Komplex (*Iro-C*), im *Drosophila*-Genom vorliegen. Die Anordnung der *Iro-C* Gene ähnelt sehr der komplexen genomischen Organisation anderer Homöoboxgene, z.B. der Gene des *Bithorax-Antennapedia*-Komplexes. Die genomische Organisation des *Bithorax-Antennapedia*-Komplexes wurde während der Evolution sehr hoch konserviert und ist auch im Genom der Wirbeltiere mit den homologen *Hox*-Genkomplexen immer noch erhalten. Allerdings kam es im Verlauf der Evolution zu einer Vervielfältigung des ursprünglichen Komplexes, so dass Säugetiere beispielsweise über insgesamt vier *Hox*-Genkomplexe verfügen (Übersicht bei Holland & Garcia-Fernández, 1996; Finnerty & Martindale, 1998).

Sollte in Analogie zu den *Hox*-Genkomplexen die gekoppelte Anordnung der *Iro-C* Gene ebenfalls während der Evolution erhalten geblieben sein, so konnten neben *Irx3* noch weitere eng verwandte *Irx* Gene durch die *Ft*-Mutation betroffen sein. Diese These wurde durch die

Hybridisierung der hoch konservierten Homöobox des *Irx3* Gens mit der genomischen DNA von homozygoten *Ft*-Embryonen unter niedrig stringenten Bedingungen unterstützt. Dabei zeigte sich, dass neben *Irx3* auch die Homöoboxen von zwei weiteren *Irx* Genen durch die *Ft*-Mutation verlorengegangen sind (Abb. 3.14). Da jedoch die bis zu diesem Zeitpunkt von anderen Arbeitsgruppen beschriebenen Mausgene *Irx1*, *Irx2* und *Irx4* im Genom von homozygoten *Ft*-Embryonen nachgewiesen werden konnten (Abb. 3.17), musste es sich dabei um bisher unbekannte Mitglieder der *Irx*-Genfamilie der Maus handeln.



3.5.1 Die Mitglieder der *Iroquois* Genfamilie des Menschen und der Maus

Wie bereits erwähnt konnten bisher mit *Irx1*, *Irx2*, *Irx3* und *Irx4* insgesamt vier Mitglieder der *Irx*-Genfamilie der Maus beschrieben werden (Bosse *et al.*, 1997; Bruneau *et al.*, 2000). Im Gegensatz dazu waren im Menschen jedoch zu diesem Zeitpunkt schon die Teilsequenzen von insgesamt fünf *IRX* Genen (*IRX1-5*) identifiziert worden (Bosse *et al.*, 1997; Lewis *et al.*, 1999). Durch die Aufklärung der verwandtschaftlichen Beziehungen zwischen den bekannten *Irx* Genen sollte zunächst ein umfassender Überblick über die *Irx*-Genfamilie des Menschen und der Maus geschaffen werden. Ausgehend von dem Verständnis der phylogenetischen Verwandtschaftsgrade sollten später in Zusammenarbeit mit Dr. Renate Dildrop weitere Mitglieder der *Irx*-Genfamilie identifiziert werden. Allerdings enthielten die bis dahin veröffentlichten Sequenzen der *Irx* Gene von Mensch und Maus häufig nur Bereiche, die

jeweils für die hoch-konservierte Homöodomäne kodieren. Da die Homöodomäne bei allen *Irx* Genen bis auf wenige Basenaustausche fast vollkommen identisch ist, konnten diese Teilsequenzen für einen aussagekräftigen phylogenetischen Vergleich nicht herangezogen werden. Daher war es zunächst erforderlich die bestehenden Sequenzen mit Hilfe überlappender EST-Sequenzen *in silico* zu verlängern (Tab. 3.3). Die Gegenüberstellung der orthologen *Irx* Gene offenbarte, dass die bisherige Nomenklatur nicht den tatsächlichen verwandtschaftlichen Beziehungen entspricht. So wurden z.T. orthologe Gene im Menschen und in der Maus mit unterschiedlichen Ziffern gekennzeichnet (Tab. 3.3). Allerdings lieferte dieser Vergleich einen ersten Hinweis auf das Vorhandensein eines weiteren menschlichen *IRX* Gens (Ortholog von *Irx2*) sowie zweier weiterer *Irx* Gene der Maus (Orthologe von *IRX2* und *IRX3*).

orthologe Gruppe	Maus	Mensch
1	<i>Irx1</i> , Y15002 (Bosse <i>et al.</i> , 1997) <i>in silico</i> verlängert durch ESTs AW227409, AI597351, AW230496, AA060558, AA537433	<i>IRX5</i> , U90307 (Lewis <i>et al.</i> , 1999) <i>in silico</i> verlängert durch ESTs AI193273, AI565799, AI193794, AI871044
2	<i>Irx2</i> , Y15000 (Bosse <i>et al.</i> , 1997) <i>in silico</i> verlängert durch ESTs AI154442, AI508269, AA798304	(Dr. Renate Dildrop, diese Arbeit) identifizierte ESTs AI613079, AI654035 AI831283, R50645
3	<i>Irx3</i> , Y15001 (Bosse <i>et al.</i> , 1997) <i>in silico</i> verlängert durch EST AW913488	<i>IRX1</i> , U90308 (Lewis <i>et al.</i> , 1999) <i>in silico</i> verlängert durch ESTs AI308059, AA969519, AI217994
4	<i>Irx4</i> , AF124732 (Bruneau <i>et al.</i> , 2000)	<i>IRX4</i> , AF124733 (Bruneau <i>et al.</i> , 2000)
5	<i>Irx5</i> , (diese Arbeit) identifizierte ESTs W54596, W74908, W89314, AI527498 <i>in silico</i> verlängert durch ESTs BE995453, AV230325	<i>IRX2</i> , U90304 (Lewis <i>et al.</i> , 1999) <i>in silico</i> verlängert durch EST R46202
6	<i>Irx6</i> , (Dr. Renate Dildrop, diese Arbeit) identifizierter EST AA709522 vollständig sequenzierter Klon AJ271055	<i>IRX3</i> , U90305 (Lewis <i>et al.</i> , 1999) <i>in silico</i> verlängert durch die Genomsequenz (BAC-Klon) AC009165

Tab. 3.3: Die *Irx* Gene der Maus und des Menschen. In der NCBI-GenBank konnten Teilsequenzen für insgesamt jeweils sechs verschiedene *Irx* Gene der Maus und des Menschen identifiziert werden. Die Sequenzeinträge von bereits beschriebenen *Irx* cDNAs sind durch die Angabe der entsprechenden Referenz markiert. Alle anderen Sequenzeinträge, die z.T. zu einer beträchtlichen Verlängerung der publizierten *Irx* cDNA-Sequenzen führten, wurden in Zusammenarbeit mit Dr. Renate Dildrop *in silico* in dieser Arbeit identifiziert.

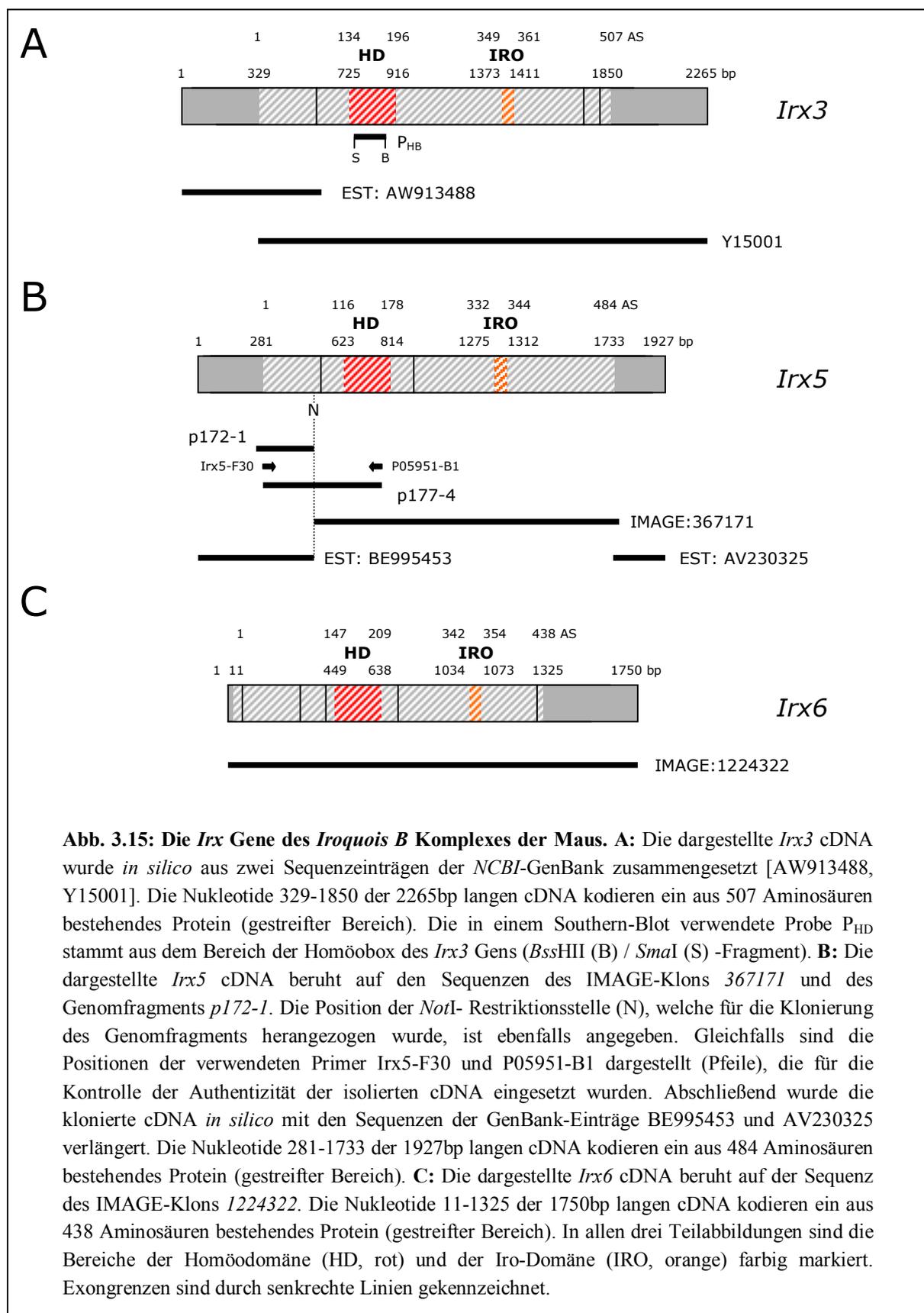
3.5.2 Die Identifizierung von *Irx5*, ein zuvor unbekanntes Mitglied der *Irx*-Genfamilie der Maus

Um die zusätzlichen Mitglieder der *Irx*-Genfamilie der Maus zu beschreiben, wurde die NCBI-GenBank nach weiteren homologen Mausequenzen durchsucht. Dabei konnten zunächst vier überlappende EST-Sequenzen [W54596, W74908, W89314, AI527498] erkannt werden, deren cDNA-Sequenz Ähnlichkeiten mit der menschlichen *IRX2* cDNA [U90304] aufwies. Die *in silico* Translation der Maus EST-Sequenz W54596 lieferte zwei kurze Leserahmen mit einer Länge von 63 bzw. 55 Aminosäuren, deren Sequenz zu 94% identisch mit der C-terminalen Region der abgeleiteten Aminosäuresequenz der menschlichen *IRX2* cDNA war. Da es sich somit eindeutig um die Teilsequenz eines weiteren Maus *Irx* Gens handelte, wurde dieses zuvor unbekannte Gen *Irx5* genannt.

Der dem EST W54596 zugrundeliegende cDNA-Klon *IMAGE:367171* war über das IMAGE-Konsortium (Lennon *et al.*, 1996) erhältlich. Die Sequenzanalyse des gesamten cDNA-Fragments offenbarte einen langen durchgehenden Leserahmen (410 AS), der jedoch kein vermeintliches Startcodon enthielt (Abb. 3.15-B). Der untersuchte cDNA-Klon bestand somit nur aus einer unvollständigen cDNA [AJ271054]. Das Fehlen der 5' cDNA-Sequenz konnte später auf das Vorhandensein einer internen *NotI* Schnittstelle in der *Irx5* cDNA zurückgeführt werden. Diese führte höchstwahrscheinlich zu einer Verkürzung der klonierten cDNA, da für die Klonierung der entsprechenden cDNA-Bibliothek (Soares 13,5-14,5 dpc Embryo NbME13.5-14.5) die Restriktionsenzyme *NotI* und *EcoRI* verwendet worden waren.

Da in der NCBI-GenBank keine weiteren 5'-Sequenzen erkennbar waren, wurde das 5'-Ende des *Irx5* Gens schließlich aus genomischer DNA isoliert. Dazu wurde der für *Irx5* positive PAC-Klon 73 (Abb. 3.18; Ausmeier, 2000) mit den Restriktionsenzymen *NotI* und *NheI* fragmentiert. Die erhaltenen Bruchstücke wurden kloniert und sequenziert. Die Sequenzen der genomischen Bruchstücke wurden *in silico* translatiert und virtuell mit der Aminosäuresequenz anderer *Irx* Proteine verglichen. Aufgrund der engen Verwandtschaft von *Irx5* mit *Irx2* (s. Abschnitt 3.9.1, Abb. 3.24) konnte ein Genomfragment (*p172-1*) identifiziert werden, dessen *in silico* translatierte Sequenz eine sehr hohe Ähnlichkeit (77%) zum N-terminalen Ende des *Irx2* Proteins aufwies [AJ271053]. Die Authentizität dieser vermeintlichen 5'-Sequenz des *Irx5* Gens wurde mittels einer RT-PCR mit den Primern *Irx5*-F30 (abgeleitet aus der genomischen Sequenz *p172-1*) und P05951-B1 (abgeleitet aus der cDNA Sequenz *IMAGE:367171*) überprüft (Abb. 3.15-B). Wie erwartet konnte dabei ein 445bp langes cDNA-Fragment (*p177-4*) aus der RNA von 11.5 p.c. Tage alten Mausembryonen isoliert werden (ohne Abb.). Die Sequenzierung der erhaltenen cDNA bestätigte nochmals die korrekte Anordnung der beiden Ursprungssequenzen (*p172-1* und *IMAGE:367171*).

Abschließend wurde das 5'-Ende bzw. 3'-Ende der bisher isolierten cDNA durch zwei EST-Sequenzen [BE995453 bzw. AV230325] *in silico* verlängert (Abb. 3.15-B).



Die zusammengesetzte *Irx5* cDNA hat eine Gesamtlänge von 1927bp. Die Sequenz enthält ein Startcodon an Position 281, einen 1449bp langen offenen Leserahmen und ein Stoppcodon an Position 1730. Das vermeintliche *Irx5* Protein besteht aus 483 Aminosäuren und ist in einem vergleichbaren Bereich von 401 Aminosäuren zu 90% identisch mit der abgeleiteten partiellen Aminosäuresequenz des menschlichen *IRX2* Gens [U90304]. Neben einer 63 Aminosäuren langen TALE-Homöodomäne (AS 116-178) enthält *Irx5* die für *Irx* Proteine charakteristische Iro-Domäne (AS 332-344). Parallel zu dieser Arbeit wurde *Irx5* in der Maus von zwei weiteren Arbeitsgruppen beschrieben (Bosse *et al.*, 2000; Cohen *et al.*, 2000).

3.5.3 Die Identifizierung von *Irx6*, ein zuvor unbekanntes Mitglied der *Irx*-Genfamilie der Maus

Das Durchsuchen der *NCBI*-GenBank mit der *in silico* translatierten cDNA eines 5' Segments des menschlichen *IRX3* Gens [U90305] führte zu der Identifizierung der Maus EST-Sequenz AA709522. Die von beiden Sequenzen abgeleiteten kurzen Leserahmen waren in ihrem überlappenden Bereich (33 Aminosäuren) zu 66% identisch. Der dem EST AA709522 zugrundeliegende cDNA-Klon *IMAGE:1224322* war über das IMAGE-Konsortium (Lennon *et al.*, 1996) erhältlich. Die enthaltene cDNA hat eine Gesamtlänge von 1750bp und beinhaltet einen einzigen langen Leserahmen mit einem Startcodon an Position 11 und einem Stoppcodon an Position 1324 (Abb. 3.15-C). Das von diesem Leserahmen kodierte vermeintliche Protein besteht aus 438 Aminosäuren und enthält sowohl eine TALE-Homöodomäne (Aminosäuren 147-209) als auch eine Iro-Domäne (Aminosäuren 342-354). Es ist in einem überlappenden Bereich von 192 Aminosäuren zu 83% identisch mit der abgeleiteten Aminosäuresequenz der partiellen menschlichen *IRX3* cDNA (Tab. 3.3). Da es sich um ein weiteres Mitglied der *Irx* Familie handelt, wurde dieses zuvor unbekannte Gen *Irx6* genannt.

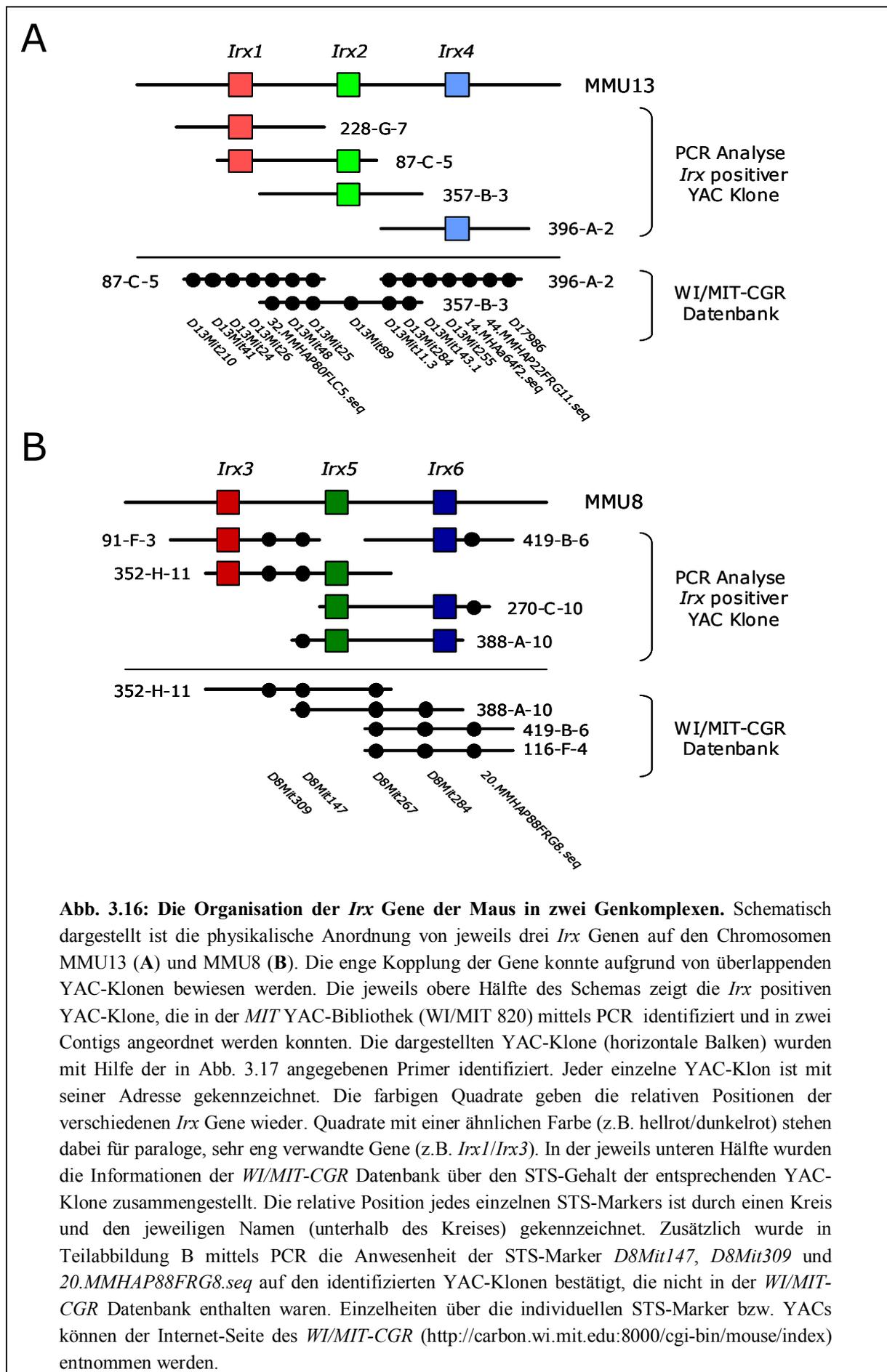
3.5.4 Die Anordnung der *Irx* Gene der Maus in zwei Genkomplexen

Wie bereits an anderer Stelle (Abschnitt 3.5) erwähnt sind die drei *Irx* Gene der Fruchtfliege in einem Genkomplex angeordnet. Die Tatsache, dass neben *Irx3* sehr wahrscheinlich noch zwei weitere *Irx* Gene von der *Ft*-Mutation betroffen sind (Abb. 3.14), sprach für eine Bewahrung dieser komplexen genomischen Organisation während der Evolution. Diese Hypothese sollte durch die Kartierung der *Irx* Gene im Genom der Maus überprüft werden. Dabei wurde die Kartierung der Maus *Irx* Gene durch die Verfügbarkeit einer auf YAC-

Klonen basierenden physikalischen Karte des Mausgenoms (Nusbaum *et al.*, 1999) erheblich erleichtert. Für die Erstellung dieser Karte am *Whitehead Institute for Biomedical Research / MIT Center for Genome Research (WI/MIT-CGR)* wurden sequenzmarkierte Stellen (STS) per PCR auf ausgewählten YAC-Klonen einer *MIT* YAC-Bibliothek (Haldi *et al.*, 1996) lokalisiert. Die verwendete YAC-Bibliothek zeichnet sich vor allem durch die besondere Größe der enthaltenen Genom-Fragmente von durchschnittlich 820kb aus. Da die kartierten STS-Marker vor allem auf bereits etablierten genetischen Markern beruhen, konnte die erstellte physikalische Karte zusammen mit den bereits bestehenden genetischen Karten (Dietrich *et al.*, 1996; Rhodes *et al.*, 1998) in einer hochauflösenden physikalischen Karte des Mausgenoms vereint werden. Das Aufspüren von YAC-Klonen, die in dieser Karte enthalten sind und ein zu untersuchendes Gen tragen, erlaubt somit eine schnelle und genaue Lokalisierung eines Gens im Mausgenom.

Das Durchsuchen der *MIT* YAC-Bibliothek mittels PCR führte zu der Identifizierung von mehreren *Irx* positiven YAC-Klonen (Abb. 3.16). Dabei konnten zunächst YACs für die Gene *Irx1* (228-G-7 und 87-C-5), *Irx2* (87-C-5 und 357-B-3) und *Irx4* (396-A-2) isoliert werden. Schon der Befund, dass *Irx1* und *Irx2* beide auf dem YAC 87-C-5 lokalisiert sind, ließ die enge Kopplung dieser beiden Gene erkennen (Abb. 3.16-A). Da die YAC-Klone 87-C-5 und 357-B-3 (positiv für *Irx2*) in der beschriebenen integrierten genetischen und physikalischen Karte enthalten sind, konnten über die *WI/MIT-CGR*-Datenbank ausführliche Informationen über die physikalische Kartierung dieser YACs abgerufen werden. Beide YACs sind in dem Contig WC-321 enthalten, der bei 25cM auf dem Chromosom MMU13 kartiert worden war. Die Datenbankinformation über diesen Contig verriet eine physikalische Überschneidung der YACs 87-C-5 und 357-B-3 und bestätigte somit die enge Kopplung beider Gene. Interessanterweise gab die *WI/MIT-CGR* Datenbank Auskunft über die physikalische Überlappung des *Irx2* positiven YACs 357-B-3 mit dem *Irx4* positiven YAC 396-A-2 (Abb. 3.16-A). Beide Gene liegen also ebenfalls eng gekoppelt im Genom der Maus vor. Folglich sind die Gene *Irx1*, *Irx2* und *Irx4* in dieser linearen Abfolge in einem Genkomplex bei 25cM auf MMU13 angeordnet. Dieser Genkomplex wurde demnach *IroquoisA (IrxA)* Komplex genannt.

Gleichermaßen konnten positive YACs für die Gene *Irx3* (91-F-3 und 352-H-11), *Irx5* (352-H-11, 270-C-10 und 388-A-10) und *Irx6* (116-F-4, 419-B-6, 270-C-10 und 388-A-10) isoliert werden (Abb. 3.16-B). Einige dieser YACs sind in den Contigs WC-1037 und WC-1038 der *WI/MIT-CGR* Datenbank zusammengefasst. Beide Contigs waren bei 43cM auf Chromosom MMU8 kartiert worden. Die Analyse des YACs 352-H-11, der die zwei Gene *Irx3* und *Irx5* enthält, deckte die enge Kopplung dieser beiden Gene auf. In ähnlicher Weise konnte die enge Kopplung der Gene *Irx5* und *Irx6* nachgewiesen werden, da beide Gene jeweils auf den YACs 270-C-10 und 388-A-10 lokalisiert waren. Die bestehenden physikalischen Überschneidungen



Beide Gene werden während der Embryonalentwicklung in einigen der betroffenen Gewebe exprimiert (Bosse *et al.*, 2000; Cohen *et al.*, 2000; Mummenhoff *et al.*, 2001; Houweling *et al.*, 2001). *Irx5* und *Irx6* kommen daher grundsätzlich als Kandidatengene für den *Ft*-Phänotyp in Betracht.

3.6 Die Identifizierung von Kandidatengenen in den flankierenden Genomabschnitten

Nicht nur Mutationen im transkribierten Bereich (Exon- und Intronsequenzen) können zum Funktionsausfall eines Gens und damit zur Ausprägung von krankheitsrelevanten Phänotypen führen. Ebenso kann eine Mutation bzw. die Deletion von *cis*-aktiven Signalstrukturen (Promotor, Enhancer, Silencer) zu einer Veränderung oder sogar zu einem Verlust der gewebe- oder stadienspezifischen Expression des betroffenen Gens führen. Diese Signalstrukturen können oft weit entfernt vom transkribierten Bereich des kontrollierten Gens lokalisiert sein. So können Enhancer-Elemente noch über eine Distanz von vielen Kilobasen wirken. Dabei müssen diese Elemente nicht unbedingt im 5'-Bereich des entsprechenden Gens angesiedelt sein, sondern können auch im 3'-Bereich liegen oder sogar Teil einer Intronsequenz sein. Entsprechend können auch die Signalstrukturen des in den *Ft*-Locus integrierten Transgens (β -Interferon-Promotor) einen Einfluss auf die Expression der nun benachbarten Gene haben und diese entscheidend stören (Wilson *et al.*, 1990). Es ist somit nicht ausgeschlossen, dass auch Gene in der Nähe der *Ft*-Deletion an der Ausprägung der *Ft*-Phänotypen beteiligt sein können. Daher sollten in dieser Arbeit auch diejenigen Gene identifiziert werden (neben *Rbl2*), die sich in der unmittelbaren Nachbarschaft des betroffenen Genomabschnitts befinden.

In einem ersten Schritt sollte dafür zunächst ein YAC-Contig erstellt werden, der sich über den gesamten *Ft*-Locus einschließlich der benachbarten Regionen erstreckt. Als Grundlage diene dazu der bereits in Abschnitt 3.5.4 beschriebene Contig, der mit der genomischen Region der Gene *Irx3*, *Irx5* und *Irx6* einen großen Teil des gesuchten Bereichs abdeckt (Abb. 3.16-B). Ausgehend von diesem Contig wurden durch das Durchmustern der MIT YAC-Bibliothek mittels PCR weitere miteinander überlappende YACs identifiziert. Vorrangig wurden dazu insgesamt zehn neue genspezifische STS-Marker eingesetzt, die aus den Sequenzen der bisher isolierten Gene abgeleitet worden waren (Abb. 3.18). Die so identifizierten YACs wurden zusätzlich auf das Vorhandensein weiterer STS-Marker des *WI/MIT-CGR* aus dieser chromosomalen Region (43cM, MMU8) getestet. Unabhängig davon konnte der STS-Gehalt einiger YACs der *WI/MIT-CGR* Datenbank (Contigs WC8.29 und

WC8.30) entnommen werden. Insgesamt wurden auf diese Weise 13 YAC-Klone für die *Ft*-Region identifiziert und in einem Contig zusammengefasst. Die Anordnung der YACs basiert auf dem jeweiligen STS-Gehalt, d.h. YACs, die denselben STS-Marker enthalten, wurden als überschneidend betrachtet und entsprechend angeordnet. Diese Analyse ergab, dass die *Ft*-Deletion durch die genetischen STS-Marker *D8Mit250*, *D8Mit251*, *D8Mit252* auf der einen Seite und *D8Mit240* auf der anderen Seite begrenzt wird. Ebenso wurden die STS-Marker des *WI/MIT-CGR* aus dem Bereich des YAC-Contigs in einen parallel zu dieser Arbeit erstellten BAC/PAC-Contig übertragen (Ausmeier, 2000). Somit wurden in der vorliegenden Arbeit alle genomischen Fragmente, STS-Marker und Gene integriert und in einer hochauflösenden physikalischen Genkarte dieser chromosomalen Region zusammengefasst (Abb. 3.18).

Nächste Seite:

Abb. 3.18: Die physikalische Genkarte des *Fused toes* Locus auf MMU8. Der obere Bereich der Karte zeigt die Gene des *Ft*-Locus, welche durch die Positionsklonierung sowie mit Hilfe der vergleichenden Genomanalyse identifiziert werden konnten. Die Integrationsstelle des Transgens (Tg) ist durch ein dunkelblaues Dreieck markiert. Der graue Pfeil überspannt den in *Ft*-Mäusen verlorengegangenen Genomabschnitt mit einer Größe von ungefähr 1,65Mb. Unterhalb der durchgehenden horizontalen Linie, die das MMU8 repräsentiert, sind die BAC, PAC und Cosmid-Klone gezeigt (horizontale Balken), die in einer parallelen Arbeit in einem '*chromosomal walk*' über insgesamt 1,8Mb identifiziert und in einem Contig angeordnet wurden (Ausmeier, 2000). *NotI*-Restriktionsstellen (N), die für die Bestätigung der Contig-Anordnung herangezogen wurden, sind rot gekennzeichnet (Ausmeier, 2000). Die grünen Vierecke und die dazugehörigen grünen Nummern markieren die wichtigsten molekularen Sonden, die von den entsprechenden Klonen isoliert wurden. Vertikale grün-gestrichelte Pfeile kennzeichnen überlappende Genomfragmente und geben die Verlaufsrichtung des '*chromosomal walks*' wieder. Die genaue Bezeichnung der Sonden und Klone sowie weitere Einzelheiten können der bereits erwähnten Dissertation (Ausmeier, 2000) entnommen werden. Im unteren Bereich sind die YAC-Klone (horizontale Balken) gezeigt, die im Verlauf der vorliegenden Arbeit mit Hilfe von genspezifischen STS-Markern (blaue Dreiecke) identifiziert und in einem Contig zusammengefasst werden konnten. Jeder einzelne YAC ist mit seiner Adresse gekennzeichnet. Die neu generierten genspezifischen STS-Marker wurden in einer PCR mit den folgenden Primern nachgewiesen: *KIAA0308-5'* (A): KIAA0308-F2, KIAA0308-B2; *KIAA0308-3'* (B): KIAA0308-F3, KIAA0308-B3; *Fts* (C): Wt, Ft/Wt; *Ftm* (D): F14-F1, B63.2; *Fto-5'* (E): Ft3-5'F1, Ft3-5'B1; *Fto-3'* (F): Ft3-F3000, Ft3-B3400; *Irx3* (G): Irx3-F1, Irx3-B1; *Irx5* (H): P05951-F1, P05951-B1; *Irx6* (I): Irx6-F1530, Irx6-B1700; *Mmp2-5'* (J): Mmp2-F1, Mmp2-B1; *Mmp2-3'* (K): Mmp2-F2, Mmp2-B2. Vertikale blau-gestrichelte Linien überspannen überlappende Genomfragmente. Orange Kreise markieren STS-Marker, die den Contigs WC8.29 und WC8.30 der *WI/MIT-CGR* Datenbank entnommen und in den bestehenden BAC/PAC-Contig bzw. in den erstellten YAC-Contig integriert wurden. Die orange-farbenen Zahlen unterhalb der Position des jeweiligen STS-Markers stehen für folgende Bezeichnungen: *D8Mit107* (1), *D8Mit266* (2), *19.MHAA58b7.seq* (3), *D8Mit250* (4), *D8Mit251* (5), *D8Mit252* (6), *43.MMHAP10FRG10.seq* (7), *D8Mit196* (8), *D8Mit309* (9), *D8Mit147* (10), *D8Mit267* (11), *D8Mit248* (12), *D8Mit284* (13), *D8Mit240* (14), *20.MMHAP88FRG8.seq* (15). STS-Marker, deren exakten Lokalisationen in dieser Studie nicht voneinander getrennt werden konnten, sind in Klammern zusammengefasst. Die Länge der BAC und PAC Klone ist maßstabsgetreu wiedergegeben. Die relative Mindestlänge der YAC Klone ergibt sich aus ihrem STS-Gehalt und wurde nicht experimentell bestätigt. Daher ist die Länge der eingezeichneten YAC-Klone nicht maßstabsgetreu.

Der Nachweis von einigen STS-Markern (z.B. *D8Mit107* oder *20.MMHAP88FRG8.seq*, Abb. 3.18) im Genom von homozygoten *Ft*-Embryonen offenbarte, dass der erstellte YAC-Contig ebenfalls weite Bereiche enthält, die den in *Ft*-Mäusen deletierten Genomabschnitt flankieren. Im weiteren Verlauf dieser Arbeit sollten nun Gene, die im proximalen bzw. distalen Bereich des YAC-Contigs lokalisiert sind und somit in der unmittelbaren Nähe der *Ft*-Deletion liegen, identifiziert werden. Dazu wurde wiederum auf die *GeneBridge4* Karte der menschlichen GeneMap'99 zurückgegriffen (s. Abschnitt 3.4). Diese zeigte, dass sowohl das Gen *MMP2* als auch die bisher nicht näher charakterisierte cDNA-Sequenz *KIAA0308* in der Nachbarschaft von *FTO* und *RBL2* lokalisiert worden waren (Tab. 3.2). Daher sollten beide Gene mit Bezug auf den *Ft*-Locus im Mausgenom kartiert werden.

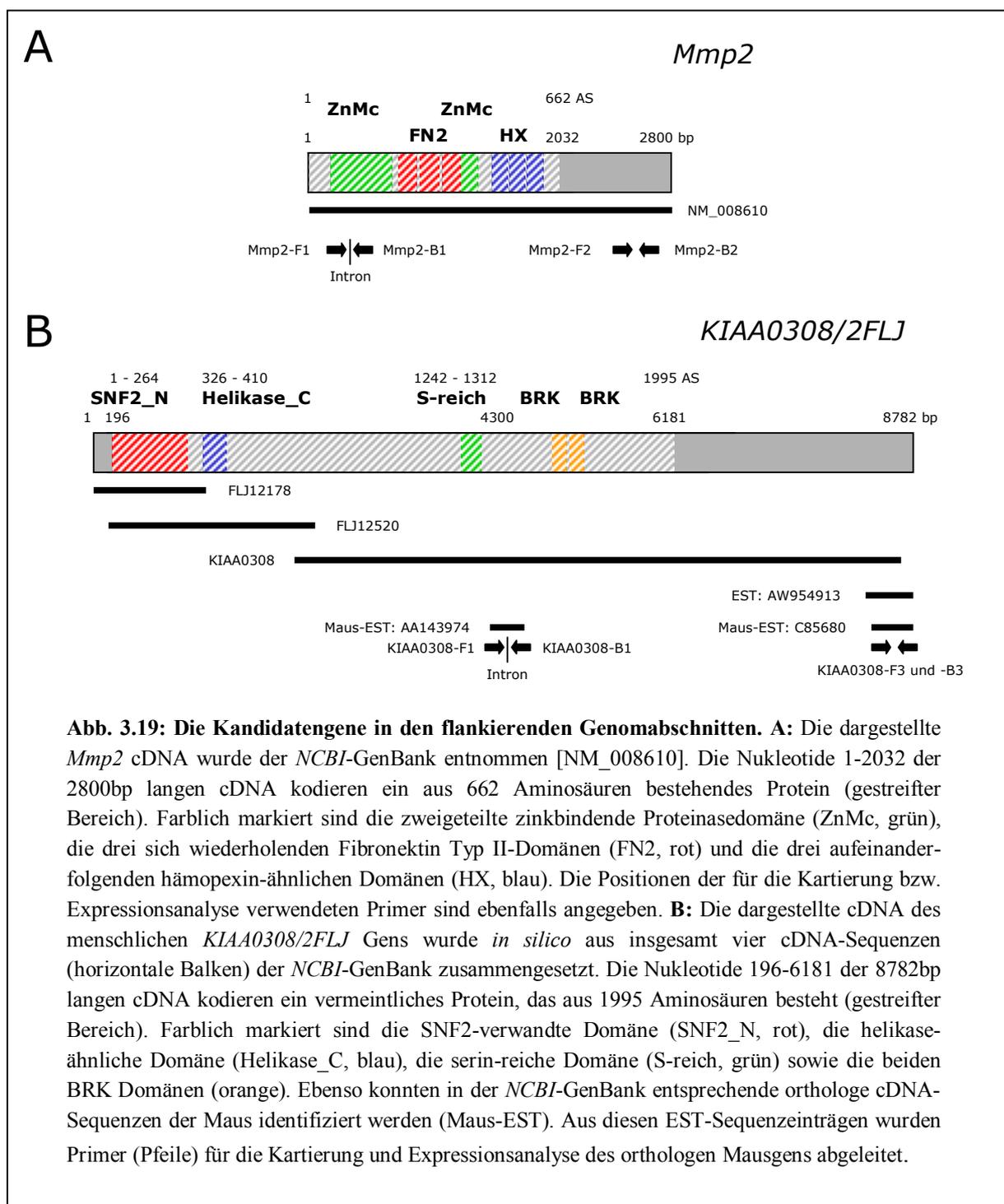
3.6.1 Die Kartierung des *Mmp2* Gens im distalen flankierenden Genomabschnitt

Das Matrix Metalloproteinase2-Protein (*Mmp2*) gehört zu einer Familie von strukturell verwandten Zn^{2+} - und Ca^{2+} -abhängigen neutralen Endopeptidasen, deren Aufgabe es ist, Proteine der extrazellulären Matrix (z.B. Kollagen, Elastin, Laminin und Fibronectin) proteolytisch zu degradieren (Yu *et al.*, 1999). *Mmp* Proteine spielen daher eine wichtige Rolle bei der Umwandlung der extrazellulären Matrix während der Embryonalentwicklung, der Morphogenese, der Resorption und Umgestaltung von Geweben, der Zellwanderung und der Angiogenese. Im Gegensatz zu diesen physiologischen Prozessen ist jedoch auch die Beteiligung dieser Proteine an pathologischen Prozessen wie z.B. rheumatische Arthritis, Osteoarthritis, Atherosklerose oder bei der Invasion und Metastasierung von Krebszellen bekannt (Woodhouse *et al.*, 1997).

Wie alle *Mmp* Proteine ist auch *Mmp2* aus folgenden charakteristischen Domänen aufgebaut: eine zweigeteilte konservierte zinkbindende katalytische Proteinasedomäne, drei sich wiederholende Fibronectin Typ II-Domänen und drei aufeinanderfolgende hämopexin-ähnliche Domänen (Abb. 3.19).

Die *Mmp2* Gene des Menschen und der Maus wurden bereits durch verschiedene Arbeiten ausführlich charakterisiert (Itoh *et al.*, 1997; Martignetti *et al.*, 2001). Für die genaue Kartierung des *Mmp2* Gens der Maus wurden genspezifische STS-Marker aus dem 5'- und 3'-Bereich der veröffentlichten *Mmp2* cDNA [NM_008610] abgeleitet und auf den isolierten Genomfragmenten des YAC bzw. BAC/PAC-Contigs getestet. Dabei konnten die beiden verwendeten STS-Marker nur auf den YACs 116-F-4, 270-C-10 und 419-B-6 nachgewiesen werden (Abb. 3.18). Daneben wurde der STS-Marker *Mmp2-5'* auch auf dem YAC 388-A-10 kartiert. In der physikalischen Genkarte des *Ft*-Locus ist *Mmp2* somit im distalen Bereich des YAC-Contigs, 'downstream' von *Irx6* und außerhalb der Deletion lokalisiert (Abb. 3.18,

3.20). Demnach ist der transkribierte Bereich von *Mmp2* nicht von der *Ft*-Deletion betroffen. Aufgrund der Orientierung des Gens befindet sich die Promotorregion vom *Mmp2* allerdings in unmittelbarer Nähe des betroffenen Genomabschnitts. Daher sollte durch eine RT-PCR-Analyse ein vermeintlicher Unterschied in der Qualität der *Mmp2* Expression in MEF-Zellen aus Wildtyp- bzw. aus homozygoten *Ft*-Embryonen aufgedeckt werden. Diese Analyse zeigte jedoch, dass *Mmp2* in beiden Zelltypen exprimiert wird und somit die generelle Expression von *Mmp2* in homozygoten *Ft*-Zellen gewährleistet ist (Abb. 3.20).



3.6.2 Die Kartierung des *KIAA0308/2FLJ* Gens im proximalen flankierenden Genomabschnitt

Die Teilsequenz *KIAA0308* [AB002306] eines bisher nicht näher charakterisierten menschlichen Gens wurde im Rahmen des bereits erwähnten *Kazusa* cDNA Sequenzierungsprojekts isoliert. Mit Hilfe der *NCBI*-GenBank konnte diese Sequenz *in silico* durch weitere cDNA-Sequenzen verlängert werden [AK022240, AK022582, AW954913]. Aufgrund der drei langen Teilsequenzen *KIAA0308*, *FLJ1250* und *FLJ12178* wurde dieses menschliche Gen in dieser Arbeit *KIAA0308/2FLJ* genannt. Die zusammengesetzte cDNA weist eine Gesamtlänge von 8782bp auf und beinhaltet einen offenen Leserahmen (5985bp), der für ein vermeintliches Protein von 1995 Aminosäuren kodiert. Die computergestützte Analyse der Aminosäuresequenz (Abb. 3.19) ließ am *N*-terminalen Ende des vermeintlichen *KIAA0308/2FLJ* Proteins eine ATPase Domäne erkennen (SNF2_N), die eine große Ähnlichkeit mit der ATPase Untereinheit des Chromatin-Remodellierungskomplexes *SWI2/SNF2* der Hefe *S. cerevisiae* zeigt (Tamkun *et al.*, 1992). Weiterhin enthält die abgeleitete Aminosäuresequenz zwei sogenannte BRK Domänen, die auch in den mit SNF2 verwandten Chromatin-Remodellierungsfaktoren *brahma* und *kismet* in *Drosophila* gefunden werden können (Daubresse *et al.*, 1999; Therrien *et al.*, 2000). Ebenso auffällig sind eine weitere helikase-ähnliche Domäne (Helikase_C), die ebenfalls als ATP- und nukleinsäurebindende Domäne oder als Helikase fungieren könnte, sowie eine 71 Aminosäure lange serinreiche Domäne (Abb. 3.19). Insgesamt zeigt *KIAA0308/2FLJ* eine beachtliche Ähnlichkeit (57% auf Proteinebene) zum bereits erwähnten *Drosophila*-Protein *kismet*. Allerdings fehlen *KIAA0308/2FLJ* die für *kismet* charakteristischen Chromodomänen (Daubresse *et al.*, 1999; Therrien *et al.*, 2000).

Die Suche nach orthologen Maus cDNA-Sequenzen in der *NCBI*-GenBank führte zu der Identifizierung mehrerer EST-Sequenzen, die jedoch nicht zu einer vollständigen cDNA zusammengesetzt werden konnten. Da für die weiteren Analysen eine vollständige Sequenz der Maus cDNA nicht notwendig war, wurde auf die Isolierung der gesamten cDNA in dieser Arbeit verzichtet. Aus den EST-Sequenzen AA143974 und C85680 wurden genspezifische STS-Marker für die Kartierung des *KIAA0308/2FLJ* Gens der Maus abgeleitet. Die PCR-Analyse der Genomfragmente des YAC bzw. BAC/PAC-Contigs zeigte, dass beide STS-Marker auf dem BAC MB1 und auf den YACs 156-E-1, 390-A-11, 390-C-11 und 401-G-10 lokalisiert sind (Abb. 3.18). Der STS-Marker *KIAA0308-5'* konnte zusätzlich auf dem YAC 223-A-9 nachgewiesen werden. In der physikalischen Genkarte des *Ft*-Locus ist *KIAA0308/2FLJ* somit im proximalen Bereich des YAC-Contigs, 'upstream' von *Rbl2* und außerhalb der Deletion lokalisiert (Abb. 3.18, 3.20). Der transkribierte Bereich von

KIAA0308/2FLJ ist demnach nicht von der *Ft*-Deletion betroffen (Abb. 3.18). Ebenso ließ die RT-PCR-Analyse der RNA aus MEF-Zellen von Wildtyp- bzw. homozygoten *Ft*-Embryonen keinen Unterschied in der Qualität der *KIAA0308/2FLJ*-Expression erkennen (Abb. 3.20).

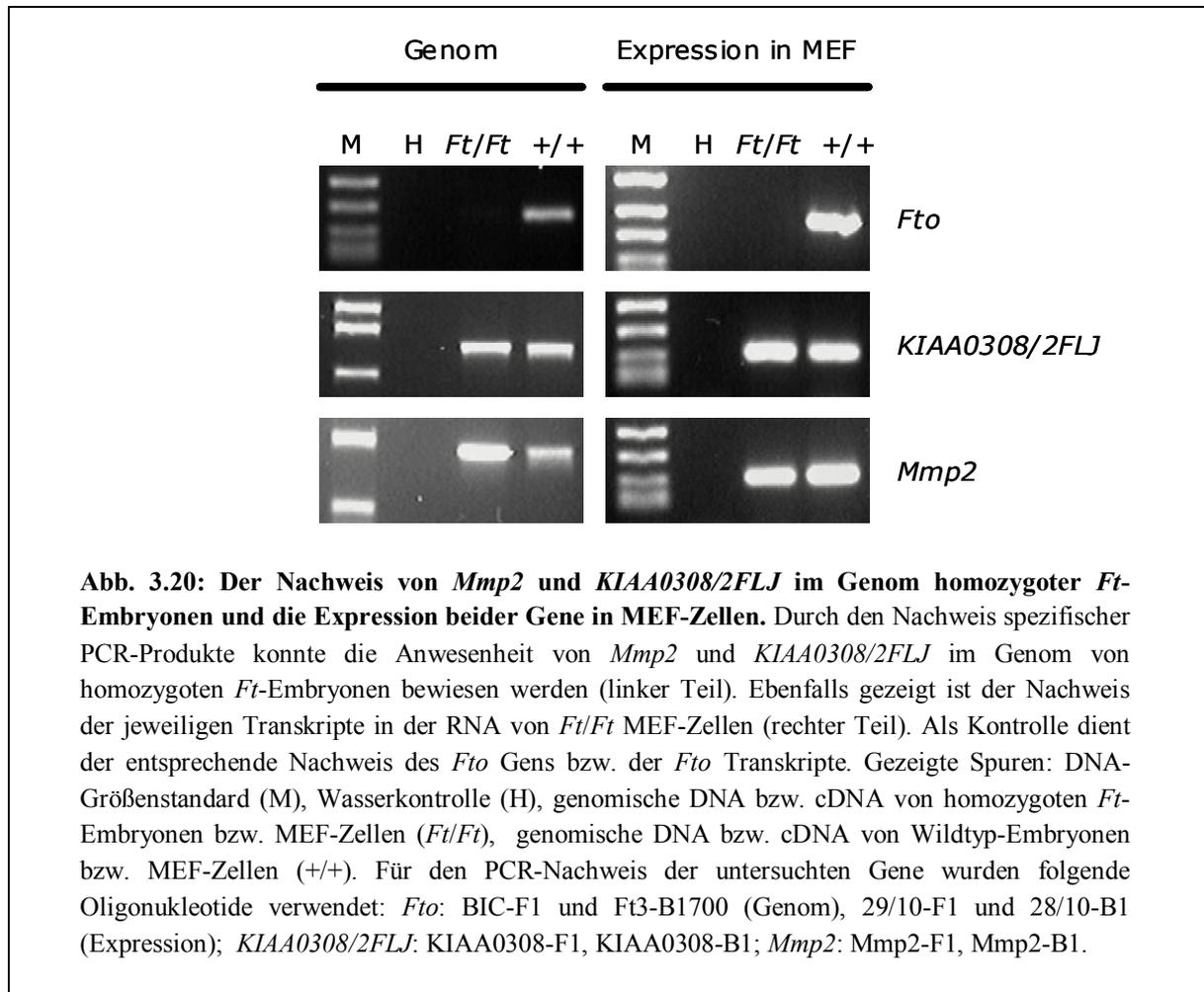


Abb. 3.20: Der Nachweis von *Mmp2* und *KIAA0308/2FLJ* im Genom homozygoter *Ft*-Embryonen und die Expression beider Gene in MEF-Zellen. Durch den Nachweis spezifischer PCR-Produkte konnte die Anwesenheit von *Mmp2* und *KIAA0308/2FLJ* im Genom von homozygoten *Ft*-Embryonen bewiesen werden (linker Teil). Ebenfalls gezeigt ist der Nachweis der jeweiligen Transkripte in der RNA von *Ft/Ft* MEF-Zellen (rechter Teil). Als Kontrolle dient der entsprechende Nachweis des *Fto* Gens bzw. der *Fto* Transkripte. Gezeigte Spuren: DNA-Größenstandard (M), Wasserkontrolle (H), genomische DNA bzw. cDNA von homozygoten *Ft*-Embryonen bzw. MEF-Zellen (*Ft/Ft*), genomische DNA bzw. cDNA von Wildtyp-Embryonen bzw. MEF-Zellen (+/+). Für den PCR-Nachweis der untersuchten Gene wurden folgende Oligonukleotide verwendet: *Fto*: BIC-F1 und Ft3-B1700 (Genom), 29/10-F1 und 28/10-B1 (Expression); *KIAA0308/2FLJ*: KIAA0308-F1, KIAA0308-B1; *Mmp2*: Mmp2-F1, Mmp2-B1.

3.7 Menschliche Erbkrankheiten im Kandidatenbereich HSA 16q12-q13

Neben der Auswertung von physikalischen und genetischen Karten kann auch der Vergleich des *Ft*-Phänotyps zu bekannten genetisch bedingten Erkrankungen des Menschen für die Identifizierung von Kandidatengen herangezogen werden. Sowohl zytogenetische Befunde von Chromosomen-Aberrationen als auch umfangreiche Segregationsanalysen führten zu der Kartierung des Kandidatenbereichs verschiedener menschlicher Erbkrankheiten im Bereich von HSA 16q12-q13 (OMIM-Datenbank, Tab. 3.4).

Erbkrankheit	Klinische Ausprägung	MIM-Nr.	Erbgang	Kandidatenbereich	mutiertes Gen
Townes-Brocks-Syndrom	Anomalien der Hände, Ohren und des Anus	107480	dominant	16q12.1	<i>SALL1</i>
Hypodontie	Fehlen von Zähnen	602639	rezessiv	16q12.1	unbekannt
Morbus Crohn	Entzündung der Darmwandschichten, Haut, Augen und Gelenke	266600	komplex	16q12	<i>NOD2</i>
Orthostatische Intoleranz	adrenerge Symptome: Schwindel- und Ohnmachtsanfälle, Müdigkeit	604715	dominant	16q12.2	<i>SLC6A2</i> (Kandidat)
Spiegler-Brooke-Syndrom	Cylindrome, Trichoepitheliome	605041	dominant	16q12-q13	unbekannt
Familiäre Cylindromatose	Cylindrome	132700	rezessiv, LOH	16q12-q13	<i>CYLD1</i>
Phosphorylasekinase-Defizienz der Leber und Muskeln	Hepatomegalie, Akkumulation von Glykogen	261750	rezessiv	16q12-q13	<i>PHKB</i>
Nodulosis-Arthropathie-Osteolyse-Syndrom	rheumatoide Nodulis, distale Arthropathie ideopathische Osteolyse	605156	rezessiv	16q13	<i>MMP2</i>
Gitelman-Syndrom	Muskelschwäche, chronische Dermatitis, Schädigung der Nieren	263800	rezessiv	16q13	<i>SLC12A3</i>
Retinitis pigmentosa	Netzhauterkrankungen, Blindheit	268000	rezessiv	16q13-q21	<i>CNGB1</i> (Kandidat)
Anisomastia	asymmetr. Brustwarzen, geistige Behinderung, dysmorphes Gesicht, Finger/Zehen-Anomalien	605746		16q13-q21	

Tab. 3.4: Menschliche Erbkrankheiten im Kandidatenbereich HSA 16q12-q13. MIM: Mendelian Inheritance in Man; LOH: Loss of Heterozygosity. Detaillierte Informationen über die aufgeführten Erbkrankheiten können der OMIM-Datenbank der John Hopkins Universität (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>) entnommen werden.

Insbesondere die klinischen Ausprägungen des Townes-Brocks-Syndroms, des Nodulosis-Arthropathie-Osteolyse-Syndroms und der Anisomastia erinnern teilweise an den *Ft*-Phänotyp. Da allerdings die molekulare Grundlage der Anisomastia noch nicht bekannt ist, konnte in diesem Fall kein Kandidatengen gezielt verfolgt werden. Dem Nodulosis-Arthropathie-Osteolyse-Syndrom liegen Mutationen im *MMP2* Gen zugrunde. Das orthologe Mausgen *Mmp2* wurde, wie bereits in Abschnitt 3.6.1 erwähnt, in unmittelbarer Nachbarschaft der *Ft*-Deletion lokalisiert.

Das Townes-Brocks-Syndrom (TBS) ist eine autosomal dominante Erkrankung, die durch Fehlbildungen der Hände (präaxiale Polydaktylie, triphalangeale Daumen), der Ohren

(besonders der oberen Helix) und der Analregion (ungeöffneter Anus, anterior vorgelagerter Anus) gekennzeichnet ist. Die klinische Ausprägung dieser Erkrankung ist sehr variabel und umfasst weiterhin Nierenfehlbildungen und Schwerhörigkeit. In einigen seltenen Fällen konnten auch Fehlbildungen des Herzens und leichte geistige Behinderungen festgestellt werden (Übersicht bei Powell & Michaelis, 1999). Die klinische Präsentation der Handfehlbildungen von TBS Patienten zeigt einige interessante Parallelen zum Gliedmaßenphänotyp (Syndaktylie) der heterozygoten *Ft*-Mäuse. So wurden bei TBS Patienten neben den bereits erwähnten triphalangealen Daumen und der präaxialen Polydaktylie auch in einigen Fällen eine Syndaktylie der Finger und eine Verkrümmung der Daumen beobachtet. Außerdem zeigen die distalen Phalangen der Daumen von TBS Patienten oft eine breite Erscheinung und partielle Duplikationen. Fußfehlbildungen, die bei heterozygoten *Ft*-Mäusen nicht festgestellt werden können, kommen jedoch ebenfalls in Zusammenhang mit TBS vor. Dabei wurden z.B. einander überschneidende Zehen, Syndaktylien oder das Fehlen von einzelnen Zehen beschrieben (Powell & Michaelis, 1999).

Aufgrund von zytogenetischen Anomalien (Translokationen, Inversionen und Deletionen) der Chromosomen einiger Patienten konnte der TBS Locus auf HSA16 im distalen Bereich des Bandes 16q12.1 kartiert werden (Friedman *et al.*, 1987; Serville *et al.*, 1993; Powell *et al.*, 1995; Michaelis *et al.*, 1996). In dieser Region ist auch *SALL1*, ein Gen, welches der *Drosophila spalt*-ähnlichen Genfamilie angehört und das für einen Zinkfinger-Transkriptionsfaktor kodiert, lokalisiert (Kohlhase *et al.*, 1996; Abb. 3.13). In der Tat konnte in einem Kandidatengenansatz bewiesen werden, dass Mutationen in *SALL1* für das Auftreten von TBS verantwortlich sind (Kohlhase *et al.*, 1998). Es wird angenommen, dass die meisten Mutationen die charakteristische Anordnung der Doppelzinkfingerdomänen (Abb. 3.21) von *SALL1* zerstören und so zu verkürzten Proteinen ohne DNA-bindende Fähigkeiten führen (Kohlhase, 2000). Jedoch existieren auch einige Fälle, in denen nicht unmittelbar eine Mutation im proteinkodierenden Bereich von *SALL1* nachgewiesen werden konnte. In diesen Patienten könnte TBS daher auf Mutationen im Promotorbereich oder in Intronbereichen von *SALL1* bzw. auf größere Deletionen oder sogar auf Mutationen von anderen Genen zurückzuführen sein (Kohlhase, 2000). So zeigt zum Beispiel der Fall einer Translokation t(5;16) (p15.3;q12.1), dass auch Mutationen in einer Entfernung von 180kb von *SALL1* TBS erzeugen können (Marlin *et al.*, 1999).

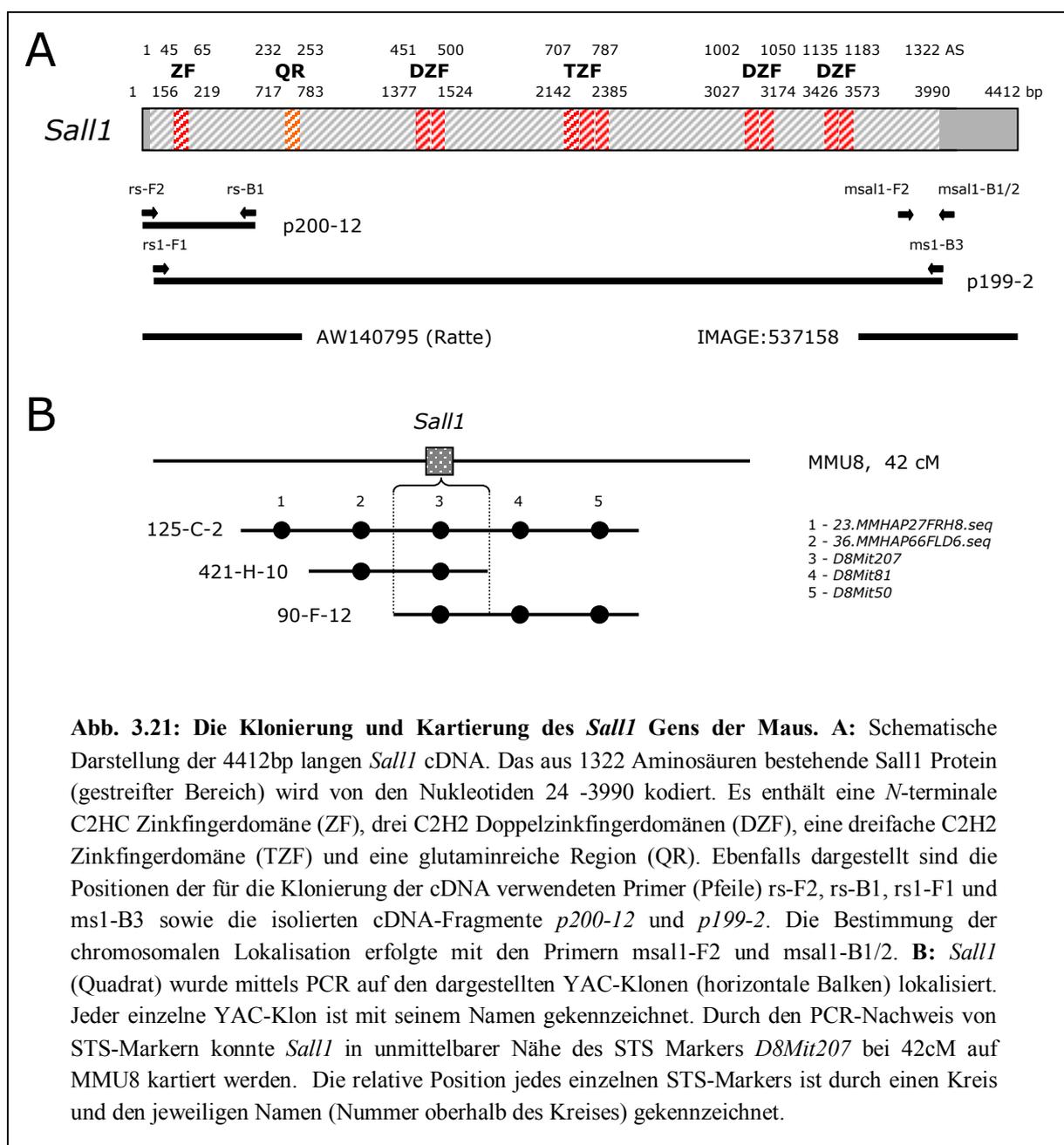
3.7.1 Die Klonierung des *Sall1* Gens der Maus

Wie bereits erwähnt, stimmen einige Ausprägungsformen der Handfehlbildungen von TBS Patienten (Syndaktylien) mit dem Gliedmaßenphänotyp der *Ft*-Mäuse überein. Daher sollte untersucht werden, ob das *Sall1* Gen der Maus als Kandidatengen für den *Ft*-Phänotyp in Betracht kommt. Da bis zum Beginn dieser Analyse das Ortholog des menschlichen *SALL1* Gens noch nicht beschrieben worden war, musste zunächst die entsprechende cDNA in der Maus isoliert werden. In der NCBI-GenBank wurden dementsprechend homologe EST-Sequenzen aus der Ratte [AW140795] bzw. der Maus [AA117993, AI426947] für das 5'- bzw. 3'-Ende der menschlichen *SALL1* cDNA identifiziert (Abb. 3.21). Mit Hilfe von Primern, die aus diesen EST-Sequenzen abgeleitet wurden, konnte danach der proteinkodierende Bereich des Maus *Sall1* Gens aus der RNA von 9.5 Tage alten Embryonen isoliert werden (*p200-12*, *p199-2*; Abb. 3.21). Schließlich wurde die auf diese Weise isolierte *Sall1* cDNA am 3'-Ende durch die vollständige Sequenz des cDNA Klons *IMAGE:537158*, dem die oben erwähnten Maus EST-Sequenzeinträge zugrunde liegen, auf insgesamt 4,4kb verlängert (Abb. 3.21). Die Sequenzanalyse der zu 88% mit der menschlichen *SALL1*-Sequenz identischen cDNA offenbarte einen langen Leserahmen (3966bp), der für ein vermeintliches Protein von 1322 Aminosäuren kodiert. Das Protein enthält die für die *spalt*-ähnliche Familie charakteristischen Domänen: eine N-terminale C2HC Zinkfingerdomäne, drei C2H2 Doppelzinkfingerdomänen, eine dreifache C2H2 Zinkfingerdomäne sowie eine glutamin-reiche Region (Abb. 3.21-A). Insgesamt ist das abgeleitete Protein zu 90% identisch mit dem menschlichen *SALL1* Protein. Dieser Vergleich bestätigte, dass die isolierte cDNA dem gesuchten Maus-Ortholog des menschlichen *SALL1* Gens entspricht.

3.7.2 Die Kartierung des *Sall1* Gens der Maus

Die PCR-Analyse der genomischen DNA homozygoter *Ft*-Embryonen zeigte jedoch, dass der proteinkodierende Bereich von *Sall1* nicht unmittelbar von der *Ft*-Deletion betroffen ist (ohne Abb.). Wie bereits erwähnt, können jedoch auch Mutationen außerhalb des proteinkodierenden Bereichs von *SALL1* (z.B. im Promotorbereich) zur Ausprägung von TBS führen (Marlin *et al.*, 1999). Deshalb sollte die physikalische Nähe von *Sall1* zum *Ft*-Locus durch die Kartierung von *Sall1* im Mausgenom bestimmt werden. Dies geschah wiederum mit Hilfe der schon beschriebenen integrierten physikalischen und genetischen Karte *des WI/MIT-CGR* (s. Abschnitt 3.5.4). Das Durchsuchen der *MIT* YAC-Bibliothek mittels PCR führte in diesem Fall zu der Identifizierung von drei YACs (90-F-12, 125-C-2 und 421-G-10), die allesamt

positiv für *Sall1* waren. Nur einer dieser YACs (125-C-2) war in der *WI/MIT-CGR* Karte enthalten. Die *WI/MIT-CGR* Datenbank verrät, dass dieser YAC aufgrund seiner Zuordnung zum Contig WC-1033 bei 42cM auf MMU8 kartiert worden war. Unabhängig davon wurden die drei YACs auf die Gegenwart einiger STS-Marker aus dieser Region getestet. Der PCR-Nachweis von insgesamt fünf STS-Markern bestätigte die chromosomale Lokalisation und die physikalische Überschneidung der drei YACs. Zudem konnte der *Sall1* Locus aufgrund des spezifischen Musters der Überschneidungen auf die unmittelbare Umgebung des Markers *D8Mit207* eingegrenzt werden (Abb. 3.21-B). *Sall1* ist somit in der Nähe von *D8Mit207* bei 42cM auf MMU8 lokalisiert und daher ungefähr 1cM bzw. 1-2 Mb vom *Ft*-Locus entfernt.



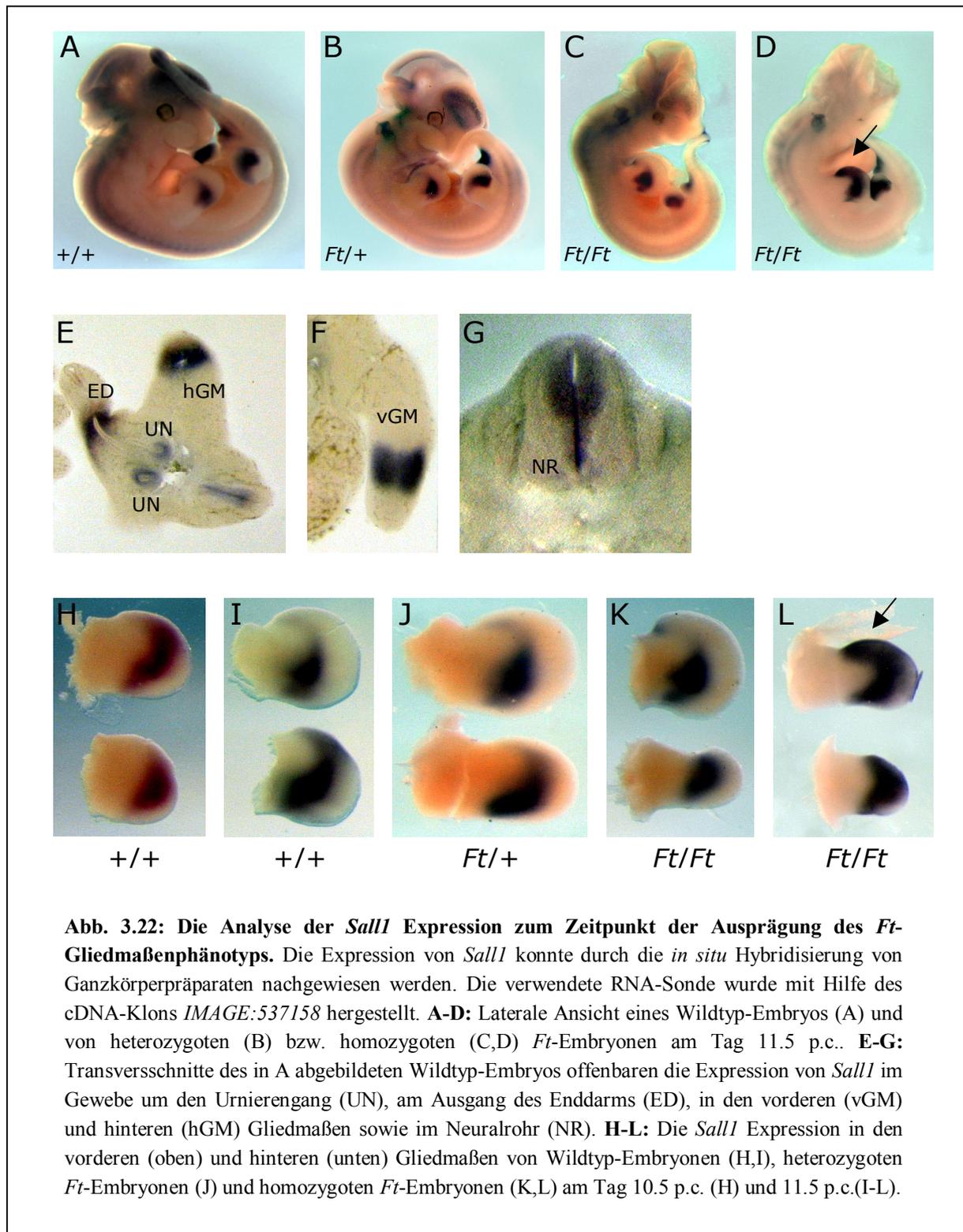
3.7.3 Die Expression von *Sall1* zum Zeitpunkt der Ausprägung des *Ft*-Gliedmaßenphänotyps

Da jedoch auch Mutationen in einer weiten Entfernung von *Sall1* TBS verursachen können (eventuell durch Mutationen von Enhancer-Elementen), sollte auch die Expression von *Sall1* in *Ft*-Embryonen zum Zeitpunkt des Auftretens des Gliedmaßenphänotyps überprüft werden. Dies geschah durch die *in situ* Hybridisierung von vollständigen Embryonen ('whole mount') mit einer spezifischen RNA-Sonde, die mit Hilfe des *IMAGE:537158* Klons aus dem 3'-Ende der *Sall1* cDNA (Nukleotide 3619-4412) hergestellt worden war. Es zeigte sich, dass *Sall1* sehr stark in den sich entwickelnden Gliedmaßen exprimiert wird (Abb. 3.22). Am Tag 10.5 p.c. ist die Expression eindeutig in den Arm- und Beinanlagen erkennbar (Abb. 3.22-A). Zu diesem Zeitpunkt nimmt die *Sall1* Expression noch fast den gesamten distalen Bereich der Beinknospe ein. Währenddessen zeigen die Armplatten bereits eine Verlagerung der Expression von distal nach proximal (Abb. 3.22-H). Am Tag 11.5 p.c. ist sowohl in den vorderen als auch in den hinteren Gliedmaßen nur noch eine proximale *Sall1* Expression nachweisbar, die sich zudem immer mehr in den posterioren Teil der Knospen zurückzieht (Abb. 3.22-I). In dieser proximalen Domäne wird *Sall1* gleichmäßig im dorsalen und im ventralen Gewebe der Gliedmaßen exprimiert (Abb. 3.22-F). Im distalen Bereich ist zu diesem Zeitpunkt keine signifikante *Sall1* Expression mehr zu erkennen. Ab Tag 12.5 p.c. wird die *Sall1* Expression in der posteroproximalen Region der Gliedmaßen immer schwächer (ohne Abb.), so dass am Tag 13.5 p.c. mit der verwendeten Sonde keine *Sall1* Expression in den Gliedmaßen mehr erkennbar war.

Auch in anderen Geweben des Embryos ließ sich die Expression von *Sall1* nachweisen. So konnten z.B. *Sall1* Transkripte im Gewebe um den Urnierengang (Abb. 3.22-E), am Ausgang des Enddarms (zu diesem Zeitpunkt Kloake; Abb. 3.22-E) und in der Flügelplatte des embryonalen Rückenmarks entdeckt werden (Abb. 3.22-G). Insgesamt wird *Sall1* in den Geweben exprimiert, die auch in der menschlichen TBS Erkrankung betroffen sind.

Die Analyse der *Sall1* Expression in heterozygoten und homozygoten *Ft*-Embryonen am Tag 11.5 p.c., also kurz bevor die ersten Veränderungen in der Expression von bestimmten Markergenen (s. Abschnitt 1.4.1) auftreten, ließ auf Transverssschnitten und im Ganzkörperpräparat in einem Großteil der analysierten Embryonen insbesondere in den Gliedmaßenplatten keine Abweichungen vom Wildtypmuster erkennen (Abb. 3.22-B,C,D,J,K). Dennoch zeigten einige wenige homozygote *Ft*-Embryonen eine ektopische *Sall1*-Expression im anterioren Bereich der Extremitäten sowie eine generelle zeitliche Verzögerung der dynamischen Expression in den Extremitätenknospen (Abb. 3.22-D,L). Allerdings traten diese Veränderungen nur bei homozygoten *Ft*-Embryonen auf, die insgesamt einen sehr hohen

Grad an schwerwiegenden Fehlbildungen und ein verlangsamtes Wachstum aufwiesen (Abb. 3.22-D). Daher handelt es sich bei den beobachteten Abweichungen in der *Sall1*-Expression sehr wahrscheinlich um Sekundäreffekte, die nicht in einem direkten kausalen Zusammenhang mit der Ausprägung der auftretenden Fehlbildungen stehen. *Sall1* ist daher weder in seiner genomischen Struktur noch in seiner Expression direkt von der *Ft*-Mutation betroffen und scheidet daher als Kandidatengen für den *Ft*-Phänotyp aus.



3.8 Die Auswertung der Konsensussequenz des relevanten menschlichen Genomsegments

Wie bereits an anderer Stelle dieser Arbeit erwähnt, ließ die vergleichende Genomanalyse erkennen, dass der größte Teil des distalen Bereichs des MMU8 homolog zu der Region des langen Arms des HSA16 ist (Abb. 3.13). Zudem bewies eine erste hochauflösende Kopplungsanalyse dieses konservierten Segments in der Maus, dass die Genanordnung innerhalb dieses Genomabschnitts außerordentlich gut konserviert ist (Becker-Follmann *et al.*, 1997). In dieser Analyse wurde eine Kopplungskarte für insgesamt 25 orthologe Loci über einen Bereich von 28cM aufgestellt. Der untersuchte Bereich reicht von *Mmp2* bis *Mc1r* und entspricht somit fast dem gesamten HSA16q. Der *Ft*-Locus grenzt also direkt an das von Becker-Follmann *et al.* analysierte chromosomale Segment bzw. ist sogar in diesem enthalten. Im Februar 2001 wurde basierend auf den Arbeiten des *Internationalen Humanen Genomprojekts* und des US-Unternehmens *Celera Genomics* eine erste Konsensus-Sequenz des menschlichen Genoms veröffentlicht. Dadurch bestand die Möglichkeit, den entsprechenden homologen Bereich auf HSA16q nach Kandidatengenen für die *Ft*-Mutation abzusuchen und mit der in dieser Studie erstellten Genkarte des *Ft*-Locus auf MMU8 zu vergleichen. Dazu wurden zunächst die orthologen Gene des *Ft*-Locus im menschlichen Genom identifiziert und in einer Karte zusammengefasst (Abb. 3.23).

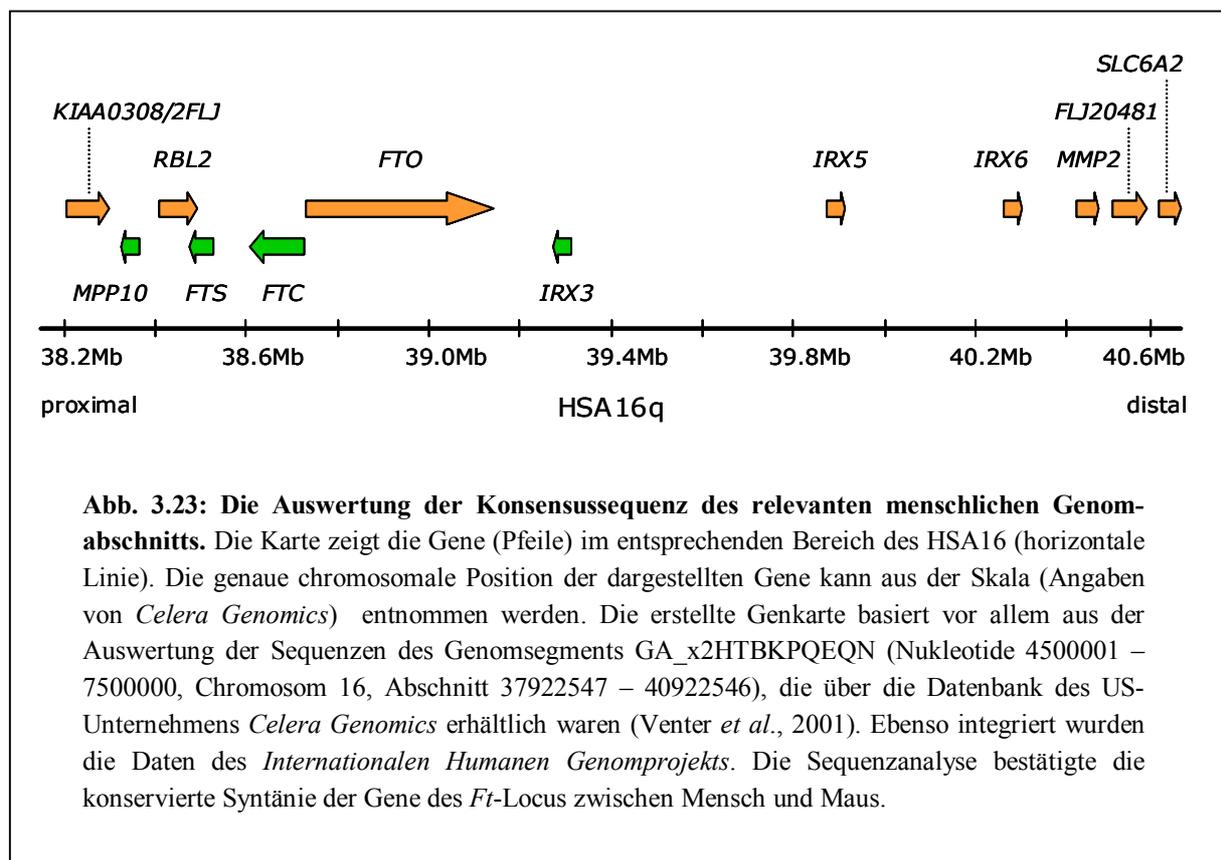


Abb. 3.23: Die Auswertung der Konsensussequenz des relevanten menschlichen Genomabschnitts. Die Karte zeigt die Gene (Pfeile) im entsprechenden Bereich des HSA16 (horizontale Linie). Die genaue chromosomale Position der dargestellten Gene kann aus der Skala (Angaben von *Celera Genomics*) entnommen werden. Die erstellte Genkarte basiert vor allem aus der Auswertung der Sequenzen des Genomsegments GA_x2HTBKPQEQN (Nukleotide 4500001 – 7500000, Chromosom 16, Abschnitt 37922547 – 40922546), die über die Datenbank des US-Unternehmens *Celera Genomics* erhältlich waren (Venter *et al.*, 2001). Ebenso integriert wurden die Daten des *Internationalen Humanen Genomprojekts*. Die Sequenzanalyse bestätigte die konservierte Syntanie der Gene des *Ft*-Locus zwischen Mensch und Maus.

In der Tat offenbarte diese Karte, dass die lineare Anordnung und auch die Orientierung der Gene des *Ft*-Locus zwischen Mensch und Maus fast vollkommen erhalten geblieben ist. Die einzige Ausnahme bildet das menschliche Gen *MPP10*, da das orthologe *Mpp10* Gen der Maus nicht wie eigentlich erwartet mittels PCR in dem Genomabschnitt zwischen *KIAA0308/2FLJ* und *Rbl2* kartiert werden konnte (ohne Abb.). Die chromosomale Lokalisation des *Mpp10* Gens befindet sich vielmehr außerhalb des erstellten YAC-Contigs. Da *Mpp10* somit nicht im unmittelbaren Bereich der *Ft*-Deletion liegt, wurde auf die genaue Kartierung dieses Gens im Mausgenom verzichtet. Aus demselben Grund wurden auch die Gene *FLJ20481* und *SLC6A2*, die sich im menschlichen Genom im distalen Bereich an *Mmp2* anschließen, in der Maus nicht näher analysiert.

Um weitere bisher unentdeckte Gene im Bereich des *Ft*-Locus zu identifizieren, wurden ebenfalls die Positionen aller cDNA-Sequenzen des korrespondierenden Bereichs der bereits erwähnten *Genebridge4*-Karte der GeneMap'99 im menschlichen Genom bestimmt. Tatsächlich konnten die auf cDNA-Sequenzen basierenden STS-Marker *SGC38139*, *sts-AA005225*, *stSG16320* und *R07176* im Bereich des relevanten menschlichen Genomabschnitts lokalisiert werden. Die den vier STS-Markern zugrundeliegenden kurzen cDNA Sequenzen beruhten jedoch jeweils nur auf ein oder zwei EST-Sequenzeinträgen und ließen sich nicht einer längeren cDNA zuordnen. Weiterhin enthielten diese vier cDNAs keine Intronsequenzen, sondern stammten jeweils von einer zusammenhängenden Genomsequenz ab. Zudem enthielten die cDNAs *SGC38139* und *stSG16320* auch repetitive Sequenzen. Insgesamt ist daher die Existenz dieser angeblichen cDNA-Sequenzen sehr wahrscheinlich auf Artefakte (z.B. bei der cDNA-Synthese) oder auf fehlerhafte Kartierungen zurückzuführen (Cirera *et al.*, 2000). Alle anderen STS-Marker aus diesem Bereich der *Genebridge4*-Karte befanden sich außerhalb des relevanten Genomabschnitts.

Zusätzlich wurde überprüft, ob die Sequenz des menschlichen Genomabschnitts eventuell weitere EST-Sequenzen der NCBI-GenBank enthält. Wiederum konnten nur einige einzelne verstreute ESTs lokalisiert werden, die aufgrund der oben erwähnten Charakteristika ebenfalls sehr wahrscheinlich auf Artefakten basieren. Sämtliche im Bereich des menschlichen *Ft*-Locus lokalisierten kurzen cDNA-Sequenzen beruhen also anscheinend nicht auf weiteren bisher unentdeckten Genen.

Abschließend wurden auch die vom *Internationalen Humanen Genomprojekt* bzw. von *Celera Genomics* veröffentlichten Karten des menschlichen Genoms ausgewertet. Auch diese mit Hilfe verschiedener Computerprogramme aufgestellten Karten gaben keinen Hinweis auf das Vorhandensein weiterer bisher unentdeckter Gene im untersuchten Genomabschnitt. Insgesamt bestätigte die Auswertung der Konsensussequenz des menschlichen Genoms also die für das Genom der Maus erstellte physikalische Genkarte des *Ft*-Locus. Ebenso lieferte diese vergleichende Genomanalyse neben den schon in der Maus identifizierten Genen keinen

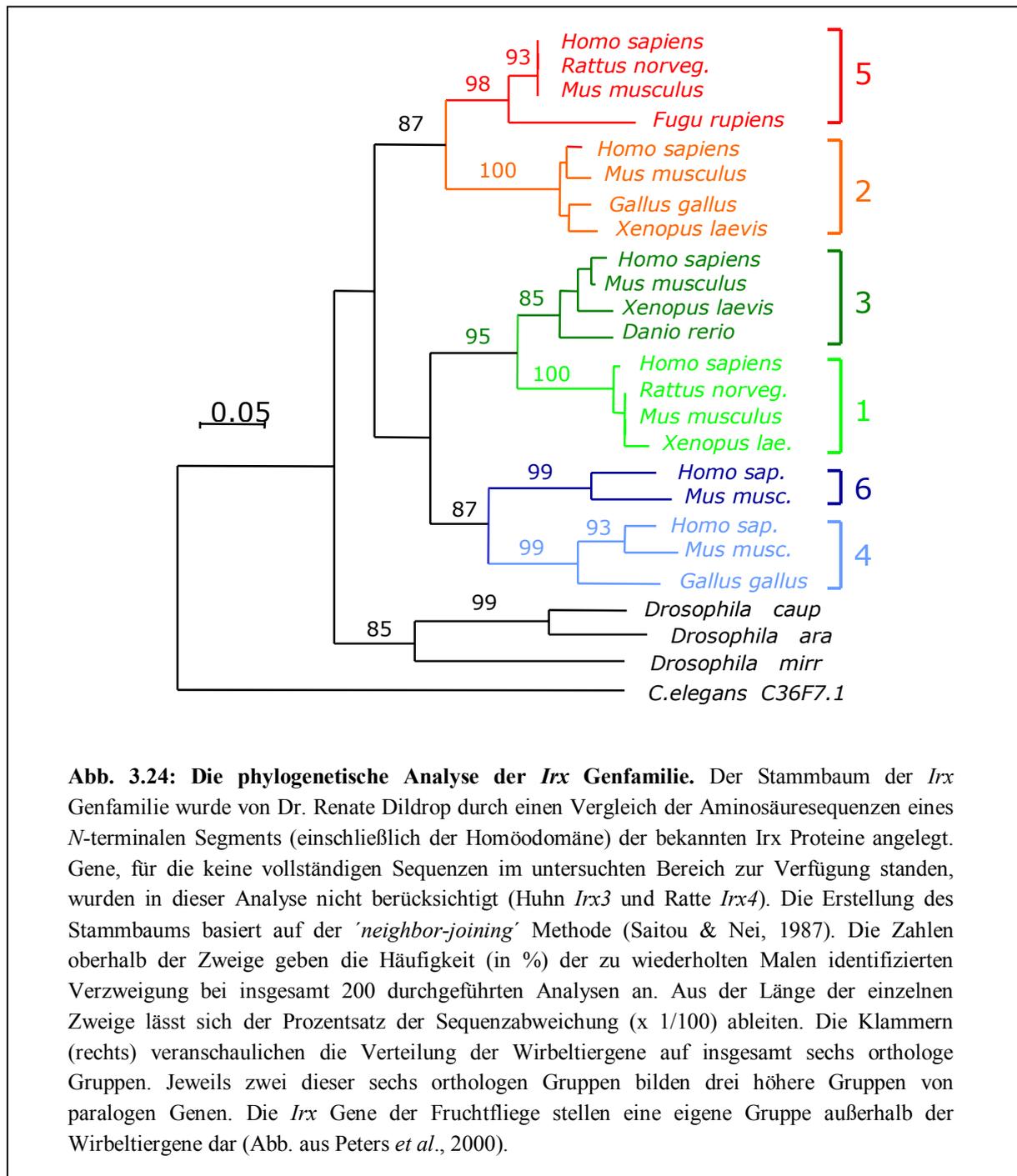
Anhaltspunkt für das Vorhandensein weiterer Gene im relevanten Genomabschnitt. Zusammenfassend legen daher die in dieser Arbeit erzielten Ergebnisse nahe, dass alle Kandidatengene im Bereich des *Ft*-Locus identifiziert worden sind.

3.9 Die Betrachtung des *Fused toes* Locus unter evolutionären Aspekten

Die phylogenetische Analyse der *Iroquois* Genfamilie ließ einige aufschlussreiche Zusammenhänge hinsichtlich der Evolution der *Irx* Gene der Wirbeltiere erkennen. Da aus dieser Studie auch Schlussfolgerungen bezüglich der Evolution des gesamten *Ft*-Locus abgeleitet werden konnten, ist die von Dr. Renate Dildrop durchgeführte Analyse im nun folgenden Abschnitt 3.9.1 kurz beschrieben.

3.9.1 Die phylogenetische Analyse der *Iroquois* Genfamilie

In der *NCBI*-GenBank konnten von Dr. Renate Dildrop weitere *Irx* Gene auch in der Ratte (orthologe Gene zu Maus *Irx1*, *Irx3*, *Irx4* und *Irx5*) und im Pufferfisch (orthologe Gene zu Maus *Irx3* und *Irx5*) identifiziert werden (Peters *et al.*, 2000). Die von diesen cDNA-Sequenzen abgeleiteten Proteinsequenzen wurden zusammen mit den beschriebenen *Irx* Proteinen des Zebrafischblinds (*Ziro3*; Tan *et al.*, 1999), des Krallenfroschs (*Xiro1*, *Xiro2*, *Xiro3*; Gómez-Skarmeta *et al.*, 1998; Bellefroid *et al.*, 1998) und des Huhns (*Irx2*, *Irx3* und *Irx4*; Goriely *et al.*, 1999; Funayama *et al.*, 1999; Bao *et al.*, 1999) sowie der Fruchtfliege (*ara*, *caup* und *mirr*; Gómez-Skarmeta *et al.*, 1996; McNeill *et al.*, 1997) und des Fadenwurms (*C36F7.1*; Bürglin, 1997) für eine phylogenetische Betrachtung der *Irx* Genfamilie verwendet. Durch diese Analyse konnten die evolutionären Beziehungen innerhalb der *Irx*-Familie aufgedeckt und orthologe (homologe Gene zwischen verschiedenen Spezies) bzw. paraloge (homologe Gene innerhalb eines Organismus) Gene definiert werden. Die molekulare Phylogenie der *Irx* Gene offenbarte einige bemerkenswerte Zusammenhänge (Abb. 3.24; Peters *et al.*, 2000). Die *Irx* Gene der Wirbeltiere verteilen sich auf sechs Gruppen von jeweils orthologen Genen. Die Gene in einer Gruppe stammen demnach alle von ein und demselben Vorläufergen ab. Jeweils zwei dieser sechs orthologen Gruppen (Gruppen 5/2, Gruppen 3/1, Gruppen 6/4) bilden drei höhere Gruppen von paralogenen Genen. Diese sehr eng miteinander verwandten Gene (z.B. Maus *Irx1/Irx3*, *Irx2/Irx5* oder *Irx4/Irx6*) stammen also von einem gemeinsamen noch älteren Urogen ab. Die *Irx* Gene der Fruchtfliege *ara*, *caup* und *mirr* bilden eine eigene Gruppe außerhalb der Wirbeltiergene. Die Genduplikationen, die zu

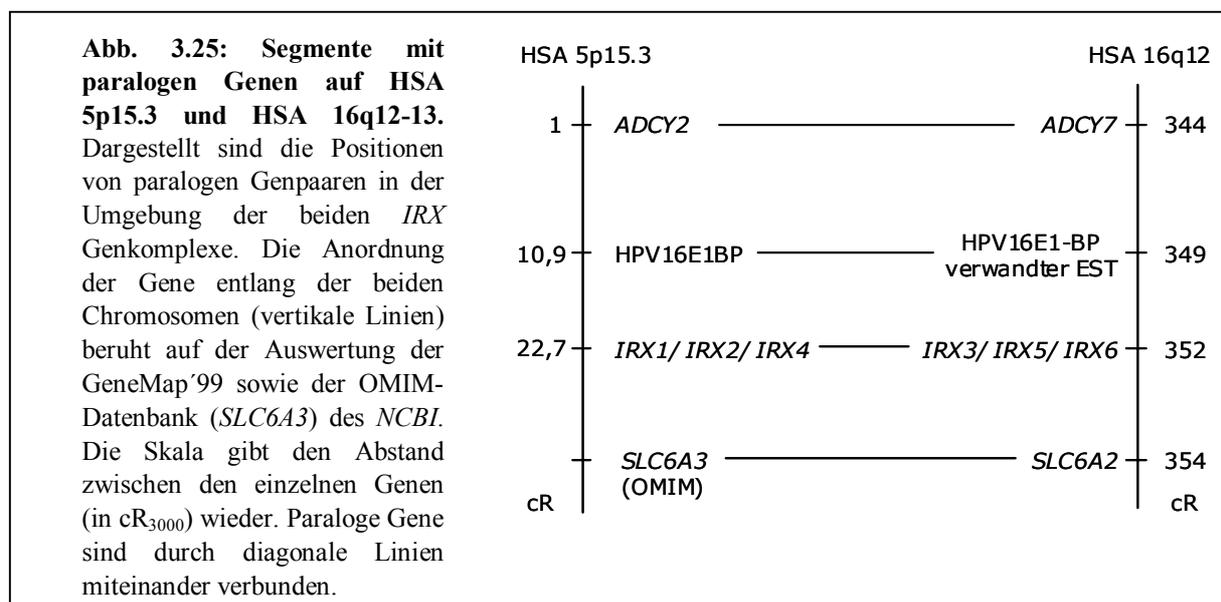


den drei Genen der Fruchtfliege und den sechs Genen der Wirbeltiere führte, sind daher sehr wahrscheinlich auf unabhängige Ereignisse zurückzuführen (Peters *et al.*, 2000). Allerdings zeigt die phylogenetische Analyse der *Irx* Familie, dass den drei Genen der Fruchtfliege keine direkten eng verwandten Wirbeltiergene zugeordnet werden können. Der *Irx* Genkomplex der Fruchtfliege scheint sich vielmehr aufgrund von zwei lokalen Genduplikationen eines gemeinsam mit dem *Irx* Gen des Fadenwurms verwandten Urgens gebildet zu haben. Die verwandtschaftlichen Beziehungen innerhalb der *Irx* Gene der Wirbeltiere lassen sich am

einfachsten durch ein ähnliches Szenario erklären. Dabei könnte sich zunächst durch zwei lokale Genduplikationen eines mit dem *Irx* Gen des Fadenwurms und dem Urgen der Fruchtfliege verwandten Urgens der Wirbeltiere ein Genkomplex bestehend aus drei Wirbeltier *Irx* Genen gebildet haben. Eine spätere zweite Duplikation des gesamten Genkomplexes könnte im weiteren Verlauf der Evolution zu zwei eng verwandten Genkomplexen von jeweils drei *Irx* Genen geführt haben, die bis heute erhalten geblieben sind.

3.9.2 Die Existenz der beiden *Irx* Genkomplexe beruht auf der Duplikation eines ausgedehnten Genomsegments

Interessanterweise enthalten die in der vorliegenden Arbeit identifizierten *Irx* Genkomplexe der Maus jeweils ein Mitglied der drei höheren phylogenetischen Gruppen von paralogen Genen (Abb. 3.16 und 3.24). Ebenso offenbarte die lineare Abfolge der *Irx* Gene in beiden Genkomplexen, dass die relative Position von paralogen Genen (z.B. *Irx1* und *Irx3*) innerhalb des jeweiligen Komplexes während der Evolution streng konserviert worden ist. Sowohl die von Dr. Renate Dildrop untersuchten phylogenetischen Beziehungen als auch die physikalische Anordnung der *Irx* Gene legten somit die Schlussfolgerung nahe, dass beide Genkomplexe aus der Duplikation eines ursprünglichen Urkomplexes hervorgegangen sind. Dieses Ereignis hat sich sehr wahrscheinlich vor der Aufspaltung der verschiedenen Wirbeltierfamilien zugetragen, da mittlerweile gezeigt werden konnte, dass neben der Maus auch das Huhn, der Mensch und der Zebrafisch über zwei *Irx* Genkomplexe verfügen (Ogura *et al.*, 2001; Wang *et al.*, 2001).



Es wird angenommen, dass sich während der Evolution der Wirbeltiere mehrfach Duplikationen von ausgedehnten Genomabschnitten bzw. des gesamten Genoms ereignet haben (Ruddle *et al.*, 1994; Skrabanek & Wolfe, 1998; Aparicio 2000). Sollte die Existenz von zwei *Irx* Genkomplexen in den Wirbeltieren auf der Duplikation eines größeren Genomsegments begründet sein, so müsste es auch zu einer Duplikation von benachbarten Genen gekommen sein, die im Verlauf der Evolution eventuell erhalten geblieben sind.

Daher sollten die zum damaligen Zeitpunkt verfügbaren genetischen und physikalischen Karten des menschlichen Genoms (z.B. GeneMap'99, Online Mendelian Inheritance in Man (OMIM) -Datenbank des *NCBI*) nach weiteren Paaren von paralogen Genen in der Nähe der beiden *IRX* Genkomplexe durchsucht werden. Tatsächlich konnten im Bereich des menschlichen *IRXA* Komplexes auf HSA 5p15.3 mehrere Gene (z.B. *ADCY2*, *SLC6A3*) identifiziert werden, deren paraloge Partner in der Umgebung des *IRXB* Komplexes auf HSA 16q12-q13 lokalisiert sind (Abb. 3.25). Die Verteilung dieser Gene im menschlichen Genom bestätigte somit die Hypothese, dass die Existenz der beiden *Irx* Genkomplexe auf der Duplikation eines größeren Genomsegments beruht.

Gene aus dem Bereich des <i>Ft</i> Locus auf HSA16q12	paraloge Partnergene/Sequenzen	chromosomale Lokalisation auf HSA
<i>KIAA0308/2FLJ</i>	keine paraloge Sequenz	-
<i>RBL2</i>	<i>RB1</i> <i>RBL1/p107</i>	13q14.2 20q11.23
<i>FTS</i>	bisher unbekannte Sequenz bisher unbekannte Sequenz bisher unbekannte Sequenz bisher unbekannte Sequenz bisher unbekannte Sequenz	5p14.1 3p21.1 11q21 14q32.31 4q27
<i>FTM</i>	<i>RPGRIP</i>	14q11
<i>FTO</i>	bisher unbekannte Sequenz	2p21
<i>IRXB</i>	<i>IRXA</i>	5p15.3
<i>MMP2</i>	<i>MMP9</i>	20q13.11

Tab. 3.5: Paraloge Partnergene/Sequenzen der Gene aus dem Bereich des *Ft* Locus. Eng verwandte Sequenzen und deren chromosomale Lokalisation wurden durch das Durchsuchen der Konsensussequenz des menschlichen Genoms (ENSEMBL Human Genome Server) identifiziert.

Dieser evolutionäre Vorgang müsste demnach auch zu einer Duplikation aller anderen Gene im Bereich des *Ft* Locus geführt haben. Um diese Hypothese zu überprüfen, wurde die Konsensussequenz des menschlichen Genoms nach entsprechenden paralogen Sequenzen durchsucht. Tatsächlich konnten für fast alle Gene im *Ft* Kandidatenbereich (s. Abb. 3.18) paraloge Partnersequenzen im menschlichen Genom identifiziert werden (Tab. 3.5). Die einzige Ausnahme bildete das Gen *KIAA0308/2FLJ*, für das auf diese Weise keine eng verwandten Sequenzen gefunden werden konnten. Die chromosomalen Positionen dieser Sequenzen zeigten jedoch an, dass die jeweiligen paraloge Partner verstreut über das gesamte menschliche Genom vorliegen und damit nicht, wie eigentlich erwartet, in enger Nachbarschaft zum *IRXA* Komplex im verwandten Genomsegment auf HSA 5p15.3 lokalisiert sind (Tab. 3.5).

4 Diskussion

Mit der Publikation eines ersten Entwurfs der Konsensussequenz des menschlichen Genoms steht mittlerweile fast die gesamte Sequenz der menschlichen DNA in öffentlichen Datenbanken zur Verfügung. Die größte Herausforderung liegt nun darin herauszufinden, was sich hinter dieser Genomsequenz verbirgt. Dazu müssen zunächst zwei fundamentale Fragen beantwortet werden: Wo sind die einzelnen Gene im menschlichen Genom lokalisiert und welche biologischen Aufgaben erfüllt jedes einzelne dieser Gene?

Die derzeitigen Arbeiten über das Mausgenom werden helfen, diese Fragen zu beantworten. Im Prinzip eignet sich jeder Organismus, dessen Genom ein eindeutig identifizierbares Äquivalent eines menschlichen Gens (Ortholog) enthält, als genetisches Modell für den Menschen. Allerdings steht die Maus dem Menschen sowohl genetisch als auch physiologisch sehr nahe. Diese biologische Ähnlichkeit und auch das konservierte Genom zwischen Mensch und Maus machen die Maus zu einem idealen Modellorganismus, um die Funktionen der Gene im Säugetiergenom und damit auch die genetische Grundlage vieler menschlicher Erbkrankheiten wie z.B. Asthma, Osteoporose oder Fettleibigkeit aufzudecken. Es wird erwartet, dass der größte Teil des Mausgenoms bis Oktober 2002 vollständig sequenziert vorliegt (Graham *et al.*, 2001). Jedoch muss das Wissen um die Sequenz durch funktionelle Analysen erweitert werden. Ein internationales Konsortium hat es sich somit zum Ziel gesetzt, die biologischen Funktionen eines jeden Mausgens und damit, aufgrund der bestehenden Homologie, auch der orthologen menschlichen Gene aufzuklären (*The International Mouse Mutagenesis Consortium*, 2001). Das Studium von Mausmutanten gehört dabei zu einer der bedeutendsten Vorgehensweisen.

Die Aufklärung der molekularen Grundlage der *Fused toes* Mausmutation verspricht die Identifizierung von Genen, die vor allem in entwicklungsbiologischen Prozessen eine entscheidende Rolle spielen. Der Phänotyp heterozygoter *Ft*-Mäuse äußert sich durch eine partielle Syndaktylie an den Vorderpfoten sowie eine massive Thymushyperplasie. In beiden Geweben ist anscheinend eine Verminderung der kontrollierten Aktivierung des programmierten Zelltods für die Ausprägung dieses Phänotyps verantwortlich (van der Hoeven *et al.*, 1994; Volkmann *et al.*, 1996). Homozygote *Ft*-Embryonen sterben aufgrund schwerer Fehlbildungen bereits während der Embryonalentwicklung. Diese Embryonen zeigen neben einem deutlich verlangsamten Wachstum besonders schwerwiegende Missbildungen im Gesichts- und im vorderen Gehirnbereich. Ebenfalls von Missbildungen betroffen sind die sich entwickelnden vorderen als auch hinteren Extremitäten. Weiterhin zeugt die zufällig ausgerichtete Richtung der Bewegung des primitiven Herzschlauchs bzw. der embryonalen Rotation von einem Verlust über die genetische Kontrolle der Etablierung

der Rechts-Links-Asymmetrie der Körperachse. Nicht zuletzt ist auch die Entwicklung des zentralen Nervensystems dieser Embryonen aufgrund der Desorganisation der ventralen Region des Neuralrohrs durch die *Ft*-Mutation erheblich beeinträchtigt.

Die molekulare Grundlage der *Ft*-Mutation ist auf der Integration eines Transgens und der damit zusammenhängenden Deletion eines 1650kb großen Genomsegments der Region D des Chromosoms MMU8 begründet (van der Hoeven *et al.*, 1994; Ausmeier, 2000). Die erhebliche Ausdehnung der Deletion sowie die Vielfalt der dadurch hervorgerufenen Fehlbildungen sprechen für eine Beteiligung von mehreren Genen an der Ausprägung des beobachteten Phänotyps. In vorangegangenen Studien konnten mit *Rbl2* (*p130*) und *Fts* (*Ft1*) bereits zwei Kandidatengene im betroffenen Genomabschnitt identifiziert werden (Lesche *et al.*, 1997). Allerdings ließen umfangreiche funktionelle Analysen sowie Expressionsstudien keinerlei Korrelation von *Fts* bzw. *Rbl2* mit der Ausprägung des beschriebenen *Ft*-Phänotyps erkennen (Peetz, 1996; Lesche, 1997). Demnach müssen im betroffenen Genomabschnitt noch weitere Gene lokalisiert sein, die wichtige biologische Aufgaben während der Entwicklung der Extremitäten, des Gesichtsbereichs und des zentralen Nervensystems und auch bei der Kontrolle des programmierten Zelltods sowie der Etablierung der Rechts-Links-Asymmetrie der Körperachse erfüllen. Diese Kandidatengene wurden in der vorliegenden Arbeit identifiziert, im Fall von bisher unbekanntem Genen weitestgehend charakterisiert und in einer hochauflösenden Genkarte des *Ft*-Locus zusammengefasst.

4.1 Die Kandidatengene im Bereich des deletierten Genomabschnitts

Unter Anwendung der positionellen Klonierung, der vergleichenden Genomanalyse und auch durch phylogenetische Analysen konnten mit *Ftm*, *Fto*, *Irx3*, *Irx5* und *Irx6* insgesamt fünf weitere Gene im deletierten Genomabschnitt identifiziert werden. Die Charakterisierung dieser Gene ermöglicht eine erste Beurteilung der Beteiligung dieser Gene an der Ausprägung der in den *Ft*-Mäusen auftretenden Fehlbildungen.

Durch die Verfügbarkeit einer vollständigen Genomsequenz bewegt sich die derzeitige Vorgehensweise bei der Identifizierung von phänotyp-assoziierten Kandidatengenen weg von den aufwendigen Arbeitsschritten im Labor und hin zur direkten Überprüfung und Bestätigung von Kandidatengenen. Die Verschiebung dieser Schwerpunkte lässt sich sehr gut anhand des Aufbaus der vorliegenden Arbeit beobachten. Lag der Schwerpunkt zu Beginn noch auf der Positionsklonierung von Kandidatengenen mit Hilfe von aufwendigen Techniken (z.B. 'Exon trapping'), so wurden im weiteren Verlauf immer mehr die Informationen der aktuellen Gen- bzw. Genomdatenbanken für die Identifizierung von Kandidatengenen

herangezogen. Tatsächlich wurde durch die Nutzung der verfügbaren Datenbanken die physikalische Kartierung des *Ft* Locus erheblich vorangetrieben. Auf diese Weise war es möglich, eine detaillierte hochauflösende physikalische Genkarte eines ca. 2000kb großen Segments des Mausgenoms zu erstellen, deren Auflösung in den aktuellen öffentlich verfügbaren Genom- und Transkriptkarten (Hudson *et al.*, 2001; Avner *et al.*, 2001, Cai *et al.*, 2001) nicht erreicht wird. Trotz alledem wird die korrekte physikalische Kartierung des untersuchten Genomsegments durch eine aktuell veröffentlichte, auf SSLP-Markern bzw. einem BAC-Gerüst basierenden physikalischen Karte des Mausgenoms bestätigt (Cai *et al.*, 2001). Allerdings konnten in dieser Karte bisher erst fünf der in dieser Arbeit verwendeten Mikrosatellitenmarkern (*D8Mit251*, *D8Mit250*, *D8Mit309*, *D8Mit147*, *D8Mit284*, *D8Mit240*) in ihrer korrekten physikalischen Abfolge angeordnet werden. Alle anderen in der vorliegenden Arbeit erwähnten Marker bzw. identifizierten Gene blieben in dieser aktuellen Genomkarte bisher unberücksichtigt.

4.1.1 *Fantom* (*Ftm*)

Über die positionelle Klonierung wurde das zuvor unbekannte Gen *Fantom* (*Ftm*) im Bereich des in der *Ft*-Mutation verlorengegangenen Genomabschnitts identifiziert. Die komplexe Klonierungsstrategie, bei der teilweise auf orthologe Sequenzen des Menschen und des Pufferfischs zurückgegriffen werden musste, führte zu der Isolierung einer vollständigen, 7kb langen *Ftm* cDNA. Da keine orthologen Sequenzen außerhalb des Unterstammes der Wirbeltiere, beispielsweise im Genom der Hefe *S. cerevisiae*, der Fruchtfliege *D. melanogaster* und des Fadenwurms *C. elegans*, identifiziert wurden, handelt es sich sehr wahrscheinlich um ein wirbeltierspezifisches Gen. Aus der isolierten cDNA-Sequenz konnte ein offener Leserahmen abgeleitet werden, der für ein 145kDa großes Protein bestehend aus 1264 Aminosäuren kodiert.

Ftm ist verwandt mit dem zilien-assoziierten Protein RPGRIP

Der modulartige Aufbau und auch die Aminosäuresequenz des vermeintlichen *Ftm* Proteins ließ eine signifikante Ähnlichkeit mit dem *Retinitis Pigmentosa* GTPase Regulator-interagierenden Protein (RPGRIP) erkennen. Wie bei RPGRIP besteht die *N*-terminale Hälfte des *Ftm* Proteins aus drei aufeinanderfolgenden sogenannten 'coiled-coil' Domänen, die sehr wahrscheinlich eine wichtige Rolle bei der Ausbildung von Protein-Protein-Interaktionen

spielen. Ebenso enthält das abgeleitete Ftm Protein in Übereinstimmung mit RPGRIP eine C2-artige Domäne sowie eine vermeintliche RPGR-Interaktionsdomäne. Verwandte C2-artige Domänen wurden bisher vor allem in phospholipid-bindenden Proteinen identifiziert und vermitteln dort sehr wahrscheinlich gerade eben die Bindung von Phospholipiden, zu denen beispielsweise auch das sekundäre Botenmolekül (*'second messenger'*) Inositol-1,4,5-triphosphat gehören kann. Innerhalb des RPGRIP Proteins ist die RPGR-Interaktionsdomäne, wie der Name schon sagt, notwendig für die Interaktion mit RPGR (*Retinitis pigmentosa* GTPase Regulator). Mutationen im *RPGR* Gen verursachen im Menschen häufig eine sehr schwere Form der Netzhautdegeneration (Meindl *et al.*, 1996; Roepman *et al.*, 1996; Vervoort *et al.*, 2000), die sogenannte X-gekoppelte *Retinitis pigmentosa* (Typ RP3). Dabei kann die Degeneration der Photorezeptorzellen eine vollständige Erblindung des Patienten zur Folge haben.

Die Funktion von RPGRIP in den Photorezeptorzellen der Netzhaut

Die Photorezeptorzellen der Netzhaut, die sogenannten Stäbchen und Zapfen, bestehen aus zwei morphologisch und funktionell völlig verschiedenen Bestandteilen, dem inneren und dem äußeren Segment. Im inneren Segment sind all diejenigen Zellorganellen enthalten, die an den biosynthetischen und metabolischen Prozessen der Zelle beteiligt sind. Im Gegensatz dazu stellt das äußere Segment mit seiner membranreichen Struktur ein hochspezialisiertes Organell für den Prozess der Phototransduktion dar, welches einer ständigen Regeneration unterliegt, bei der ältere Membranstapel an der Spitze des Segments abgestoßen und durch neue Membranen an der Basis ersetzt werden (Guerin *et al.*, 1993). Beide Segmente werden in einem verengten Übergangsbereich durch einige wenige unbewegliche Cilien miteinander verbunden, deren Achsenfaden (Axonema) im Gegensatz zum Axonema beweglicher Cilien (neun äußere Mikrotubulipaare und zwei einzelne zentrale Mikrotubuli – 9+2 Konfiguration) nur aus neun äußeren Mikrotubulipaaren besteht (9+0 Konfiguration). Über diese verbindenden Zilien erfolgt der vesikuläre Transport einer ungeheuer großen Menge von Proteinen (z.B. von Rhodopsin und anderen Komponenten der Phototransduktion) entgegen eines steilen Konzentrationsgradienten (Sung & Tai, 2000).

RPGRIP ist zusammen mit RPGR im Bereich der verbindenden Zilien und im äußeren Segment der Photorezeptorzellen lokalisiert (Hong *et al.*, 2000; Roepman *et al.*, 2000; Hong *et al.*, 2001). Aufgrund seiner stabilen Assoziation mit den verbindenden Zilien, stellt RPGRIP sehr wahrscheinlich eine strukturelle Komponente des ziliaren Axonemas dar und könnte damit als molekulares Gerüst für die Verankerung eines makromolekularen regulatorischen Komplexes einschließlich RPGR dienen, der eine Schlüsselrolle im

dynamischen Transport von Proteinen zum oder vom äußeren Segment der Photorezeptorzellen einnimmt (Roepmann *et al.*, 2000; Hong *et al.*, 2001).

Aufgrund des hohen Verwandtschaftsgrades zu RPGRIP könnte *Ftm* eine ähnliche Rolle in einem vergleichbaren oder sogar in demselben Prozess spielen. Tatsächlich wird *Ftm* sehr stark in den Augen von ausgewachsenen Mäusen exprimiert (eigene Beobachtung). Allerdings konnte eine Degeneration der Netzhaut in heterozygoten *Ft*-Mäusen bisher nicht beobachtet werden. Die Sterblichkeit der Embryonen verhindert entsprechende Studien in homozygoten *Ft*-Tieren. Interessanterweise wurde im Menschen in unmittelbarer Nähe der orthologen *Ft*-Kandidatengene ein Locus kartiert (16q13-q21), der eine schwere autosomal rezessive Form der *Retinitis pigmentosa* verursacht (Tab. 3.4). Allerdings scheint in diesem Fall eine 'missense' Mutation (Austausch eines konservierten Glycinrestes) in *CNGB1*, einem Gen, das für die β -Untereinheit eines cGMP-gesteuerten Kanals der Stäbchen kodiert, für das Auftreten dieser Krankheit verantwortlich zu sein (Bareil *et al.*, 2001).

Die Monozilien der ventralen Zellen des Primitivknoten sind maßgeblich an der Etablierung der Rechts-Links-Asymmetrie der Körperachse beteiligt

Neben den Zapfen und Stäbchen der Netzhaut können Zilien, wenn auch häufig nur in einer rudimentären Form, auch in anderen Zelltypen der Wirbeltiere gefunden werden. Dazu gehören beispielsweise Spermien, die Flimmerepithelien im Atemtrakt und in den Eileitern, die Haarzellen des Innenohrs und auch die Zellen des Primitivknotens im Primitivstreifen im frühen Embryo (im Neurula-Stadium). Letztere besitzen an ihrer Zelloberfläche eine einzelne Zilie, die bis in den extrazellulären Raum hinein ragt (Sulik *et al.*, 1994; Bellomo *et al.*, 1996). Diese sogenannten Monozilien weisen genau wie die unbeweglichen Zilien der Photorezeptorzellen eine 9+0 Mikrotubulstruktur auf und unterscheiden sich daher von der Standard 9+2 Konfiguration beweglicher Zilien. Aufgrund dieses Umstands wurde lange Zeit an der Beweglichkeit dieser Zilien gezweifelt. Allerdings konnte mit Hilfe von Mikrolatexkügelchen gezeigt werden, dass eine gegen den Uhrzeigersinn erfolgende Bewegung der Monozilien eine linksgerichtete Strömung der extrazellulären Flüssigkeit im Primitivknoten bewirkt (Nonaka *et al.*, 1998). Es wird angenommen, dass diese Primitivknotenströmung ('nodal flow') zu einer asymmetrischen Verteilung eines bisher noch nicht identifizierten Rechts-Links-Morphogens führt und somit die Grundlage für die Initiierung der Rechts-Links-Signalkaskade darstellt (Nonaka *et al.*, 1998). Diese Hypothese wurde durch umfangreiche Studien von Mausmutanten mit einer gestörten Entwicklung der Rechts-Links-Asymmetrie der Körperachse untermauert. Beispielsweise bewirkte die gezielte Inaktivierung der Gene *Kif3a* bzw. *Kif3b*, dass keine Monozilien an der Oberfläche der Zellen im

Primitivknoten ausgebildet werden (Nonaka *et al.*, 1998; Marszalek *et al.*, 1999; Takeda *et al.*, 1999). Beide Gene kodieren für Teilkomponenten eines Mikrotubuli-Motors, dem Kinesin-Komplex KIF3. Die gleiche Konsequenz hatte die gezielte Mutation des *Hfh4* Gens (*hepatocyte nuclear factor/forkhead homolog 4*) sowie des Gens *Polaris* (*Tg737*) zur Folge (Chen *et al.*, 1998; Murcia *et al.*, 2000). Des Weiteren führte der spontane (*lrd^{lv}*) bzw. der gezielte (*lrd^{ΔPl}*) Funktionsverlust eines axonemalen Dyneins (Links-Rechts Dynein - *Lrd*), ebenfalls ein auf Mikrotubuli basierendes Motorprotein, zu einer verminderten Primitivknotenströmung, die auf der Unbeweglichkeit der Monozilien beruht (Supp *et al.*, 1997; Okada *et al.*, 1999; Supp *et al.*, 1999). Auch beim Menschen wurden defekte Zilien mit einer zufallsgesteuerten Festlegung der Rechts-Links-Asymmetrie der Körperachse in Verbindung gebracht (*Immotile-cilia* Syndrom; Übersicht bei Afzelius, 1999). Zusammengefasst bewirkt also eine Fehlfunktion der Monozilien in den ventralen Zellen des Primitivknotens eine veränderte Primitivknotenströmung und führt damit zu einer gestörten Entwicklung der Rechts-Links-Asymmetrie der Körperachse, ein Phänotyp, der auch in homozygoten *Ft*-Embryonen beobachtet werden kann (Heymer *et al.*, 1997). Die Funktionalität der Monozilien in den Zellen des Primitivknotens wurde in homozygoten *Ft*-Embryonen bisher noch nicht überprüft. Sollten sich Defekte in dieser Struktur bzw. in der Strömung des Primitivknotens nachweisen lassen, dann muss *Ftm* aufgrund der engen Verwandtschaft mit dem zilien-assoziierten Protein RPGRIP als sehr gutes Kandidatengen für den Rechts-Links-Phänotyp der homozygoten *Ft*-Embryonen in Betracht gezogen werden (Tab 4.1). Auch der Nachweis der Expression von *Ftm* in den ventralen Zellen des Primitivknotens bzw. der Assoziation von *Ftm* mit den entsprechenden Monozilien würde diese Hypothese bekräftigen.

***Ftm* wird während der Embryogenese zum Zeitpunkt der Ausprägung des *Ft*-Phänotyps exprimiert**

Die Expression von *Ftm* ließ sich während der Embryogenese in allen untersuchten Stadien (Tag 8.5 p.c. bis Tag 13.5 p.c.) nachweisen. Eine besonders ausgeprägte Expression war dabei am Tag 10.5 p.c. erkennbar. Aufgrund der beobachteten Expression in differenzierenden embryonalen Stammzellen ist anzunehmen, dass *Ftm* sogar schon in früheren Stadien der Embryogenese exprimiert wird. Allerdings konnte die räumliche Verteilung von *Ftm* Transkripten im Embryo, sehr wahrscheinlich aufgrund des schwachen Expressionsniveaus, auch unter Zuhilfenahme verschiedener RNA-Sonden nicht näher aufgeklärt werden. In der ausgewachsenen Maus wird *Ftm* vor allem im Hoden und, wenn auch nur schwach, in den Nieren und im Gehirn exprimiert. In allen anderen untersuchten Geweben war unter den gewählten PCR-Bedingungen keinerlei Expression von *Ftm* nachweisbar. Entsprechende

Expressionsanalysen am *Kazusa DNA Research Institute* belegen, dass auch das menschliche *FTM* Gen (*KIAA1005*) in den meisten untersuchten Geweben gar nicht oder nur sehr schwach exprimiert wird. Eine signifikante Expression weist einzig und allein das Gehirn vor allem die Region der Amygdala bzw. des kaudalen Nukleus auf.

Das erstellte Expressionsprofil zeigt, dass *Ftm* durchaus zum Zeitpunkt der Ausprägung der beschriebenen Missbildungen während der Embryogenese exprimiert wird. Demnach muss *Ftm* weiterhin als Kandidatengen für den *Ft*-Phänotyp berücksichtigt werden (Tab. 4.1).

4.1.2 *Fatso (Fto)*

Mit *Fatso (Fto)* wurde ein weiteres zuvor unbekanntes Gen im deletierten Genomabschnitt identifiziert. Die isolierte 3,6kb lange vollständige *Fto* cDNA kodiert für ein Protein, welches aus 502 Aminosäureresten besteht und eine Größe von 58kDa aufweist.

***Fto* kodiert ein bisher unbekanntes nukleäres Protein**

Da die abgeleitete Aminosäuresequenz keinerlei bekannte funktionelle Domänen beinhaltet, können zum gegenwärtigen Zeitpunkt keine Aussagen hinsichtlich der Funktion des *Fto* Proteins getroffen werden. Allerdings lässt die Präsenz eines zweiteiligen nukleären Lokalisationssignals im *N*-terminalen Bereich des *Fto* Proteins zusammen mit dem Nachweis der zellkernspezifischen Lokalisation verschiedener *Fto*-Fusionsproteine eine biologische Aufgabe im Zellkern vermuten.

Interessanterweise offenbarte ein Sequenzvergleich mit der *NCBI*-GenBank, dass ein Sequenzabschnitt im 5'-Bereich der *Fto* cDNA (kodiert von Exon zwei und drei des *Fto* Gens) fast vollkommen identisch mit der entgegengesetzt komplementären Sequenz eines alternativen 5'-UTRs der *Somatostatinrezeptor2 (Sst2)* cDNA ist. Im Gegensatz zu *Fto* wurde *Sst2* auf dem Chromosom MMU11 kartiert, so dass eine Überlappung der beiden Gene ausgeschlossen werden konnte (Brinkmeier & Camper, 1997). Umfangreiche RT-PCR- und Northern-Blot-Analysen bewiesen, dass diese identischen Sequenzen einzig und allein Bestandteil des *Fto* Gens sind. Demnach beruht die zuvor beschriebene *Sst2* cDNA sehr wahrscheinlich auf der Sequenzanalyse eines cDNA-Hybrids (Vanetti *et al.*, 1992).

Die Klonierung eines orthologen menschlichen *FTO* Gens bestätigte die durchgeführten Sequenzanalysen. Weitere orthologe Gene wurden auch in der Ratte, im Rind, im Krallenfrosch und im Pufferfisch nachgewiesen. Im Gegensatz dazu wurden keine orthologen

Gene außerhalb des Unterstammes der Wirbeltiere beispielsweise im Genom von *S. cerevisiae*, *D. melanogaster* und *C. elegans* identifiziert. Sehr wahrscheinlich handelt es sich daher auch bei *Fto* um ein wirbeltierspezifisches Gen.

Die Kartierung der orthologen *Fto* Gene im Genom des Menschen und der Maus zeigte, dass die gesamte Genstruktur zwischen beiden Spezies vollständig erhalten geblieben ist. Obwohl die *Fto* cDNA nur eine Gesamtlänge von 3,6kb (Maus) bzw. 4,1kb (Mensch) aufweist, nimmt die Organisation des jeweiligen Gens in beiden Genomen einen Bereich von mehreren hundert kb (Mensch:410kb, Maus: mehr als 350kb) und damit fast ein Viertel des von der *Ft*-Mutation betroffenen Genomsegments in Anspruch. Dieser Umstand führte auch zur Namensgebung des Gens.

***Fto* wird in den betroffenen Geweben und während der Embryogenese exprimiert**

In der ausgewachsenen Maus wird *Fto* weitestgehend in allen Organen exprimiert. Ebenso ließ sich die Expression von *Fto* während der Embryonalentwicklung in allen untersuchten Stadien (Tage 8.5 bis 13.5 p.c.) nachweisen. Aufgrund der Expression in undifferenzierten embryonalen Stammzellen ist anzunehmen, dass *Fto* sogar schon sehr früh während der Embryogenese exprimiert wird. Die Verwendung verschiedener RNA-Sonden in der *in situ* Hybridisierung von Ganzkörperpräparaten ließ keine spezifische räumliche Verteilung von *Fto* Transkripten im Embryo erkennen, sondern erweckte vielmehr den Eindruck einer nahezu ubiquitären Expression. Das beobachtete Expressionsprofil trifft nicht nur für die Maus zu. Wie Expressionsanalysen am *Kazusa DNA Research Institute* belegen, lässt sich auch für das menschliche *FTO* Gen (*KIAA1752*) eine starke Expression in allen untersuchten Organen sowie in fötalen Geweben feststellen.

Insgesamt zeigt das erstellte Expressionsprofil, dass *Fto* in der ausgewachsenen Maus und während der Embryogenese genau zum richtigen Zeitpunkt in den betroffenen Geweben nämlich vor und während der Ausprägung der verschiedenen Missbildungen exprimiert wird. Obwohl bisher keine asymmetrische Expression von *Fto* festgestellt werden konnte, besteht dennoch die Möglichkeit, dass *Fto* auch an der Etablierung der Rechts-Links-Asymmetrie der Körperachse beteiligt sein kann. Beispielsweise bewies die gezielte Inaktivierung des *Sil* Gens in der Maus, dass auch der Funktionsverlust von ubiquitär exprimierten Genen axiale Defekte in der embryonalen Mittellinie und damit das Auftreten einer willkürlich festgelegten Rechts-Links-Asymmetrie der Körperachse zur Folge haben kann (Izraeli *et al.*, 1999). Zusammenfassend kann *Fto* demnach für das Auftreten aller beobachteten Missbildungen verantwortlich sein und kommt somit als Kandidatengen in Betracht (Tab. 4.1).

Kandidatengen	(vermeintlich) kodiertes Protein	identifiziert durch	Hinweise für eine Beteiligung an (bezogen auf den <i>Ft</i> -Phänotyp)				
			Musterbildung und Entwicklung der Extremitäten	Kontrolle des programmierten Zelltods	Entwicklung des Gesichtsbereichs	Etablierung der Rechts-Links-Asymmetrie der Körperachse	Spezifizierung des ventralen Neuralrohrs
<i>KIAA0308/2FLJ</i>	Chromatin-Remodellierungsfaktor	vergleichende Genomanalyse					
<i>Rbl2 (p130)</i>	Zellzyklusregulator	Transgenintegrationsstelle (Lesche <i>et al.</i> , 1997)					
<i>Fts (Ft1)</i>	Zellkernprotein	Transgenintegrationsstelle (Lesche <i>et al.</i> , 1997)					
<i>Ftm</i>	unbekannt	positionelle Klonierung					
<i>Fto</i>	Zellkernprotein	positionelle Klonierung					
<i>Irx3</i>	Transkriptionsfaktor	vergleichende Genomanalyse					
<i>Irx5</i>	Transkriptionsfaktor	phylogenetische Analyse					
<i>Irx6</i>	Transkriptionsfaktor	phylogenetische Analyse					
<i>Mmp2</i>	Matrix Metalloproteinase	vergleichende Genomanalyse					

Tab. 4.1: Beurteilung der Kandidatengene im Bereich des *Ft*-Locus. Die Tabelle gibt eine Einschätzung der Beteiligung aller identifizierter Kandidatengene an den in *Ft*-Mäusen gestört ablaufenden entwicklungsbiologischen Prozessen bzw. an der Ausprägung der auftretenden Fehlbildungen wieder. Die Beurteilung basiert auf der Zusammenfassung der in dieser Arbeit und anderen relevanten Studien gewonnenen Erkenntnisse (z.B. Proteinaufbau (funktionelle Domänen), Expressionsmuster oder bereits in verschiedenen Spezies durchgeführte Funktionsanalysen). Grün: sehr guter Kandidat; Gelb: eine Beteiligung kann weder befürwortet noch ausgeschlossen werden; Gelb/Rot: schlechter Kandidat, höchst wahrscheinlich nicht beteiligt; Rot: ein direkter kausaler Zusammenhang ist aufgrund von bereits vorliegenden funktionellen Studien nicht erkennbar, allerdings kann eine Beteiligung in Kombination mit dem Funktionsverlust anderer Kandidatengene nicht ausgeschlossen werden.

4.1.3 Die *Iroquois* Gene *Irx3*, *Irx5* und *Irx6*

Die distale Region des Mauschromosoms MMU8 ist einschließlich des *Fused toes* Locus homolog zu einem ausgedehnten Genomsegment auf dem langen Arm des menschlichen Chromosoms HSA16 (16q12-16q24). Zudem zeigten erste umfangreiche Kopplungsanalysen, dass die Anordnung der in einer anfänglichen Studie untersuchten Gene auf beiden Chromosomen im Verlauf der Evolution fast vollständig erhalten geblieben ist (Becker-Follmann *et al.*, 1997). Über die Auswertung der verfügbaren genetischen und physikalischen Karten des relevanten Genomabschnitts auf HSA16q konnte mit dem *Iroquois* verwandten Homöobox Gen *Irx3* (Bosse *et al.*, 1997) ein weiteres Gen im verlorengegangenen Bereich des *Ft*-Locus kartiert werden. Zudem lieferte die Analyse der genomischen DNA von homozygoten *Ft*-Embryonen einen klaren Hinweis, dass neben *Irx3* zwei weitere Mitglieder der *Iroquois* Genfamilie von der *Ft*-Mutation betroffen sind. Diese beiden zuvor unbekannt Genen *Irx5* und *Irx6* wurden in der vorliegenden Arbeit identifiziert und der proteinkodierende Bereich der jeweiligen cDNA vollständig kloniert. Des Weiteren zeigten die durchgeführten Analysen, dass sowohl die Maus als auch der Mensch über insgesamt sechs *Irx* Gene verfügen. Diese sechs Gene sind in zwei Gruppen von jeweils drei Genen auf verschiedenen Chromosomen angeordnet (*Irx1*, *Irx2* und *Irx4*: MMU13, HSA5; *Irx3*, *Irx5* und *Irx6*: MMU8, HSA16).

Die ersten Mitglieder der *Iroquois* Genfamilie wurden infolge einer Genommutagenese in *Drosophila melanogaster* identifiziert (Leyns *et al.*, 1996). Dort führten Mutationen im Bereich des *Iroquois*-Komplexes (*Iro-C*) mit den drei eng gekoppelten und verwandten Genen *araucan* (*ara*), *caupolican* (*caup*) und *mirror* (*mirr*) unter anderem zu einem Verlust der sensorischen Borsten in den lateralen Regionen des dorsalen Mesothorax, so dass in den betroffenen Fliegen nur ein Band von sensorischen Borsten in der zentralen Region des Notums ausgebildet wurde (Leyns *et al.*, 1996; Gómez-Skarmeta *et al.*, 1996; McNeill *et al.*, 1997; Grillenzoni *et al.*, 1998; Kehl *et al.*, 1998). Dieser Phänotyp erinnerte stark an die Haartracht der Irokesen (Iroquois), einem Indianerstamm in Nordamerika und war somit ausschlaggebend für die Namensgebung des identifizierten Locus. Weitere Mitglieder dieser Genfamilie wurden bisher auch in einigen anderen Wirbeltierarten, wie z.B. dem Krallenfrosch (*Xenopus laevis*), dem Zebrafisch (*Danio rerio*), dem Huhn (*Gallus gallus*), und daneben auch im Nematoden *C. elegans* beschrieben (Bürglin, 1997; Gómez-Skarmeta *et al.*, 1998; Bellefroid *et al.*, 1998; Bao *et al.*, 1999; Funayama *et al.*, 1999; Goriely *et al.*, 1999; Tan *et al.*, 1999; Wang *et al.*, 2001; Ogura *et al.*, 2001; Garriock *et al.*, 2001; Cheng *et al.*, 2001). Dabei konnten mit Ausnahme von *Ziro7*, einem erst kürzlich identifizierten divergenten *Irx* Gen (Lecaudey *et al.*, 2001), alle beschriebenen *Irx* Gene den sechs in dieser

Arbeit aufgestellten orthologen Gruppen zugeordnet werden. Irx Proteine zeigen alle ein und denselben strukturellen Aufbau. Dabei folgt auf die DNA-bindende Homöodomäne des TALE-Typus (*'three amino acid loop extension'*) mit der Iro-box ein kurzes für die *Iroquois* Familie charakteristisches Proteinmotiv. Die Aminosäurezusammensetzung der Iro-box ähnelt dem zentralen Teil des z.B. im Notch-Rezeptor Protein wiederholt vorkommenden EGF-Motivs (EGF - *'epidermal growth factor-like'*). Daher wird vermutet, dass die Iro-box in Analogie zum EGF-Motiv ebenfalls der Ausbildung von Protein-Protein-Interaktionen dienen könnte (Bürglin, 1997).

Irx Gene erfüllen wichtige entwicklungsbiologische Aufgaben

Die Funktionen der einzelnen *Iroquois* Gene der Wirbeltiere wurden bisher nur ansatzweise beschrieben (Bao *et al.*, 1999; Bruneau *et al.*, 2001; Gómez-Skarmeta *et al.*, 2001; Kudoh & David, 2001; Zülch, 2001). Allerdings sprechen sowohl die bisherigen funktionellen Studien in *Drosophila* als auch die Expressionsprofile der *Irx* Gene der Wirbeltiere für einen wesentlichen Beitrag dieser Gene bei der Abgrenzung bzw. Definition bestimmter Gewebe im frühen Embryo (z.B. das neurale Ektoderm (*Xenopus*), Regionen des dorsalen Mesothorax (*Drosophila*)). Diese Gewebe werden später nochmals unter Einwirkung der *Irx* Gene in kleinere Territorien unterteilt (z.B. das Neuralrohr und Herz der Wirbeltiere bzw. die Flügelvenen oder auch die sensorische Borsten des Notums von *Drosophila*; Übersicht bei Cavodeassi *et al.*, 2001). Die vermutliche Wirkungsweise wird anhand der Spezifizierung der dorsoventralen Achse der Augen in *Drosophila* im folgenden Abschnitt verdeutlicht.

Irx Gene und der Notch-Signaltransduktionsweg

Zu Beginn der Augenentwicklung werden die drei *Iro-C* Gene *ara*, *caup* und *mirr* nur in der dorsalen Hälfte der Augen-Imaginalscheibe exprimiert (McNeill *et al.*, 1997; Dominguez & de Celis, 1998; Cavodeassi *et al.*, 1999). Durch die Aktivität von *Iro-C* wird dabei die Expression des Gens *fringe* (*fng*) in den dorsalen Zellen unterdrückt, so dass *fng* nur in der ventralen Hälfte der Augen-Imaginalscheibe exprimiert werden kann (Cho & Choi, 1998; Dominguez & de Celis, 1998; Papayannopoulos *et al.*, 1998; Cavodeassi *et al.*, 1999). Auf diese Weise entsteht eine scharfe Grenzfläche zwischen den dorsalen *Iro-C⁺/fng⁻* exprimierenden und den ventralen *Iro-C/fng⁺* exprimierenden Zellen, entlang derer sich aufgrund der von *fng* modulierten Aktivierung des Transmembranrezeptors Notch ein dorsoventraler Organisator ausbildet (Dominguez & de Celis, 1998; Cho & Choi, 1998). Die

ursprüngliche Position dieses Organisators ist später im ausgewachsenen Auge mit dem sogenannten Äquator gut zu erkennen. Der Äquator unterteilt das Auge in eine dorsale und eine ventrale Hälfte und entspricht damit einer dorsoventralen Symmetrieachse, die sich in einer spiegelbildlichen Orientierung der Omatiden der beiden Augenhälften äußert. Implantiert man nun Zellklone mit einer defizitären *Iro-C* Aktivität in der dorsalen Augenhälfte, wird aufgrund des Entstehens einer zusätzlichen Grenze zwischen *Iro-C⁺/fng⁻* und *Iro-C/fng⁺* Zellen ein zusätzlicher Äquator oder sogar ein komplettes ektopisches Auge induziert (McNeill *et al.*, 1997; Cho & Choi, 1998; Dominguez & de Celis, 1998). In einem komplementären Experiment führte die universelle Expression von *Iro-C* in der Augen-Imaginalscheibe zu einem vorzeitigen Abbruch der Augenentwicklung, da Notch durch die nun in der gesamten Imaginalscheibe unterdrückten Expression von *fng* nicht aktiviert werden konnte (Cho & Choi, 1998; Dominguez & de Celis, 1998). Die Aktivität von Notch, moduliert durch die von *Iro-C* gesteuerte Ausbildung einer Grenzfläche zwischen *fng⁻* und *fng⁺* Zellen, spielt daher eine wesentliche Rolle bei der Musterbildung und Entwicklung des Auges.

Der Notch-Signaltransduktionsweg ist ein evolutionär hoch konservierter Mechanismus, der das Entwicklungsschicksal einzelner Zellen durch lokale Zellinteraktionen bestimmt (Übersicht bei Artavanis-Tsakonas *et al.*, 1999). Umfangreiche Studien bewiesen, dass dieser Signalmechanismus eine außergewöhnlich große Anzahl von Entwicklungsvorgängen wie z.B. Zellproliferation, Migration, Wachstum, Differenzierung und den programmierten Zelltod kontrolliert (Übersicht bei Gridley, 1997). Daher ist es nicht weiter verwunderlich, dass Störungen dieses Signalweges zu einem breiten Spektrum von Fehlbildungen während der Embryonalentwicklung führen. Einige der dabei beobachteten morphologischen Veränderungen zeigen Parallelen zum *Ft*-Phänotyp. Beispielsweise bewirkte die gezielte Inaktivierung von *Jagged2*, einem Liganden für die Notch-Rezeptorfamilie der Wirbeltiere, Missbildungen im Gesichtsbereich (z.B. Gaumenspalte), eine veränderte Morphologie des Thymus sowie die Ausbildung von Syndaktylien an den Vorder- und Hinterpfoten der betroffenen homozygoten Mäuse (Jiang *et al.*, 1998). Gleichermäßen ist das Auftreten einer dorsoventralen Verdickung der Extremitätenknospe und die damit verbundene Ausprägung einer Syndaktylie in der spontanen Mausmutante *syndactylism* auf der Mutation von *Serrate2*, ebenfalls ein Notch-Ligand, begründet (Sidow *et al.*, 1997). Ebenso interessant ist, dass die gemeinsame Inaktivierung der beiden Presenilin Gene *PS1* und *PS2*, die für die proteolytische Spaltung von Notch und damit für die Freigabe eines aktiven intrazellulären Notch-Fragments verantwortlich sind, neben anderen Defekten eine Desorganisation des ventralen Neuralrohrs hervorruft. Allerdings ist dieser Phänotyp genau wie bei *Ft* nur auf den Rumpf beschränkt (Donoviel *et al.*, 1999).

***Irx3* und die Spezifizierung des ventralen Neuralrohrs**

Die Rolle von *Irx3* bei der Spezifizierung von einzelnen Vorläuferregionen im ventralen Neuralrohr konnte bereits durch erste funktionelle Studien im Hühnchen beschrieben werden. Diese Analysen zeigten, dass *Irx3* zusammen mit anderen Homöobox Genen in Abhängigkeit eines von der *Chorda dorsalis* und der Bodenplatte ausgehenden *Shh* Gradienten entlang der dorsoventralen Achse des Neuralrohrs exprimiert wird (Briscoe *et al.*, 2000). Auf diese Weise werden über die teilweise überlappenden Expressionsdomänen der einzelnen Homöobox Gene fünf spezifische ventrale Vorläuferregionen des Neuralrohrs definiert, aus denen später fünf unterschiedliche Arten von Neuronen hervorgehen (von dorsal nach ventral: V0-, V1-, V2-Interneuronen, Motorneuronen, V3-Interneuronen, Bodenplatte). Eine überlappende Expression von *Irx3* und *Nkx6.1* scheint dabei für die Spezifizierung der V2 Interneuronen verantwortlich zu sein, während die alleinige Expression von *Nkx6.1* zu einer Differenzierung in Richtung Motorneuronen führt (Übersicht bei Briscoe & Ericson, 2001). Wird *Irx3* in dieser Vorläuferregion der Motorneuronen ektopisch exprimiert, nehmen die Zellen in diesem Bereich aufgrund der nun überlappenden Expression von *Irx3* und *Nkx6.1* die Identität von V2-Interneuronen an (Briscoe *et al.*, 2000). Demnach sollte der Verlust der *Irx3* Expression in der Vorläuferregion der V2-Interneuronen durch die verbleibende Expression von *Nkx6.1* eigentlich zu einer Transformation der Zellidentität in Richtung Motorneuronen führen. Dieser Effekt lässt sich im Neuralrohr von homozygoten *Ft*-Embryonen nicht feststellen (Dr. J. Briscoe, persönliche Mitteilung). Vielmehr scheinen in einigen Abschnitten der kaudalen Körperregion die noch weiter ventral gelegenen V3-Interneuronen die Identität von Motorneuronen anzunehmen, während die Spezifizierung der V2-Interneuronen ganz normal abläuft (Abb. 1.2). Allerdings ist dieser Effekt sehr wahrscheinlich auf das Fehlen der *Shh* Expression in der Bodenplatte zurückzuführen. Da die Abwesenheit von *Irx3* im ventralen Neuralrohr nicht zu den erwarteten Konsequenzen führt, muss die Rolle von *Irx3* bei der Spezifizierung der Vorläuferregionen im ventralen Neuralrohr noch einmal überdacht und durch weitere Experimente hinterfragt werden.

***Irx* Gene und der Bmp- bzw. Activin-Signaltransduktionsweg**

Über welchen Mechanismus *Irx* Proteine ihre Funktionen umsetzen, liegt bisher noch nahezu im Verborgenen. Erste Hinweise lieferte kürzlich eine Studie über die Rolle von *Xiro1*, ein Ortholog des *Irx1* Gens der Maus, bei der Spezifizierung der Neuralplatte in *Xenopus laevis* (Gómez-Skarmeta *et al.*, 2001). Dabei zeigten die durchgeführten Analysen, dass während dieses Prozesses die Expression von *Xiro1* zunächst durch den Wnt-Signaltransduktionsweg

im dorsalen Ektoderm, dem Bereich der zukünftigen Neuralplatte, aktiviert wird. Anschließend kommt es aufgrund einer wechselseitigen Inhibition zwischen *Bmp4* und *Xiro1* zu einer Unterdrückung der zuvor universellen *Bmp4* Expression im dorsalen Ektoderm des Embryos und somit zur Definition der zukünftigen Neuralplatte. Anscheinend bindet *Xiro1* dabei an den *Bmp4* Promotor und agiert somit als direkter transkriptioneller Repressor von *Bmp4* (Gómez-Skarmeta *et al.*, 2001). Im Gegensatz dazu scheinen jedoch andere *Irx* Proteine (*Irx4*) eine Funktion als transkriptioneller Aktivator zu erfüllen (Bao *et al.*, 1999).

Aus diesen Studien lässt sich ableiten, dass der Funktionsverlust eines oder mehrerer *Irx* Gene, wie es bei der *Ft*-Mutation der Fall ist, zu Störungen des Bmp-Signaltransduktionsweg führen könnte. Bmp Proteine, insbesondere *Bmp4*, sind fast an allen entwicklungsbiologischen Prozessen beteiligt, die bei der Ausbildung der in den *Ft*-Tieren missgebildeten Gewebe eine maßgebliche Rolle spielen. Beispielsweise ist aus mehreren Studien ersichtlich, dass über einen Bmp/anti-Nodal Signaltransduktionsweg in der rechten Körperhälfte sowie einen Nodal/anti-Bmp Signaltransduktionsweg in der linken Körperhälfte die Rechts-Links-Asymmetrie der Körperachse definiert wird (Übersicht bei Yost, 2001). Ebenso konnte kürzlich gezeigt werden, dass *Bmp4* in der rechten Hälfte des Primitivknotens des Embryos exprimiert wird und dort die Expression von *Shh* inhibiert. Auf diese Weise wird vermutlich über *Shh* die asymmetrische Expression von Genen (z.B. Nodal) in der linken Körperhälfte gewährleistet (Monsoro-Burq & Le Douarin, 2001). Unbestritten ist die Bedeutung der Bmp-Signalkaskade für die Steuerung der Ausbildung von Knorpel- und Knochenstrukturen in den Extremitäten und auch im Gesichtsbereich (Satokata & Maas, 1994; Hogan, 1996). Beispielsweise nimmt *Bmp4* eine Schlüsselposition bei der Regulation des programmierten Zelltods im Zwischenfingerewebe der sich entwickelnden Gliedmaßen ein (Merino *et al.*, 1999a; Tang *et al.*, 2000). Demnach werden die interzellulären Signale für die Induktion des programmierten Zelltods in diesem Bereich vor allem über den Bmp-Signaltransduktionsweg vermittelt (Chen & Zhao, 1998, Merino *et al.*, 1998; Merino *et al.*, 1999a). Eine Störung dieses Signalweges konnte durch die veränderte Expression von *Bmp4*, *Msx2* und *Noggin* in den Extremitäten von heterozygoten *Ft*-Embryonen bereits nachgewiesen werden (Heymer & Rütter, 1999). Zukünftige Experimente werden zeigen, ob ein Zusammenspiel von *Irx* und *Bmp* Genen ausschlaggebend für die in den *Ft*-Tieren beobachteten Missbildungen ist.

Dass *Irx* Gene in wichtige, entwicklungsbiologisch relevante Signaltransduktionswege eingebunden sind, wird auch durch eine weitere Studie bekräftigt. Dort konnte gezeigt werden, dass die Expression von *Irx1* teilweise über die Activin-Signalkaskade kontrolliert wird (Ferguson *et al.*, 2001). Wie die Bmp Proteine sind auch die Activine Mitglieder der Tgf β -Superfamilie. Neben vielen anderen morphoregulatorischen Prozessen wurde den Activinen bereits eine maßgebliche Beteiligung an der Etablierung der Rechts-Links-Asymmetrie der Körperachse, an der Spezifizierung der Fingerstrahlen sowie der Ausbildung

der Skelettelemente, Gelenke und Sehnen der Extremitäten und auch an der Entwicklung kraniofazialer Gewebe zugeschrieben (Nomura & Li, 1998; Merino *et al.*, 1999b; Ferguson *et al.*, 2001, Yost, 2001). Der Ablauf dieser Prozesse ist, wie bereits an anderer Stelle erwähnt, in den *Ft*-Mäusen erheblich gestört. Diese Fehlentwicklungen könnten somit durchaus auf einer Beeinträchtigung dieses Signaltransduktionsweges beruhen, die über den Funktionsverlust der betroffenen *Irx* Gene vermittelt wird (Tab. 4.1).

***Irx3*, *Irx5* und *Irx6* werden in den betroffenen Geweben exprimiert**

Nicht nur die erwähnten Zusammenhänge mit entwicklungsbiologisch relevanten Signalkaskaden lassen die *Irx* Gene als gute Kandidaten für den *Ft*-Phänotyp erscheinen. Ebenso zeigen die bisherigen Expressionsstudien, dass vor allem *Irx3* und *Irx5* in den betroffenen Geweben zum Zeitpunkt des Auftretens der morphologischen Veränderungen exprimiert werden. Interessant ist dabei, dass *Irx3* und *Irx5* oft ein übereinstimmendes Expressionsprofil zeigen, während sich das Expressionsmuster von *Irx6* deutlich davon unterscheidet (Houweling *et al.*, 2001). Im Gegensatz zu *Irx3* und *Irx5* tritt die Expression von *Irx6* häufig erst später und dann auch nur auf einem niedrigen Niveau in Erscheinung (Mummenhoff *et al.*, 2001; Houweling *et al.*, 2001). Im Neuralrohr werden *Irx3* und *Irx5* in der ventrikulären Matrixzone der Flügel- und Grundplatte, nicht aber in der Boden- und Deckplatte exprimiert. Ebenfalls konnte eine Expression beider Gene in der Vorläuferregion der Motorneuronen sowie in der *Chorda dorsalis* nachgewiesen werden (Houweling *et al.*, 2001). Im Vergleich dazu war die Expression von *Irx6* nur in den Motorneuronen erkennbar (Mummenhoff *et al.*, 2001). In den sich entwickelnden Gliedmaßen werden *Irx3*, *Irx5* und *Irx6* zunächst gemeinsam in einer proximoanterioren Domäne exprimiert. Interessanterweise verändert sich dieses Expressionsmuster im Verlauf der Gliedmaßenentwicklung, so dass die Expression aller drei Gene später nur noch auf das Zwischenfingerewebe beschränkt ist (Houweling *et al.*, 2001). Auch im späteren Gesichtsbereich zeigen *Irx3* und *Irx5* vor allem in der Region der Augen und der Kiemenbögen eine ausgeprägte Expression (Bosse *et al.*, 1997; Bosse *et al.*, 2000). Weitere Expressionsorte von *Irx3*, *Irx5* und *Irx6* sind im Gehirn, im Herzen, der Lunge, der Haut, den Zähnen und in den Augen zu finden (Bosse *et al.*, 1997; Bosse *et al.*, 2000; Mummenhoff *et al.*, 2001; Houweling *et al.*, 2001).

Da sich die Aminosäurezusammensetzung der DNA-bindenden Homöodomäne zwischen den einzelnen *Iroquois* Proteinen nicht sonderlich unterscheidet, ist durchaus anzunehmen, dass *Iroquois* Proteine ähnliche oder sogar identische DNA-Motive erkennen. Es ist daher möglich, dass die verschiedenen *Iroquois* Proteine ein ähnliches oder sogar dasselbe genetische Programm steuern und sich daher teilweise funktionell ersetzen können

(funktionelle Redundanz). Diese Hypothese wird auch durch die ähnlichen Expressionsmuster einiger *Irx* Gene (z.B. *Irx3* und *Irx5*; Houweling *et al.*, 2001) nochmals bekräftigt. In der Tat konnte die funktionelle Redundanz von *Iroquois* Genen bereits in einigen Studien in *Drosophila* nachgewiesen werden (Diez del Corral *et al.*, 1999). Die gezielte Inaktivierung individueller oder mehrerer *Irx* Gene in der Maus wird zeigen, ob diese Situation auch auf die *Iroquois* Genfamilie der Wirbeltiere zutrifft.

Alles in allem machen die vorliegenden Expressionsdaten und auch die vermeintlichen funktionellen Zusammenhänge von *Irx* Genen mit bedeutenden Signaltransduktionswegen insbesondere *Irx3* und *Irx5* zu sehr guten Kandidatengenen für den *Ft*-Phänotyp (Tab 4.1). Dagegen scheint *Irx6* aufgrund der erst spät auftretenden Expression häufig erst nach Erscheinen der morphologischen Veränderungen eine eher untergeordnete Rolle bei der Ausprägung des *Ft*-Phänotyps zu spielen (Tab. 4.1).

4.2 Die Kandidatengene in den flankierenden Genomabschnitten

Basierend auf der Zusammenstellung und STS-Kartierung eines ausgedehnten YAC-Contigs wurden mit *KIAA0308/2FLJ* und *Mmp2* zwei weitere vermeintliche Kandidatengene identifiziert. Beide Gene wurden in der unmittelbaren Nähe des in der *Ft*-Mutation verlorengegangenen Genomsegments kartiert. *KIAA0308/2FLJ* und *Mmp2* können daher aufgrund einer Beeinträchtigung von wichtigen Steuerungselementen (Promotor, Enhancer oder Silencer), die unter Umständen im verlorengegangenen Genomsegment lokalisiert sind, ebenfalls an der Ausprägung des *Ft*-Phänotyps beteiligt sein. Im Gegensatz dazu trägt *Sall1*, das aufgrund einer teilweisen Ähnlichkeit des *Ft*-Phänotyps mit der klinischen Präsentation des Townes-Brocks-Syndroms untersucht wurde, anscheinend nicht zum Auftreten der Fehlbildungen in den *Ft*-Mäusen bei.

4.2.1 *KIAA0308/2FLJ*

Mit *KIAA0308/2FLJ* wurde in dieser Arbeit ein bisher nicht näher beschriebenes Gen im proximalen Bereich des erstellten YAC-Contigs kartiert. Aufgrund der Tatsache, dass *KIAA0308/2FLJ* zusammen mit *Rbl2* und *Fts* auf dem BAC-Klon MB1 lokalisiert ist, beträgt der Abstand dieses Gens zum proximalen Deletionsendpunkt sehr wahrscheinlich nicht mehr als 150-200kb (Abb. 3.18). Im menschlichen Genom konnte eine Distanz von 165kb zwischen *KIAA0308/2FLJ* und *Fts* ermittelt werden (Abb. 3.23; Venter *et al.*, 2001).

KIAA0308/2FLJ zeigt eine Ähnlichkeit mit Chromatin-Remodellierungsfaktoren

Die *in silico* zusammengesetzte menschliche *KIAA0308/2FLJ* cDNA kodiert für ein Protein, das aus 1995 Aminosäureresten besteht. Die abgeleitete Aminosäuresequenz und die darin enthaltenen funktionellen Domänen lassen eine Verwandtschaft von KIAA0308/2FLJ mit dem *Drosophila*-Protein *kismet* (*kis*) erkennen, für das bereits eine Funktion als Chromatin-Remodellierungsfaktor vorgeschlagen wurde (Daubresse *et al.*, 1999, Therrien *et al.*, 2000). Erste funktionelle Analysen in *Drosophila* zeigten, dass der Funktionsverlust von *kis* zu homöotischen Transformationen führt. Ebenso wird die Funktion von *kis* für die ordnungsgemäße Segmentierung der *Drosophila* Larve benötigt (Daubresse *et al.*, 1999; Therrien *et al.*, 2000).

Aufgrund der erwähnten Verwandtschaft mit *kis* ist es plausibel anzunehmen, dass KIAA0308/2FLJ ebenfalls als Chromatin-Remodellierungsfaktor agiert. Diese Proteine nutzen mit Hilfe ihrer ATPase Untereinheit die Energie der ATP-Hydrolyse für die Erzeugung superhelikaler Verdrehungen der DNA. Auf diese Weise verändern diese Faktoren die Topologie der DNA und unterbrechen Interaktionen zwischen der DNA und Histonen. Die Chromatinstruktur bestimmt, ob ein Gen zugänglich für die Transkriptionsmaschinerie ist und damit aktiviert werden kann oder nicht (Wolffe & Hayes, 1999; Flaus & Owen-Hughes, 2001). Während der Embryogenese und Differenzierung werden in einer Zelle Gene nach einem zeitlich genau festgelegten Muster exprimiert. Dieses Expressionsprofil wird an die Tochterzellen weitergegeben, so dass dieselben Gene auch in Abwesenheit eines induktiven Signals weiterhin exprimiert werden. Chromatin-Remodellierungsfaktoren sind Bestandteil dieses epigenetischen Regulationsmechanismus und erhalten bzw. unterdrücken somit die Expression bestimmter Gene in den Abkömmlingen einer Zelle (Übersicht bei Müller & Leutz, 2001). Allerdings sind diese Faktoren auch an einigen anderen Prozessen wie z.B. der homologen Rekombination, der DNA-Reparatur, der Trennung der Schwesterchromatiden während der Mitose oder auch an der Histon-Deacetylierung und der Aufrechterhaltung des jeweiligen DNA-Methylierungsstatus beteiligt (Übersicht bei Fyodorov & Kadonaga, 2001).

Umfangreiche Studien belegen, dass Chromatin-Remodellierungsfaktoren bedeutende entwicklungsbiologische Aufgaben erfüllen (Müller & Leutz, 2001). In *Drosophila* sind diese Faktoren beispielsweise für die Aufrechterhaltung der spezifischen Expressionsmuster der homöotischen Gene (*Hox* Gene der Wirbeltiere) zuständig (Übersicht bei Flaus & Owen-Hughes, 2001). Auch in den Wirbeltieren konnte bereits für die mit *kis* und KIAA0308/2FLJ verwandten Gene *Brg1* (*SNF2 β*) und *Brahma* (*SNF2 α*) wichtige funktionelle Aufgaben während der Embryogenese nachgewiesen werden (Reyes *et al.*, 1998; Bultman *et al.*, 2000).

Die räumliche und zeitliche Expression von KIAA0308/2FLJ in der ausgewachsenen Maus bzw. im Embryo wurde bisher nicht näher untersucht. Auch die Datenbank des *Kazusa DNA*

Research Institute enthält nur einige wenige Informationen über die Expression von *KIAA0308/2FLJ* in menschlichen Geweben (Niere, Leber, Lunge und z.T. auch Herz und Gehirn). Zum gegenwärtigen Zeitpunkt lässt sich daher die Rolle von *KIAA0308/2FLJ* bei der Ausprägung des *Ft*-Phänotyps nur sehr schwer beurteilen. Allerdings zeigte eine erste in dieser Arbeit durchgeführte Expressionsanalyse, dass die generelle Expression von *KIAA0308/2FLJ* durch die *Ft*-Mutation anscheinend nicht beeinträchtigt wird. Obwohl *KIAA0308/2FLJ* in etwa 150-200kb vom deletierten Genomabschnitt entfernt ist, kann eine veränderte Expression dieses Gens durchaus für die Ausprägung der beobachteten Missbildungen verantwortlich sein (Tab. 4.1). Zukünftige Expressionsanalysen werden sicherlich den Verdacht einer Beteiligung von *KIAA0308/2FLJ* verstärken oder aber entkräften.

4.2.2 *Mmp2*

Im Gegensatz zu *KIAA0308/2FLJ* wurde das zuvor bereits bekannte Gen *Mmp2* im distalen Bereich des zusammengestellten YAC-Contigs kartiert (Abb. 3.18). Allerdings lässt sich die Entfernung zum deletierten Genomabschnitt anhand der erstellten physikalischen Genkarte des *Ft*-Locus nur sehr ungenau ermitteln. Im Menschen beträgt der Abstand von *MMP2* zu *IRX6* ca. 150kb (Abb. 3.23; Venter *et al.*, 2001). Überträgt man diese Distanz auch auf das Mausgenom, dann sollte *Mmp2* ungefähr 50-100kb vom distalen Deletionsendpunkt entfernt liegen.

Matrix Metalloproteinasen regulieren die Zusammensetzung der extrazellulären Matrix

Matrix Metalloproteinasen (Mmps) bilden eine große Familie von hauptsächlich sekretierten aber auch membrangebundenen zink- und kalzium-abhängigen Endopeptidasen (Sternlicht & Werb, 2001). Gemeinschaftlich sind die Mitglieder der Mmp-Familie dazu befähigt, sämtliche Bestandteile der extrazellulären Matrix (EZM) abzubauen. Dabei zeichnet sich jedes dieser Enzyme durch seine individuelle Substratspezifität aus (Yu *et al.*, 1999). Die EZM besteht vorrangig aus Kollagen, Ankerproteinen (z.B. Fibronectin, Elastin, Laminin) und raumfüllenden Proteoglykanen (z.B. Hyaluronsäure). Eine je nach Entwicklungsstadium notwendige Umwandlung der EZM bedarf einer genauen Regulation, da eine unkontrollierte Proteolyse zu einer abnormalen Entwicklung und somit zur Ausbildung pathologischer Phänotypen (z.B. Entzündungen, Arthritis, Krebs) führen kann (Chambers & Matrisian, 1997; Yu *et al.*, 1997; Chandler *et al.*, 1997; Yong *et al.*, 1998). *Mmp2* (Synonyme: Gelatinase-A,

72kDa Typ IV Kollagenase) ist dabei vor allem für den Abbau von Kollagen (Typ IV, V, VII, X und XI), Gelatine, Elastin, Laminin, Fibronectin, Aggrekan, Fibrillin, Vitronectin, Tenascin und auch für die Aktivierung des Wachstumsfaktors Tg β zuständig (Yu *et al.*, 1999). Tg β wird von den produzierenden Zellen zunächst in die umgebende EZM sekretiert, wo er aufgrund einer proteolytischen Spaltung durch Mmp2 aktiviert wird. Der Tg β -Signaltransduktionsweg spielt beispielsweise eine wichtige Rolle für die Ausbildung von Knochen und Knorpeln und auch bei der Entstehung einiger menschlicher Erkrankungen wie z.B. Osteoporose, Atherosklerose oder auch Krebs (Ganan *et al.*, 1996; Blobel *et al.*, 2000; Miyazono *et al.*, 2001).

Die wichtigste Aufgabe der EZM ist die Förderung des Zusammenschlusses individueller Zellen zu komplexen funktionellen Einheiten wie Geweben und Organen. Ebenso erhält die EZM deren Beschaffenheit, Struktur und Integrität. Zudem ist die EZM ausschlaggebend an der Bildung von Knorpeln, Knochen, Gelenken und Sehnen beteiligt und dient auch als Pfad für die Orientierung von wandernden Zellen (DeLise *et al.*, 2000). Ebenso werden über Interaktionen mit der umgebenden EZM viele fundamentale Prozesse einer Zelle wie z.B. Wachstum, Überleben, Differenzierung, Mobilität und auch Signaltransduktionswege reguliert (Werb, 1997; Capdevila & Izpisua Belmonte, 1999; Larraín *et al.*, 2000). Vor allem bei der enchondralen Ossifikation, der Knochenbildung über ein später verknöchernendes Knorpelmodell, konnte die Rolle der EZM bereits sehr gut aufgeklärt werden (Übersicht bei DeLise *et al.*, 2000). Dieser Prozess beginnt zunächst mit der Kondensation von locker gepackten mesenchymalen Zellen (Thorogood & Hinchliffe, 1975). Während der Kondensation ist die Kommunikation zwischen den benachbarten Zellen über Adhäsions- und EZM-Moleküle ausschlaggebend für die zeitliche und räumliche Regulation der Chondrogenese. Dabei fungiert die EZM nicht nur als adhäsive Matrix sondern reguliert auch die Verbreitung bzw. Weiterleitung und auch die Aktivität von Wachstumsfaktoren und anderen Morphogenen (Arteaga-Solis *et al.*, 2001). Später wird die Differenzierung von mesenchymalen Zellen in Chondrozyten von einer stetigen dynamischen Umwandlung der EZM begleitet und auch gesteuert (Übersicht bei DeLise *et al.*, 2000).

Auch bei der Gesichtsentwicklung, insbesondere bei der Bildung der dortigen Knorpel, Knochen, Zähne und der Skelettmuskulatur (Zunge) sind EZM-Interaktionen beteiligt (Hall, 1991; Noden, 1991). Beispielsweise ermöglichen Umwandlungen der EZM das Einwandern von Zellen aus der Neuralleiste. Aus diesen Zellen geht später ein Großteil des Kopfmesenchyms hervor. Daher ist dieser Prozess absolut notwendig für den Aufbau der einzelnen Kiemenbögen, aus denen später die Hauptbestandteile des Gesichts und auch der Ohren hervorgehen (Noden, 1986; LeDouarin & Kalcheim, 1999). Erste Expressionsanalysen lieferten bereits einen guten Hinweis auf die Beteiligung von Mmps, auch von *Mmp2*, an der Entwicklung des Mandibularbogens (1. Kiemenbogen), aus dem später Ober- und Unterkiefer

gebildet werden (Chin & Werb, 1997). Insgesamt wird *Mmp2* während der Embryogenese vor allem in mesenchymalen Geweben exprimiert (Kanwar *et al.*, 1999; Karelina *et al.*, 2000). Später erfolgt die Expression von *Mmp2* insbesondere bei regenerativen Prozessen (Yu *et al.*, 1999).

Der Funktionsverlust von *Mmp2* führt zu Knochenfehlbildungen

Kürzlich konnte in einer Studie gezeigt werden, dass der Funktionsverlust von *MMP2* im Menschen zu einem verstärkten Knochenabbau, Arthritis, charakteristischen Fehlbildungen im Gesichtsbereich und zu subkutanen Knoten im Bindegewebe führen kann (Martignetti *et al.*, 2001). Dagegen scheint der gezielte Funktionsverlust von *Mmp2* in Mäusen bis auf ein verlangsamtes Wachstum keinerlei Entwicklungsdefekte hervorzurufen (Itoh *et al.*, 1997). Sehr wahrscheinlich ist dieser Umstand auf die funktionelle Redundanz mit anderen eng verwandten Mitgliedern dieser Genfamilie zurückzuführen. Dass *Mmp2* auch in der Maus wichtige entwicklungsbiologische Aufgaben erfüllt, beweist die gezielte Inaktivierung von *Mt1-Mmp* (*Mmp14*), einem Aktivator von *Mmp2*. Zusammen mit dem spezifischen Inhibitor *Timp2* (*Timp* - 'tissue inhibitor of metalloproteinases') bilden *Mt1-Mmp* und die unprozessierte und damit inaktive Form von *Mmp2* (pro*Mmp2*) an der Zelloberfläche einen trimeren Proteinkomplex, über den die Freisetzung und somit die Konzentration von prozessierten, aktiven *Mmp2* reguliert wird (Strongin *et al.*, 1995; Butler *et al.*, 1998). Der Funktionsverlust von *Mt1-Mmp* führt in den betroffenen Mäusen zu einer Verringerung des Kollagenabbaus, einer reduzierten Knochenbildung und zu einer Zunahme der Anzahl von Osteoclasten. Diese Defekte verursachen die Ausprägung von Knochen- und Muskelschwund, Arthritis sowie Fehlbildungen im Gesichtsbereich (Holmbeck *et al.*, 1999).

Aufgrund ihrer Regulation der Zusammensetzung der EZM sind *Mmps* an wichtigen entwicklungsbiologischen Prozessen wie z.B. Zellproliferation, Differenzierung und Zelltod beteiligt. Zudem haben Störungen der *Mmp2* Funktion sowohl im Menschen als auch in der Maus eine Unausgewogenheit im geregelten Auf- und Abbau von Knochen und Knorpeln zur Folge. Dementsprechend ist es plausibel, Veränderungen in der *Mmp2* Expression mit den in den *Ft*-Mäusen beobachteten Missbildungen in Zusammenhang zu bringen. Besonders die dysmorphe Entwicklung des Gesichtsbereichs von homozygoten *Ft*-Embryonen und auch die an den Pfoten heterozygoter *Ft*-Mäuse auftretenden Knochenfehlbildungen legen eine Beteiligung von *Mmp2* nahe (Tab. 4.1). Allerdings konnte in Fibroblastenzellen, die zuvor aus homozygoten *Ft*-Embryonen isoliert worden waren, eine Expression von *Mmp2* nachgewiesen werden. Daher wird die generelle Expression von *Mmp2* durch die *Ft*-Mutation nicht beeinträchtigt. Dennoch kann zum gegenwärtigen Zeitpunkt eine ektopische Expression oder

auch der Ausfall der *Mmp2* Expression in einigen bestimmten Geweben nicht ausgeschlossen werden. Eine vermeintlich veränderte Expression könnte in zukünftigen Experimenten beispielsweise durch die *in situ* Hybridisierung der *Mmp2* mRNA in homozygoten *Ft*-Embryonen nachgewiesen werden. Ebenso besteht die Möglichkeit mit Hilfe einer Gelatine-Zymographie die enzymatische Aktivität von *Mmp2* in Fibroblasten von homozygoten *Ft*-Embryonen zu quantifizieren. Durch beide Experimente könnte eine Beteiligung von *Mmp2* an der Ausprägung der beobachteten *Ft*-Phänotypen bestätigt oder ausgeschlossen werden.

4.2.3 *Sall1*

Mit *SALL1* befindet sich im Bereich des relevanten menschlichen Genomsegments auf HSA 16q12.1 ein Gen, dessen Mutation die klinische Ausprägung des Townes-Brocks-Syndroms, einer autosomal dominanten Erkrankung des Menschen, zur Folge hat (Powell & Michaelis, 1999; Kohlhasse, 2000). Die chromosomale Lokalisation des Gens sowie die in TBS Patienten beschriebenen Handfehlbildungen ließen *Sall1* in den Kreis der Kandidatengene für die *Ft*-Mutation aufrücken. Auch die bisherigen funktionellen Analysen des eng verwandten *Drosophila*-Homologs *Spalt* (*sal*) zeigten, dass diese Transkriptionsfaktoren wichtige Regulatoren von entwicklungsbiologischen Prozessen sind (Kühnlein *et al.*, 1994). Zum Beispiel ist *sal* für die Spezifizierung des Kopf- und des Hinterteilsegments notwendig (Jürgens, 1988). Ebenso spielt *sal* eine entscheidende Rolle bei der Entwicklung des Tracheensystems der *Drosophila*-Larve und auch bei der Morphogenese der Flügel (Kühnlein & Schuh, 1996; Sturtevant *et al.*, 1997; de Celis & Barrio, 2000). Dabei wird die Expression von *sal* in den Flügel-Imaginalscheiben durch *Dpp*, dem *Drosophila*-Ortholog von *Bmp4*, und somit über den entwicklungsbiologisch bedeutenden Hedgehog-Signaltransduktionsweg reguliert (de Celis *et al.*, 1996; Nellen *et al.*, 1996).

Eine veränderte Expression von *Sall1* steht nicht im direkten kausalen Zusammenhang mit der Ausprägung der Fehlbildungen in den Extremitäten von *Ft*-Embryonen

Im Verlauf dieser Arbeit wurde das *Sall1* Gen der Maus kloniert und charakterisiert. Parallel dazu wurde in zwei unabhängigen Studien ebenfalls über die Identifizierung von *Sall1* berichtet (Buck *et al.*, 2000; Nishinakamura *et al.*, 2001). Die in der vorliegenden Arbeit isolierte *Sall1* cDNA unterscheidet sich dabei nur in einigen wenigen Nukleotiden von den dort beschriebenen Sequenzen. Die Kartierung des Gens im Mausgenom ergab, dass *Sall1* eng an die genetischen Marker *D8Mit207* und *D8Mit81* gekoppelt ist. Aus einer aktuellen

physikalischen Karte ist ersichtlich, dass damit die Distanz zwischen *D8Mit81* bzw. *Sall1* und dem genetischen Marker *D8Mit250* bzw. dem proximalen Endpunkt der *Ft*-Deletion ca. 2Mb beträgt (Cai *et al.*, 2001). Allerdings beweist der Fall einer reziproken Translokation (t(5;16)(p15.3;q12.1), dass auch Positionsdefekte in einer Entfernung von 180kb vom *Sall1* Locus zur klinischen Manifestation des Townes-Brocks-Syndroms führen können (Serville *et al.*, 1993; Marlin *et al.*, 1999). Interessanterweise scheint sich diese Translokation genau zwischen den in Abschnitt 4.3 näher beschriebenen verwandten Genomsegmenten zugetragen zu haben. Um Veränderungen in der *Sall1* Expression als Ursache für die Ausprägung des *Ft*-Gliedmaßenphänotyps ausschließen zu können, wurde die räumliche und zeitliche Expression von *Sall1* in den sich entwickelnden Extremitäten von hetero- und homozygoten *Ft*-Embryonen im relevanten Zeitraum hinterfragt. In einigen wenigen homozygoten *Ft*-Embryonen konnte dabei eine ektopische *Sall1*-Expression im anterioren Bereich der Extremitätenknospen festgestellt werden. Da jedoch der Großteil der analysierten Embryonen keinerlei Abweichungen im Expressionsprofil von *Sall1* zeigte, kann diese nur zum Teil auftretende Veränderung in der *Sall1*-Expression nicht als Ursache für die Ausprägung der beschriebenen Fehlbildungen betrachtet werden. Insgesamt ist daher anzunehmen, dass *Sall1* aufgrund der beträchtlichen Entfernung vom *Ft*-Locus und der zum größten Teil unveränderten Expression in den Extremitäten von *Ft*-Embryonen nicht zur Ausprägung des *Ft*-Phänotyps beiträgt. Diese Hypothese wird auch durch eine kürzlich veröffentlichte Studie über die gezielte Inaktivierung des *Sall1* Gens bekräftigt (Nishinakamura *et al.*, 2001). In den entsprechenden hetero- und homozygoten Mäusen kam es dabei zu keinerlei Missbildungen der Extremitäten, der Ohren oder der Analregion und somit entgegen den eigentlich Erwartungen nicht zu einer Ausprägung eines mit TBS vergleichbaren Phänotyps. Parallelen zu TBS lassen sich in diesen homozygoten Mäusen, die kurz nach der Geburt sterben, nur aufgrund einer Fehlentwicklung der Nieren erkennen (Nishinakamura *et al.*, 2001).

4.3 Die Evolution der *Irx* Genfamilie der Wirbeltiere

Die von Dr. Renate Dildrop aufgestellte molekulare Phylogenie der *Irx* Gene zeigte, dass sich die *Irx* Gene der Wirbeltiere bisher auf insgesamt sechs orthologe Gruppen verteilen. Weiterhin bilden jeweils zwei dieser orthologen Gruppen eine höhere Gruppe von paralogenen Genen (Peters *et al.*, 2000). Während der Evolution sind daher die eng miteinander verwandten Gene *Irx1* und *Irx3* sehr wahrscheinlich aus der Duplikation eines gemeinsamen Vorläufergens hervorgegangen. Dasselbe Szenario kann auch für die Gene *Irx2/Irx5* und *Irx4/Irx6* angenommen werden, die jeweils ein zweites bzw. drittes Paar von eng verwandten, paralogenen Genen bilden. In Zusammenhang mit der physikalischen Kartierung der *Irx* Gene

der Maus, die eine enge Kopplung von *Irx3*, *Irx5* und *Irx6* (MMU8) und auch von *Irx1*, *Irx2* und *Irx4* (MMU13) aufdeckte, lässt sich somit schließen, dass die Entstehung dieser drei Duplikationen das Ergebnis eines einzigen evolutionären Vorgangs, z.B. der Verdopplung eines prototypischen Urkomplexes ist. Mittlerweile konnte die enge Kopplung von *Irx* Genen aus unterschiedlichen paralogenen Gruppen über zytogenetische Kartierungen der orthologen Gene im Menschen, im Hühnchen und auch im Zebrafisch bestätigt werden (Ogura *et al.*, 2001; Wang *et al.*, 2001). Die *Irx* Gene der Säugetiere, Vögel und Fische weisen somit eine identische genomische Organisation auf. Da sich die *Irx* Gene der Amphibien gleichermaßen den sechs orthologen und den drei höheren paralogenen Gruppen zuordnen lassen, kann auch für diese Mitglieder der Wirbeltierfamilie eine entsprechende Anordnung der *Irx* Gene im jeweiligen Genom angenommen werden (Peters *et al.*, 2000). Damit lässt sich der Zeitpunkt des Duplikationsereignisses, welches zu der Entstehung von zwei verwandten *Irx* Genkomplexen geführt haben kann, bis zu einem gemeinsamen Urahn der Wirbeltiere zurückverfolgen, der noch vor der Aufspaltung der Linie in die einzelnen Klassen (Fische, Amphibien, Reptilien, Vögel und Säugetiere) existiert haben muss.

Die Entstehung der *Irx* Genkomplexe beruht auf einer *en masse* Duplikation eines ausgedehnten Genomsegments

Möglicherweise erfolgte diese Duplikation im Zusammenhang mit Verdopplungen des kompletten Genoms, die sich bereits sehr früh in der Evolution der Wirbeltiere ereignet haben sollen (Ohno, 1970; Übersicht bei Skrabanek & Wolfe, 1998; Aparicio 2000). Diese als 2R-Hypothese bezeichnete Theorie wird derzeit durch eine breite Akzeptanz befürwortet und ist auf insgesamt zwei genomweiten Duplikationen begründet: eine unmittelbar vor und eine unmittelbar nach der Abspaltung der Überklasse der *Agnatha* (Kieferlose; Skrabanek & Wolfe, 1998). Besonders veranschaulicht wurde dieser Mechanismus durch umfangreiche Studien über die Evolution der *Hox*-Genkomplexe. Die bisherigen Analysen auf diesem Gebiet lassen vermuten, dass die insgesamt vier *Hox*-Genkomplexe der Säugetiere aus der Vervierfachung eines prototypischen Genkomplexes ähnlich dem *Antennapedia-Bithorax*-Komplex von *Drosophila* hervorgegangen sind (Übersicht bei Holland & Garcia-Fernández, 1996, Ferrier & Holland, 2001). Dabei kam es auch zu einer Duplikation von ausgedehnten benachbarten Genomabschnitten, die beispielsweise Mitglieder der *Distal-less*, *Integrin*, *Paired box* und *Wingless* Genfamilien enthalten (Ruddle *et al.*, 1994, Holland & Garcia-Fernández, 1996; Stock *et al.*, 1996). Da auch die Auswertung der Konsensussequenz des menschlichen Genoms bestätigte, dass die vier Chromosomen HSA2, 7, 12 und 17, welche jeweils einen der vier *Hox* Genkomplexe kodieren, eine Vielzahl von vervierfachten Genen

enthalten, werden genomweite Duplikation gemäß der 2R-Hypothese für diese Konstellation verantwortlich gemacht. Den genomweiten Duplikationen folgte vermutlich der Verlust von einzelnen Genen sowie auch zahlreiche Umordnungen von Genomregionen (Ruddle *et al.*, 1994, Nadeau & Sankoff, 1997). Aufgrund dieser Ereignisse ist es heutzutage sehr schwer, zwischen einem Szenario zweier genomweiter Duplikation und einem Szenario mit zahlreichen vereinzelt Duplikationen von kleineren Segmenten zu unterscheiden. Entsprechend werden beide Evolutionsszenarien ausgiebig in der gegenwärtigen Literatur diskutiert und auch von ihren jeweiligen Verfechtern befürwortet (Hughes, 1999a; Hughes, 1999b; Martin, 2001; Wolfe, 2001; Friedman & Hughes, 2001).

Überträgt man die 2R-Hypothese auch auf die Evolution der *Irx* Gene, so müssten dementsprechend die Genome von einigen Vertretern der Wirbeltierklassen zwei weitere *Irx* Genkomplexe beinhalten, die allerdings im weiteren Verlauf der Evolution bei den höheren Wirbeltieren verlorengegangen sind. Tatsächlich konnte mit dem leicht abweichenden *Irx* Gen *Ziro7* bereits ein *Irx* Gen des Zebrabärblings identifiziert werden, das sich nicht den sechs orthologen Gruppen der *Irx* Gene der Wirbeltiere zuordnen ließ (Lecaudey *et al.*, 2001). Ob dieser Befund allerdings auch im Zusammenhang mit einer in weiten Kreisen akzeptierten zusätzlichen genomweiten Duplikation während der Evolution der Fische gesehen werden muss oder aber auf einer lokalen Duplikation beruht, bleibt zu beweisen (Aparicio, 2000). Diese Problematik könnte möglicherweise durch das Studium der Organisation der *Irx* Gene im Genom von Mitgliedern der Überklasse der *Agnatha* bzw. der Klasse der *Chondrichthyes* (Knorpelfische) aufgeklärt werden.

Ein Modell für die Evolution der *Irx* Gene der Wirbeltiere

Berücksichtigt man die phylogenetischen Beziehungen zwischen den *Irx* Genen der Fruchtfliege *Drosophila melanogaster* und der Wirbeltiere, so zeigt sich, dass sich der *Irx* Genkomplex der Fruchtfliege und die Genkomplexe der Wirbeltiere aller Wahrscheinlichkeit nach unabhängig voneinander entwickelt haben. Das Entstehen des *Drosophila* Genkomplexes kann am einfachsten mit Hilfe zweier lokaler *Cis*-Duplikation erklärt werden, die zu einer tandemartigen Anordnung von drei verwandten Genen geführt haben (z.B. durch eine ungleichförmige Rekombination - '*unequal crossing over*'). Ein ähnliches Szenario lässt sich auch für die Entstehung der *Irx* Genkomplexe der Wirbeltiere ableiten (Abb. 4.1). Unterstützt wird dieses im folgenden näher erläuterte Szenario durch den Vergleich der stadien- und gewebespezifischen Expression der individuellen *Irx* Gene der Maus (Houweling *et al.*, 2001). Wie bereits erwähnt, verfügen beispielsweise die Mitglieder des *IrxA* Genkomplexes

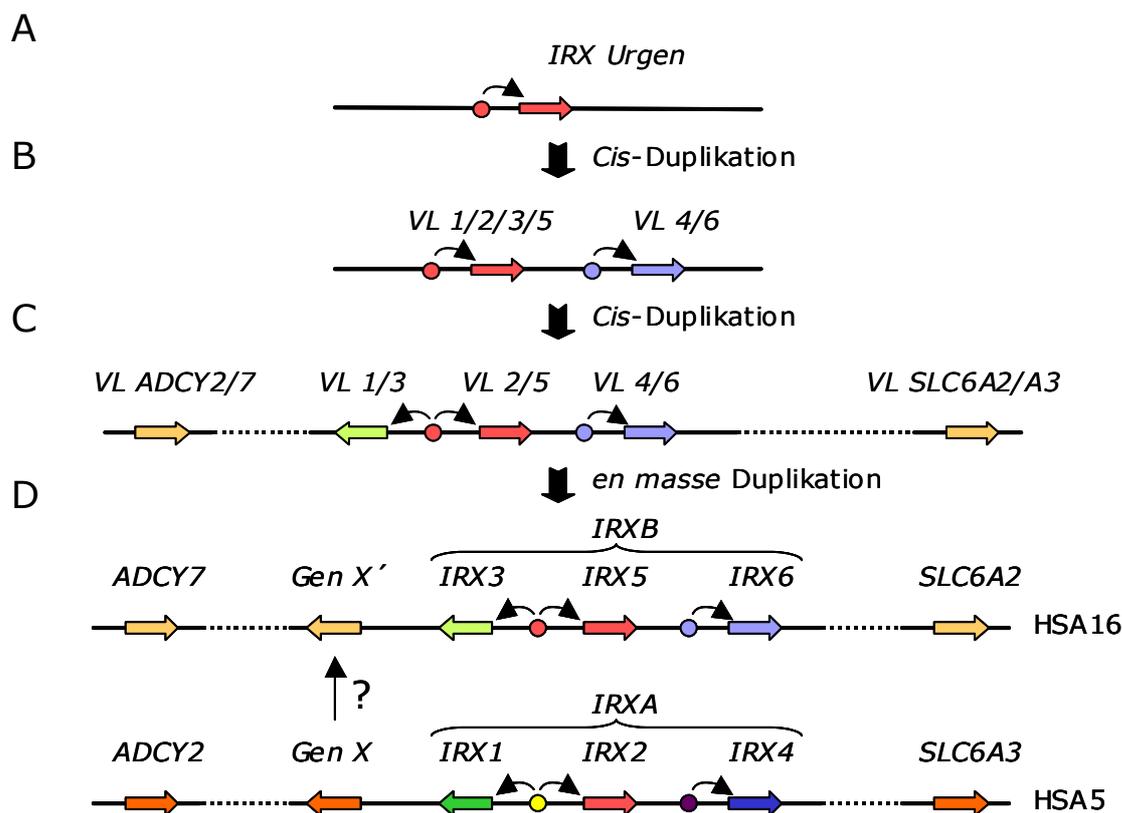


Abb. 4.1: Ein mögliches Szenario für die Evolution der *Irx* Genfamilie und des *Ft*-Locus. Ein einzelnes *Irx Urgen* (A) wurde zunächst mitsamt der dazugehörigen Regulatorregion (Kreis) in einer ersten *Cis*-Duplikation verdoppelt und ließ somit die Vorläufer (VL) von *Irx1/2/3/5* und *Irx4/6* entstehen (B). Aus dem Vorläufer von *Irx1/2/3/5* gingen in einer zweiten *Cis*-Duplikation die Vorläufer von *Irx1/3* und *Irx2/5* hervor, deren Expression jedoch auch danach von ein und derselben Regulatorregion gesteuert wird (C). Durch die *en masse* Duplikation eines ausgedehnten Genomsegments entstanden Paare von paralogen Genen (D). Aufgrund der Auseinanderentwicklung der duplizierten Regulatorregionen veränderten sich die Expressionsprofile der paralogen *Irx* Gene. Allerdings wird die Expression der Gene *Irx1/2* bzw. *Irx3/5* immer noch von einer gemeinsamen Regulatorregion kontrolliert, so dass die jeweils benachbarten Gene eine ähnliche Expression aufweisen, paraloge Partner jedoch unterschiedlich exprimiert werden. Im paralogem Genomsegment auf HSA5/MMU13 könnten weitere Gene lokalisiert sein (*Gen X*), deren paraloge Partner (*Gen X'*) entweder ebenfalls im Bereich des *Ft*-Locus auf HSA16/MMU8 erhalten geblieben oder im Laufe der Evolution verlorengegangen sind. Auf diese Weise könnte die Sequenzanalyse des paralogem Genomsegments zur Identifizierung von bisher unentdeckten *Ft*-Kandidatengenen beitragen. Allerdings geben die bisher verfügbaren Sequenzen des menschlichen Genoms keinen Hinweis auf das Vorhandensein weiterer Gene in beiden verwandten Genomsegmenten. Die jeweilige Position der Gene im paralogem Genomsegment auf HSA5 wurde der geringauflösenden Strahlungshybridkarte GeneMap'99 entnommen. Die Konsensussequenz dieses Genomabschnitts liegt in den öffentlichen Datenbanken bisher nur lückenhaft vor. Daher könnte die exakte lineare Abfolge dieser Gene im menschlichen Genom unter Umständen geringfügig von der hier dargestellten Form abweichen.

Irx1 und *Irx2* über ähnliche Expressionsmuster, während *Irx4* ein abweichendes Expressionsprofil aufweist. Gleiches gilt für die Mitglieder des *IrxB* Genkomplexes, wo für die Gene *Irx3* und *Irx5* ein z.T. übereinstimmendes Expressionsmuster sowie für *Irx6* ein davon abweichendes Expressionsprofil festgestellt werden konnte (Houweling *et al.*, 2001). Folglich könnte zunächst die *Cis*-Duplikation eines einzelnen mit dem *Irx* Gen von *C. elegans* verwandten Urgens (Abb. 4.1-A) und der dazugehörigen Regulatorsequenzen zur Entstehung eines Vorläufers von *Irx4/6* sowie eines Vorläufers von *Irx1/3/2/5* geführt haben, deren Expressionsprofile aufgrund der fortschreitenden Evolution mittlerweile stark voneinander abweichen (Abb. 4.1-B). Eine weitere, relativ junge *Cis*-Duplikation ließ aus dem Vorläufer von *Irx1/2/3/5* die Vorläufer von *Irx1/3* und *Irx2/5* hervorgehen, deren Expression jedoch auch danach noch von ein und derselben Regulatorregion gesteuert wurde (Abb. 4.1-C). Auf diese Weise entstand somit zunächst ein Urgenkomplex von drei verwandten Genen. Anschließend führte die *en masse* Duplikation eines ausgedehnten Genomsegments zur Formation des *IrxA* und des *IrxB* Genkomplexes (Abb. 4.1-D). Im weiteren Verlauf der Evolution müsste es demnach zu einer Auseinanderentwicklung der ebenfalls duplizierten Regulatorelemente gekommen sein, so dass die benachbarten Gene *Irx3* und *Irx5* bzw. *Irx1* und *Irx2* eine ähnliche Expression, paraloge Partner (z.B. *Irx3* und *Irx1*) jedoch eine unterschiedliche Expression aufweisen. Dieses Evolutionsszenario führt die beobachteten Expressionsprofile und auch die phylogenetischen Beziehungen der *Irx* Gene der Wirbeltiere zusammen. Möglicherweise können die postulierten Regulatorelemente, die bei der Steuerung der Expression der *Irx* Gene eine wesentliche Rolle spielen, z.B. durch gezielte Deletionen von kurzen Genomabschnitten im Bereich der *Irx* Genkomplexe oder auch über Reportergenkonstrukte in zukünftigen Experimenten identifiziert werden (Pennacchio & Rubin, 2001).

4.4 Die Konsequenzen der *en masse* Duplikation für den *Ft*-Locus

Es wird angenommen, dass Duplikationen von vereinzelten Genomsegmenten bzw. genomweite Duplikationen die entscheidende treibende Kraft bei der Evolution der Wirbeltiergenome darstellen (Ohno, 1970; Ohno, 1999; Meyer & Scharl, 1999). Jüngste Auswertungen der Konsensussequenz des menschlichen Genoms zeigten, dass etwa 10,6% des menschlichen Genoms aus duplizierten und damit paralogen Segmenten bestehen (Bailey *et al.*, 2001). Da Sequenzvergleiche zwischen paralogen Segmenten offenbarten, dass sowohl Exon- als auch Intronsequenzen hoch konserviert worden sind (z.T. wurden weniger als 6% Nukleotidabweichung zwischen duplizierten Regionen festgestellt), schließt man auf ein relativ junges Alter dieser Duplikationen (IHGSC, 2001).

Durch die systematische Betrachtung der Gene in unmittelbarer Nähe der *IRX* Genkomplexe des Menschen auf HSA5 und HSA16 konnten in der vorliegenden Arbeit weitere Paare von paralogen Genen identifiziert werden (z.B. *ADCY2/ADCY7*, *SLC6A3/SLC6A2*). Die Existenz der zwei verwandten *Irx* Genkomplexe und somit auch der zwei identifizierten paraloge Segmente kann damit ebenfalls am einfachsten durch die *en masse* Duplikation eines größeren Genomsegments erklärt werden (Abb. 4.1).

Segmentale Duplikationen können in zwei verschiedene Klassen unterteilt werden: In interchromosomale und intrachromosomale Duplikationen. Interchromosomale Duplikationen, also Segmente die zwischen zwei nicht-homologen Chromosomen dupliziert wurden, können dabei sehr häufig in den Bereichen der Centromere (pericentromere Region) und Telomere (subtelomere Region) gefunden werden (IHGSC, 2001). Beispielsweise bewiesen Sequenzanalysen des Chromosoms HSA22, dass sowohl die pericentromere Region (1,5Mb) als auch die subtelomere Region (50kb) dieses Chromosoms fast ausschließlich aus interchromosomalen Duplikationen bestehen (Dunham *et al.*, 1999; IHGSC, 2001). Eine Erklärung für diesen Umstand wird zur Zeit in einem Schadenskontrollmechanismus gesucht, bei dem chromosomale Bruchstücke bevorzugt an die Centromer- oder auch an die Telomerregion angeheftet werden (IHGSC, 2001). Die in der vorliegenden Arbeit identifizierten paraloge Segmente zeigen eine ähnliche Verteilung im menschlichen Genom. So wurde das den *Ft*-Locus kodierende Segment im pericentromeren Bereich von HSA16 lokalisiert, während das paraloge Segment in der subtelomeren Region von HSA5 kartiert werden konnte. Allerdings zeigt die Anordnung der entsprechenden Loci im Mausgenom (*Ft*-Locus: MMU8, 43cM ; paraloges Segment: MMU13, 25cM) nicht diese charakteristische Verteilung.

Geht man nun davon aus, dass die Gene des *Ft*-Locus bereits in frühen Vertebraten gekoppelt vorlagen, dann müsste dieser evolutionäre Vorgang auch zu einer Verdopplung aller anderen Mitglieder des *Ft*-Locus geführt haben. In der Tat existieren im menschlichen Genom paraloge Partner für fast alle identifizierten *Ft* Gene. Nur dem Gen *KIAA0308/2FLJ* konnte keine paraloge Sequenz zugeordnet werden. Allerdings sind die identifizierten, jeweils paraloge Partnergene bzw. Sequenzen nicht, wie eigentlich erwartet, in enger Nachbarschaft zum *IRXA* Komplex im verwandten Genomsegment auf HSA5p15 lokalisiert, sondern wurden verstreut über das gesamte menschliche Genom kartiert. Beispielsweise sind *RB1* und *RBL1* (*p107*), die paraloge Partner von *RBL2* (*p130*), auf HSA13q14 bzw. HSA20q11 lokalisiert. Mit *RPGRIP* liegt der paraloge Partner von *FTM* auf HSA14q11. Gleichmaßen wurde *MMP9*, der vermeintliche paraloge Partner von *MMP2* aus der Multigenfamilie der Metalloproteinasen, auf HSA20q13 kartiert (Massova *et al.*, 1998). Dem Gen *FTO* konnte eine bisher nicht näher charakterisierte paraloge Sequenz auf HSA2p21 zugeordnet werden. *FTS* scheint sogar über mehrere bisher unbekannte paraloge Partnersequenzen zu verfügen. Eine dieser Sequenzen wurde sogar auf HSA5p14, also in der Nachbarschaft des *IRXA* Genkomplexes

(HSA5p15) kartiert. Allerdings werden beide Loci immer noch durch mehr als 23Mb voneinander getrennt (Information des *ENSEMBL Human Genome Server*).

Die Verteilung der jeweiligen paralogen Partnergene im menschlichen Genom lässt sich auf unterschiedliche Art und Weise begründen. Zum einen könnte es im Verlauf der Evolution aufgrund von Genomumordnungen zu einer Auflösung der ursprünglichen Kopplung der gemeinsam duplizierten Gene gekommen sein. Ebenso besteht allerdings die Möglichkeit, dass die Gene im Bereich des *Ft*-Locus, mit Ausnahme der Mitglieder des *IrxB* Genkomplexes, erst später in dieses Genomsegment immigriert sind, während ihre paralogen Partner ihre ursprüngliche Position im Genom beibehalten haben. Ebenfalls erlaubt ist die Aussage, dass die betrachteten Genpaare eventuell gar keine direkten paralogen Partner darstellen. Zwar belegen Sequenzvergleiche die enge Verwandtschaft der untersuchten Sequenzen, allerdings kann diese Analyse keinen Aufschluss über die eventuelle Existenz von weiteren paralogen Genen in den Urahnen der heutigen Wirbeltiere geben. Die eigentlichen paralogen Partner könnten somit im Laufe der Evolution verlorengegangen sein.

Führt man sich den Mechanismus der *en masse* Duplikation eines größeren Genomsegments als grundlegendes Szenario für die Entstehung der zwei *Irx* Genkomplexe und damit auch des *Ft*-Locus noch einmal vor Augen, dann kann auch die Sequenzanalyse des verwandten Genomsegments auf HSA5/MMU13 eventuell für die Identifizierung weiterer, bisher unentdeckter Kandidatengene im Bereich des *Ft*-Locus herangezogen werden. Demnach besteht die Möglichkeit, dass sich aufgrund von Genen, die sich in unmittelbarer Nachbarschaft des *IrxA* Komplexes (MMU13) befinden (z.B. Gen *X*, Abb. 4.1), entsprechende paraloge Partner dieser Gene im Bereich des *IrxB* Komplexes bzw. *Ft*-Locus (MMU8) identifizieren lassen (z.B. Gen *X'*). Die Sequenz der entsprechend relevanten Region des menschlichen Genoms auf HSA5p15 liegt bisher nur lückenhaft vor. Allerdings ließ die Analyse der verfügbaren Sequenzen bisher keine weiteren vermeintlichen Gene in diesem Genomabschnitt erkennen (Venter *et al.*, 2001; IHGSC, 2001). Nichtsdestoweniger sollte die Verfügbarkeit und Analyse einer vollständigen Genomsequenz in naher Zukunft Klarheit über die Evolution der beiden verwandten Genomregionen schaffen.

4.5 Die konservierte Syntanie zwischen Mensch und Maus im Bereich des *Ft*-Locus

Die vergleichende Genomanalyse der relevanten Genomabschnitte zwischen Mensch und Maus ließ erkennen, dass nicht nur die lineare Anordnung sondern auch die Orientierung der Gene im Bereich des *Ft*-Locus fast vollkommen erhalten geblieben ist. Eine Abweichung war

nur für die chromosomale Lokalisation des *Mpp10* Gens feststellbar, das im Gegensatz zum menschlichen Genom nicht im Bereich des *Ft*-Locus auf MMU8 kartiert werden konnte.

In der vergleichenden Genomanalyse lässt sich grundsätzlich zwischen zwei verschiedenen Qualitäten einer konservierten Genkopplung unterscheiden. Man spricht von einer konservierten Syntänie, wenn mindestens zwei Gene, die auf ein und demselben Chromosom einer bestimmten Spezies lokalisiert sind, auch bei einer anderen Spezies auf einem gemeinsamen Chromosom gefunden werden können. Im Gegensatz dazu ist in einem konservierten Segment nicht nur die chromosomale Position sondern auch die lineare Anordnung der jeweiligen Gene ohne Unterbrechungen durch chromosomale Umordnungen erhalten (Nadeau, 1989; IHGSC, 2001).

Obwohl Mensch und Maus seit ihrer Abspaltung aus einem gemeinsamen Urahn vor ca. 90 Millionen sehr unterschiedliche Evolutionswege beschritten haben, ist eine signifikante Anzahl von Genen beider Spezies immer noch in konservierten Segmenten angeordnet (Nadeau & Taylor, 1984). Aktuelle Einschätzungen beziffern die Anzahl dieser orthologen Genomsegmente auf 190 bis 230. Dabei reicht das Längenspektrum der 183 bisher eindeutig identifizierten konservierten Segmente von 24kb bis 90,5Mb. Dies entspricht einer Durchschnittslänge von ca.15,4Mb (IHGSC, 2001). Außergewöhnlich stark konserviert scheint die Genanordnung beispielsweise auf dem Chromosom HSA20 zu sein. Dieses Chromosom ist mit Ausnahme von zwei kurzen Sequenzabschnitten über seine gesamte Länge ortholog zum Mauschromosom MMU2 (IHGSC, 2001). Aufgrund des sehr hohen Grads der konservierten Syntänie zwischen Mensch und Maus ist es somit teilweise möglich, die Position eines bestimmten Gens im Mausgenom aus der chromosomalen Position des orthologen menschlichen Gens abzuleiten. Aus diesem Grund wird die derzeitige Annotierung des vollständigen Mausgenoms durch die Nutzung der Positionsinformationen von orthologen Genen im menschlichen Genoms erheblich erleichtert (Rogers & Bradley, 2001). Konservierte Syntänien bzw. Segmente wurden jedoch nicht nur zwischen Mensch und Maus sondern auch im Vergleich mit entfernt verwandten Spezies identifiziert. So konnten durch den Vergleich zum menschlichen Genom konservierte Segmente im Genom des Pufferfischs, des Zebrafischs, der Fruchtfliege und sogar im Genom des Fadenwurms ausgemacht werden (Trachtulec *et al.*, 1997; McLysaght *et al.*, 2000; Barbazuk *et al.*, 2000).

Der *Ft*-Locus ist inmitten eines konservierten Genomsegments lokalisiert

Die Auswertung der Konsensussequenz des menschlichen Genoms zeigte, dass es sich, sieht man einmal von der Umordnung von *Mpp10* ab, im Falle des untersuchten Genomabschnitts um ein konserviertes Segment handelt. Innerhalb dieses Segments ist die lineare Anordnung

und jeweilige Orientierung der Gene sowie auch der relative Abstand zwischen den Genen zwischen Mensch und Maus absolut erhalten geblieben. Die Konsensussequenz des menschlichen Genoms lieferte keine weiteren Anhaltspunkte für das Vorhandensein zusätzlicher, bisher unentdeckter Gene. Der Vergleich mit dem menschlichen Genom lässt daher den Schluss zu, dass in der vorliegenden Arbeit alle Kandidatengene im relevanten Genomabschnitt identifiziert worden sind. Dabei ist es nicht verwunderlich, dass in einem Bereich von 2Mb insgesamt nur neun Gene kartiert werden konnten. Wie die Auswertung der Konsensussequenz des menschlichen Genoms zeigt, sind solche genarmen Regionen (Genwüsten) durchaus keine Seltenheit. So weisen die beiden genärmsten Chromosomen HSA13 und Y nur eine Gendichte von fünf Genen pro Mb auf (IHGSC, 2001). Anscheinend enthalten insgesamt 605Mb des menschlichen Genom keinerlei nachweisbare Genstrukturen, wobei die Größe der bisher längsten identifizierten intergenischen Region ca. 3Mb beträgt (IHGSC, 2001). Der endgültige Beweis für die Hypothese der Identifizierung aller vermeintlichen Kandidatengene im Bereich des *Ft*-Locus wird allerdings erst durch die Verfügbarkeit einer vollständigen Genomsequenz der Maus geliefert.

4.6 Ein Ausblick auf die weitere Vorgehensweise

Die bisher vorliegenden Informationen über den strukturellen Aufbau oder die subzelluläre Lokalisation der kodierten Proteine, über die vermeintliche Teilnahme an relevanten Signaltransduktionswegen sowie auch die beobachteten Expressionsprofile sprechen alle für eine Beteiligung von einigen der identifizierten Kandidatengene an der Ausprägung der in den *Ft*-Mäusen beobachteten Missbildungen (Tab. 4.1). Allerdings dürfen diese Erkenntnisse nur als erste Indizien in Hinblick auf die Funktion der identifizierten Kandidatengene angesehen werden. Folglich muss in einem nächsten konsequenten Schritt das Wissen um die Kandidatengene im relevanten Genomabschnitt durch entsprechende funktionelle Analysen erweitert werden.

Die Technik der homologen Rekombination in embryonalen Stammzellen (ES-Zellen) ermöglicht das gezielte Einbringen von gewünschten Mutationen aller Art in das Mausgenom, inklusive Null- oder Punktmutationen (Übersicht bei Hogan *et al.*, 1994; Capecchi, 2001). In einer ersten Studie könnten auf diese Weise zunächst die identifizierten Kandidatengene individuell inaktiviert werden. Der Phänotyp einer solchen Maus, die eine Mutation in einem einzelnen Kandidatengenen trägt, sollte Hinweise auf mindestens eine Funktion des entsprechenden Gens liefern. Die dadurch gewonnenen Erkenntnisse könnten in einem nächsten Schritt durch die Erzeugung eines hypomorphen oder eines dominant negativen Allels weiter bekräftigt werden. Ebenso trägt auch die Erzeugung einer ganzen Serie von

allelischen Mutationen zur Aufklärung der Aufgaben der jeweiligen funktionellen Domänen des kodierten Proteins bei (Nagy *et al.*, 1998).

Häufig hat die Inaktivierung eines einzelnen Gens umfangreiche zellautonome, nichtzellautonome oder auch systemische Effekte zur Folge, die somit zu weitreichenden Veränderungen im ganzen Tier führen und damit die Analyse des beobachteten Phänotyps erheblich erschweren (z.B. frühe Sterblichkeit des Embryos). In diesem Fall lässt sich die gewebe- bzw. stadienspezifische Funktion eines Gens durch die Technik der konditionellen Geninaktivierung mit Hilfe des *Cre/loxP* oder *Flp/FRT* ortspezifischen Rekombinationssystems aufklären (Übersicht bei Lewandoski, 2001). Diese Möglichkeit der räumlichen und zeitlichen Kontrolle der Inaktivierung eines endogenen Gens erlaubt zusammen mit einer allelogenen Strategie (Serie von verschiedenen Allelen) eine höchst präzise Aufklärung der biologischen Funktion des untersuchten Gens (Gu *et al.*, 1996; Meyers *et al.*, 1998; Partanen *et al.*, 1998).

Wie z.B. die gezielte Inaktivierung von *Fts* zeigte (Lesche, 1997), führt die oben beschriebene Vorgehensweise nicht immer zur Aufklärung der Genfunktion, da die funktionellen Aufgaben unter Umständen auch von anderen, meist eng verwandten Genen übernommen werden können (funktionelle Redundanz). In diesem Fall führt nur die gemeinsame Inaktivierung der redundanten Gene zur Ausprägung eines informativen Phänotyps (Donoviel *et al.*, 1999). Die bisherige Charakterisierung der beiden verwandten Gene *Irx3* und *Irx5* lässt aufgrund des ähnlichen strukturellen Aufbaus der kodierten Proteine (hochkonservierte Homöodomäne und Iro-box) sowie der teilweise übereinstimmenden Expressionsmuster eine funktionelle Redundanz dieser beiden Gene vermuten. Zum Zweck des Studiums der Konsequenzen eines gemeinsamen Funktionsverlusts lassen sich individuell erzeugte inaktivierte Allele beider Gene, wenn überhaupt, nur durch einen erheblichen Kreuzungsaufwand zusammenführen. Da beide Gene sehr eng gekoppelt im Mausgenom vorliegen (Distanz von ca. 550kb, Abb. 3.18), ist die Wahrscheinlichkeit einer Rekombination zwischen den beiden Loci sehr gering. Unter Anwendung des *Cre/loxP* oder *Flp/FRT* ortspezifischen Rekombinationssystems kann diese Problematik elegant umgangen werden. Neben der bereits erwähnten konditionellen Geninaktivierung ermöglicht dieses System auch die Erzeugung von definierten chromosomalen Umordnungen, Deletionen, Duplikationen und Inversionen (Ramirez-Solis *et al.*, 1995; Übersicht bei Yu & Bradley, 2001). Auf diese Weise kann ein über seine Endpunkte definiertes chromosomales Segment mit einer Größe von bis zu 22cM aus dem Genom einer Maus entfernt und somit die Funktion der in diesem Segment enthaltenen Gene aufgehoben werden (Li *et al.*, 1996; Zheng *et al.*, 2000). Durch diese Vorgehensweise lassen sich auch die eng benachbarten und eventuell funktionell redundanten Gene *Irx3* und *Irx5* gemeinsam in einem Organismus inaktivieren. Natürlich kann mit Hilfe dieser Technik auch die Regulation der identifizierten Kandidatengene durch die Deletion und Inaktivierung von vermeintlichen

Steuerungselementen in den intergenischen Regionen studiert werden (Kondo & Duboule, 1999). Ebenso ermöglicht diese Technik auch die schnelle Erzeugung einer ganzen Serie von Deletionen mit einem unterschiedlichen Ausmaß von bis zu mehreren cM (LePage *et al.*, 2000; Lindsay, 2001). Somit lässt sich beispielsweise das gemeinsame Zusammenwirken aller identifizierter Kandidatengene untersuchen. Dabei sollte jedoch nicht das eigentliche Ziel, nämlich die Aufklärung der individuellen Genfunktionen und nicht die Rekonstruktion der gesamten Deletion, aus den Augen verloren werden.

Viele der genannten Vorzüge werden in einer Weiterentwicklung des *Cre/loxP*-Systems, dem transchromosomalen Rekombinationssystem *TAMERE* (*'trans-allelic targeted meiotic recombination'*), vereint. Diese Methode bietet die Möglichkeit, Mäuse, die jeweils eine Mutation in einem einzelnen Gen tragen, für die Erzeugung von gezielten chromosomalen Umordnungen zu verwenden (Hérault *et al.*, 1998). Beispielsweise können dabei die Gene *Irx3* und *Irx5* zunächst individuell inaktiviert und die eventuell auftretenden Phänotypen jeweils für sich untersucht werden. Um die redundanten Funktionen der beiden Gene aufzudecken, werden in einem nächsten Schritt die jeweiligen Mauslinien miteinander gekreuzt, so dass die beiden inaktivierten Allele schließlich auf unterschiedlichen homologen Chromosomen vorliegen. Schließlich wird während der Meiose eine über das *Cre/loxP*-System vermittelte Rekombination zwischen den homologen Chromosomen induziert. Diese hat auf einem Chromosom die Deletion der zwischen den ursprünglich gezielt mutierten Loci gelegenen chromosomalen Region zur Folge, während der gleiche Vorgang auf dem anderen Chromosom zu einer Duplikation der entsprechenden Region führt.

Deletionen im Bereich des *Ft*-Locus können aber auch auf eine andere Art und Weise induziert werden. Beispielsweise besteht in ES-Zellen die Möglichkeit, chromosomale Deletionen durch eine Behandlung mit Chemikalien oder auch vor allem durch radioaktive Strahlung zu erzeugen (You *et al.*, 1997). Ein ähnlicher Effekt wurde auch bei der Selektion von transgen-homozygoten ES-Zellen (z.B. homozygot für das Neomycin-Resistenzgen) beobachtet (Lefebvre *et al.*, 2001). Dazu müssen zunächst ES-Zellen erzeugt werden, die infolge einer gezielten homologen Rekombination ein Markergen (Neomycin-Resistenz) in einem bestimmten Locus tragen. Für den *Ft*-Locus wurden derartige ES-Zellen im Zuge der gezielten Inaktivierung von *Fts* bereits etabliert (Lesche, 1997). Nach der radioaktiven Bestrahlung lassen sich diejenigen ES-Zellen, die eine Deletion im gewünschten Genomabschnitt aufweisen, sehr einfach selektieren (Verlust der Neomycin-Resistenz) und für die Herstellung entsprechender Mäuse verwenden. Auf diese Weise können in einem einzigen Experiment zahlreiche Deletionen unterschiedlichen Ausmaßes an einem spezifischen Locus hergestellt werden. Das Ausmaß jeder einzelnen Deletion lässt sich dabei, wie natürlich auch bei den *Cre/loxP* vermittelten Deletionen, anhand der in dieser Arbeit beschriebenen physikalischen Karte des *Ft*-Locus sehr einfach bestimmen.

Die Funktionen der identifizierten Kandidatengenene können auch durch Komplementierungsstudien bestimmt werden. Im Zuge der physikalischen Kartierung des *Ft*-Locus wurde eine ganze Reihe von Genomfragmenten (BACs, PACs und YACs) charakterisiert (Ausmeier, 2000; diese Arbeit). Mit transgenen Mäusen, die Kopien dieser Fragmente in ihrem Genom tragen, ist es möglich die ursprüngliche Mutation eines Gens zu komplementieren ('*transgenic rescue*'; Schedl *et al.*, 1993; Probst *et al.*, 1998; Zhu *et al.*, 2000). Allerdings unterliegt dieses System einer gewissen Limitierung, da Sequenzen, die zwingend für die Expression eines Gens notwendig sind, auch in einer weiten Entfernung vom eigentlichen Gen lokalisiert sein können. In diesem Zusammenhang ist auch eine erste Studie zur Wiederherstellung der *Irx3* Expression in *Ft*-Mäusen zu erwähnen. Dabei konnten in transgenen Mäusen, die ein 90kb großes BAC-Fragment tragen, in dem *Irx3* zu beiden Seiten von 45kb genomischer Sequenzen flankiert wird, nicht die endogene Expression des Gens erzielt werden (Ausmeier, 2000). Erschwerend kommt hinzu, dass der *Ft*-Phänotyp aufgrund der Vielfalt der beobachteten Missbildungen vermutlich auf der Beteiligung von mehreren Genen beruht. Der Funktionsausfall von anderen Genen könnte somit die Komplementierung des untersuchten Gens überschatten (z.B. Sterblichkeit der Embryonen). Insgesamt scheint dieser Ansatz daher nicht die geeignete Vorgehensweise für eine präzise Aufklärung der Funktion der identifizierten Kandidatengene darzustellen.

Natürlich tragen auch biochemische Analysen zur funktionellen Charakterisierung der Kandidatengene bei. Beispielsweise lassen sich die identifizierten Genen über die Ermittlung von vermeintlichen Interaktionspartnern (z.B. '*yeast two hybrid*'-System; Fields & Song, 1989) in bestimmte physiologische Prozesse oder Signaltransduktionswege einordnen. Gleiches könnte durch die Darstellung der Bindung an bestimmte DNA-Motive (z.B. im Fall der *Irx* Transkriptionsfaktoren) erreicht werden. Weiterhin bestärkt eventuell die Bestimmung der Lokalisation des jeweiligen Proteins innerhalb der Zelle (z.B. über GFP-Fusionsproteine) bestehende Vermutungen über die Funktion eines Kandidatengens (z.B. Assoziation von *Ftm* mit den Monozilien der ventralen Zellen im Primitivknoten).

Letztendlich wird die funktionelle Analyse der identifizierten Kandidatengene zur Aufklärung der Organisation und Kontrolle der in den *Ft*-Mäusen betroffenen genetischen Signalwege beisteuern. Auf diese Weise liefert die Identifizierung und Charakterisierung der Kandidatengene in der vorliegenden Arbeit einen Beitrag zur Erforschung wichtiger biologischer Prozesse, die beispielsweise an der Entwicklung der Extremitäten, des Gesichtsbereichs und des zentralen Nervensystems und auch an der Kontrolle des programmierten Zelltods sowie der Etablierung der Rechts-Links-Asymmetrie der Körperachse beteiligt sind.

5 Zusammenfassung

Die *Fused toes (Ft)* Mausmutation beruht auf der Integration eines Transgens und einer damit verbundenen Deletion eines 1650kb großen Genomsegments des Chromosoms MMU8. Der komplexe Phänotyp heterozygoter (Syndaktylie, Thymushyperplasie) und homozygoter *Ft*-Mäuse (embryonale Sterblichkeit, schwerwiegende Missbildungen) deutet darauf hin, dass die von der Mutation betroffenen Gene eine Schlüsselposition bei der Entwicklung der Extremitäten, des Gesichtsbereichs, des zentralen Nervensystems und auch bei der Kontrolle des programmierten Zelltods sowie der Etablierung der Rechts-Links-Asymmetrie der Körperachse einnehmen. Jedoch konnten bisher noch nicht diejenigen Gene beschrieben werden, deren Mutation in einem direkten kausalen Zusammenhang mit der Ausprägung der beobachteten Fehlbildungen steht.

Diese Kandidatengene wurden in der vorliegenden Arbeit identifiziert und charakterisiert. Dabei wurden mit der Positionsklonierung und der vergleichenden Genomanalyse zwei unterschiedliche Ansätze verfolgt. Innerhalb des deletierten Genomabschnitts konnten mit *Ftm* und *Fto* zwei zuvor unbekannte Gene isoliert werden. Während *Fto* für ein unbekanntes Zellkernprotein kodiert, ist *Ftm* eng mit dem zilien-assoziierten Protein RPGRIP verwandt. Die durchgeführten Expressionsanalysen offenbarten, dass beide Gene während der Embryogenese zum Zeitpunkt der Ausprägung der komplexen Fehlbildungen exprimiert werden. Ebenfalls im betroffenen Genomsegment wurden mit *Irx3* und den zuvor unbekanntenen Genen *Irx5* und *Irx6* drei verwandte Homöoboxgene identifiziert, die der *Iroquois* Familie von Transkriptionsfaktoren angehören. Im Zuge dieser Analyse wurden die *Irx* Gene des Menschen und der Maus in sechs orthologe Gruppen sowie in drei höhere Gruppen von paralogenen Genen eingeordnet. Zudem offenbarte die physikalische Kartierung der *Irx* Gene die Organisation von jeweils drei Genen aus verschiedenen paralogenen Gruppen in zwei eng verwandten Genkomplexen auf den Chromosomen MMU13 (*IrxA*-Komplex) und MMU8 (*IrxB*-Komplex). In der unmittelbaren Umgebung des in den *Ft*-Mäusen deletierten Genomabschnitts wurden in den flankierenden Regionen mit *Mmp2* und *KIAA0308/2FLJ* zwei weitere Kandidatengene identifiziert. Die Sequenzanalyse des zuvor noch unbekanntenen Gens *KIAA0308/2FLJ* ließ eine signifikante Ähnlichkeit mit Chromatin-Remodellierungsfaktoren erkennen. *Mmp2* ist als Mitglied der Familie der Matrix Metalloproteinasen, dessen Funktionsverlust zu Knochenfehlbildungen führt, bereits aus anderen umfangreichen Studien bekannt. Die Charakterisierung der identifizierten Gene zeigte, dass alle bisher erwähnten Kandidatengene an der Ausprägung der beobachteten Fehlbildungen beteiligt sein können. Im Gegensatz dazu konnte eine Beteiligung von *Sall1*, ein Gen, das in der Nähe des *Ft*-Locus lokalisiert wurde und dessen Mutation im Menschen die klinische Ausprägung des Townes-

Brocks-Syndroms, einer dominanten Erkrankung mit Ähnlichkeiten zu den Fehlbildungen der Gliedmaßen heterozygoter *Ft*-Mäuse, zur Folge hat, ausgeschlossen werden.

Alle Kandidatengene wurden in einer auf einem YAC-Contig basierenden hochauflösenden Genkarte des *Ft*-Locus zusammengefasst. Dadurch konnte der *Ft*-Locus in die bestehenden genetischen und physikalischen Karten des Mausgenoms integriert werden. Die Auswertung der Konsensussequenz des menschlichen Genoms offenbarte, dass der *Ft*-Locus inmitten eines konservierten Segments lokalisiert ist. Innerhalb dieses Genomsegments ist die Anordnung, Orientierung und auch der relative Abstand zwischen den enthaltenen Genen nahezu vollständig erhalten geblieben. Zudem konnte durch diese Analyse ein Modell für die Evolution des *Ft*-Locus und der *Irx* Gene der Wirbeltiere durch die *en masse* Duplikation eines ausgedehnten Genomsegments aufgestellt werden. Vor allem die Ergebnisse der vergleichenden Genomanalyse weisen darauf hin, dass in der vorliegenden Arbeit alle Kandidatengene im relevanten Genomabschnitt beschrieben worden sind. Die Identifizierung der betroffenen Kandidatengene bildet somit die Grundlage für die Aufklärung der in den *Ft*-Mäusen gestörten biologischen Prozesse und genetischen Signalwege.

6 Literaturverzeichnis

- Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., et al.** (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**: 1651-1656.
- Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al.** (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185-2195.
- Afzelius, B.A.** (1999). Asymmetry of cilia and of mice and men. *Int. J. Dev. Biol.* **43**: 283-286.
- Altschul, S.F., Boguski, M.S., Gish, W., Wootton, J.C.** (1994). Issues in searching molecular sequence databases. *Nat. Genet.* **6**: 119-129.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J.** (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389-3402.
- Aparicio, S.** (2000). Vertebrate evolution: recent perspectives from fish. *Trends Genet.* **16**: 54-56.
- Artavanis-Tsakonas, S., Rand, M.D., Lake, R.J.** (1999) Notch signaling: cell fate control and signal integration in development. *Science* **284**: 770-776.
- Arteaga-Solis, E., Gayraud, B., Lee, S.Y., Shum, L., Sakai, L., Ramirez, F.** (2001). Regulation of limb patterning by extracellular microfibrils. *J. Cell Biol.* **154**: 275-281.
- Ausmeier, K.** (1997). Charakterisierung der Deletion in der *Fused toes* Mausmutante. Diplomarbeit, angefertigt am Fachbereich Biologie, Universität Hannover.
- Ausmeier, K.** (2000). Physikalische Kartierung des *Fused toes* Locus und funktionelle Analyse des *Iroquois 3* (*Irx3*) Gens in der Maus. Dissertation, angefertigt am Fachbereich Biologie, Universität Hannover.
- Avner, P., Bruls, T., Poras, I., Eley, L., Gas, S., Ruiz, P., Wiles, M.V., Sousa-Nunes, R., Kettleborough, R., Rana, A., et al.** (2001). A radiation hybrid transcript map of the mouse genome. *Nat. Genet.* **29**: 194-200.
- Azzi, A., Boscoboinik, D., Hensey, C.** (1992). The protein kinase C family. *Eur. J. Biochem.* **208**: 547-557.
- Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J., Eichler, E.E.** (2001). Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.* **11**: 1005-1017.
- Bao, Z.Z., Bruneau, B.G., Seidman, J.G., Seidman, C.E., Cepko, C.L.** (1999). Regulation of chamber-specific gene expression in the developing heart by *Irx4*. *Science* **283**: 1161-1164.
- Barbazuk, W.B., Korf, I., Kadavi, C., Heyen, J., Tate, S., Wun, E., Bedell, J.A., McPherson, J.D., Johnson, S.L.** (2000). The syntenic relationship of the zebrafish and human genomes. *Genome Res.* **10**: 1351-1358.
- Bareil, C., Hamel, C.P., Delague, V., Arnaud, B., Demaille, J., Claustres, M.** (2001). Segregation of a mutation in *CNGB1* encoding the beta-subunit of the rod cGMP-gated channel in a family with autosomal recessive *retinitis pigmentosa*. *Hum. Genet.* **108**: 328-334.
- Becker-Follmann, J., Gaa, A., Bausch, E., Natt, E., Scherer, G., von Deimling, O.** (1997). High-resolution mapping of a linkage group on mouse chromosome 8 conserved on human chromosome 16Q. *Mamm. Genome* **8**: 172-177.
- Bellefroid, E.J., Kobbe, A., Gruss, P., Pieler, T., Gurdon, J.B., Papalopulu, N.** (1998). *Xiro3* encodes a *Xenopus* homolog of the *Drosophila Iroquois* genes and functions in neural specification. *EMBO J.* **17**: 191-203.
- Bellomo, D., Lander, A., Harragan, I., Brown, N.A.** (1996). Cell proliferation in mammalian gastrulation: the ventral node and notochord are relatively quiescent. *Dev. Dyn.* **205**: 471-485.

- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A., Wheeler, D.L.** (2000). GenBank. *Nucleic Acids Res.* **28**: 15-18.
- Blobe, G.C., Schiemann, W.P., Lodish, H.F.** (2000). Role of transforming growth factor beta in human disease. *N. Engl. J. Med.* **342**:1350-8.
- Boguski, M.S., Tolstoshev, C.M., Bassett, D.E. Jr.** (1994). Gene discovery in dbEST. *Science* **265**: 1993-1994.
- Bosse, A., Zülch, A., Becker, M.B., Torres, M., Gómez-Skarmeta, J.L., Modolell, J., Gruss, P.** (1997). Identification of the vertebrate *Iroquois* homeobox gene family with overlapping expression during early development of the nervous system. *Mech. Dev.* **69**: 169-181.
- Bosse, A., Stoykova, A., Nieselt-Struwe, K., Chowdhury, K., Copeland, N.G., Jenkins, N.A., Gruss, P.** (2000). Identification of a novel mouse *Iroquois* homeobox gene, *Irx5*, and chromosomal localisation of all members of the mouse *Iroquois* gene family. *Dev. Dyn.* **218**: 160-174.
- Boylan, J.P., Wright, A.F.** (2000). Identification of a novel protein interacting with RPGR. *Hum. Mol. Genet.* **9**: 2085-2093.
- Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B., Aparicio, S.** (1993). Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. *Nature* **366**: 265-268.
- Brinkmeier, M.L., Camper, S.A.** (1997). Localization of somatostatin receptor genes on mouse chromosomes 2, 11, 12, 15, and 17: correlation with growth QTLs. *Genomics* **43**: 9-14.
- Briscoe, J., Pierani, A., Jessell, T.M., Ericson, J.** (2000). A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube. *Cell* **101**: 435-445.
- Briscoe, J., Ericson, J.** (2001). Specification of neuronal fates in the ventral neural tube. *Curr. Opin. Neurobiol.* **11**: 43-49.
- Bruneau, B.G., Bao, Z.Z., Tanaka, M., Schott, J.J., Izumo, S., Cepko, C.L., Seidman, J.G., Seidman, C.E.** (2000). Cardiac expression of the ventricle-specific homeobox gene *Irx4* is modulated by *Nkx2-5* and *dHand*. *Dev. Biol.* **217**: 266-277.
- Bruneau, B.G., Bao, Z.Z., Fatkin, D., Xavier-Neto, J., Georgakopoulos, D., Maguire, C.T., Berul, C.I., Kass, D.A., Kuroski-de Bold, M.L., de Bold, A.J., et al.** (2001). Cardiomyopathy in *Irx4*-deficient mice is preceded by abnormal ventricular gene expression. *Mol. Cell Biol.* **21**: 1730-1736.
- Buck, A., Archangelo, L., Dixkens, C., Kohlhase, J.** (2000). Molecular cloning, chromosomal localization, and expression of the murine *SALL1* ortholog *Sall1*. *Cytogenet. Cell Genet.* **89**: 150-153.
- Buckler, A.J., Chang, D.D., Graw, S.L., Brook, J.D., Haber, D.A., Sharp, P.A., Housman, D.E.** (1991). Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci. USA* **88**: 4005-4009.
- Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., Randazzo, F., Metzger, D., Chambon, P., Crabtree, G., Magnuson, T.** (2000). A *Brg1* null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Mol. Cell* **6**: 1287-1295.
- Buonassisi, V., Sato, G., Cohen, A.I.** (1962). Hormone-producing cultures of adrenal and pituitary tumor origin. *Proc. Natl. Acad. Sci. USA* **48**: 1184-1190.
- Bürglin, T.R.** (1997). Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, *Iroquois*, TGIF) reveals a novel domain conserved between plants and animals. *Nucleic Acids Res.* **25**: 4173-4180.
- Butler, G.S., Butler, M.J., Atkinson, S.J., Will, H., Tamura, T., van Westrum, S.S., Crabbe, T., Clements, J., d'Ortho, M.P., Murphy, G.** (1998). The *TIMP2* membrane type 1 metalloproteinase "receptor" regulates the concentration and efficient activation of progelatinase A. A kinetic study. *J. Biol. Chem.* **273**: 871-880.
- Cai, W.W., Chow, C.W., Damani, S., Gregory, S.G., Marra, M., Bradley, A.** (2001). An SSLP marker-anchored BAC framework map of the mouse genome. *Nat. Genet.* **29**: 201-205.

- Capdevila, J., Belmonte, J.C.** (1999). Extracellular modulation of the Hedgehog, Wnt and TGF-beta signalling pathways during embryonic development. *Curr. Opin. Genet. Dev.* **9**: 427-433.
- Capecchi, M.R.** (1989). Altering the genome by homologous recombination. *Science* **244**: 1288-1292.
- Capecchi, M.R.** (2001). Generating mice with targeted mutations. *Nat. Med.* **7**: 1086-1090.
- Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C.R., Lim, E.P., Kalyanaraman, N., et al.** (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* **22**: 231-238.
- Carver, E.A., Stubbs, L.** (1997). Zooming in on the human-mouse comparative map: genome conservation re-examined on a high-resolution scale. *Genome Res.* **7**: 1123-1137.
- Cavodeassi, F., Diez Del Corral, R., Campuzano, S., Dominguez, M.** (1999). Compartments and organising boundaries in the *Drosophila* eye: the role of the homeodomain *Iroquois* proteins. *Development* **126**: 4933-4942.
- Cavodeassi, F., Modolell, J., Gómez-Skarmeta, J.L.** (2001). The *Iroquois* family of genes: from body building to neural patterning. *Development* **128**: 2847-2855.
- Ceci, J.D.** (1994). Mouse chromosome 8. *Mamm. Genome* **5** (Spezialausgabe) : S124-138.
- Chambers, A.F., Matrisian, L.M.** (1997). Changing views of the role of matrix metalloproteinases in metastasis. *J. Natl. Cancer Inst.* **89**: 1260-1270.
- Chandler, S., Miller, K.M., Clements, J.M., Lury, J., Corkill, D., Anthony, D.C., Adams, S.E., Gearing, A.J.** (1997). Matrix metalloproteinases, tumor necrosis factor and multiple sclerosis: an overview. *J. Neuroimmunol.* **72**: 155-161.
- Chen, J., Knowles, H.J., Hebert, J.L., Hackett, B.P.** (1998). Mutation of the mouse hepatocyte nuclear factor/forkhead homologue 4 gene results in an absence of cilia and random left-right asymmetry. *J. Clin. Invest.* **102**: 1077-1082.
- Chen, Y., Zhao, X.** (1998). Shaping limbs by apoptosis. *J. Exp. Zool.* **282**: 691-702.
- Cheng, C.W., Hui, C.C., Strähle, U., Cheng, S.H.** (2001). Identification and expression of zebrafish *Iroquois* homeobox gene *irx1*. *Dev. Genes Evol.* **211**: 442-444.
- Chin, J.R., Werb, Z.** (1997). Matrix metalloproteinases regulate morphogenesis, migration and remodeling of epithelium, tongue skeletal muscle and cartilage in the mandibular arch. *Development* **124**: 1519-1530.
- Cho, K.O., Choi, K.W.** (1998). *Fringe* is essential for mirror symmetry and morphogenesis in the *Drosophila* eye. *Nature* **396**: 272-276.
- Chomczynski, P., Sacchi, N.** (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**: 156-159.
- Cirera, S., Wintero, A.K., Fredholm, M.** (2000). Why do we still find anonymous ESTs? *Mamm. Genome* **11**: 689-693.
- Cohen, D.R., Cheng, C.W., Cheng, S.H., Hui, C.C.** (2000). Expression of two novel mouse *Iroquois* homeobox genes during neurogenesis. *Mech. Dev.* **91**: 317-321.
- Colgan, D.F., Manley, J.L.** (1997). Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* **11**: 2755-2766.
- Copeland, N.G., Zelenetz, A.D., Cooper, G.M.** (1979). Transformation of *NIH/3T3* mouse cells by DNA of *Rous sarcoma* virus. *Cell* **17**: 993-1002.
- Daubresse, G., Deuring, R., Moore, L., Papoulas, O., Zakrajsek, I., Waldrip, W.R., Scott, M.P., Kennison, J.A., Tamkun, J.W.** (1999). The *Drosophila* *kismet* gene is related to chromatin-remodeling factors and is required for both segmentation and segment identity. *Development* **126**: 1175-1187.

- de Celis, J.F., Barrio, R., Kafatos, F.C.** (1996). A gene complex acting downstream of *dpp* in *Drosophila* wing morphogenesis. *Nature* **381**: 421-424.
- de Celis, J.F., Barrio, R.** (2000). Function of the *spalt/spalt*-related gene complex in positioning the veins in the *Drosophila* wing. *Mech. Dev.* **91**: 31-41.
- DeLise, A.M., Fischer, L., Tuan, R.S.** (2000). Cellular interactions and signaling in cartilage development. *Osteoarthritis Cartilage* **8**: 309-334.
- Deloukas, P., Schuler, G.D., Gyapay, G., Beasley, E.M., Soderlund, C., Rodriguez-Tome, P., Hui, L., Matise, T.C., McKusick, K.B., Beckmann, J.S., et al.** (1998). A physical map of 30,000 human genes. *Science* **282**: 744-746.
- Denny, P., Justice, M.J.** (2000). Mouse as the measure of man? *Trends Genet.* **16**: 283-287.
- Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., et al.** (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**: 152-154.
- Dietrich, W.F., Miller, J., Steen, R., Merchant, M.A., Damron-Boles, D., Husain, Z., Dredge, R., Daly, M.J., Ingalls, K.A., O'Connor, T.J., et al.** (1996). A comprehensive genetic map of the mouse genome. *Nature* **380**: 149-152.
- Diez del Corral, R., Aroca, P., Gómez-Skarmeta, J.L., Cavodeassi, F., Modolell, J.** (1999). The *Iroquois* homeodomain proteins are required to specify body wall identity in *Drosophila*. *Genes Dev.* **13**: 1754-1761.
- Dominguez, M., de Celis, J.F.** (1998). A dorsal/ventral boundary established by *Notch* controls growth and polarity in the *Drosophila* eye. *Nature* **396**: 276-278.
- Donoviel, D.B., Hadjantonakis, A.K., Ikeda, M., Zheng, H., Hyslop, P.S., Bernstein, A.** (1999). Mice lacking both presenilin genes exhibit early embryonic patterning defects. *Genes Dev.* **13**: 2801-2810.
- Dunham, I., Shimizu, N., Roe, B.A., Chissoe, S., Hunt, A.R., Collins, J.E., Bruskiewich, R., Beare, D.M., Clamp, M., Smink, L.J., et al.** (1999). The DNA sequence of human chromosome 22. *Nature* **402**: 489-495.
- Durick, K., Mendlein, J., Xanthopoulos, K.G.** (1999). Hunting with traps: genome-wide strategies for gene discovery and functional analysis. *Genome Res.* **9**: 1019-1025.
- Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., Tuschl, T.** (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**: 494-498.
- Ferguson, C.A., Tucker, A.S., Heikinheimo, K., Nomura, M., Oh, P., Li, E., Sharpe, P.T.** (2001). The role of effectors of the activin signalling pathway, activin receptors IIA and IIB, and Smad2, in patterning of tooth development. *Development* **128**: 4605-4613.
- Ferrier, D.E.K., Holland, P.W.H.** (2001). Ancient origins of the *Hox* gene cluster. *Nat. Rev. Genet.* **2**: 33-38.
- Fields, S., Song, O.** (1989). A novel genetic system to detect protein-protein interactions. *Nature* **340**: 245-246.
- Finnerty, J.R., Martindale, M.Q.** (1998). The evolution of the *Hox* cluster: insights from outgroups. *Curr. Opin. Genet. Dev.* **8**: 681-687.
- Flaus, A., Owen-Hughes, T.** (2001). Mechanisms for ATP-dependent chromatin remodelling. *Curr. Opin. Genet. Dev.* **11**: 148-154.
- Fraser, A.G., Kamath, R.S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., Ahringer, J.** (2000). Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**: 325-330.
- Friedman, P.A., Rao, K.W., Aylsworth, A.S.** (1987). Six patients with the Townes-Brocks syndrome including five familial cases and an association with a pericentric inversion of chromosome 16. *Am. J. Hum. Genet. Suppl.* **41**: A60.

- Friedman, R., Hughes, A.L.** (2001). Pattern and timing of gene duplication in animal genomes. *Genome Res.* **11**: 1842-1847.
- Funayama, N., Sato, Y., Matsumoto, K., Ogura, T., Takahashi, Y.** (1999). Coelom formation: binary decision of the lateral plate mesoderm is controlled by the ectoderm. *Development* **126**: 4129-4138.
- Furth, J., Upton, A.C.** (1953). ACTH secreting transplantable pituitary tumors. *Proc. Soc. Exp. Biol. Med.* **84**: 253-254.
- Fyodorov, D.V., Kadonaga, J.T.** (2001). The many faces of chromatin remodeling: SWItching beyond transcription. *Cell* **106**: 523-525.
- Ganan, Y., Macias, D., Duterque-Coquillaud, M., Ros, M.A., Hurle, J.M.** (1996). Role of TGF beta s and BMPs as signals controlling the position of the digits and the areas of interdigital cell death in the developing chick limb autopod. *Development* **122**: 2349-2357.
- Garriock, R.J., Vokes, S.A., Small, E.M., Larson, R., Krieg, P.A.** (2001). Developmental expression of the *Xenopus Iroquois*-family homeobox genes, *Irx4* and *Irx5*. *Dev. Genes. Evol.* **211**: 257-260.
- Geisler, R., Rauch, G.J., Baier, H., van Bebber, F., Brobeta, L., Dekens, M.P., Finger, K., Fricke, C., Gates, M.A., Geiger, H., et al.** (1999). A radiation hybrid map of the zebrafish genome. *Nat. Genet.* **23**: 86-89.
- Gey, G.O., Coffman, W.D., Kubicek, M.T.** (1952). Tissue culture studies of the proliferative capacity of cervical carcinoma and normal epithelium. *Cancer Res.* **12**: 264-265.
- Giannelli, F., Choo, K.H., Rees, D.J., Boyd, Y., Rizza, C.R., Brownlee, G.G.** (1983). Gene deletions in patients with haemophilia B and anti-factor IX antibodies. *Nature* **303**: 181-182.
- Giraldo, P., Montoliu, L.** (2001). Size matters: use of YACs, BACs and PACs in transgenic animals. *Transgenic Res.* **10**: 83-103.
- Gluzman, Y.** (1981). SV40-transformed simian cells support the replication of early SV40 mutants. *Cell* **23**: 175-182.
- Gönczy, P., Echeverri, G., Oegema, K., Coulson, A., Jones, S.J., Copley, R.R., Duperon, J., Oegema, J., Brehm, M., Cassin, E., et al.** (2000). Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**: 331-336.
- Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., et al.** (1996). Life with 6000 genes. *Science* **274**: 546, 563-67.
- Gómez-Skarmeta, J.L., del Corral, R.D., de la Calle-Mustienes, E., Ferre-Marco, D., Modolell, J.** (1996). *Araucan* and *caupolican*, two members of the novel *iroquois* complex, encode homeoproteins that control proneural and vein-forming genes. *Cell* **85**: 95-105.
- Gómez-Skarmeta, J.L., Glavic, A., de la Calle-Mustienes, E., Modolell, J., Mayor, R.** (1998). *Xiro*, a *Xenopus* homolog of the *Drosophila Iroquois* complex genes, controls development at the neural plate. *EMBO J.* **17**: 181-190.
- Gómez-Skarmeta, J.L., de La Calle-Mustienes, E., Modolell, J.** (2001). The Wnt-activated *Xiro1* gene encodes a repressor that is essential for neural development and downregulates *Bmp4*. *Development* **128**: 551-560.
- Goriely, A., Diez del Corral, R., Storey, K.G.** (1999). *c-Irx2* expression reveals an early subdivision of the neural plate in the chick embryo. *Mech. Dev.* **87**: 203-206.
- Graham, B., Battey, J., Jordan, E.** (2001). Report of second follow-up workshop on priority setting for mouse genomics. *Mamm. Genome* **12**: 1-2.
- Graham, F.L., van der Eb, A.J.** (1973). Transformation of rat cells by DNA of human adenovirus 5. *Virology* **54**: 536-539.

- Gridley, T.** (1997). Notch signaling in vertebrate development and disease. *Mol. Cell. Neurosci.* **9**: 103-108.
- Grillenzoni, N., van Helden, J., Dambly-Chaudiere, C., Ghysen, A.** (1998). The *iroquois* complex controls the somatotopy of *Drosophila* notum mechanosensory projections. *Development* **125**: 3563-3569.
- Grotewold, L.** (2001). Molekulare Mechanismen der Gliedmaßenentwicklung: Analyse der *Fused toes* Mausmutante und des Dickkopf-1 Gens. Dissertation, angefertigt an den Mathematisch-Naturwissenschaftliche Fakultäten, Heinrich-Heine-Universität Düsseldorf.
- Gu, H., Marth, J.D., Orban, P.C., Mossmann, H., Rajewsky, K.** (1994). Deletion of a DNA polymerase B gene segment in T cells using cell type-specific gene targeting. *Science* **265**: 103-106.
- Guerin, C.J., Lewis, G.P., Fisher, S.K., Anderson, D.H.** (1993). Recovery of photoreceptor outer segment length and analysis of membrane assembly rates in regenerating primate photoreceptor outer segments. *Invest. Ophthalmol. Vis. Sci.* **34**: 175-83.
- Haffter, P., Granato, M., Brand, M., Mullins, M.C., Hammerschmidt, M., Kane, D.A., Odenthal, J., van Eeden, F.J., Jiang, Y.J., Heisenberg, C.P., et al.** (1996). The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* **123**: 1-36.
- Haldi, M.L., Strickland, C., Lim, P., VanBerkel, V., Chen, X., Noya, D., Korenberg, J.R., Husain, Z., Miller, J., Lander, E.S.** (1996). A comprehensive large-insert yeast artificial chromosome library for physical mapping of the mouse genome. *Mamm. Genome* **7**: 767-769.
- Hall, B.K.** (1991). Cellular interactions during cartilage and bone development. *J. Craniofac. Genet. Dev. Biol.* **11**: 238-50.
- Hebinck, A., Dalski, A., Engel, H., Mattei, M., Hawken, R., Schwinger, E., Zuhlke, C.** (2000). Assignment of transcription factor NFAT5 to human chromosome 16q22.1, murine chromosome 8D and porcine chromosome 6p1.4 and comparison of the polyglutamine domains. *Cytogenet. Cell. Genet.* **90**: 68-70.
- Hérault Y., Rassoulzadegan, M., Cuzin, F., Duboule, D.** (1998). Engineering chromosomes in mice through targeted meiotic recombination (TAMERE). *Nat. Genet.* **20**: 381-384.
- Heymer, J., Kuehn, M., Rüther, U.** (1997). The expression pattern of nodal and lefty in the mouse mutant *Ft* suggests a function in the establishment of handedness. *Mech. Dev.* **66**: 5-11.
- Heymer, J., Rüther, U.** (1999). Syndactyly of *Ft*⁺ mice correlates with an imbalance in *bmp4* and *fgf8* expression. *Mech. Dev.* **88**: 173-181.
- Hogan, B.L.** (1994). Manipulating the mouse embryo. Cold Spring Harbor Press, New York
- Hogan, B.L.** (1996). Bone morphogenetic proteins: multifunctional regulators of vertebrate development. *Genes Dev.* **10**: 1580-1594.
- Holland, P.W., Garcia-Fernández, J.** (1996). *Hox* genes and chordate evolution. *Dev. Biol.* **173**: 382-395.
- Holmbeck, K., Bianco, P., Caterina, J., Yamada, S., Kromer, M., Kuznetsov, S.A., Mankani, M., Robey, P.G., Poole, A.R., Pidoux, I., Ward, J.M., Birkedal-Hansen, H.** (1999). MT1-MMP-deficient mice develop dwarfism, osteopenia, arthritis, and connective tissue disease due to inadequate collagen turnover. *Cell* **99**: 81-92.
- Hong, D.H., Pawlyk, B.S., Shang, J., Sandberg, M.A., Berson, E.L., Li, T.** (2000). A *retinitis pigmentosa* GTPase regulator (RPGR)-deficient mouse model for X-linked *retinitis pigmentosa* (RP3). *Proc. Natl. Acad. Sci. USA* **97**: 3649-3654.
- Hong, D.H., Yue, G., Adamian, M., Li, T.** (2001). *Retinitis pigmentosa* GTPase regulator (RPGR)-interacting protein is stably associated with the photoreceptor ciliary axoneme and anchors RPGR to the connecting cilium. *J. Biol. Chem.* **276**: 12091-12099.

- Hoskins, R.A., Phan, A.C., Naemuddin, M., Mapa, F.A., Ruddy, D.A., Ryan, J.J., Young, L.M., Wells, T., Kopczyński, C., Ellis, M.C. (2001). Single nucleotide polymorphism markers for genetic mapping in *Drosophila melanogaster*. *Genome Res.* **11**: 1100-1113.
- Houweling, A.C., Dildrop, R., Peters, T., Mummenhoff, J., Moorman, A.F., Rütter, U., Christoffels, V.M. (2001). Gene and cluster-specific expression of the *Iroquois* family members during mouse development. *Mech. Dev.* **107**: 169-174.
- Hrabé de Angelis, M.H., Flaswinkel, H., Fuchs, H., Rathkolb, B., Soewarto, D., Marschall, S., Heffner, S., Pargent, W., Wuensch, K., Jung, M., *et al.* (2000). Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat. Genet.* **25**: 444-447.
- Hudson, T.J., Stein, L.D., Gerety, S.S., Ma, J., Castle, A.B., Silva, J., Slonim, D.K., Baptista, R., Kruglyak, L., Xu, S.H., *et al.* (1995). An STS-based map of the human genome. *Science* **270**: 1945-1954
- Hudson, T.J., Church, D.M., Greenaway, S., Nguyen, H., Cook, A., Steen, R.G., Van Etten, W.J., Castle, A.B., Strivens, M.A., Trickett, P., *et al.* (2001). A radiation hybrid map of mouse genes. *Nat. Genet.* **29**: 201-205.
- Hughes, A.L. (1999a). Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. *J. Mol. Evol.* **48**: 565-576.
- Hughes, A.L. (1999b). Adaptive Evolution of Genes and Genomes. Oxford University Press, New York.
- Inoue, H., Nojima, H., Okayama, H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**: 23-28.
- Itoh, T., Ikeda, T., Gomi, H., Nakao, S., Suzuki, T., Itohara, S. (1997). Unaltered secretion of beta-amyloid precursor protein in gelatinase A (matrix metalloproteinase 2)-deficient mice. *J. Biol. Chem.* **272**: 22389-22392.
- Izraeli, S., Lowe, L.A., Bertness, V.L., Good, D.J., Dorward, D.W., Kirsch, I.R., Kuehn, M.R. (1999). The *SIL* gene is required for mouse embryonic axial development and left-right specification. *Nature* **399**: 691-694.
- Jiang, R., Lan, Y., Chapman, H.D., Shawber, C., Norton, C.R., Serreze, D.V., Weinmaster, G., Gridley, T. (1998). Defects in limb, craniofacial, and thymic development in *Jagged2* mutant mice. *Genes Dev.* **12**: 1046-1057.
- Joyner, A.L. (1993). Gene Targeting - A practical approach. Oxford University Press, Oxford.
- Jürgens, G. (1988). Head and tail development of the *Drosophila* embryo involves *spalt*, a novel homeotic gene. *EMBO J.* **7**: 189-196.
- Kanwar, Y.S., Ota, K., Yang, Q., Wada, J., Kashihara, N., Tian, Y., Wallner, E.I. (1999). Role of membrane-type matrix metalloproteinase 1 (MT-1-MMP), *MMP-2*, and its inhibitor in nephrogenesis. *Am. J. Physiol.* **277**: F934-947.
- Karelina, T.V., Bannikov, G.A., Eisen, A.Z. (2000). Basement membrane zone remodeling during appendageal development in human fetal skin. The absence of type VII collagen is associated with gelatinase-A (*MMP2*) activity. *J. Invest. Dermatol.* **114**: 371-375.
- Kehl, B.T., Cho, K.O., Choi, K.W. (1998). *mirror*, a *Drosophila* homeobox gene in the *Iroquois* complex, is required for sensory organ and alula formation. *Development* **125**: 1217-1227.
- Kikuno, R., Nagase, T., Suyama, M., Waki, M., Hirosawa, M., Ohara, O. (2000). HUGE: a database for human large proteins identified in the Kazusa cDNA sequencing project. *Nucleic Acids Res.* **28**: 331-332.
- Koch, R., van Luenen, H.G., van der Horst, M., Thijssen, K.L., Plasterk, R.H. (2000). Single nucleotide polymorphisms in wild isolates of *Caenorhabditis elegans*. *Genome Res.* **10**: 1690-1696.

- Kohlhase, J., Schuh, R., Dowe, G., Kühnlein, R.P., Jäckle, H., Schroeder, B., Schulz-Schaeffer, W., Kretzschmar, H.A., Kohler, A., Müller, U., Raab-Vetter, M., Burkhardt, E., Engel, W., Stick, R.** (1996). Isolation, characterization, and organ-specific expression of two novel human zinc finger genes related to the *Drosophila* gene *spalt*. *Genomics* **38**: 291-298.
- Kohlhase, J., Wischermann, A., Reichenbach, H., Froster, U., Engel, W.** (1998). Mutations in the *SALL1* putative transcription factor gene cause Townes-Brocks syndrome. *Nat. Genet.* **18**: 81-83.
- Kohlhase, J.** (2000). *SALL1* mutations in Townes-Brocks syndrome and related disorders. *Hum. Mutat.* **16**: 460-466.
- Kondo, T., Duboule, D.** (1999). Breaking colinearity in the mouse *HoxD* complex. *Cell* **97**: 407-417.
- Kozak, M.** (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**: 8125-8148.
- Kraus, J., Wöltje, M., Schönwetter, N., Höllt, V.** (1998). Alternative promoter usage and tissue specific expression of the mouse somatostatin receptor 2 gene. *FEBS Lett.* **428**: 165-170.
- Krizman, D.B., Berget, S.M.** (1993). Efficient selection of 3'-terminal exons from vertebrate DNA. *Nucleic Acids Res.* **21**: 5198-5202.
- Kudoh, T., Dawid, I.B.** (2001). Role of the *iroquois3* homeobox gene in organizer formation. *Proc. Natl. Acad. Sci. U S A* **98**: 7852-7857.
- Kühnlein, R.P., Frommer, G., Friedrich, M., Gonzalez-Gaitan, M., Weber, A., Wagner-Bernholz, J.F., Gehring, W.J., Jäckle, H., Schuh, R.** (1994). *spalt* encodes an evolutionarily conserved zinc finger protein of novel structure which provides homeotic gene function in the head and tail region of the *Drosophila* embryo. *EMBO J.* **13**: 168-179.
- Kühnlein, R.P., Schuh, R.** (1996). Dual function of the region-specific homeotic gene *spalt* during *Drosophila* tracheal system development. *Development* **122**: 2215-2223.
- Laemmli, U.K.** (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.
- Laird, P.W., Zijderveld, A., Linders, K., Rudnicki, M.A., Jaenisch, R., Berns, A.** (1991). Simplified mammalian DNA isolation procedure. *Nucleic Acids Res.* **19**: 4293.
- Lange, R., Christoph, A., Thiesen, H.J., Vopper, G., Johnson, K.R., Lemaire, L., Plomann, M., Cremer, H., Barthels, D., Heinlein, U.A.** (1995). Developmentally regulated mouse gene NK10 encodes a zinc finger repressor protein with differential DNA-binding domains. *DNA Cell. Biol.* **14**: 971-981.
- Larrain, J., Bachiller, D., Lu, B., Agius, E., Piccolo, S., De Robertis, E.M.** (2000). BMP-binding modules in chordin: a model for signalling regulation in the extracellular space. *Development* **127**: 821-830.
- Lecaudey, V., Thisse, C., Thisse, B., Schneider-Maunoury, S.** (2001). Sequence and expression pattern of *ziro7*, a novel, divergent zebrafish *iroquois* homeobox gene. *Mech. Dev.* **109**: 383-388.
- LeDouarin, N., Kalcheim, C.** (1999). The Neural Crest, 2nd edition; Teil der *Developmental and Cell Biology* Serie, Cambridge University Press.
- Lefebvre, L., Dionne, N., Karaskova, J., Squire, J.A., Nagy, A.** (2001). Selection for transgene homozygosity in embryonic stem cells results in extensive loss of heterozygosity. *Nat. Genet.* **27**: 257-258.
- Lennon, G., Auffray, C., Polymeropoulos, M., Soares, M.B.** (1996). The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics* **33**: 151-152.
- LePage, D.F., Church, D.M., Millie, E., Hassold, T.J., Conlon, R.A.** (2000). Rapid generation of nested chromosomal deletions on mouse chromosome 2. *Proc. Natl. Acad. Sci. U S A* **97**: 10471-10476.

- Lesche, R.** (1997). Molekulare und funktionelle Analyse des *Ft1*-Gens in der Maus. Dissertation, angefertigt am Fachbereich Biologie, Universität Hannover.
- Lesche, R., Peetz, A., van der Hoeven, F., Rütter, U.** (1997). *Ft1*, a novel gene related to ubiquitin-conjugating enzymes, is deleted in the *Fused toes* mouse mutation. *Mamm. Genome* **8**: 879-883.
- Lesche, R., Rütter, U.** (1998). Close linkage of *p130* and *Ft1* is conserved among mammals. *Mamm. Genome* **9**: 253-255.
- Lewandoski, M.** (2001). Conditional control of gene expression in the mouse. *Nat. Rev. Genet.* **2**: 743-755.
- Lewis, M.T., Ross, S., Strickland, P.A., Snyder, C.J., Daniel, C.W.** (1999). Regulated expression patterns of *IRX-2*, an *Iroquois*-class homeobox gene, in the human breast. *Cell Tissue Res.* **296**: 549-554.
- Leys, L., Gómez-Skarmeta, J.L., Dambly-Chaudiere, C.** (1996). *iroquois*: a prepattern gene that controls the formation of bristles on the thorax of *Drosophila*. *Mech. Dev.* **59**: 63-72.
- Li, Z.W., Stark, G., Gotz, J., Rulicke, T., Gschwind, M., Huber, G., Müller, U., Weissmann, C.** (1996). Generation of mice with a 200-kb amyloid precursor protein gene deletion by Cre recombinase-mediated site-specific recombination in embryonic stem cells. *Proc. Natl. Acad. Sci. U S A* **93**: 6158-6162.
- Lindblad-Toh, K., Winchester, E., Daly, M.J., Wang, D.G., Hirschhorn, J.N., Laviolette, J.P., Ardlie, K., Reich, D.E., Robinson, E., Sklar, P., et al.** (2000). Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat. Genet.* **24**: 381-386.
- Lindsay, E.A.** (2001). Chromosomal microdeletions: dissecting del22q11 syndrome. *Nat. Rev. Genet.* **2**: 858-868.
- Lovett, M., Kere, J., Hinton, L.M.** (1991). Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci. U S A* **88**: 9628-9632.
- Lupas, A.** (1996). Prediction and analysis of coiled-coil structures. *Methods Enzymol.* **266**: 513-525.
- Marlin, S., Blanchard, S., Slim, R., Lacombe, D., Denoyelle, F., Alessandri, J.L., Calzolari, E., Drouin-Garraud, V., Ferraz, F.G., Fourmaintraux, A., et al.** (1999). Townes-Brocks syndrome: detection of a *SALL1* mutation hot spot and evidence for a position effect in one patient. *Hum. Mutat.* **14**: 377-386.
- Marszalek, J.R., Ruiz-Lozano, P., Roberts, E., Chien, K.R., Goldstein, L.S.** (1999). *Situs inversus* and embryonic ciliary morphogenesis defects in mouse mutants lacking the KIF3A subunit of kinesin-II. *Proc. Natl. Acad. Sci. U S A* **96**: 5043-5048.
- Martignetti, J.A., Aqeel, A.A., Sewairi, W.A., Boumah, C.E., Kambouris, M., Mayouf, S.A., Sheth, K.V., Eid, W.A., Dowling, O., Harris, J., et al.** (2001). Mutation of the matrix metalloproteinase 2 gene (*MMP2*) causes a multicentric osteolysis and arthritis syndrome. *Nat. Genet.* **28**: 261-265.
- Martin, A.P.** (2001). Is tetralogy true? Lack of support for the 'one-to-four' rule. *Molecular Biology and Evolution* **18**: 89-93.
- Massova, I., Kotra, L.P., Fridman, R., Mobashery, S.** (1998). Matrix metalloproteinases: structures, evolution, and diversification. *FASEB J.* **12**: 1075-1095.
- Matthias, P., Müller, M.M., Schreiber, E., Rusconi, S., Schaffner, W.** (1989). Eukaryotic expression vectors for the analysis of mutant proteins. *Nucleic Acids Res.* **17**: 6418.
- McLysaght, A., Enright, A.J., Skrabanek, L., Wolfe, K.H.** (2000). Estimation of synteny conservation and genome compaction between pufferfish (*Fugu*) and human. *Yeast* **17**: 22-36.
- McNeill, H., Yang, C.H., Brodsky, M., Ungos, J., Simon, M.A.** (1997). *mirror* encodes a novel PBX-class homeoprotein that functions in the definition of the dorsal-ventral border in the *Drosophila* eye. *Genes Dev.* **11**: 1073-1082.

- Meindl, A., Dry, K., Herrmann, K., Manson, F., Ciccodicola, A., Edgar, A., Carvalho, M.R., Achatz, H., Hellebrand, H., Lennon, A., et al.** (1996). A gene (RPGR) with homology to the RCC1 guanine nucleotide exchange factor is mutated in X-linked *retinitis pigmentosa* (RP3). *Nat. Genet.* **13**: 35-42.
- Merino, R., Ganan, Y., Macias, D., Economides, A.N., Sampath, K.T., Hurle, J.M.** (1998). Morphogenesis of digits in the avian limb is controlled by FGFs, TGFbetas, and noggin through BMP signaling. *Dev. Biol.* **200**: 35-45.
- Merino, R., Ganan, Y., Macias, D., Rodriguez-Leon, J., Hurle, J.M.** (1999a). Bone morphogenetic proteins regulate interdigital cell death in the avian embryo. *Ann. N. Y. Acad. Sci.* **887**: 120-132.
- Merino, R., Macias, D., Ganan, Y., Rodriguez-Leon, J., Economides, A.N., Rodriguez-Esteban, C., Izpisua-Belmonte, J.C., Hurle, J.M.** (1999). Control of digit formation by activin signalling. *Development* **126**: 2161-2170.
- Meyer, A., Schartl, M.** (1999). Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr. Opin. Cell. Biol.* **11**: 699-704.
- Meyers, E.N., Lewandoski, M., Martin, G.R.** (1998). An *Fgf8* mutant allelic series generated by Cre- and Flp-mediated recombination. *Nat. Genet.* **18**: 136-141.
- Michaelis, R.C., Kaiser-Rogers, K.A., Reitnauer, P.J., Rao, K.W., Powell, C.M.** (1996). Refinement of the critical region for Townes-Brocks syndrome. *Am. J. Hum. Genet.* **59**: A228.
- Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., Ding, W., et al.** (1994). A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science* **266**: 66-71.
- Misquitta, L., Paterson, B.M.** (1999). Targeted disruption of gene function in *Drosophila* by RNA interference (RNA-i): a role for nautilus in embryonic somatic muscle formation. *Proc. Natl. Acad. Sci. U S A* **96**: 1451-1456.
- Miyazono, K., Kusanagi, K., Inoue, H.** (2001). Divergence and convergence of TGF-beta/BMP signaling. *J. Cell. Physiol.* **187**: 265-276.
- Monsoro-Burq, A., Le Douarin, N.M.** (2001). BMP4 plays a key role in left-right patterning in chick embryos by maintaining Sonic Hedgehog asymmetry. *Mol. Cell* **7**: 789-799.
- Müller, C., Leutz, A.** (2001). Chromatin remodeling in development and differentiation. *Curr. Opin. Genet. Dev.* **11**: 167-174.
- Mummenhoff, J., Houweling, A.C., Peters, T., Christoffels, V.M., Rütther, U.** (2001). Expression of *Irx6* during mouse morphogenesis. *Mech. Dev.* **103**: 193-195.
- Murcia, N.S., Richards, W.G., Yoder, B.K., Mucenski, M.L., Dunlap, J.R., Woychik, R.P.** (2000). The Oak Ridge Polycystic Kidney (*orpk*) disease gene is required for left-right axis determination. *Development* **127**: 2347-2355.
- Murray, J.C., Buetow, K.H., Weber, J.L., Ludwigsen, S., Scherpbier-Heddema, T., Manion, F., Quillen, J., Sheffield, V.C., Sunden, S., Duyk, G.M., et al.** (1994) A comprehensive human linkage map with centimorgan density. Cooperative Human Linkage Center (CHLC). *Science* **265**: 2049-2054.
- Murphy, W.J., Sun, S., Chen, Z.Q., Pecon-Slattery, J., O'Brien, S.J.** (1999). Extensive conservation of sex chromosome organization between cat and human revealed by parallel radiation hybrid mapping. *Genome Res.* **9**: 1223-1230.
- Nadeau, J.H., Taylor, B.A.** (1984). Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc. Natl. Acad. Sci. U S A* **81**: 814-818.
- Nadeau, J.H.** (1989). Maps of linkage and synteny homologies between mouse and man. *Trends Genet.* **5**: 82-86.

- Nadeau, J.H., Sankoff, D.** (1997). Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution. *Genetics* **147**: 1259-1266.
- Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Hirosawa, M., Miyajima, N., Tanaka, A., Kotani H., Nomura, N., Ohara, O.** (1999). Prediction of the coding sequences of unidentified human genes. XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res.* **6**: 63-70.
- Nagase, T., Kikuno, R., Hattori, A., Kondo, Y., Okumura, K., Ohara, O.** (2000). Prediction of the coding sequences of unidentified human genes. XIX. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res.* **7**: 347-355.
- Nagy, A., Moens, C., Ivanyi, E., Pawling, J., Gertsenstein, M., Hadjantonakis, A.K., Pirity, M., Rossant, J.** (1998). Dissecting the role of N-myc in development using a single targeting vector to generate a series of alleles. *Curr. Biol.* **8**: 661-664.
- Nasevicius, A., Ekker, S.C.** (2000). Effective targeted gene 'knockdown' in zebrafish. *Nat. Genet.* **26**: 216-220.
- Nellen, D., Burke, R., Struhl, G., Basler, K.** (1996). Direct and long-range action of a DPP morphogen gradient. *Cell* **85**: 357-368.
- Nishinakamura, R., Matsumoto, Y., Nakao, K., Nakamura, K., Sato, A., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Scully, S., Lacey, D.L., et al.** (2001). Murine homolog of *SALL1* is essential for ureteric bud invasion in kidney development. *Development* **128**: 3105-3115.
- Noden, D.M.** (1986). Origins and patterning of craniofacial mesenchymal tissues. *J. Craniofac. Genet. Dev. Biol. Suppl.* **2**: 15-31.
- Noden, D.M.** (1991). Cell movements and control of patterned tissue assembly during craniofacial development. *J. Craniofac. Genet. Dev. Biol.* **11**: 192-213.
- Nolan, P.M., Peters, J., Strivens, M., Rogers, D., Hagan, J., Spurr, N., Gray, I.C., Vitor, L., Brooker, D., Whitehill, E., et al.,** (2000) A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat. Genet.* **25**: 440-443.
- Nomura, N., Miyajima, N., Sazuka, T., Tanaka, A., Kawarabayasi, Y., Sato, S., Nagase, T., Seki, N., Ishikawa, K., Tabata, S.** (1994). Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1. *DNA Res.* **1**: 27-35.
- Nomura, M., Li, E.** (1998). Smad2 role in mesoderm formation, left-right patterning and craniofacial development. *Nature* **393**: 786-790.
- Nonaka, S., Tanaka, Y., Okada, Y., Takeda, S., Harada, A., Kanai, Y., Kido, M., Hirokawa, N.** (1998). Randomization of left-right asymmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein. *Cell* **95**: 829-837.
- Nusbaum, C., Slonim, D.K., Harris, K.L., Birren, B.W., Steen, R.G., Stein, L.D., Miller, J., Dietrich, W.F., Nahf, R., Wang, V., et al.,** (1999). A YAC-based physical map of the mouse genome. *Nat. Genet.* **22**: 388-393.
- Nüsslein-Volhard, C., Wieschaus, E.** (1980). Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**: 795-801.
- Ogura, K., Matsumoto, K., Kuroiwa, A., Isobe, T., Otaguro, T., Jurecic, V., Baldini, A., Matsuda, Y., Ogura, T.** (2001). Cloning and chromosome mapping of human and chicken *Iroquois (IRX)* genes. *Cytogenet. Cell Genet.* **92**: 320-325.
- Ohara, O., Nagase, T., Ishikawa, K., Nakajima, D., Ohira, M., Seki, N., Nomura, N.** (1997). Construction and characterization of human brain cDNA libraries suitable for analysis of cDNA clones encoding relatively large proteins. *DNA Res.* **4**: 53-59.
- Ohno, S.** (1970). Evolution by gene duplication. Springer-Verlag, New York.

- Ohno, S.** (1999). Gene duplication and the uniqueness of vertebrate genomes circa 1970-1999. *Semin. Cell. Dev. Biol.* **10**: 517-522.
- Okada, Y., Nonaka, S., Tanaka, Y., Saijoh, Y., Hamada, H., Hirokawa, N.** (1999). Abnormal nodal flow precedes *situs inversus* in *iv* and *inv* mice. *Mol. Cell* **4**: 459-468.
- Papayannopoulos, V., Tomlinson, A., Panin, V.M., Rauskolb, C., Irvine, K.D.** (1998). Dorsal-ventral signaling in the *Drosophila* eye. *Science* **281**: 2031-2034.
- Partanen, J., Schwartz, L., Rossant, J.** (1998). Opposite phenotypes of hypomorphic and Y766 phosphorylation site mutations reveal a function for Fgfr1 in anteroposterior patterning of mouse embryos. *Genes Dev.* **12**: 2332-2344.
- Peetz, A.** (1997) Untersuchungen der Mausmutante *Ft* unter besonderer Berücksichtigung des Gens *p130*. Dissertation, angefertigt am Fachbereich 2 Biologie/Chemie; Universität Bremen
- Pennacchio, L.A., Rubin, E.M.** (2001). Genomic strategies to identify mammalian regulatory sequences. *Nat. Rev. Genet.* **2**: 100-109.
- Peters, T., Dildrop, R., Ausmeier, K., Rütther, U.** (2000). Organization of mouse *Iroquois* homeobox genes in two clusters suggests a conserved regulation and function in vertebrate development. *Genome Res.* **10**: 1453-1462.
- Petes, T.D.** (2001). Meiotic recombination hot spots and cold spots. *Nat. Rev. Genet.* **2**: 360-369.
- Philippsen, P., Stotz, A., Scherf, C.** (1991). DNA of *Saccharomyces cerevisiae*. *Methods Enzymol.* **194**: 169-182.
- Powell, C.M., Reitnauer, P.J., Kaiser-Rogers, K.A., Rao, K.W.** (1995). A paracentric inversion of 16q in a patient with anus, hand, and ear anomalies: further evidence for a Townes-Brocks syndrome gene at 16q12.1. *Am. J. Hum. Genet.* **57**: A100
- Powell, C.M., Michaelis, R.C.** (1999). Townes-Brocks syndrome. *J. Med. Genet.* **36**: 89-93.
- Probst, F.J., Fridell, R.A., Raphael, Y., Saunders, T.L., Wang, A., Liang, Y., Morell, R.J., Touchman, J.W., Lyons, R.H., Noben-Trauth, K., Friedman, T.B., Camper, S.A.** (1998). Correction of deafness in *shaker-2* mice by an unconventional myosin in a BAC transgene. *Science* **280**: 1444-1447.
- Raap, A.K.** (1998). Advances in fluorescence in situ hybridization. *Mutat. Res.* **400**: 287-298.
- Rajewsky, K., Gu, H., Kühn, R., Betz, U.A., Müller, W., Roes, J., Schwenk, F.** (1996). Conditional gene targeting. *J. Clin. Invest.* **98**: 600-603.
- Ramirez-Solis, R., Liu, P., Bradley, A.** (1995). Chromosome engineering in mice. *Nature* **378**: 720-724.
- Reyes, J.C., Barra, J., Muchardt, C., Camus, A., Babinet, C., Yaniv, M.** (1998). Altered control of cellular proliferation in the absence of mammalian *brahma* (SNF2alpha). *EMBO J.* **17**: 6979-6991.
- Riley, J.H., Allan, C.J., Lai, E., Roses, A.** (2000). The use of single nucleotide polymorphisms in the isolation of common disease genes. *Pharmacogenomics* **1**: 39-47.
- Rhodes, M., Straw, R., Fernando, S., Evans, A., Lacey, T., Dearlove, A., Greystrom, J., Walker, J., Watson, P., Weston P., et al.** (1998) A high-resolution microsatellite map of the mouse genome. *Genome Res.* **8**: 531-542.
- Roepman, R., van Duijnhoven, G., Rosenberg, T., Pinckers, A.J., Bleeker-Wagemakers, L.M., Bergen, A.A., Post, J., Beck, A., Reinhardt, R., Ropers, H.H., et al.** (1996). Positional cloning of the gene for X-linked *retinitis pigmentosa* 3: homology with the guanine-nucleotide-exchange factor RCC1. *Hum. Mol. Genet.* **5**: 1035-1041.

- Roepman, R., Bernoud-Hubac, N., Schick, D.E., Maugeri, A., Berger, W., Ropers, H.H., Cremers, F.P., Ferreira, P.A.** (2000). The *retinitis pigmentosa* GTPase regulator (RPGR) interacts with novel transport-like proteins in the outer segments of rod photoreceptors. *Hum. Mol. Genet.* **9**: 2095-2105.
- Rogers, J., Bradley, A.** (2001). The mouse genome sequence: status and prospects. *Genomics* **77**: 117-118.
- Ruddle, F.H., Bentley, K.L., Murtha, M.T., Risch, N.** (1994). Gene loss and gain in the evolution of the vertebrates. *Dev Suppl.* **1994**: 155-161.
- Sage, J., Yuan, L., Martin, L., Mattei, M.G., Guenet, J.L., Liu, J.G., Hoog, C., Rassoulzadegan, M., Cuzin, F.** (1997). The *Sycp1* loci of the mouse genome: successive retropositions of a meiotic gene during the recent evolution of the genus. *Genomics* **44**: 118-126.
- Saitou, N., Nei, M.** (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406-425.
- Sambrook, J., Fritsch, E.F., Maniatis, T.** (1989). *Molecular Cloning - A Laboratory Manual*, 2nd Edition. Cold Spring Harbor Laboratory Press, New York.
- Satokata, I., Maas, R.** (1994). *Msx1* deficient mice exhibit cleft palate and abnormalities of craniofacial and tooth development. *Nat. Genet.* **6**: 348-356.
- Saunders, J.W., Fallon, J.F.** (1967). Cell death in morphogenesis, in *Major problems in developmental biology* (herausgegeben von Locke, M.) 289-314, Academic Press, New York und London.
- Schedl, A., Larin, Z., Montoliu, L., Thies, E., Kelsey, G., Lehrach, H., Schütz, G.** (1993). A method for the generation of YAC transgenic mice by pronuclear microinjection. *Nucleic Acids Res.* **21**: 4783-4787.
- Scherer, W.F., Syverton, J.T., Gey, G.O.** (1953). Studies on the propagation in vitro of poliomyelitis viruses. *J. Exp. Med.*, **97**: 695-710.
- Schork, N.J., Fallin, D., Lanchbury, J.S.** (2000). Single nucleotide polymorphisms and the future of genetic epidemiology. *Clin. Genet.* **58**: 250-264.
- Schuler, G.D., Boguski, M.S., Stewart, E.A., Stein, L.D., Gyapay, G., Rice, K., White, R.E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E., et al.** (1996). A gene map of the human genome. *Science* **274**: 540-546
- Serville, F., Lacombe, D., Saura, R., Billeaud, C., Sergent, M.P.** (1993). Townes-Brocks syndrome in an infant with translocation t(5;16). *Genet. Couns.* **4**: 109-112.
- Shimoda, N., Knapik, E.W., Ziniti, J., Sim, C., Yamada, E., Kaplan, S., Jackson, D., de Sauvage, F., Jacob, H., Fishman, M.C.** (1999). Zebrafish genetic map with 2000 microsatellite markers. *Genomics* **58**: 219-232.
- Sidow, A., Bulotsky, M.S., Kerrebrock, A.W., Bronson, R.T., Daly, M.J., Reeve, M.P., Hawkins, T.L., Birren, B.W., Jaenisch, R., Lander, E.S.** (1997). *Serrate2* is disrupted in the mouse limb-development mutant *syndactylism*. *Nature* **389**: 722-725.
- Skrabanek, L., Wolfe, K.H.** (1998). Eukaryote genome duplication - where's the evidence? *Curr. Opin. Genet. Dev.* **8**: 694-700.
- Stephens, J.C., Cavanaugh, M.L., Gradie, M.I., Mador, M.L., Kidd, K.K.** (1990). Mapping the human genome: current status. *Science* **250**: 237-244.
- Sternlicht, M.D., Werb, Z.** (2001). How matrix metalloproteinases regulate cell behavior. *Annu. Rev. Cell. Dev. Biol.* **17**: 463-516.
- Stock, D.W., Ellies, D.L., Zhao, Z., Ekker, M., Ruddle, F.H., Weiss, K.M.** (1996). The evolution of the vertebrate *Dlx* gene family. *Proc. Natl. Acad. Sci. U S A* **93**: 10858-10863.
- Strongin, A.Y., Collier, I., Bannikov, G., Marmer, B.L., Grant, G.A., Goldberg, G.I.** (1995). Mechanism of cell surface activation of 72-kDa type IV collagenase. Isolation of the activated form of the membrane metalloprotease. *J. Biol. Chem.* **270**: 5331-5338.

- Sturtevant, M.A., Biehs, B., Marin, E., Bier, E.** (1997). The *spalt* gene links the A/P compartment boundary to a linear adult structure in the *Drosophila* wing. *Development* **124**: 21-32.
- Sulik, K., Dehart, D.B., Iangaki, T., Carson, J.L., Vrablic, T., Gesteland, K., Schoenwolf, G.C.** (1994). Morphogenesis of the murine node and notochordal plate. *Dev. Dyn.* **201**: 260-278.
- Sung, C.H., Tai, A.W.** (2000). Rhodopsin trafficking and its role in retinal dystrophies. *Int. Rev. Cytol.* **195**: 215-267.
- Supp, D.M., Witte, D.P., Potter, S.S., Brueckner, M.** (1997). Mutation of an axonemal dynein affects left-right asymmetry in *inversus viscerum* mice. *Nature* **389**: 963-966.
- Supp, D.M., Brueckner, M., Kuehn, M.R., Witte, D.P., Lowe, L.A., McGrath, J., Corrales, J., Potter, S.S.** (1999). Targeted deletion of the ATP binding domain of left-right dynein confirms its role in specifying development of left-right asymmetries. *Development* **126**: 5495-5504.
- Sutton, R.B., Davletov, B.A., Berghuis, A.M., Sudhof, T.C., Sprang, S.R.** (1995). Structure of the first C2 domain of synaptotagmin I: a novel Ca²⁺/phospholipid-binding fold. *Cell* **80**: 929-938.
- Takeda, S., Yonekawa, Y., Tanaka, Y., Okada, Y., Nonaka, S., Hirokawa, N.** (1999). Left-right asymmetry and kinesin superfamily protein KIF3A: new insights in determination of laterality and mesoderm induction by *kif3A*^{-/-} mice analysis. *J. Cell. Biol.* **145**: 825-836.
- Tamkun, J.W., Deuring, R., Scott, M.P., Kissinger, M., Pattatucci, A.M., Kaufman, T.C., Kennison, J.A.** (1992). *brahma*: a regulator of *Drosophila* homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell* **68**: 561-572.
- Tan, J.T., Korzh, V., Gong, Z.** (1999). Expression of a zebrafish *iroquois* homeobox gene, *Ziro3*, in the midline axial structures and central nervous system. *Mech. Dev.* **87**: 165-168.
- Tang, M.K., Leung, A.K., Kwong, W.H., Chow, P.H., Chan, J.Y., Ngo-Muller, V., Li, M., Lee, K.K.** (2000). Bmp-4 requires the presence of the digits to initiate programmed cell death in limb interdigital tissues. *Dev. Biol.* **218**: 89-98.
- The *C. elegans* Sequencing Consortium** (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**: 2012-2018.
- The Huntington's Disease Collaborative Research Group** (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**: 971-983.
- The International Human Genome Sequencing Consortium** (2001). Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.
- The International Mouse Mutagenesis Consortium** (2001). Sequence interpretation. Functional annotation of mouse genome sequences. *Science* **291**: 1251-1255.
- Therrien, M., Morrison, D.K., Wong, A.M., Rubin, G.M.** (2000). A genetic screen for modifiers of a kinase suppressor of Ras-dependent rough eye phenotype in *Drosophila*. *Genetics* **156**: 1231-1242.
- Thorogood, P.V., Hinchliffe, J.R.** (1975). An analysis of the condensation process during chondrogenesis in the embryonic chick hind limb. *J. Embryol. Exp. Morphol.* **33**: 581-606.
- Trachtulec, Z., Hamvas, R.M., Forejt, J., Lehrach, H.R., Vincek, V., Klein, J.** (1997). Linkage of TATA-binding protein and proteasome subunit C5 genes in mice and humans reveals synteny conserved between mammals and invertebrates. *Genomics* **44**: 1-7.
- van der Hoeven, F., Schimmang, T., Volkmann, A., Mattei, M.G., Kyewski, B., Rüther, U.** (1994). Programmed cell death is affected in the novel mouse mutant *Fused toes (Ft)*. *Development* **120**: 2601-2607.
- Van Etten, W.J., Steen, R.G., Nguyen, H., Castle, A.B., Slonim, D.K., Ge, B., Nusbaum, C., Schuler, G.D., Lander, E.S., Hudson, T.J.** (1999). Radiation hybrid map of the mouse genome. *Nat. Genet.* **22**: 384-387.

- Vanetti, M., Kouba, M., Wang, X., Vogt, G., Höllt, V.** (1992). Cloning and expression of a novel mouse somatostatin receptor (*SSTR2B*). *FEBS Lett.* **311**: 290-294.
- Venkatesh, B., Gilligan, P., Brenner, S.** (2000). *Fugu*: a compact vertebrate reference genome. *FEBS Lett.* **476**: 3-7.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al.** (2001). The sequence of the human genome. *Science* **291**: 1304-1351.
- Vervoort, R., Lennon, A., Bird, A.C., Tulloch, B., Axton, R., Miano, M.G., Meindl, A., Meitinger, T., Ciccodicola, A., Wright, A.F.** (2000). Mutational hot spot within a new RPGR exon in X-linked *retinitis pigmentosa*. *Nat. Genet.* **25**: 462-466.
- Vogelstein, B., Gillespie, D.** (1979). Preparative and analytical purification of DNA from agarose. *Proc. Natl. Acad. Sci. USA* **76**: 615-619.
- Volkman, A., Doffinger, R., Rütger, U., Kyewski, B.A.** (1996). Insertional mutagenesis affecting programmed cell death leads to thymic hyperplasia and altered thymopoiesis. *J. Immunol.* **156**: 136-145.
- Wang, D.G., Fan, J.B., Siao, C.J., Berne, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., et al.** (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**: 1077-1082.
- Wang, X., Emelyanov, A., Sleptsova-Friedrich, I., Korzh, V., Gong, Z.** (2001). Expression of two novel zebrafish *iroquois* homologues (*ziro1* and *ziro5*) during early development of axial structures and central nervous system. *Mech. Dev.* **105**: 191-195.
- Werb, Z.** (1997). ECM and cell surface proteolysis: regulating cellular ecology. *Cell* **91**: 439-442.
- Wianny, F., Zernicka-Goetz, M.** (2000). Specific interference with gene function by double-stranded RNA in early mouse development. *Nat. Cell Biol.* **2**: 70-75.
- Wilkinson, D.G., Nieto, M.A.** (1993). Detection of messenger RNA by *in situ* hybridization to tissue sections and whole mounts. *Methods Enzymol.* **225**: 361-373.
- Wilson, C., Bellen, H.J., Gehring, W.J.** (1990). Position effects on eukaryotic gene expression. *Annu. Rev. Cell. Biol.* **6**: 679-714.
- Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, V., Meinhardt, T., Pruss, M., Reuter, I., Schacherer, F.** (2000). TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res.* **28**: 316-319.
- Wolfe, K.H.** (2001). Yesterday's polyploids and the mystery of diploidization. *Nat. Rev. Genet.* **2**: 333-341.
- Wolffe, A.P., Hayes, J.J.** (1999). Chromatin disruption and modification. *Nucleic Acids Res.* **27**: 711-720.
- Woodhouse, E.C., Chuaqui, R.F., Liotta, L.A.** (1997). General mechanisms of metastasis. *Cancer* **80**: 1529-1537.
- Yeung, R.S., Bell, D.W., Testa, J.R., Mayol, X., Baldi, A., Grana, X., Klinga-Levan, K., Knudson, A.G., Giordano, A.** (1993). The retinoblastoma-related gene, *RB2*, maps to human chromosome 16q12 and rat chromosome 19. *Oncogene* **8**: 3465-3468.
- Yong, V.W., Krekoski, C.A., Forsyth, P.A., Bell, R., Edwards, D.R.** (1998). Matrix metalloproteinases and diseases of the CNS. *Trends Neurosci.* **21**: 75-80.
- Yost, H.J.** (2001). Establishment of left-right asymmetry. *Int. Rev. Cytol.* **203**: 357-381.
- You, Y., Bergström, R., Klemm, M., Lederman, B., Nelson, H., Ticknor, C., Jaenisch, R., Schimenti, J.** (1997). Chromosomal deletion complexes in mice by radiation of embryonic stem cells. *Nat. Genet.* **15**: 285-288.

- Yu, A.E., Hewitt, R.E., Connor, E.W., Stetler-Stevenson, W.G.** (1997). Matrix metalloproteinases. Novel targets for directed cancer therapy. *Drugs Aging* **11**: 229-244.
- Yu, A.E., Murphy, A.N., Stetler-Stevenson, W.G.** (1999). in Matrix Metalloproteinases (herausgegeben von Parks W.C. & Mecham R.P.) 85-113, Academic Press, San Diego.
- Yu, Y., Bradley, A.** (2001). Engineering chromosomal rearrangements in mice. *Nat. Rev. Genet.* **2**: 780-790.
- Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L., Friedman, J.M.** (1994). Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**: 425-432.
- Zheng, B., Sage, M., Sheppard, E.A., Jurecic, V., Bradley, A.** (2000). Engineering mouse chromosomes with Cre-loxP: range, efficiency, and somatic applications. *Mol. Cell. Biol.* **20**: 648-655.
- Zhu, Y., Jong, M.C., Frazer, K.A., Gong, E., Krauss, R.M., Cheng, J.F., Boffelli, D., Rubin, E.M.** (2000). Genomic interval engineering of mice identifies a novel modulator of triglyceride production. *Proc. Natl. Acad. Sci. USA* **97**: 1137-1142.
- Zülch, A.U.** (2001). Funktionsanalyse des *Irx1*-Gens. *Irx1*- ein Vormustergen essentiell für die murine Gastrulation. Dissertation, angefertigt an den Mathematisch-Naturwissenschaftliche Fakultäten, Georg-August Universität Göttingen.

7 Anhang

7.1 Abkürzungen

α -...	anti-...	E_{xxx}	Extinktion bei xxx nm
Abb.	Abbildung		Wellenlänge
abs.	absolut	<i>E. coli</i>	<i>Escherichia Coli</i>
AER	<i>apical ectodermal ridge</i>	EDTA	Ethylendiamintetraacetat
ALP	Alkalische Phosphatase	EGF	<i>epidermal growth factor</i>
AP	Ankerprimer	ELISA	<i>enzyme linked immunosorbent assay</i>
APS	Ammoniumpersulfat		
AS	Aminosäure	ENU	Äthylnitrosoharnstoff
ATP	Adenosintri-phosphat	EST	<i>expressed sequence tag</i>
ATPase	Adenosintri-phosphatase	ES-Zellen	embryonale Stammzellen
ATCC	<i>American Type Culture Collection</i>	<i>et al.</i>	<i>et alteri</i> (und andere)
BAC	<i>bacterial artificial chromosome</i>	EZM	extrazelluläre Matrix
BCIP	5-Brom-4-Chlor-3-Indolyl-Phosphat	FCS	fötale Kälberserum
		Fgf	<i>fibroblast growth factor</i>
Bez.	Bezeichnung	<i>Ft</i>	<i>Fused toes</i>
Bmp	<i>bone morphogenetic protein</i>	g	Gramm
bp	Basenpaare	y xg	y-fache Erdbeschleunigung
bzw.	beziehungsweise		g = 9,81 m/s
°C	Temperatur in Grad Celsius	GAPDH	Glycerinaldehyd 3-phosphat Dehydrogenase
ca.	circa	GB4	<i>Genebridge4</i>
cDNA	<i>complementary DNA</i>		Strahlungshybridkarte
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>	GBF	Gesellschaft für
Chr.	Chromosom		biotechnologische Forschung
Ci	Curie	GFP	grün-fluoreszierendes Protein
cM	centiMorgan	GSP	genspezifischer Primer
CMV	Cytomegalievirus	HA	Hämagglutinin
cpm	gemessener radioaktiver Zerfall pro Minute	HEPES	N-[2-Hydroxyethyl]piperazin-N'-[2-Ethansulfonsäure]
cR	centiRay	HPRT	Hypoxanthin-Guanin-Phosphoribosyl-Transferase
dCTP	Desoxy-Cytidintri-phosphat	HSA	<i>Homo sapiens</i> Chromosom
d	Tag	HUGE	<i>Database of Human Unidentified Gene-Encoded Large Proteins</i>
DEPC	Diethylpyrocarbonat		
<i>D. rerio</i>	<i>Danio rerio</i>	IHGSC	<i>International Human Genome Sequencing Consortium</i>
dest.	destilliert		
DIG	Digoxigenin	IMAGE	<i>Integrated Molecular Analysis of Genome Expression</i>
<i>D. melanogaster</i>	<i>Drosophila melanogaster</i>		
DMSO	Dimethylsulfoxid	kb	Kilo-Basenpaare
DNA	Desoxyribonukleinsäure	kDa	Kilo-Dalton
DNase	Desoxyribonuklease	konz.	konzentriert
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen	μ Ci	Mikrocurie
DTT	1,4-Dithiothreitol	μ l	Mikroliter
dNTP	Desoxyribonukleotid-Triphosphat-Gemisch	μ g	Mikrogramm

M	Molar	SDS	Natriumdodecylsulfat
Mb	Megabasenpaare	s.	siehe
mCi	Millicurie	³⁵ S	Schwefelisotop
MEF	Mausembryo Fibroblasten	<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
mg	Milligramm	sek	Sekunden
MGI	<i>Mouse Genome Informatics</i>	SNP	<i>single nucleotide polymorphism</i>
min	Minuten	SSCP	<i>single strand conformational polymorphism</i>
ml	Milliliter	SSLP	<i>simple sequence length polymorphism</i>
mM	Millimolar		
MMU	<i>Mus musculus</i> Chromosom	std	Stunden
MOPS	3-N-Morpholinopropan-Sulfonsäure	STS	<i>sequence tagged site</i>
mRNA	<i>messenger</i> -RNA	SV40	Simian Virus 40
NaAc	Natriumacetat	Tab.	Tabelle
NBT	4-Nitrotetrazolium Chloridblau	TAE	Tris-Acetat-EDTA
NCBI	<i>National Center for Biotechnology Information</i>	TALE	<i>three amino acid loop extension</i>
ng	Nanogramm	TBS	Townes-Brocks-Syndrom
NLS	nukleäres Lokalisationssignal	TE	Tris-EDTA
nm	Nanometer	TEMED	Tetramethyl-Ethyldiamin
OD _{xxx}	optische Dichte bei xxx nm Wellenlänge	tgf	<i>transforming growth factor</i>
OMIM	<i>Online Mendelian Inheritance in Man</i>	TnT	Transkription und Translation
³² P	Phosphorisotop	Tris	Tris-(hydroxymethyl)aminoethan
PAC	<i>P1 derived artificial chromosome</i>	U	<i>Unit</i>
PBS	phosphatgepufferte NaCl-Lösung	US	<i>United States</i>
p.c.	<i>post coitum</i>	UTR	untranslatierte Region
PCR	Polymerasekettenreaktion	UV	ultraviolettes Licht
PFA	Paraformaldehyd	V	Volt
PIPES	Piperazin-1,4,bis-(2-ethansulfonsäure)	Vol.	Volumen
RACE	<i>rapid amplification of cDNA ends</i>	v/v	Volumen pro Volumen
RFLP	Restriktionsfragmentlängenpolymorphismus	Wdh.	Wiederholungen
RH	radiation hybrid	WI/MIT- CGR	<i>Whitehead Institute for Biomedical Research / Massachusetts Institute of Technology, Center for Genome Research</i>
RNA	Ribonukleinsäure	w/v	Gewicht pro Volumen
RNAse	Ribonuklease	<i>X. laevis</i>	<i>Xenopus laevis</i>
RT	reverse Transkription	YAC	<i>yeast artificial chromosome</i>
RZPD	Ressourcenzentrum / Primärdatenbank im Deutschen Humangenomprojekt	z.B.	zum Beispiel
		z.T.	zum Teil

7.2 Sequenzen

7.2.1 *Ftm: Mus musculus* Fantom

NCBI-GenBank: AJ344253

```

1 GCAGAGGCTCGCGCCCTGGGCGCTATAGCAACCGGTAGCTAGCTGAGGGCGAACCGGGAATAGACCATTGCTGGCCCAAGTGACGAGAC      8
91 AGCAGGAGACCTGCCGGTGAAGATACAGGTCTAAACCTCTTTGGAGTTGGAGGGTTACAAAGAACCTCAACTCGGGCACGGTGAAGAC      8
  A G D L P V K D T G L N L F G V G G L Q E T S T A R T V K T
181 TCGCCAGGCAGTGTACGTTGTCAGCCGTGAAGAACGTGAAGACAGATTTTTGGCTTGCATGATGAGAACATTTTACTTAAACAGCATGC      38
  R Q A V S R V S R E E L E D R F L R L H D E N I L L K Q H A
271 CCGCAAGCAAGAGGATAAAATTAAGAAATGGCCACCAAGTTAATACGGCTAGTTAATGACAAGAAAAGATATGAGCGGGTTGGTGGCGG      68
  R K Q E D K I K R M A T K L I R L V N D K K R Y E R V G G G
361 CCCCAGCGGCTGGGACGAGATGTGAAATGGAAGAAATGATTGAGCAGCTGCAAGAAAAGTTCATGAGCTTGAAGACAAAATGAAGT      98
  P K R L G R D V E M E E M I E Q L Q E K V H E L E R Q N E V
451 TCTCAAAAACAGGCTGATATCAGCCAAACAGCAGCTCCAAGTCCAGGGACACAGGCAGACGTCGTACAGTCGAGTGCAGGCTCGAGTTAA      128
  L K N R L I S A K Q Q L Q V Q G H R Q T S Y S R V Q A R V N
541 CACCGGCCCGGAGAGCAAGTGCAGTGCAGGCTCGCAGGAATGCCCTGGGAAAAGTCTAAGATTTCAAAATGTGGATGAAGCAGAAAC      158
  T G R R R A S A S A G S Q E C P G K G L R F Q N V D E A E T
631 TGTACAACCCAGCTCACAAAATACAGCAATAGTTTACTTGAAGAGCCAGAGGAAAATAAGAAAATTTGGAAAATGTTATTCAGTCACA      188
  V Q P T L T K Y S N S L L E E A R G E I R N L E N V I Q S Q
721 GAGAGCCAGATAGAAAGATTAGAGCACTTGGCAGAGATCCGAAAACCTCACTTGAAGAGAAAAGAAAATGAAATCGAGCTATCTCTCT      218
  R G Q I E E L E H L A E I L K T Q L K R K E N E I E L S L T
811 TCAGCTTCGAGAACAGCAGGCTACAGATCAGAGGTCAAAATATTCGGGCAATGTAGAAACGATTAAGCTTCATAAACACTAGTGGAGAA      248
  Q L R E Q Q A T D Q R S N I R D N V E T I K L H K Q L V E K
901 AAGCAATGCTCTTTCAGTCATAGAAGGAAAATTTATTCAGCTTCAAGAGAAAGCAAACTCAGGATCAGCCATGATGCCCTTGATGGC      278
  S N A L S V A I E G K F I Q L Q E K Q R T L R I S H D A L M A
991 CAATGGAGATGAGTGAACCAAGCAGCTTBAAGGAGCAGGCTTGAAGTCTGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG      308
  N G D E L N K Q L K E Q R L L K C C S L E K Q L H S V R F S E
1081 AAGCAGATGAGAGGAGCTCAGGATGAAATTAATGATTTAGAAAAGGAGCGGGAACATTAAGAAAAGGAGGAGGAGGAGGAGGAGGAGGAG      338
  R R V E L Q D R I N D L E K E R E L L K E N Y D K L Y N S Q
1171 CGCTTCAGTGTGCCCATGAAGAGCAGTGAAGTTAAAGGAGCAGCAGATGAAAGTGCAGATTCAGTTCAGTTCAGTTCAGTTCAGTTCAG      368
  A F S A A H E E Q W K L E Q Q M K V Q I A Q L E T A L K S
1261 TGATCTTACAGACAAAACCGAAGTCTTGCACAACTAAAGACTGAAAGAAGACCAAAATGAGAACTGGTTCAGAGAAATAGAGATCTACA      398
  D L T D K T E V L D K L K T E R D Q N E K L V Q E N R D L Q
1351 GCTGACAGTGTCTGACGAAAAGCAGCGGCTGCATGAGCTTCAGAGCCGCTGAAGTTCCTCAACCAAGGAGAGTACATCAATGATGATGA      428
  L Q C L Q Q K Q R L H E L Q S R L K F F N Q E S D I N A D D
1441 CCTCAGTGAAGCCTCTGCTTATAAAAGCTCAAAAAGAACAAAANAATGGAGATCTTTCATTTTGAAGAAAAGTGGACAGCAAAATCAA      458
  L S E A L L I K A L Q K E L R N G D L S F L E K V I K I N
1531 TAAAGATCTAGACCGATCCATGAAAGAGCTTCAAGCAACTCAGTACAGAAACAGTACAAAGAACTGAAAAGACAGAAATATGCTAAATAT      488
  K D L D R Q L M E R Q L A T H A E T V Q E L E K T R N M L I M
1621 GCAACATAAAATAAATAAGACTATCAAGTGAAGTGAACAGTACTCAGAAAATGAAAATTTGCAACAAGATTATGAACCTAAAGT      518
  Q H K I N K D Y Q M E V E T V T Q K M E N L Q Q D Y E L K V
1711 GGAACAAATATGTTCTCTCGATATCAGGGCTGACGCTCAAAAAGCTTGAAGCCCAAGTAAAGGATATTCGCTATGGCAGTAAAGCA      548
  E Q A Y V H L L D I R A A R I Q K L E A Q L K D I A Y G T K Q
1801 GTACAAATTAACACAGAAATCATGCCAGATGACTCAGTTGAATTTGATGAAACCTCCACCTAGAACGAGGAGAAATCTCTTTGA      578
  Y K F K P E I M P D D S V D E F D E T I H L E R G E N L F E
1891 AATCCACATCAACAAAGTCACTTTCTCTGAGTCTCTCGGCACTTCGTTGATAAAGAACTTGTCACTTCTGTACCTACGCTTCTTA      608
  I H I N G V T F S K E V L R A S G D K E L V T F C T Y A A C Q
1981 TGACTTTGAACTGCAGACTACTCCATAGTGCAGGCGCTTACCAGGAGTATAACTTCACTTCTCAGTATCTGGTCCAGTCAATGACTT      668
  D F E L Q T T P I V R G L Y P E Y N F T S Q Y L V H V N D L
2071 ATTTTTCAGTACATTCAGAAGAACACTGTCACGCTCCCAAGCCACAGCACAGATTCAGAAAACAAATTCAGCGTGTCACTTTCAGT      698
  F L Q Y I Q K N T V T L E L H Q A H S T D Y E T I A A C Q
2161 GAGATTCATGAGATCTAGAAAAGTGTGAAATTTTTCAGCAACAAAGCTTGAAGTGAACATAAGGAGACATCCCAAAATTTGGCAG      728
  R F H E I L E K S G R I F C T T S L V G T K G D I P N F G T
2251 AGTAGAATACTGGTTCGATTAAGAGTTCCTCATGATCAAGCTATTTCGACTTTATCGTGAGCGGGGAAAAGCATTAGAGATACATAAC      758
  V E Y W F R L R V P M D G A I R L Y R E R A K A L G Y I T S
2341 AAGTTTAAAGAGCCAGAGAAAATGCAATTTGTCAGTTCAGCAAGCAGCCACTGCTCAGATCAGCCAGCAGAGTCCACAGACGGCAA      788
  S F K K P E K M Q L P S S Q Q A T A C T A Q I S P A E S T D G N
2431 CTTGAATGAACCTTCAATGTTACAGTAAGATGCTGCACCGGCTCCAGTCCCGAGCGAGCTACCTGCAGCCACACGCATATGTTGGTATACAA      818
  L N E L H V T V R C C T G C G L Q S R A S Y L Q P H A Y V V Y K
2521 GTTTTTCAGTTCAGGACCATGATCCCGCTCTTCCAGTAGCAATGATCTCAAGAAATTTGATGACCAATGTTTCCAGTCCCGGAT      848
  F F D F P D H D T A I V P S S N D P Q F D D H M C F P V P M
2611 GAATATGGATTGGATTCGCTACCTAAGTTCAGAGTCTCTGAGTTTCTATGTTTTCAGTATGATGATACCCAGGAGAAATTTACATGGG      878
  N M D L D R Y L K S E S L S F Y V F D D S D T Q E N I Y M G
2701 AAAAGTCAATGTGCCCTGATTTCCCTGGCAGATGACAAGTGTATCTCAAGAAATTTGAGTTGATGGATAAGGAAAAGCATGCTGCAGG      908
  K V N V P L I S L A H D K C I S G I F E L M D K E K H A A G
2791 AACTATCCAAGTTATATTGAAATGAAAATTCACCTACCTTCCCAAGTGGATCAATAACTACGAAAGCTTGGAAAGTTCGTTGCGAG      938
  T I Q V I L K W K F T Y L P P S G S I T T E D L G K F V C R
2881 AGAAGAGCCAGGCTGTTTCAGAGACTACCTCCAAGTCTCAGATGTTTACATCAGTTGTAGCACCAAACTAAACCAAGCAACAGTTC      968
  E E P E A V Q R L P P K S S D V T S V V A P K P K P R Q R L
2971 AACATTTGGACAAAAGTCTCTTTTGCAGATACCATATCACCCAGTCCCGAGACTTCTCTCCCTCCCTTAAAGATATFAAAGACAG      998
  T F V D A D K K V S F A D T I S H P S P E T S P P P K D I K D S
3061 TTCTCCAGAAAGTGGGACAAAACAGAAAATGGTCTGTCAGCTGTTGCGCTATCTTCCAAGGAAAGCGGTTGGCTAAGGTPAGAAGAAA      1028
  S P E V G P K P E N G L S A V A Y P S K X S G V A K V A E E N
3151 TGTGGGGGAAATGCAACAAGGAAAAGAACGACATCTCTTCTTATCTGAGGCGCAGCTTGCCTCAGGAAAGTGGCTTCTCTGAAAG      1058
  V G E M Q Q G K E D D I S F L S E G Q L A S S V A S S E D
3241 TGAACAGAAAATAACTGAGAATGGAACCAAGAAATGAAGACAGATCAGCTTCAGACAGTGCATGACTGATATCATTCCAAGTTCTGTCTC      1088
  E T E I T A E E L E P E D E D R S A S D S D D C I I P S S V S
3331 CACAAATACCAACAGCCATCAGAAAGAAATTCGAGATGAGATCATAGCTTGAACCTTAATGATTTCTCAAAATACCAAGGAGGATAT      1118
  T N T A C K Q P S E E I R R I E I I A L N L N D S Q I T R E D A T I
3421 TCAGCAGCTCTTCAATGAGTGGCGGTTCTACAGTCTTCCCGCTGAGGAGACTCCTATGTCCTGCCAAACCCAGAGTGGACAGTGGGT      1148
  Q R L F I E C R F Y S L P A E E T P M S L P K P Q S S G Q W V
3511 CTACTATAACTATTCCAATGATGATCTACCTGGATAAAGAAAACCAACCGGCTGAGAGACTCTTAAAAGCCATCTGCAAAAGGCGAGA      1178
  Y Y N Y S N V I Y L D K A E N N P A V R D I L K A I L Q R R E
3601 ACTGCTCAGAGAAAGTGTGCGCTTCCCGTGGTCACTGATCTCCCGAGGACGAGCAGGACCTAGAATGTGAGGACATTTGGGTTGCCCA      1208
  L P H R S V R F T V V S D P P E D E Q D L E C E D I G V A H

```

3691	TGTTGACCTTGCTGACCTGTTTCAGAAAGGAAGAGACATCATAGAGCAAGATATCGATGTTCTCGATGCCCGAACAGACGGTGGCACCAT	1238
	V D L A D L F Q K G R D I I E Q D I D V L D A R T D G G T I	
3781	CGGCAACTGAAGGTGACTGTGGGAAGCATTCAATGCCCTTACGGTCTGCTACGAGCAAAATAGAAAGGACTTGGAGCCCTGAAAGAACCT	1264
	G K L K V T V E A L H A L R S V Y E Q N R K D L E A *	
3871	GTTCCAGAGCGACCCGCTTTAAAGAAAATGCTCGCATGAAAGCTCAAGGATGAGATTTTGAGTAAATCACCTCGATGTTGTGTAATCTATAA	
3961	TTTCATCGGTTGCTGGGAGGGAGCCGCTGGGTAGACTGAAATGTTCAAGTTTTGTATAGGTTTGTGTTTGTATTTTTCATCCAAAT	
4051	CTCCCCATGACTGCCACCCATAGAGCTTCAGTGTCTCCAAAGTGGGCTACCTTATCAAATAATTTATAGCAACATGATAGCATGAAGCAT	
4141	CTATTTGCATTTGCCACATTTTCTTTATGAAGAACTAGGAGTAGTCCCATTTTATTTTTCATAAGTAAAAGCTAAAATCTTAGGAAAC	
4231	CAATGAAAAGAAAATGCTTTTATATGTTATCCCTAAATTTGCTCCCAATAAAGGAGCCCTCGATAGGCTCATTAAAGACTTTTATACCTGTG	
4411	AAAGCTCAGATATGCATTTTCTTTGATGCATACAGTTTCTATCAATTAATAATAAAGAAAGGAGACTTGGCACCACCTTAAAGTTACCTCC	
4501	CAGTCCATTTTGTGTTCTTTATAAAACATTTGAGTTGTTCCAGGTTATGTTGGTATCAGTGCATTTTGTATGCTGATGCTGGATCCCAAGGTT	
4591	GTATTCATATAAAGATGGATGGATGGATGGATAGATAGATAGATAGATGCACAGGAAATCTATACTTTAAAAATTAAGTGTATTTGTATAGAAA	
4681	AAAAATTAAGATATGGCTCAAATTTCTAAAATGCTTAGATATGAAAAGAGGCCCCTTATCAATTTTAAACTTACATCAATCAAGCTTATTTCA	
4771	GAGCTGGGATGTATCTGGATATCCAGACCCCTCAGAAAGCAGCTCCAAAAGTAAATTTCTCGCTTTTACGCACTCCGCTGAAGGTTTAA	
4861	TTGATATTTTAAATGGAGCTGTGAACTCAGTGTCTGCAAGGAAATGCTCTCTGGGGGCTGAGATATAATGCATTTCAATTTTAAATTTACT	
4951	TTTTTATTTGTCATTTTCTTCAAAGGATTTTTATAGGCTGTGGGTTTTCCTGTTTGCAAAACAGGCTATTCCTTTTAAAGCATATGTAAA	
5041	GATTCAGGGAGATAAGTAATTTATTAATAATCTACCTAAATTTGCCGTGCATATTAATCTCTGCTTCAGTGATCACAGACCAATTTCAAT	
5131	TAGAGAAATAGGCATTTCCCTCTTTGTGTGATTAAGTCTCGAAAATGAGTCTTTGCTCCTATTTGTGAAACATTTCAAAGCCTGCTAATG	
5221	GCAGGAAGATGGAACAGATTTAGAGACAATTTACAGTTTCAACAGGGGAAAGCAGCTATAATGCAAAAATGCAAGTATGAATATCA	
5311	CGGTATAGCTGAACCATCTGTGCTCCAACAGCTTCGGTGTAGTCTCTCTTGGGAGGCTTTGAAATGACATCCGTACGATATTAATTTG	
5401	TTGGCCACTCTGTGGGACCAGTAGAGTATGATCTGACCACAGAACTCTATAAATCTGCTGCTGACCACTGGGCAACCCCTGCTCCAGGC	
5491	CTGCCATCTCCAAGAAGAATCTTGTCTCATCAGACTGAAAAATCTTCATAACAACATCAATTCCTCGAACCTCCAGTGACCCACAAACA	
5581	TCTGACAGACAAGGCTCTCCAGTCTTAGTAAACAACACATGTTGTTGTCACCTTACGCTGGCTGAGCTCATCTTTGGCCTCCCTT	
5671	CTAGCTTTCTTTGGCTTTGTGTCAGCTACAGTGTCTAATAAGACCCATAAATTTCCCTGACGTAATTTAAAACAGCCAGTTATACCCTAG	
5761	GATTAGCCGAGTGGGTTTGAGACATTTATTTGTTGTTGCTCTTAGGAGGAGGCAATTTATGTTTAAAAATGTTATATCCAAAACATAAATA	
5851	TGCTTTCTCTGGGGGGGGGGGGTCTGACTTCTTAGCGTCTGTAAGTGGAGACAGGAAGTTTCAATCCAAAAGCCCTTACACGTTGAG	
5941	TCAGCTGTGTATATTTTATACCTCAGGTAAGATTTGGATTTCCAGGCTTCTCTTATACTACTAGCTTTAAAAGTTCTCAGCCTCAGAC	
6031	TCCTGCCAGCGTCCCGGCTCCTGACTTCACATGTACCGTGTACCTAACGGTCTCTGAAGAGAGAGATCAAGTGTGCTTCCACTTGAAGT	
6121	GCACCTGGCATCTTGTCTAGATGTGTGTTAACAGAGTTCATTTGAAAGTTCACCTCCACTGGTTTAAAGTGCCTGCTGCTCTGTGA	
6211	ACCAGGCTTGCCTGTAGAGGATGGCTGGGAAGTGGGGGTTCCACTCACACAGACACAGCCCTGGAGGAGAACTTGCAGCTCTCAGTAAT	
6301	TGCACAGTACACTCAGCACTCTAGAGCTGTAATGCTTTTGTGTTGGAATAATAAATAGAGATGTGATGTGGCGTGTGGGACAGTGTATTA	
6391	TACATGGTTTTTTTGGCTTACTGTAGTAGAAAATGTCAAATTTCCAGCCACTGTCTATGAGAAGCAGTTGACAAAAATAAGTTTATGTTTAA	
6481	CAGTGTGGTTTACGTAGAGAACAATTTCTCTTCACTCTCGGTTTCAACATGTTTCTGAAATATCATGGTTTGTGTCATTTAACTTAAAT	
6571	TAGTTTTTTTTAAGATAAATAGAGTCTATGTTTAAATCAATAAAGAAAGGCCACATGAAATCTGCTGTCTGCCATTCATTTGTTA	
6661	CCACAGAACAATAAATTTCTATGTGTCATGAAGAAGTGTGGCTGAGACCTGGTCTATGTGTTTGGGAGACTCCGCTGCACCAATGT	
6751	TCCCAACAACAATAAATTTATTTAGTAACTTGACATCAAGAGCTGTGACAAATGAAAACAGCTCTTACACATGTTCTAACAGAAATC	
6841	TGATTAACATTTTGTGGATATGAAAATAAATTTTATGTAAGCTGGCACTGTTGCTGAAATTTAGAATGATTTCTTCTATGTACAGCA	
6931	TTGCAAAATGTGAATCATGTCTGTGAACAAATAAGAGATAAAGTGTGATCTTTGTGTAACATAAAACCCCAAAAAAAAAAAAAAAAAA	

Gezeigt ist die vollständige *Ftm* cDNA Sequenz (linke Randleiste: Nukleotid-Positionen) und die daraus abgeleitete Aminosäuresequenz des *Ftm* Proteins (rechte Randleiste: Aminosäure-Positionen). Farblich hervorgehoben sind die drei *coiled-coil* Domänen (blau), die C2-artige Domäne (grün) sowie die vermeintliche RPGR-Interaktionsdomäne (orange) des *Ftm* Proteins. Exongrenzen sind ebenfalls farblich markiert sind (grün: erste Base eines Exons; rot: letzte Base eines Exons). Drei klassische Polyadenylierungssignale (AATAAA) sind unterstrichen. Die Sequenzen der 'Exon-trapping'-Fragmente F1 und F4 sind durch eine gestrichelte Markierung hervorgehoben.

7.2.2 *Fto*: *Mus musculus* Fatso

NCBI-GenBank: AJ27917

1	CACGGCCACGGTGGCGAAGCGGCTTTAGTAGCAGCATGAAGCGCGTCCAGACCGCGGAGGAACGAGAGCGGGAAGCTAAGAAACTGAG	
	M K R V Q T A E E R E R E A K K L R	18
91	GCTCCTTGAGGAGCTTGAAGACACTTGGCTTCTACCTGACCCCAAGATGATGAGTCTATCAGCACTGGCAGCTGAAATACCCCTAA	
	L L E L E D T W L P Y L T P K D D E F Y Q Q W Q L K Y P K	48
181	ACTGGTTTTCCGAGAGGCCGGCAGCATACAGAGGAGCTGCATTAAGGAGTCCCCGAGGCTTTCTCACACTGCATAAGCATGGCTGCTT	
	L V F R E E A G S I P E E L H K E V P E A F L T L H K H G C L	78
271	GTTTCCGGGAGTGGTGGATCCCAAGCAAGATGTGCTACCCAGTGTCTCGCATCCGAGGACCCAGGCTGCACCTGACATGACAGTA	
	F R D V V R I Q G K D V L T P V S R I L I G D P G C T Y K Y	108
361	CFTGAACACAGACTCTTACGGTGCCTTGAAGGGCTGACCAAGTACACAGAGGCTGAGATCGCCGCTGCATGTCAGAC	
	L N T R L F T V P W P V K G C T V K Y T E A E I A A A C Q T	138
451	CFTTCAAAGCTCAATGACTACCTCAGGTGGAGACCTCCAGGCTTGGAAAGAACTGGCTGTGAGAGAAAGCCAAATGAAGACGCTGT	
	F L K L N D Y L Q V E T I Q A L E E L A V R E K A N E D A V	168
541	GCCACTGTGCATGGCAGAGTTCCCGAGGCGGGCTGTCGCGATGATGAGTGGACCTTAAAGAGCAGAGCAGCTACAAGCT	
	P L C M A E F P R A G V G P S C D D E V D L K S R A A Y N V	198
631	GACTTTGCTAAACTTTCATGGATCCTCAGAAAGATGCCCTACTTGAAGAGAGGACCCATTTTCGGCATGGGGAAGATGGCGGTGAGTGGCA	
	T L L N F M D P Q K M P Y L K E E P Y F G M G K M A V S W H	228
721	TCACGATGAGAACCTGGTGGACAGTCCAGCCGTGGCAGTGTACAGCTATAGCTGCGAAGGCTCTGAGGATGAAAGTGGACCGAGTCCAG	
	H D E N G L V D R S A V A V Y S Y S C E G S E D E S E D E S	258
811	CTTCGAAGCGAGATCTCTGATACTTGGCATGTTGGTTTAAAGATCTCATGGGACATCGAGACACCCAGGATTAACAATCTCCCTTACCA	
	F E G R D P D T W H V F F K I S W D I E T P G L T I P L H Q	288
901	GGGAGACTGCTATTTTACGTCTGATGACCTCAAAGCCACCACAGCACTGTGTTTTGGCTGGCTCACAGCCTCGGTTTGTTCACCTCA	
	G D C Y F M L D D L N A T H Q H C V L A G S Q P R F S S T H	318
991	CCGTGTGGCAGAGTGTCTCAACAGGCACCTTGGATTATATCTAGAAGCTGTGAGTGGCGCTGCAGAAATGTCCTCAATGACTCAGACGA	
	R V A E C S T G T L D Y I L E R C Q L A L Q N V L N D S D D	348
1081	TGGCAGCTCTGTTGAAATCTTTGATCTGCAAGTTTGAACAAGGAGGAAATCCATAATGAGTGGAGTTTGGAGTGGCTGAGGCA	
	G D V D S L P L K G G E I H N E V E F E V E L R Q	378
1171	GTTCTGGTTTCAAGGCACTCGATACAAACTTTCACCCGATTTGGTGTGAGCCATGACTCACCTGGAGGGGCTGTGGAAGAAGATGGA	
	F W F Q G S R Y K L C T D W W C E P M T H L E G L W K K M E	408
1261	GAGCATGCAAAATGCCGTCCGTAAGTAAAGAGAGGGGCTCCCGGTGGAACAAAGGAGTGAATTTCTGCTGCTCCATCTGGTCCC	
	S M T N A V L R E V K R E G L P V E Q R S E I L S A I L V P	438

1351	GCTCACCGTGCCGACGAACTGAGGAAGGAGTGGCATGCCA GG TGCCAGTCCCGAGTCGTCGGACTTTACCAGTACAGCAGAAACCAGA	468
	L T V R Q N L R K E W H A R C Q S R V V R T L P V Q Q K P D	
1441	CTGCGGCCATATGGGAGAAGGATGACCTTCCATGCCTCTGCCCTTTGACCTCACAGACGTGGTTCCGAGCTCAGAGGCCAGCTGCT	498
	C R P Y W E K D P S M P L P F D L T D V V S E L R G Q L L	
1531	GGAAGCAAGATCCTAGAGGAAGCACAAAGTCTTAGGTGGAGAAGGATA ATC CTTGGCTTTCTCTCAAACCTTGGACAAGGAGTAGT	502
	E A R S *	
1621	GTAGACTCTCTCAACCCAGTGGTTATAGGACCTGGTTCCAATTCAAAACAGCTAGAGAACAAGCCACAGATTCTGAAGTGTGAAAC	
1711	GTTACTGTGTTTGTAGTCTGACTTGGTAAATTTAGAGCACCCCTCCCTCCCCCAATATCTTCTACTCACCACCCCCACACACACCT	
1801	GAAGCCCTTTGTCTCCCAAGGCAGCAGCAGGGATGGGCTAAAGCTGTGACAAGCCACACTGCTCTCCCTTTGGCTGGCCACCTCAA	
1891	GACCTGGAAGTAGGCCTGGCCATAGGGTGACAGAAGCTAGAGACATCCATCCAGCTTCTCACCGTGGCTTCTAGTGGTTAGACCATCAA	
1981	GATGGCAGCCAGTGTCTCACCGCGGTGCACAGATAAAATTTAGCTCAGTTGTCATTAATAATCCATTCTCAGTTGCCCTGGCCGATG	
2071	CTGATGACTGTGCTTACAGCACATACAGATGTTTATGTGTTCCACATCTGTGAGCGAGCCTTCTCTCCAGCCCTTCTGTAGTGG	
2161	CTAAAACATGGGGGAAAACAGTGAAGGCAGTACCAGCGCTCACCTCTTTCTCTCTGAGAGGGAAGGGGCCATGAGCCCTGTGCACTG	
2251	CACACAGGAAGAGGAAGGAGCTACAAGGAGACAGCCAAACCTGTCTTTGCGGGGCTTCCCTGATCCAGTGACAGAGCCAGCCCTGGGA	
2341	CCACTGTCTGTGGAGCTAGTGTCTTCTCCTAGAATTTCCCACTAGATACTGACAGGTCGTATGGGGCTTGAAGCACCAGTTCCCA	
2431	ACCCCTGACGCTCAGAGCATCAGAGAGAGCATGGACTCTGATTTCAAAGTGCCAGGAGTAAGATCCACCCGAGCACTCCTCCATGACCT	
2521	TTGCAATGCTTGGGCTCAGTTTCTCATCTGTATGATTTTGGGAGGTACCCAAAATGCCAACTCCTAATGTGAAAATGACGTGATATG	
2611	ACAGAGCTACAGATCTAGAGTAATTTCTAGCATACCGCTGAATATCCGATACGTTCCAGTGTCTGTTACGTGAGCTGACCTGGAAGAAGCC	
2701	GCTTCTCTCCCCCCCCCCCCGGTCTGCAGTCTTTCCCGCTTGTGTCTCTGTCGGCCACATCTTTCCCTAGTGACCTGAGGCTTTT	
2791	TTGTGCTGAAAGGTTGACAAAGTTAGCCATGATCTCTGATTTGGCCGAGACCTGCAATTTGGAAGGACACAAATAGTTTGTGGTTT	
2881	TTGCTCATCCAGGATTCAGTTAAGGTGCCCTTCCGCTTAGAATCCACCATAGAAGCTGCATCCACAGCCCAAGGGCTTCTGGTG	
2971	CAGGGACCAAGGCTCTCAGACCTTGAATTTGTTCCCTTTGTCTCTGAGCCAGGACTTCTCATCAGTAGAGAGTAGGTTACATTTCTAC	
3061	GGGTTAAAATGGAACAACCTAAGTAACGGTATAGTTGGCTCATATGTCACAGACTGGCCGCTCTCCCATCCCCACCCATCTGGCTCTG	
3151	TGGTTTATTAGGCCCCGCTGATATTTCTTTGTCTATTTAATTTGGGTCGCAAGTAAATAATGAGATATTTCCACATAAAATAAAATGTC	
3241	TGCATTTCTTTAGAAGCCAGATACTCTGTAGCCCTGGGTTCTCATAATCGCATGGCGGTGGCTTGGCAATGCCCTGTGACCTTTGTGC	
3331	TCTGCTGTGACCTGTGTCTCCCTGTCCACATTTTGTAAACCTGCGTCCGTCACAGCAGCACTTCTAGAGCTGCTTGTCTGCATCTTAA	
3421	GTAGAAAGCAGTAGCGTTGAGTGTCTCAGTCCCTGCTCCTCCATTTGGTCTAAGTTTACGTGCTGCTTGTATGTCTGTGCTGTGTCATTT	
3511	GTTCATACACAATAAAGAATTAACGATGAAAAAATAAAAAA	

Gezeigt ist die vollständige *Fto* cDNA Sequenz (linke Randleiste: Nukleotid-Positionen) und die daraus abgeleitete Aminosäuresequenz des *Fto* Proteins (rechte Randleiste: Aminosäure-Positionen). Die 5'-Enden der identifizierten cDNA-Klone *J0125*, *B1598* und *N2210* befinden sich an den Nukleotid-Positionen 11, 43 und 800 (fettgedruckt und kursiv). Blau hervorgehoben sind die Aminosäurereste, die das vermeintliche zweiteilige nukleäre Lokalisationssignal am N-terminalen Ende des Proteins kodieren. Exongrenzen sind ebenfalls farbig markiert (grün: erste Base eines Exons; rot: letzte Base eines Exons). Zwei klassische Polyadenylierungssignale (AATAAA) sind ebenso wie die für die Herstellung des Derivat-Klons ΔJ verwendete *EcoRV*-Restriktionsschnittstelle (GATATC) unterstrichen. Die Sequenzen der 'Exon-trapping'-Fragmente 29/10 und 28/10 sind durch eine gestrichelte Markierung hervorgehoben.

7.2.3 *FTO*: *Homo sapiens* Fatso

NCBI-GenBank: BG258711, AB051539, U79260 (repräsentative Zugangsnummern)

1	TAGGCGGCTTTAGTAGCAGCATGAAGCGCACCCGACTGCGGAGGAACGAGAGCGCGAAGCTAA GA AAC TG AGGCTTCTTGAAGAGCTTG	24
	M K R T P T A E E R E R E A K K L R L L E E L E	
91	AAGACACTTGGCTCCCTTATCTGACCCCAAAGATGATGAATTTCTATCAGCA GT GGCAGCTGAAATATCCTAACTAATTTCTCCGAGAAG	54
	D T W L P Y L T P K D D E F Y Q Q W Q L K Y P K L I L R E A	
181	CCAGCAGTGTATCTGAGGAGCTCCATAAAGAGGTTCAAGAACCTTTCTCACACTGCACAAGCATGGCTGCTTATTTTCGGGACCTGGTTA	84
	S S V S E E L H K E V Q E A F L T L H K H G C L F R D V R	
271	GGATCCAAGGCAAGATCTGCTCACCCGGTATCTCGCATCTCATTTGGTAATCCAGGCTGCACCTACAAGTACCTGAACACCAGGCTCT	114
	I Q G K D L L T P V S R I L I G N P G C T Y K Y L N T R L F	
361	TTACGGTCCCTGGCCAGTGAAGGGTCTAATAAACAACACCCGAGGCTGAAATAGCCGCTGCTTGTGAGACCTTCTCAAGCTCAATG	144
	T V P W P V K G S N I K H T E A E I A A A C E T F L K L N D	
451	ACTACCTGCAGATAGAACCATCCAGGCTTTGGAAGAAC TT GCACCAAGAGGCTAATGAGGATGCTGTGCAATTTGTGTATGTCGTG	174
	Y L Q I E E T I Q A L E E L A A K E K A N E D A V P L C M S A	
541	CAGATTTCCCCAGGGTTGGGATGGGTTCACTACACCGGACAAGATGAAGTGGACATTAAGAGCAGAGCAGCATACAACGTAACCTTTGC	204
	D F P R R V G M G S S Y N G Q D E V D I K S R A A Y N V T L L	
631	TGAATTTTATGGATCTCAGAAAATGCCATACCTGAAAGAGGAACCTTATTTGGCATGGGGAATAAGGAGTGGCTGAGCTGAGCTCATGATG	234
	N F M D P Q K M P Y L K E E P Y F G M G K M A V S W H H D E	
721	AAAATCTGGTGGACAGGTGACGGTGGCAGTGTACAGTTATAGCTGTGA AG CCCTGAAAGAGGAAAGTGAAGATGAGCTCATCTCAGGAT	264
	N L V D R S A V A V Y S Y S G C E G P E E E S E D D S H L E G	
811	GCAGGGATCTGATATTTGGCATTTGGTTTAAGATCTCATGGACATAGAGACACCTGGTTGGCGATACCCCTTACCACAGGAGACT	294
	R D P D I W H V G F K I S W D I E T P G L A I P L H Q G D C	
901	GCTATTTTATGCTT GA TGATCTCAATGCCACCCCAACACTGTGTTTGGCCGGTTCAACCTCGGTTTAGTTCCACCCACCGAGTGG	324
	Y F M L D D L N A T H Q H C V L A G S Q P R F S S T H R V A	
991	CAGAG TC CACCAGAACTTGGATATATTTTACAACCGCTGCTCAGTTGGCTCTGAGAAATGCTGTGACGATGTGGACAATGATGATG	354
	E C S P G T L D Y I L Q R C Q L A L Q N V C D D V D D V	
1081	TCTCTTTGAAATCTTTGAGCCTGCAGTTTGAACAAGGAGAAGAAATTCATAATGA AG TCGAGTTTGTAGTGGCTGAGGCAGTTTGGT	384
	S L K S F E P A V L K Q G E E I H N E V E F E W L R Q F W F	
1171	TTCAAGGCAATCGATACAGAAAGTGCACTGGTGGTGTCAACCTGAGCTCAACTGGAAGCAAGTGGGAAAGATGGAGGGT GA	414
	Q G N R Y R K C T D W W C Q P M A Q L E A L W K K M E G V T	
1261	CAAATGCTGTGCTTCATGAAGTTAAAAGAGAGGGGCTCCCGTGGAAACAAAGGAATGAAATCTTGACTGCCATCTTGGCTCCGCTCACTG	444
	N A V L H E V K R E G L P V E Q R N E I L T A I L A S L T A	
1351	CACGCAGAACTGAGGAGGAATGGCATGCCA GG TGCCAGTACCAATTTGCCGAACAATACCTGCTGATCAGAAGCCAGAATGTGCGG	474
	R Q N L R R E W H A R C Q S R I A R T L P A D Q K P E C R P	
1441	CATAC TT GGGAAAAGGATGATGCTTCGATGCCTTGGCGTTTACCTCACAGACATCGTTTTCAGAACTCAGAGGTCAGCTTCTGGAAGCAA	504
	Y W E K D A S M P L P F D L T D I V S E L R G Q L L E A K	
1531	AACCCTAGAAGAGCACAAAGTCTCAGCGGAGGAGAAAAGAGATCGGCTTTCTCTCCAACGTTGTCATGGCTTAAAGCAAGAGACT	505
	X *	
1621	GGAGACTTCTCTTGGCCCTTAGATTGTAGCACCCGGTCCCAATTCAAAACAGCTAGGAAATGGTGCCATGAAGTTTAAATGTTTAA	
1711	AATGACCCCTGTGTTATAGTCTGATTTGGTGTAAACAGGACCTTCTTCCCCAAAATTTGTTAGATTAATAAATGTGAGCCATTCAGCC	

```

1801 CCAAGGTCCAGGGCAGGCGACAGGAACGAGCCAGCGTGTGACAAAGCCTAACCTACTTTCCCTCTTTCCCAAGCTTTTTCAGAGACTCTG
1891 GAGTGGACCCAGCCCTCTGGGGAAAGACAGAACCTAGAGACATCCCAAGTTACTCACACACCCATAGTGTGTCCAATATGGTAGCCACT
1981 AGCTAGCTGTGGCTACTTCAATTTAAATTCAGTTTAAATTTTAAATTAATAAATGCAGCTCTTCAGTCCGCCCTGGCCACATTTCAAGTGCTT
2071 AACAGCCTCATGTGGCTAGTACTGCTGTATTTGGACGGTACAGATATGGAAACATTTTCATCATCGAAGAAAGTCCATTGGACAACTT
2161 CTATAAAAAATTTGAGAGCGAAATTCATTTCCATTCGTCTGTAGCTTCTATCCCAAAGGCAAAGAACTAAAAAGAGAAATGACTCA
2251 TTGAAGATGGCCTCTTTCCCTTCTCTAAGACAAACCTAAGTAAAAGCCTGACTTTGAGTCCATATGCTCAGCCAGGGAAAGGAGATGTTAA
2341 TAATTTAAATAAAGGTGGATATCCCGTCTTTAGGGAGTTCCCTTGATCTCTGAAAGAGACACAGCCCAATTTACATTTATTTCTGGGAT
2431 TTCACCAGCATAGTATAGTTTTCCTGTAGTCCCTCATTTCTATGTAATAACAGGTGGAACAGGTTTGAAGAACCCTCAGTGGCCCA
2521 TCCTGATGACATTTGGAGACTCAAAGAGACAAGAGAGAGTAGGGTTTAAACCTGAGCTTTAAGACTCCCACCTAGCTTCGTGTCTTTGGC
2611 ATGTTAACCGTCCCTCAGTTTCCCTCATCTGTATAATGGGGATATATGAAAGGCCACAGTCCCTAAGGTGAAACATTAAGTGAAGATGATTCAG
2701 TTACAGACTTAGAACAAATTTCCAGCACAATAGTTAAATATCCAGGAAATTTCTGGTACTGTTATGTTGGGTGAGCTGACCTGGATGTAGAT
2791 GTTTTCTCTCTCTTGTGTACCCTCCGCCAGTTTGTCTGTGTATGCCATTAACACATCTCTCCCTTTCTGACCTGGCTCTGCCCAT
2881 GGTGTCCCAAGAAATCGTGAGAAATAGTTAGCCCCCTCTCCCGAGCTGTGTCTTCTCGTGTAGTTGTTTACAGTAGTTGAGAAAGTTGA
2971 AGAGCTTTTGGCTATTTGAAGGTGCATGAGATAAACTCTTCTGCCACCAGAAATGACAGTGGTTACAGGCCCTGCACCTCATTTCCATGA
3061 ATGCAGTAAATAGCCACAGAAATGTCACATTAAGCAAAGCAGCAGGGTCTCATCTGTGTTGAGACTCGAGTCTCTCAGACCTTGGATCA
3151 TTCCCTGGTGTCTTTGAGCCTCAGTTTCCCTCATTTGTTAAAGAGAAGTGAAGCAGTGTCTCACAGGGTCTATACAGAGATTAATGAAT
3241 AAATGAAATAACATAGACCAGGGGGCGTGGTGTTTAAAGTCAAGATGGGGCACCCTCGGCCATCCAGCCAGTGTTTTCTTTTAGCC
3331 CCTATGATGTTCAATTTTGTATATAATCCATTAGGTGCCCATATTTAAAAATTTGGGAGATTTACATAAAAAATTAAGGTCATTCATTTT
3421 TTTTCTCTTTCTTTTCTTTTCTTTTGTAGACACAGTCTCACTCTGTCCAGGCTAGAGTGCAGTGGCAGGATCTCAGCTCACTGCAACC
3511 TCTGCCCTCCAGGTTCAAGTAAATCTCTGCTCAGCCTCCCAAGTAGCTGGGACTACAGGCAGTGGCCACCGCCCACTAATTTTGT
3601 TATTTTACAGAGATGGGGTTTACCACATTTGACCAGGATGATCTCGATCTCAACCTCGTATCCACCCACCCTCGGCTCTCCAAAGCGC
3691 TGGGATACAGGCGTGGCCACCGCCCAAGCAAGTCTGCATTTTCTTTAGAACTCAGAACCCCAATAGTCCTAGGCCCCCATCCT
3781 CGCATGGCAGCAAGCTAAATAAGCATCTTCCCACTGCGAGTTGGGGCATGACCCAGCCTATGGTTTGGCATACTCCCTCTTTTCTCCGT
3871 TTTTTCATTAATTTGAACTGACCTGCATCACCTTTTCATGTCAGTGCCTCCAAACCTGCTTGTCTGACCCCTTAGTCGAAATATT
3961 TGTGCTTACCCAAATATGTTGTGTGACTATTGAACCTATTTCGTAGACTGCTTGTATCTAATGTCATTTGCATCATAAAAATTTTCATAT
4051 CCAATAAACATATTAAGGATGAGATAAAGAACCCGAAAAAATAAAAAAAAAAAAAAAAAA

```

Gezeigt ist die menschliche *FTO* cDNA Sequenz (linke Randleiste: Nukleotid-Positionen) und die daraus abgeleitete Aminosäuresequenz des *FTO* Proteins (rechte Randleiste: Aminosäure-Positionen). Die Aminosäurereste, die das vermeintliche zweiteilige nukleäre Lokalisationssignal am *N*-terminalen Ende des Proteins kodieren sind blau hervorgehoben. Exongrenzen sind ebenfalls farbig markiert (grün: erste Base eines Exons; rot: letzte Base eines Exons). Drei klassische Polyadenylierungssignale (AATAAA) sind unterstrichen.

7.2.4 *Irx3: Mus musculus Iroquois* Homöobox Gen 3

NCBI Genbank: AW913488, Y15001 (repräsentative Zugangsnummern)

```

1 CTGAGCGCCTATCGCAGAGCGGCTCGCCGGCCTAGAGGAGGGAGTGTGCGCAGAGAGGAGGGCTTGAGGCCCGCTAGAAATGTCAAT
91 CAGAGCCCGGAGCCAGGGAATCTCGGCCAATCTGTTCGGACCTGACCTGACTCCTCGCTGCCGCCGCTGCCGCCGCCGCCGCCGCCG
181 GGAGCAGATCAATAGCGCAACCGCGGAGCGCAGCGAGCGCGGCCCGGGCTCGGCCAGCCCGCGATGCGCTCCGGAGCCGGGGTCCGGAG
271 CTGAGCAGAGCTCGGGCTCGGGCCCGCCAGTCCGCTGCGCCCGGGCCACCAATGTCCTTCCCCAGCTCGGATACCAATACAT
M S F P Q L G Y Q Y I 11
361 CCGCCCTCTACCCCGGAACGCCCGGGAGCTCGGGCGGCGTGGAGCGGCGGAGTACGCTTGGGGCCGGAGCGGTCCCGGCCCGG
R P L Y P P E R P G A A G G G G G S S A G G R S G P G A 41
451 AGCCTCCGAGTGGCCGCTCTGGGTCCTTATCCAATGTCTTTCATCAGTGTACGGGGCACCTACGCCGCTGCCGAGCTGGCCGCCG
A S E L A A S G S L S N V L S S V Y G A P Y A A A A A A A 71
541 CGCCGCTCAGGGTTAGCGGCCCTTCTGCCCCAAGCAGGAGTGGCCATTTCCACAGCTGGCGCTCAGTATGAGCTGAAGGACAG
A A Q G Y G A F L P Y A T E L L P I F P Q L G A Q Y E L K D S 101
631 CCCTGGCGTCCAGCATCCCGCCACCGCCCGGCTTCCCGCCAGCCCGCCCTTCTACCCATATGGCCAGTACCAGTTCCGGGACCC
P G V Q H P A T A A A F P H P A F Y P Y G Q Y Q F G D P 131
721 GTCCGTCCTCCAGCAACCGCCCGGAAAGCAGGACGAGCTCAAGGCTGGCTCAACAGCAGCCCGCAAGAACCCGTACCCCAAGGG
S R P K N A T R E S T S T L K A W L N E H R K N P Y P T K G 161
811 CGAGAAGATCATGTTGGCCATCATCAAGATGACCTCACCCAGTGTCCACCTGGTTGCGCAACGCGCCCGGCCCTCAAGAAGGA
E K I M L A I I T K M T L T Q V S T W F A N A R R L K E 191
901 GAACAAGATGACGTGGCTCCCGCTAGTGCAGGACGAGGAGGGCAATGCTTATGGGAGCGAGCGGGAGGAGGAGGACGAAGAGAA
N K M T W A P R S R T D E E G N A Y G S E R E E D E E D 221
991 CGAGGAAGAGAGCAACCGCAGCTGGAGATGGAGGAGGAGGACTCGCAGGGAGAGGAGGAGGACACGGGGGGCGAGGGGCTGGCCGACG
E E E S K R E L E M E E E L A G R G G G H G G R G A G R R 251
1081 ACGACGAGATGAGGAGATCGAATTTGAAAACTTAGACAGCGCGGACCGGGTCCGAGCTGACCTGGCTGGGGCCGCGCACAGGAACCG
R R D E E I D L E N L D S A A A G S E L T L A G A A H R N G 281
1171 CGACTTCGGCCCTGGGACCCATTTCCGACTGCAAACTAGCAGCTCGGACGACAGCTCCGAAGGCTTGGAGGACCGACCCTGTCCGTCT
D F G L G P I S D C K T S D S D D S S E G L E D R P L S V L 311
1261 GAGCCTGGCCCCACCGCCCGCTGGCCAGGGCTCTGCACTCCACCTTCTCCACCTCCAGCCTGGATCCCTGCGCTCCGGCCCC
S L A P P P P V A R A P A S P P S P P S S L D P C A P A P 341
1351 GCGCCCTCTCCGCCCTCCAGAAGCCCAAGATCTGGTCACTGGCCGAGCGCCACCAGCCCGGACACCCACCGCTCCGCTCCCGG
A P S S A L Q K P K I W S L A E T A T S P D N P R R S P P G 371
1441 AGCCGAGGTTTCGCCCTCCGGCCGACCGCTTGGCCCGGACGCTGCAGCTCTCGCCCGGCGCTGCAGCGGCTGCAGCGGCACACAG
A G G S P P G A A V A P P T L Q L L S P A A A A A A A A H R 401
1531 ACTCGTTCGGCCGCTGGCAAAATTCGCCGTTGGCAACAGGCTTTCCAGGCGCGCCGCGGTCGGCCGCGACCCGCTGTC
L V S A P L G K F P A W T N R P F P G P P A G P R P H P L S 431
1621 CATGCTGGGCTCGGCCACAGCACCTGCTGGGACTTCCCGGAGCCGACAGTTCATCCCGCTGCGGGCCGAGCCGCTATGCTCGCCCTGC
M L G S A P Q H L L G L P G A A G H P A A A A A A Y A R P A 461
1711 CGAGCCGAGAGTGAACAAGTCTGCTGTAGTGCCTTGAAGTGGAGAAAGATTAACAAGACAGCTTCCAGCCGCTGCAAGGCCGCC
E P E S G T D R C S A L E V E K K L L K T A F Q P V P R P P 491
1801 ACAGAACCCTGAGGAGCTCTGCTCTTATCAGCTCTCTGCTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAA
Q N R L D A A L V L S A L S S S * 507
1891 AAACA AAAA CAAA AATCGCTCTGTATAGTGACAGATTTGCAAGATGCTGTGTATGATTACCTTAAAAAGAACTGAAACAACAAC
1981 AAACAACAACAACAACAGGAAAGGACAAAAGAAAGAAAGAAAGCCAGTCCCCCACCCTTGGCCCTCCCCAAGTCAACCCACATTTAG
2071 ATTTGTGAGTTCCTCTGTTGTGTATTTCCGGGGGGGGGAGTGGGTTTACGGAGAGAAACAGCCTGCTGACTCTTTTGTATATAC
GGGAAAATGTACATACGTGTAACCAAAATGTACAAGAAATCTATTTTTGGCTAAATAATGAGCTGTTGCCACTTTTGACTATAA
2251 AAAAAAAAAAAAAAAAAA

```

Gezeigt ist die *Irx3* cDNA Sequenz (linke Randleiste: Nukleotid-Positionen) und die Aminosäuresequenz des *Irx3* Proteins (rechte Randleiste: Aminosäure-Positionen). Farblich hervorgehoben sind die Homöodomäne (blau) und die Iro-Domäne (orange) des *Irx3* Proteins. Exongrenzen sind ebenfalls farblich markiert (grün: erste Base eines Exons; rot: letzte Base eines Exons). Ein klassisches Polyadenylierungssignal (AATAAA) ist unterstrichen.

7.2.5 *Irx5: Mus musculus Iroquois Homöobox Gen 5*

NCBI-GenBank: BE995453, AJ271053, AJ271054, AV230325 (repräsentative Zugangsnummern)

```

1 AAAGCAGAGCCCCCGGTCCAAAGAGCGAGGGAGGGGGTTGGGGGAGAGAAAGCCTAGCCCCGGGCAAGCGGGGCGCGCAGAGGAGCGGGC
91 GCGGCGGTTCGACGACAGAGGCGCGCGGGAAGCCAACCAGGAGGCGCGCGGGCCGGAGCCCGGGAGCCCGGGGCTCGAAGAGCTCGGCCCC
181 GGGCCAGCCAGAAGCCAGGTGCCCTCTCGCTCGCCCTCGCAGGGCCCGCCCGGCTCGTTGGCCCGCCGCGCGCGCCCGCCCATGCC
271 GTGTGTGGCCATGTCTACCCCGCAGGGCTACTTGTACCAGCGCTCCGCCCTCGCTGGCGCTCTACTCGTGCCCGCATACAGCACCAGCGT
M S Y P Q G Y L Y Q P S A S L A L Y S C P A Y S T S V 27
361 CATTTTCGGGGCCCCGACGGATGAGCTCGGCCGCTCTCTTCGGGCTCCGCGTCTCTCGCCCTACGCTGGCTCCACTGCCCTTCACGGCTCC
I S G P R T D E L G R S S S G S A F S P Y A G S T A F T A P 57
451 CTCGCCCGGCTACAACCTCGACCTCCAGTACGGCGCCGACCCCGGGCTCGGGCCGGCTGCCTTCTTACGTGGCTCTCCCTACGA
S P G Y N S H L Q Y G A D P A A A A A A A A F S Y V G S P Y D 87
541 CCACACTCCCGGTATGGCAGGTTCCTTTGGGATATCACCCCTTATGCAGCACCCTCTGGGCTCCTATCCTTATGGGACCCCGCTACCGGAA
H T P G M A G S L G Y H P Y A A P L G S Y P Y G D P A Y R K 117
631 GAATGCCACTCGGGACGCTACAGCCACTCTCAAGCCTGGCTCAACGAGCACCGAAGAACCCTACCCACCAAGGGAGAGAAGATCAT
N A T R D A T A T L K A W L N E H R K N P Y P T K G E K I M 147
721 GCTGGCCATATACCAAGATGACCCCTCACCCAGGTGTCCACCTGGTTTGCACACGCGCGCCGCGCCTCAAGAAAAGAGAACAAGATGAC
L A I I T K M T L T Q V S T W F A N A R R R L K K E N K M T 177
811 GTGGACGCCACGGAAACCGAGCGAGGAGGAGGAGGAGAACATCGATCTGGAGAAGAACGACGAGGACGAGCCACAGAAGCCCGA
W T P R N R S E D E E E E E N I D L E K N D E D E P Q K P E 207
901 GGACAAGGGCGACTCGAGGGCCCCGAATCAGGAGGAGCAGAGCAGGCGACTGCGGGCTGCGAACCGCTGCGAGGGGGCCGCTCAGCCC
D K G D L E G P E S G G A E Q K A T A G C E R L Q G P L S P 237
991 CGCCGGCAAGGAGACGAAGGCAGCCTCAGCGACTCGGATTTAAGGAGTCGTCTCCGAAGGCCGCGCAGATGAGCTGCCAGGCCCCC
A G K E T E G S L S D S D F K E S S S E G R H D E L P R P P 267
1081 GCGCGCGGGGAGTCTCCCGGCTGGGCGGCAACCGCTCGTCTGGCGGAAGACGCGGGTCTCACTACCCCGCAGTGTGCCGGCTCC
R A G E S S P A A G P A T A R L A E D A G P H Y P A S V P A P 297
1171 CGGCCGACCCATCAGCCGAGAGCTGCCCCCGGCTCGGGCGGATCCTGGTTCATCCACTCGCCACCCGACCTCCGCCACCCGCTCC
G P H P S A G G E L P P G S G G S S V I H S P P P P P P P P 327
1261 GGCCGTGTAGCAAGCCAAACTGTGGTCTCTGGCCGAGATCGCCACATCTCGGACAAGGTCAAGGACGGGGCGGGAGGGAGCGAGGG
A V L A K P K L W S L A E I A T S S D K V K D G G G G S E G 357
1351 CTCTCCATGCCACTGCCAGGGCCCATGGCGGGCAACCGCTTGGAGCCAGCCGCGCTCTCCGCCCCAGCGCCCGCACGCTCACC
S P C P C P G P M G Q T L G G S R A S P A P A P A P A P 387
1441 CTCGCGCAGTGTCCCTTTCGGGGCGGACGGTGTGTCCCGGCTCTTTATTACACGGCGCCCTTCTATCCGGGTACACGAACTATGG
S A Q C P F P G G T V L S R P L Y Y T A P F Y P G Y T N Y G 417
1531 CTCCTTCGGACATCTTACGGCCACCCAGGGCCAGGGCCAAAGCCTACAGCGGGTCTGGCTCTCATTTCAATGGATTAACCCAGACGGT
S F G H L H G H P G P G S P T A G P G S H F N G L N Q T V 447
1621 GTTGAATCGTGGGCTTTTGGCTAAAGACCCGAAAATGTTCGGAGCCAGTCTCAGCTAGACCTGTGCAAGACTCCCCCTATGAGTT
L N R A D V L A K D P K M L R S Q S Q L D L C K D S P Y E L 477
1711 GAAGAAGTATGTCCGACATTTAACTCTGGTCCGATATCCCGGACTTTCCTAAATTTATAAAACCATGGCCTTGGCAGTTATTTTTC
K K G M S D I * 484
1801 ATCACCAGAGAGAGAAAGAAAGAAATAAACTACCCCTTCTTATTTCAGAAGTTTATAGTTTATGGAGATGGATGGCATATAAATGTAT
1891 ACTTCTCCACACTCGAAAAAATGTTTTAGCCAAACCG

```

Gezeigt ist die *Irx5* cDNA Sequenz (linke Randleiste: Nukleotid-Positionen) und die Aminosäuresequenz des *Irx5* Proteins (rechte Randleiste: Aminosäure-Positionen). Farblich hervorgehoben sind die Homöodomäne (blau) und die Iro-Domäne (orange) des *Irx5* Proteins. Exongrenzen sind ebenfalls farblich markiert (grün: erste Base eines Exons; rot: letzte Base eines Exons). Ein klassisches Polyadenylierungssignal (AATAAA) ist unterstrichen.

7.2.6 *Irx6: Mus musculus Iroquois Homöobox Gen 6*

NCBI-GenBank: AJ271055

```

1 GGCACGAGACGCCCTGCTATGGCCTTTTACCTTTTGGACACCCGATGGCAGCAGCATCCCACTTCTGGTGTCTGCAAGTTCCAGTGGC
M A F S P F G H P Y G S T S Q F L V S A S S A 24
91 ACTTGTCTGCGAAACCGCCCTCGGTCAAGTGTGAGTGTGGCCCTCAGCCCTCCACTCTGCCTCTACTCTTTGCTGTACACCCATGACAGT
T C C E T A P R S V S D V A S A S T S A S T L C C T P Y D S 54
181 CGCTGCTGGGCAAGTCTCGGCCAGAGCTGGGAGCTGCCCTTGGGCATCTATGGAGCACCCTATGACAGTCCAGAGTTACCCCTGGGTAC
R L L G S A R P E L G A A L G I Y G A P Y A A A Q S Y P G Y 84
271 CTGACCTATGGCCAGAGCCACCCACACTGTGTGGCGCACTGAATCCTCAGTATGAGTTCAAGGATGCTGCAGGAAGCTTTGCCCGGACG
L T Y G P E P P T L C G A L N P Q Y E F K D A A G S F A P S 114
361 CTGACAGCCAGGGCCCTACTACCCCTATGAGACAACTCTGGGGCAGTATCAATACGACAAGTATGGCGGAGTGAATTTGAGCAGTGTCT
L T Q P G A Y P Y E T T L G G Q Y Q Y D R Y G G V E L S S A 144
451 GGCCGAGAAAGAAATGCCACAAGGGAGAGCACCAGCGCACTGAAGGCTGGCTCCACGAGCACCGAAGAACCCTACCCACCAAGGGC
G R R K N A T R E S T S A L K A W L H E H R K N P Y P T K G 174
541 GAGAAGATCATGCTGCCCATCATCAACAAGATGACCCCTACCCAGGTGTCCACCTGGTTCCCAACCGCGCGCCCGCCCTAAGAAGGAG
E K I M L A I I T K M T L T Q V S T W F A N A R R R L K K E 204

```

631	AACAAGATGACATGGGCGCCCAAGAACAAGGGAGGGGAGGAGAGGAAAGCAGACAGTGGAGAAGACTCTCTGGGATGCCTGAATGGTGAC	234
721	ACCAAAGATGCTACTGCCAGCCAGGAAGCCCGGGGGCTGAGGCTGAGTGCCTGGAAGACCTAGAGGAAGAGGAGGAGAGGAAGCA	
811	GAGGAGGAGGACGGGCTCAGCAGCAGCAGGCTGGCGGATTTTCAGAAGAGCAGCAACCCCTGCCTGCCTCCCTGCCTCCCGCCAG	264
901	GAGGGATGCTGGAGAGCAGAGAGTGCAGGCTACCCCGTTCTTTTACTGAGGCCCAATCAGGGGAGAGCTGACTTATC	294
991	ACAGCAGAGCCAGGCGGCCACGATGATCTTACACTACCCAAAGTGGCCACAACCCCGCATTGGTCCCTTAGCTCACACTGCGGCAGCC	324
1081	AGCGCTGTCGAAAGTCTCCCTCAACTCCGCCAGGGCACAAGTCCAGAGTGCACATGATTTCCAGACAGCCAGTTCTATCAGGCGA	354
1171	CTCCTGGTACCCAGAGACTCCGAAGGCGAAGAGGATTTCTCTGAGCCTAAAGCTTTTGGAAACTCCAGCTTCCACCTGCAGGGGCTGCCA	384
1261	CTGAAGTGCACACACAGAACCACTGAAAGTGCAGTCCAGTACCCATCTGGAGCAGAAAGTTAGTCAATGGCTGCGGT	414
1351	TTGAGAAGACTTTTGGAAATGGCTGCTCTCTGATCCACATTCATTAAGAAGTTTCCAGACCTTGGCATGGAGCTCTCACTT	438
1441	TGCCTGTGACAAACACAGAACCACTGAAAGTGCAGTCCAGTACCCATCTGGAGCAGAAAGTTAGTCAATGGCTGCGGT	
1531	AGGAGAGAAGGTTGGAGACTTGGGGGAAAGCCCTGTGTGCTCAGAGTACTTTGACCAAGTCTCTCTCTCCATGAAATGGAGGC	
1621	CCTGAGCAGCATCTCTCTGCTCACCCACCCCTGCCACCTGAGTGAACCTTTCTGCTGAAATGGTCTCCAAGGACTGCTGTG	
1711	TGTCTCTCTCTATTTCCCGGCTTTACCTTCAAAAATAAACTTTGGT	

Gezeigt ist die *Irx6* cDNA Sequenz (linke Randleiste: Nukleotid-Positionen) und die Aminosäuresequenz des *Irx6* Proteins (rechte Randleiste: Aminosäure-Positionen). Farblich hervorgehoben sind die Homöodomäne (blau) und die Iro-Domäne (orange) des *Irx6* Proteins. Exongrenzen sind ebenfalls farblich markiert (grün: erste Base eines Exons; rot: letzte Base eines Exons). Ein klassisches Polyadenylierungssignal (AATAAA) ist unterstrichen.

7.2.7 *Mmp2*: *Mus musculus* Matrix Metalloproteinase 2

NCBI-GenBank: NM_008610

1	GCCAGAGACTCAGGGTGACACGCGGAGCCCGGGAGCGCAACGATGGAGGCACGAGTGGCTGGGGAGCGCTGGCCGGACCTCTCGCGGGT	
91	TCTCTGCGTCTGTGCTGCTGTGGGCGCGCCATCGCTGCACCATCGCCATCAAGTTCCCGGGGATGTCGCCCTAAAACAGA	16
181	CAAAGAGTTGGCAGTCAATACCTGAATACTTTCTATGCTGCCCAAGGAGAGTTGCAACCTCTTTGCTGTAAGATACCTCAAGAA	46
271	GATGCAGAAGTCTTTGGGCTGCCCCAGACAGGTGACCTTGACAGAACCCATCGAGACCATGCGGAAGCCAAGATGTGGCAACCCAGA	76
361	TGTGGCCAACTACAACCTTTCCCGCAAGCCAAAGTGGGACAAGAACAGATCACATACAGGATCATGGTTACACACCTGACCTGGA	106
451	CCCTGAAACCGTGGATGATGCTTTTGGCTCGGGCTTAAAGTATGAGGACGCTCACTCCGCTGCGCTTTTCTCGAATCCATGATGGGA	136
541	GGCTGACATCATCAACTTTGGACGCTGGGAGCATGGAGTGGATACCCGTTTGTGGAAGGATGGACTCTGGCACATGCTTTGC	166
631	CCCGGCACCTGGTGTGGGGGAGATTTCTCACTTTGATGATGATGAGCTGTGGACCTGGGAGAAGGACAAGTGTCCCGTAAAGTATGG	196
721	GAAAGTGTGGCGAGTACTGCAAGTTCCCTTCCCTGTTCAACGGTCCGGAAATACAGCAGCTGTACAGACACTGGTCGAGTGTGCTT	226
811	CCTCTGGTCTCCATCACTTTGAGAAGGATGGCAAGTGGCTTCTGCCCCATGAAAGCTTGTTTACCATGAGTGGTGGCAATGC	256
901	TGATGACAGCCCTGCAAGTTTCCGTTCCGCTTCCAGGGACCTTCTACAAAGCTGTACCACCGAGGGCCGACCGACGGCTACCGCTG	286
991	GTTGGTCCACCCAGGACTATGACCGGGATAAAGAAATGAAATCTGCCCCGAGACCGCTATGTCACCTGTGGGTGGAAATCAGAAGG	316
1081	TGCCCATGTGCTTTCCCTTCACTTTCTGCGCAACAAGTATGAGAGCTGCACCAGCGCCCGCCCAACGATGGCAAGGTGTGGTGTGC	346
1171	AACCACAACCAACTACGATGATGACCGGAAGTGGGCTTCTGTCTGACCAAGGATATAGCCTATTCTCGTGCAGCCCATGAGTTCCG	376
1261	CCATGCCATGGGGCTGGAACTCTCAGGACCTGGAGCTGTGATGGCCCTATCTACACCTACACCAAGAACTTCCGATATCCCATGA	406
1351	TGACATCAAGGGATCCAGGAGCTTATGGGCCCTCCCTGATGCTGATACTGACACTGGTACTGGCCCCACCCGACATTTGGGACTGT	436
1441	CACTCCGGAGATCTGCAACAGGACATTTGCTTTGATGGCATCGCTCAGATCCGTTGGTGGATCTTCTTCTCAAGGACCGGTTTATTTG	466
1531	GGGACAGTGCACACCGTACAGCCAGGCTCCCTTGGTGGTCCACATTTCTGGCCTGAGCTCCAGAAAAGATGACCGTGTGTA	496
1621	TGAGGCCCCACAGGAGGAGAAGGCTGTGTTCTTCCGAGGGAATGAGTACTGGGCTTATTCTGCCAGCACTCTGGAGCGAGGATACCCCAA	526
1711	GCCACTGACCCAGCTGGGTTGCCCTGATGTCAGCAAGTAGAGTCTGCTTTAACTGGAGTAAAGAAAGAAAGACATACATCTTGC	556
1801	AGGAGTCAAGTTCTGGAGATACAATGAAGTGAAGAAAGAAATGGACCCGGTTTCCCTAAGCTCATCGCAGACTCTTGGAAATGCCATCCC	586
1891	TGATAACCTGGATGCTGCTGGACCTGCAGGGCGTGGTCACTGACTGATCCCTTCAAGGGTGTATTACTGAAAGTGGAGAAACCAAG	616
1981	TCTCAAGAGCGTGAAGTTTGGAAAGCATCAATCAGACTGGCTGGGCTGCTGAGCTGGCCCTGTTCCACCGGGCCCTATCATCTTATCCG	646
2071	TGCACACAGGTTGAAGGATGTGAAGCAGCTGGCGCTCTGCTCTCTGTAGTTAACCAGCCTTCTCTTCACTGGTACTTCAAGT	662
2161	TTAAGAGGTTGGCTTTTGTGCCCCAAAGAAAGGTTGCTGACTGATCCCTCTGGTGTGCTTCTCTTCTTCCCTCCCTAGGGGAT	
2251	GCTGGATTTTGCATGACCGCTACTCTGGGCTGCCCTGGTGTCTCACTTCTGTTTCTCAACATCTATGACCTTTTATGGCTTTT	
2341	CAGCACTCTCAGAGTTAATAGAGACTGGCTTAGGAGGGCACTGGTGGCCCTGTTAACAGCTGGCATGGGGCAGTGGGGTACAGGTGTGC	
2431	CAAGGTGGAATCAGAGACACTGGTTTCAACCTTTCTGCTGCCAGATACCTGCACCACCTTAACTGTGCTTTGTAAGCCCTCGCT	
2521	CGTTTCTTCAACTTTTTCAGTTTTCACCTCCACTGCATTTCCCTCCAAAGGACTCGGGTGTCTGCATCGCTGCATGATGCATCTCA	

2611 GCCCGCCTAGTGATGGTTCCCTCCCTCACTCTGTGCAGATCATGCCAGTCACTTCTCCACTGGATGGAGGAGAACCAAGTCAGTGGCT
 2701 TCCTGCTCAGCCTTCTTGTCTTCCCTTTAACAGTTCCCATGGGAAATGGCAACAAGTATAAATAAAGACACCCATTGAGTGACAAAA
 2791 AAAAAAAAAA

Gezeigt ist die *Mmp2* cDNA Sequenz (linke Randleiste: Nukleotid-Positionen) und die Aminosäuresequenz des *Mmp2* Proteins (rechte Randleiste: Aminosäure-Positionen). Farblich hervorgehoben sind die zweigeteilte zinkbindende katalytische Proteinasedomäne (grün), die drei Fibronectin Typ II-Domänen (rot) sowie die drei hämopexin-ähnlichen Domänen (blau) des *Mmp2* Proteins. Ein klassisches Polyadenylierungssignal (AATAAA) ist unterstrichen.

7.2.8 KIAA0308/2FLJ: *Homo sapiens* KIAA0308/2FLJ

NCBI-GenBank: AK022240, AK022582, AB002306, AW954913 (repräsentative Zugangsnummern)

1 GATATATGGAAGATTTAAACACCATTCAATCTTTGTTTTTCCACAAAGGACCGTCCCTCTTAATATTTGGAAGAAAAAGATCAATCC
 91 AGGGACTATAAAAATGGCAATCAACTCAGGGAATATCAACTGGAAGGGCTCAACTGGCTCTTGTTCATTTGGTACAATAGACGAAACATGCG
 181 ATCTTAGCAGATGAAATGGGCTTGGCAAAACTATTAATCAATACATTCCTCTATGAAATCCTTCTGACTGGTATAAGAGGACCTTTC
 M G L G K T I Q S I T F L Y E I L L T G I R G P F 25
 271 CTGATTATTGCTCCACTTTCTACTATTGCAAACTGGGAGAGAGAAATTCGTACGTGGACTGATATTAACGTTGGTTTATCATGGGAGC
 L I I A P L S T I A N W E R E F R T W T D I N V V V Y H G S 55
 361 CTGATTAGCAGACAAATGATACAGCAATACGAGATGTACTTCAGGATTACAGGGCGTATCATTTCAGGAGCTTACAGATTCCAAGCC
 L I S R Q M I Q Q Y E M Y F R D S Q G R I I R G A Y R F Q A 85
 451 ATCATCACCACCTTTGAAATGATCTTGGAGGCTGTGGAGAGCTTAATGCAATGGAATGGCGATGTGTGATTTGATGAAGACACATAGG
 I I T T F E M I L G G C G E L N A I E W R C V I I D E A H R 115
 541 TTAATAAATAAATAATGTAACCTTAGAGGGCTGAAACTCATGAATCTGGAACACAAGGTGCTTTTGGACTGGCACCCCTTCCAAAAT
 L K N K N C K L L E G L K L M N L E H K V L L T G T P L Q N 145
 631 ACAGTTGAAGAACTATTAGTCTTCTTCACTTTCTGAAACCCTTAAGGTTTCTCTTGAATCAACATTTATGCAAGAATTTGGGGATCTG
 T V E E L F S L L H F L E P L R F P S E S T F M Q E F G D L 175
 721 AAAACAGAGGAACAGGTACAGAATTCAGGTTATCTGAAACCAATGATGTTGAGACGATTAAGAAAGAGATGTGGAAGAAAGATTTGGCA
 K T E E Q V K Q L V I L K P M M L R R L K E D V E K K L A 205
 811 CCTAAAGAAGAAATCAATGGAAGTAGAACTTACTAATATTCAAAAGAAATACTACCGGGCTATCTTGGAAAAGAACTTTCTTTTAA
 P K E E T I I E V E L T N I Q K K Y Y R A I L E K N F S F L 235
 901 TCCAAAGGACGAGCAAACTAATGTACCTAATCTGGTCAATACCAATGATGGAGCTCAGGAAATGTTGTAATCATCCATCTTATAAAA
 S K G A G Q T N V P N L V N T M M E L R K C C N H P Y L I K 265
 991 GGTGCTGAGGAGAAATACTTGGAGAAATTTAGAGATACTTACAATCCAGCTGCTTCTGATTTTCATCTTCAAGCAATGATCCAGTCTGCT
 G A E E K I L G E F R D T Y N P A A S D F H L Q A M I Q N S A 295
 1081 GGTAATTTGGTCTTATGATAAATGCTTCCAAAATGAAAGCCGAGGTCATAAAGTCTCATCTTCTCAAAATGGTTTCGCTTGCCTT
 G K L V L I D K L L P K M K A G G H K V L I F S Q M V R C L 325
 1171 GACATCTGGAGGACTATCTCATACATAAAGATACTTATATGAGCGAATGATGGCAGAGTCAGAGGAAATCTTCGGCAAGCTGCTATA
 D I L E D Y L I H K R Y L Y E R I D G R V R G N L R Q A A I 355
 1261 GATAGATTTAGTAAACCTTCAGATAGATTTGTTTCTCTGTGTACCCGAGCTGGTGGGTTGGGCATCAACTTAACTGACAGCTGAT
 D R F T S K P D S D R F V F L L C T R A G G L G I N L T A A D 385
 1351 ACATGTATAAATTTTGTGATCTGATTTGGAATCTCAGAATGATCTTCAGGCCCCAAGCTCGTTGCCACAGAATGGTCAGAACCAAGCAGTT
 T C I I F D S D W N P D N G L Q A Q A R C H R I G Q N K A V 415
 1441 AAAGTCTACAGACTGTAACCTCGTAACCTCATATGAGAGAGAGATGTTGAGCAGGAGCTTGAAGCTGGGCTTAGATAAAGCTGTGTTA
 K V Y R L V T R N S Y E R E M F D R A S L K L G L D K A V L 445
 1531 CAGAGCATGAGTGAAGAGAAAGTAAATGTTGGTGGTATTCAGCAGCTTTCCAAAAGGAAATAGAAGATCTGCTTCCAGAGGTTGCTTAT
 Q S M S G R E E S N V G G I Q Q L S K K E I E D L R R G A Y 475
 1621 GGTCTATTATGGAGGAAGAAGATGAAGGCTTAAATCTCGAAGAGGATATCGATCAGATTTTACTACGTCGTACAAAACACTATTACA
 G A I M E E E D E G S K F C E E D I D Q I L L R R T K T I T 505
 1711 ATTGAATCAGAAGGAGTGGTCAACATTTCCAAAGGCGAGTTTGTGGCATCTGGAAACCCGGACAGATATTTCTTAGATGATCCCAAC
 I E S E G R G S T F A K A S F V A S G N R T D I S L D D P N 535
 1801 TTTGCGCAAAATGGGCTAAAAGGAGGAGAAATAGATATAGAGGCCATCAGTGGCAGAAACCGGCTGGTATTGACACTCCAAGAAATFAGG
 F W Q K W A K K A E I D I E A I S G R N G L V I D T P R I R 565
 1891 AAGCAACAGACCTTTTGTAGTCCAAAAGATGAATTTGCTGAAATATCTGAAGCTGAAAGTGAAGGAGATGAAAGCCCAACTCCGG
 K Q T R P F S A T K D E L A E S E A E S E G D E K P K L R 595
 1981 AGACCTCTGACCGTTCCAAATGGCTATGGAAGAACTGAATGCTTTAGAGTTGAGAAAAACCTGCTAGTTTATGGGTGGGGCCGATGGAGA
 R P C D R S N G Y G R T E C F R V E K N L L V Y G W G R W R 625
 2071 GAGATCTATCTCATGCGCTTTCAAAGGACGCTAAATGAACACGATGTAGAGATAAATTTGCCGAGCTCTCTTAGCATATTTGCTTGT
 E I L S H G R F K R Q L N E H D V E I I C R A L L A Y C L V 655
 2161 CACTACCGAGGAGATGAGAAGATTAAGGTTTTCATATGGGATCTCAATTACTCCAACCTGAAGATGGACAGACACGAGAGCTACAGAATCAT
 H Y R G E K I K G F I W D L I T P T E D G Q T R E I Q N H 685
 2251 CTAGGCTATCAGCTCTGACCGGCTCGAAAAGGAGAAAGTAAATACTCAAACAGCTCATTGATATACAAAAGCAGAAATGG
 L G L S A P R G R K K V K T Q T S S F D I Q K A E A W 715
 2341 CTTCGAAAATATAATCCGAGCAGCTCCTCAAGATGAAGGCTACAAAACATATAAACACACCTGTAATAAGGTTTGTCTTCTGTTG
 L R K Y N P E Q L L Q D E G Y K K H I K H H C N K V L L R V 745
 2431 AGAATGCTGTATTATCAAAGCAAGAAGTTATTGGAATGAGTGTGAGAAAGTATTTGATGGAGTTGATGCAAGTGACATTTGATGTTGG
 R M L Y Y L K Q E V I G N E C Q K V F D G V D A S D I D V W 775
 2521 GTACCAGAACCAGACCTCAGAAGTTCTGCTGAGTGGTGGATTTTGTGCTGATAAGTCACTCCTTATTGGAGTTTAAACATGGA
 V P E P D H S E V P A E W W D F D A D K S L L I G V F K H G 805
 2611 TATGAAAATATAACACTATTTCAGCAGACCCAGCATTTAGCTTCTTGGAAAGAGTGGGAAAACCTGATGAGAAAGCAGTTGCTGCTGAA
 Y E K T I R A D P A L C F L E R V G K P D E K A V A E 835
 2701 CAGAGCGAATGATATATGATGGAGTGGAGATGGAGATCCAGAATACAACCTGCCCGACGCTTTTAAAGATGATATAGAGGATGAT
 Q R A N D Y M D G D V E D P E Y K P A P A P A I F K D D I E D 865
 2791 GTTTCCTCACCAGAGATCTTGTATAGCAGATGGAGATGGTCAACTGATGGAGGGTATAAAGTATATTGGCCTACTCAAGCTTAA
 V S S P G D L V I A D G D G Q L M E G D K V Y W P T Q S S A L 895
 2881 ACCACAGCTTTGAGGCGTCTCATCTGATACCCAGCGTACTAATAAAAACAGACAAATTCACAGATACAACCCGACTTCTCGGCTCCCT
 T T R L R R L I T A Y Q R T N K N R Q I Q Q I Q P T F S V P 925
 2971 ACCAGTGTAAATGCGCCTATTATGAGGAAGCCACTCTTAATCTAAAATGGCAGCAAGATGAGAAAGACAGCAAGATGGACAAAGAA
 T S V M G P I Y E E A P T L N P K M A A K I E R Q Q R W T R R 955
 3061 GAAGAAGCTGACTTTTATAGGGTTGTATCTACATTTGGAGTGGTTTTGGACCTGACAGAGGCGCAATTTGATTGGACAAAATTTAGAGCT
 E E A D F V R V S T F G V V F D P D R G Q F D W T K F A 985
 3151 ATGGCTAGGCTACATAAGAAAATGATGATAGTTTGGAAAATATTTGTACGCATTCATGTCATGTCGGAGGGTTTGTCTTCTTCT
 M A R L H K K T D D S L E K Y L Y A F M S M C R R V C R L P 1015
 3241 TCCAAAGAAGAAATTTGGTGGATCCAAATATTTTATCCAGCCATCACAGAAGAACGTGCTTCTAGGACTTTGTATCCCATTTGAAGCTTCTA

S K E E L V D P N I F I Q P I T E E R A S R T L Y R I E L L 1045
3331 AGGAAAGTACGGGAAACAGGCCCTTCGACATCCACAGTGTGTTGAACCGCTTGAAGCTTGGCCATCCAAATCCAGATTTACAGTCTGGTGG 1075
R K V R E Q A L R H P Q L F E R L K L C H P N P D L P V W W
3421 GAATGTGGCCCTCATGATAGGATTTGCTTATTTGGTGTGCCAAACCGGGGTGAGCCGAACAGACTATCACATTTCTCGTGATCTCTGAA 1105
E C G P H D R D L L I G A A K H G V S R T D Y H I L R D P E
3511 CTCTCATATTTATGGCAGCTCAGAGAACTACAGTCAAAGTAAAGTGGCTATCAAGGACTTCTACCCCACTTCTACAGCAATATCAAGTA 1135
L S F M A A Q R N Y A C Q S K M A H S R T S T P L L Q Q V
3601 GCATTTCTGCTTCTCTCTTACCTCTCTACCTAGGCTCTAGATGCTAAAGGTATTATTTCTAGAGGAGATGAAAGTTAAAGTGAAGAAC 1165
A L S A S P L T S L P R L L D A K G I I L E E M K V K S E N
3691 CTTAAAGAGGAGCCTCAGTCTCTGAAAGAAATCTATGCTTCTGTGAAACCAGGACACTGATAAAATCTGAGCCTGTAAGTCCAAAG 1195
L K E E P Q S S E E E S M S S V E T R T L I K S E P V S P K
3781 AATGGTGTTTTACCAGGCTACTGGAGACCAGAAATCTGGTGGAAATGTGAAACAGACAGCGCATGGTTCAGCCAGAACAGAACCC 1225
N G V L P Q A T G D Q K S G G K C E T D R R M V A A R T E P
3871 CTAAC TCCAAACCAGCTTCTAAGAAACCAAGAGTCCACAAAAGGCTCAGAATCTAGTCTTCTGATTTCTGACTCAGATTTCTGAGAGACTCA 1255
L T P N P A S K K P R V H K R G G S S S S D S D S S E R S
3961 TCTTGTCTTCCAGATCATCTTCTCTCATCATCTCTTCTTCTGCTCCCCTCTCGATCAGGCTCTAGTCTTCTTCTCATCTCTCATCTGT 1285
S C S S R S S S S S S S S S S S S H S R S S G S S S S S S
4051 TCTTCAGCATCTTCTCATCTCTTCTTCCACCTCTTCT 1315
S S A S E E S D S D E E
4141 GAAGCCAAAAGAGAGAACTACTACTCATGAAAGCCCTATGATGAAGAAAGCGTCCGCTCAGTGCAGACTACCCAGGATGAGACTCAG 1345
E A Q K R E S T T H M K A Y D E E S V A S L S T T Q D E T Q
4231 GATAGTTTTCAGATGAACAATGGGACACCAGATCTGCTTATATCTTACAAGGTGGATATATGCTGGCAGCCTCGTATTGGCCAAAGGAT 1375
D S F Q M N N G T P E S A Y I L Q G G Y M L A A S Y W P K D
4321 CGTGTGATGATCAATAGGTTGGACAGATTTGTCAAACAGTCTGAAAGGAAAGTGGCTTCCAGCTAGAAGAAGTTATGATGCTTAACACA 1405
R V M I N R L D S I C Q T V L K G K W P S A R R S Y D A N T
4411 GTGGCTTCTTCTTATACCACAAAACGTGGACAGCCCTGGAGCAGCTACAGAATACAGCGAGCCAGTGTACCCACTCCCCAGGTGCC 1435
V A S F Y T T K L L D S P G A A T E Y S E P S V P T P P G A
4501 GGTGTTAAAGAAAGAAATGATCAGTCAACACAGATGTCAAAGGAAAGTGGTGTGAAGTTGACATTTTCAGAAAGCAAGGCTTGGCTCAGAAA 1465
G V K E E H D Q S T Q M S K E G G L K L T F Q K Q G L Q K Q
4591 AGACCAATTTGATGGTGAAGACGGTCTGCGGCGAGCAGTACCTCACTCGGCTTCGAGAGCTTCAAAGTGCATCAGAGACCAGCCCTC 1495
R P F D G E D G A L G E Q Q Y L T R L R E L Q S A S E T S L
4681 GTCAATTTCCAAAATCCATACCAGTATCAGGTACTTCCATCAACTCAACCCCTTGGTGCCAATGGTGTGATATTAGACAACAGCCCTATA 1525
V N F P K A S I P V S K G T S I Q C P T L G A N G V I L D N Q P I
4771 GTCAAAAGAGCGGAGGAAGGAGGAAGAATGTAGAAGGTGTGACATCTCTTTTTTAAACAGAAATAAACCCACTAATCATGTTTCTTTA 1555
V K K R R R R R K N V E G V D I F F F N R N K P P N H V S L
4861 GGCTTAACTCTCAGAGATTTCCACAGGATAAACTCCAGCAGTACTTCTTACTCAACTCAAGGAAATCTCTGATACAGAAAGTCCAGTT 1585
G L T S S Q I S T G I N P A L S Y T Q P Q G I P D T E S P V
4951 CCAGTTAATTAATTAAGAGTGAACGAGACTTGCAGGAGATGATGCCAAAGAGAAAGGATTTGGAAATGGCTTAAGGACACCCG 1615
P A V I N L K D G T R L A G D D A P K R K D L E K W L K E H P
5041 GGTATGTGGAAGATTTGGGAGCTTTATCTTAGAATGCAGCTTCAATGAGGGAAGACCCAAACAAAAGACACCGTTGCAGAAACCC 1645
G Y V E D L G A F I P R M Q L H E G R P K Q K R H R C R N P
5131 AATAAATAGATGTAAATAGTCTCAGTGGAGAAGAGCTGTCAACTGATTAACAGAAAGAAATGCTAGAAGGTTGGAGGTGCATTTGCT 1675
N K L D V N S L T G E E R V Q L I N R R N A R K V G G A F A
5221 CCCCCTTTGAAAGATTTATGTAGATTCCTAAAAGAAAATTCAGAAATATGGAGTAGCTCCGAAATGGGAGATGTGTGTTAAGCAATCTGGA 1705
P P L K A D L C R F L K E N S E Y G V A P E W G D V V K Q S G
5311 TTTCTTCCAGAGCTGTGAACAGTATCTCAGTCCGCTTGTGGAGAGGAAGTAAGCAGCGGGGAGACCGCTAAAAGTGGGA 1735
F L P E S M Y E R I L T G P V V R E E V S R R G R R P K S G
5401 ATTGCAAGGCCACAGCAGCAGCTGCTGCAATCTGCCACCAGTGTTCAGGCAATCTTTGTTAGCCAAATGGACTACTTCCAGGTGTG 1765
I A K A T A A A A A A A S A T S V S G N P L A N G L P G V
5491 GATCTCACAACTTCCAGCCCTTACAACAAAACCTCAAAAACCTGCACTGCAAGTAACTGCTGGGTGATGGGAATGCCATCCGGC 1795
D L T T L Q L Q L Q N L Q N L Q S L Q V T A G L M G M P T G
5581 CTTCTTCTGGAGGAGAAGCTAAAACATGGCTGCTATGTTCCCAATGCTGCTGTCAGGAATGGCTGGATTACCAAATCTGTTGGGCATG 1825
L P S G G E A K N M A A M F P M L L S G M A G L P N L L G M
5671 GGAGGACTCTCAGAAAGCTACGGAATCTGGGACAGAAAGCAAAAAGGAAAGTGAAGTGAAGGAGTGAAGGAAAGCAAGAAAGGACA 1855
G G L L T K P A T E S G T E D K K G S D S K E S E G K T E R T
5761 GAGAGCCAAAGTTCAGAGAAATGGTGGAGAAAACCTGTGTCAAGTCTCCTCCGCATCCTACTGCTGCATTAATACAGCTGCAGCT 1885
E S Q S S E N G G E N S V S S S P S A S S T A A L N T A A A
5851 GCCAACCCATTAGCTTTAACCCACTATTACTATCTAAATATCTTATCCAGGATGCTTCTCACTCCAGCCCTAATCTTCATATTCPCA 1915
A N P L A L N P L L S N I L T P G M L L T P G L N L H I P
5941 ACTTTGTCAGTCCAAATCTTTGATGTACAAAACAAAACAGTACTTAGGCTCGCTTAAGTCTGTAGAAGTAAAGAAAGAGATTC 1945
T L S Q S N T F D V Q N K N S D L G S S K S V E V K E E D S
6031 AGAATTAAGACTCAGAAAGCAAGGAGGAACTGAACCAAGTCTCTCAATGAAAACAGCACAGATGAGGTTTCAGAGAAAGCTGATGCT 1975
R I K D Q E D K G G T E P S P L N E N S T D E G S E K A D A
6121 TCATCTGGATCTGATAGTACATCGTCTCATCTGAGGATTCAGATCTAGTAAATGAAGACTGATTCAGACTTAAAATATGA 1995
S S G S D S T S S S S E D S D S S N E D *
6211 ACTGATTTGGATTTTCTTAAATAATTAATTTGAAATACCCAGTGTGAGTGCATCAATAACTTACTGACCGAACATTTACGTTATTT 6301
TGTTTAGAAGTGCAACTGCTTTCAGAGACTTTTGCATGTAATTTCTTAAAGTATAAGTTTCTGAACTGATGACTACAAAT 6391
ACATAAAGGTGTAATAATACAAACAAAGGCATTAATAATTTGTTGGGGTTAATTTTATGAAATTTATGCTCAATAAGAGTTGATATTT 6481
AATATATTTGCGAGTGAACACAGAACTTTATGCATATTACTGATTTTAAATTTGAAATAGTATTACAGCCCTCCCTTGACACCTTAATAAT 6571
CAGATCAAACCTCAGCAATAAATTTGGGCAGCTAATGAATGTCATGAAAGCTGTAGAATCTACATCACCATCCATTTGCTTTAATACATGA 6661
AAATGCTCTAGTGTGTGATGCACTGCTGATGTTCCAAATTCAGGTACAAGTATGTTTAAAGAAAGAAATAAGTTTCCCAATCAGCCAAAT 6751
TTAATGCGTACCTGTTACCTCAGTGAAGTTAGTTTGGAGGTTACATCTGCTTCTAAATCTATACTGTTTTCAGGGGTTTAAAC 6841
ACATCTATATATCATGCTCAATCTGGCAAGAAATATGACTTGTCTTTTGTGAGCTTAAATCAGATATCAGTAAATTAAGTCAATAAAT 6931
AATCAATGTCATGTCATTTGGCACCTTATAGACATATAGTTTTAACTTTTCAAAGTTTGGCTTCTTATTAAGAAATAATCATGCTC 7021
AGATGAGTAAATGCTGTTTCCAGGGTTTCAGAAAAGGCAAACTCATGAAATGGCACTGAAAAGAACTTTCAACACAGCACTACTGATGTA 7111
AAGAAATGTTGTTGCTTCTTGTGTAGATTTCTATTTGTTTATGTCATGGAATATTTCCAGAAATTAACAGATAATAGTGGTAA 7201
AGTAAATAGCAGATAGTCTAAATTCATTTGAGTTCAGGTGTAAGCAGACTAAATTTGGCCAGAAATCAGTGTGGGTTATCAGTTTA 7291
TATTAATATATCTAGTTGCGGCTTTTGAATAATCTCAAAGAGTGTACAAGTTATACCTGTAACACAAAAGTGA 7381
AGCCTGAGGCTCTGTTCAATTTCAATAGACTCTTTACCATTGAAAATTTGCTGATATTTGATTTGTGATACAATTTCTCTGCTAAAAG 7471
CTGCTATTTCTGACAAGGTAGAGGTCAGGTTCACTTTATATATATTTAAACAAATTAAGTACTGAAATGGACATAAAATAATAGACA 7561
TTCATAAGGAGATAATAGTTAGCATTTTCTGGTACTCAAAATAGTTAGTAAAGTCTGCAAGGCAATAAATTTAGGGGAAAGT 7651
GTCCAGTCTCTTCCACAGAAAATACTTTCCAGTATGTTTGAATAAACTGTGCTTTGTCATGAGTTAGTCAATTTAGTCTAGGTTT 7741
CCAAGACCTTTACCAGTAAATATAGTTTCTGTAATGTAATAAACCCTTATTAGAGAGACAGTGTATATGTAATTTACAAAATATATATA 7831
GTTCCATTTGGATTTGATTTTGTATTTTCCAAAATAGTACTTTGAAATGATAGTCTTTATGCAATGCTTACGAAATAGTCTCTA 7921
TAATGCCATCCAGGAGAGTGGGTAGTAATCTTCCATCATGAAATGATATATACATATTTAGTATCTTCCCTTTGCGATTTAGGACT 8011
TTTGTTTAACTAGAAATACACCTATGAGATAGCCAAAGTTTCAAACACAGTTATCTTAGTTTACCGTGGAGATTTCAAACACCAACCA 8101
TTCCTCTCTCTCTCAATTTACCCATGATCTTTATCTCTCTTCCCAATTAAGAAAAGGAAAAGGAAAAGGAAAAGGAAAAGGAAAAGG 8191
CTTGTGCACTAAAATCTAATTTATATCAAAATTTAGAGAAAGTATTTTCTCAATTTATGCTCAAAATAATTTGGTTAACTCTGATGTA 8281
TTCCTTTCTATAATAAAGCAATACAGTTTCAAAGCATTAAGTCAACATAAATTTAAACATCTCTTAGGTTTCAAAGACACTTCT 8371
ATTTAATATTCATTTGGGAAAAGTTTGTCCAGTATCAGCTTAAGCAACATGCAATAATGTTGTGTAAGTAAAGGTTAAGGAA 8461
AAAAATCAGTAGAAATACATAATACTAAAGTTGAGTGTGAAAGATATCAAGTATGTTGTGATGTTACTAAAGAAATTTAGCTGTT 8551
ATTGCCCTGTATTTATAGCCCTTGTTCAGGTTTATGATCAAGTCTTAGTCCAATCTTTCTTTGGACATTTGCAATATTTACCGATT 8641
GTGTTTGGTAGTCTGAAATTTGCTTTCTGTAGTTGACAAACCTTTAAAAGGCTAATTTGTAATTTAATAATTTCTATGATGTT 8731
CTATAGAGCAAAAGGAGTTAATATTTTATTAACATATTTTACT


```

3421 ACTATTGTAACACGTGTGGTAAACCTTCTCTTTCGAGTGCCTGCAGATCCACGAGAGAACTCACACTGGAGAGAAGCCCTTTGCTT
  Y C N T C G K T F S S S S A L Q I H E R T H T G E K P F A C 1163
3511 GCACTATCTGTGGAAGAGCATTCACAACAAAAGGCAATCTGAAGGTCCACATGGGCACCCATATGTGGAACAGCACCCCTGCACGCCGGG
  T I C G R A F T T K G N L K V H M G T H M W N S T P A R R G 1193
3601 GCCGGCGGCTCTCAGTGGATGGCCCATGACATTTCTAGGAGGCAATCCTGTCAAGTCCAGAAATGTTCCAGAAGGATCTAGCGGCGA
  R R L S V D G P M T F L G G N P V K F P E M F Q K D L A A R 1223
3691 GGTCAAGGAAGCGGGATCCGTCCAGCTTTTGAACCAGTATACAGCAGCACTGTCCAACGGGCTGGCCATGAAGCCAACGAGATCTCTG
  S G S G D P S S F W N Q Y T A A L S N G L A M K A N E I S V 1253
3781 TCATTCAGAACGGCGGCATCCCTCCAATTCCTGGAAGCCTGGGCAGTGGGAGCTCACCTATCAGTGGGCTGACTGGGAACGTGGAAAAAC
  I Q N G G I P P I P G S L G S G S S P I S G L T G N V E K L 1283
3871 TGGCAATTCGGAACCCAGTGCCTCTGGCTGGCCTGGAGAAAAATGGCGAGCAGCAGGAGAACGGAACCAACTTCCGTTTCACCCGCTTCG
  G N S E P S A P L A G L E K M A S S E N G T N F R F T R F V 1313
3961 TGGAGGACAGCAAAGAGATCGTCACAAGCTAAAGCTGCCAGAGTGCCTCTCAGAAATGTGCCCGAGGTTGCCTCCTGCCCTCCCC
  E D S K E I V T S * 1322
4051 CGCCCTTTACCTTCTGGCCCTGGGACTATGAACTACACTATGAAGACATTCCTTTGTATGTTGTTCAACCTCTAGAGTCTAAGAAAGC
4141 TTATTTATTAGTGCTATAACCTTGCCTTGCAAACAGAGTGCAGGCATTAACCTTGGTCTTCTGTATTTGGACTAAATACTAATGGACTA
4231 GAGTGTCTCTAGCTTGCTGTAACATTTATGGCAATCGCAAGTTGCCCTGCTAGGCGGCTTAAATCCGGCATTAACTTATTTTCTATATCC
4321 AGTTTAAATGAATCTGGTGTGTGATGCAATGCCTCCGTGATACGTTAGATCTCTAATAAAGTCTGTATATAAAATGTAATAAAAAAAAAA

```

Gezeigt ist die *Sall1* cDNA Sequenz (linke Randleiste: Nukleotid-Positionen) und die daraus abgeleitete Aminosäuresequenz (rechte Randleiste: Aminosäure-Positionen). Farblich hervorgehoben sind die Zinkfingerdomänen (rot) und die glutaminreiche Region (blau) des Sall1 Proteins. Ein klassisches Polyadenylierungssignal (AATAAA) ist unterstrichen.

Danksagung

Mein besonderer Dank gilt Herrn Prof. Dr. Ulrich Rüter für die Überlassung des Themas, die ständige Diskussionsbereitschaft und die anhaltende Unterstützung über den gesamten Zeitraum der Arbeit.

Bei Herrn Prof. Dr. Heinlein möchte ich mich herzlich für die bereitwillige Übernahme des Korreferats bedanken.

Besonders möchte ich mich bei Dr. Renate Dildrop für die angenehme Zusammenarbeit, entspannende Fachsimpelien, die gemeinsame Ergründung der unglaublichen Weiten der vergleichenden Genomanalyse und auch für die Durchsicht des Manuskripts bedanken.

Ein herzlicher Dank gilt auch Tanja Schöttler und Peter Sikorski, die mir mit vielen aufmunternden Worten sowie mit Rat und Tat zur Seite standen.

Ein großes Dankeschön geht ebenfalls an alle Mitarbeiter des Instituts für Entwicklungs- und Molekularbiologie der Tiere der Heinrich-Heine-Universität Düsseldorf und auch an alle ehemaligen Kollegen am Institut für Molekularbiologie der Medizinischen Hochschule Hannover, die mich ein gutes Stück meines Weges begleitet haben.

Mein ganz besonderer Dank gilt Andrea, meinen Eltern und meiner Familie, die während der gesamten Zeit immer hinter mir gestanden haben, und die es wohl kaum glauben können, dass diese Arbeit endlich ein Ende gefunden hat. Gut dass es Menschen gibt, die mich ab und zu an die vielen kleinen aber wichtigen Dinge im Leben erinnern.

Erklärung:

Hiermit versichere ich, dass ich die vorliegende Dissertation eigenständig, ohne unzulässige Hilfe und nur unter Verwendung der angegebenen Hilfsmittel und Quellen angefertigt habe. Die Dissertation wurde weder in dieser noch in einer ähnlichen Form bei einer anderen Institution vorgelegt.

Thomas Peters

Düsseldorf, 04. Februar 2002

Teile dieser Arbeit wurden bereits veröffentlicht:

Peters, T., Ausmeier, K. & Rütter U. (1999). Cloning of *Fatso (Fto)*, a novel gene deleted by the *Fused toes (Ft)* mouse mutation. *Mamm. Genome* 10: 983-986.

Peters, T., Dildrop, R., Ausmeier, K. & Rütter, U. (2000). Organization of mouse *Iroquois* homeobox genes in two clusters suggests a conserved regulation and function in vertebrate development. *Genome Res.* 10: 1453-1462.

Houweling, A.C., Dildrop, R., Peters, T., Mummenhoff, J., Moorman, A.F., Rütter, U. & Christoffels V.M. (2001). Gene and cluster-specific expression of the *Iroquois* family members during mouse development. *Mech. Dev.* 107: 169-174.

Peters, T., Ausmeier, K., Dildrop, R. & Rütter U. (2002). The mouse *Fused toes (Ft)* mutation is based on a 1.6 Mb deletion including the entire *Iroquois B* gene cluster. *Mamm. Genome* 13: 186-188.