„Vergleichende Sequenzierung und Analyse eines ca. 250 kb großen Bereiches der humanen Chromsomenregion 11p15.3 und der homologen Region der Maus“

Dissertation
zur Erlangung des Grades
„Doktor der Naturwissenschaften“

am Fachbereich Biologie
der Johannes Gutenberg-Universität
in Mainz

Andrea Cichutek
geboren in Weil am Rhein
(Baden-Württemberg)

Mainz, 2001

I Inhaltsverzeichnis

II Abbildungsverzeichnis

III Tabellenverzeichnis

IV Abkürzungsverzeichnis

A	Adenin
A. bidest	Aqua bidestillata
Abb.	Abbildung
abs.	Absolut
Ac	Acetat
Acc. Nr.	„accession number“
AS	Aminosäure
bp	Basenpaar
BSA	bovines Serumalbumin
C	Cytosin
cDNA	komplementäre DNA
CEGF1	„epidermal growth factor“-ähnliches Gen
Ci	Curie
cpm	„counts per minute“
DEPC	Diethylpyrocarbonat
DNA	Desoxyribonukleinsäure
dNTP	2´-Desoxynukleosid-5´-triphosphat
DTT	Dithiothreitol
EDTA	Ethylendiamintetraacetat
EST	„expressed sequence tag“
EtOH	Ethanol
FISH	„fluorescence *in situ*-hybridization“
G	Guanin
h	Stunde
HUSAR	„Heidelberg Unix Sequence Analysis Resources“
IPTG	Isopropyl-β-D-thio-galactopyranosid
kb	Kilobasenpaare
konz.	konzentriert
LINE	„long interspersed nuclear element“
LTR	„long terminal repeat“
M	Molar
MaLR	„mammalian LTR-retrotrasposon“
Mb	Megabasenpaare
MER	„medium reiteration frequency interspersed repeat“
min	Minute
MIR	„mammalian-wide interspersed repeat“
MOPS	3-(N-morpholin)propansulfonsäure
OD	optische Dichte

ORF	„open reading frame“ (offener Leserahmen)
PBS	„phosphate buffered saline“-Puffer
PCR	„polymerase chain reaction“
PIP	„percent identity plot“
RNA	Ribonukleinsäure
RT	reverse Transkriptase
RT-PCR	„reverse transcriptase polymerase chain reaction“
sec	Sekunde
SDS	Natriumdodecylsulfat
SINE	„short interspersed nuclear element“
SSC	„standard saline citrat“-Puffer
T	Thymin
Tab.	Tabelle
TBE	Tris-Borat-EDTA-Puffer
TEMED	N,N,N´,N´-Trismethyl-ethylendiamin
Tris	Tris(hydroxymethyl)aminomethan
TRITC	Tetramethylrhodaminisothiocyanat
U	„units“ (Enzymeinheit)
RT	Raumtemperatur
ü. N.	über Nacht
Upm	Umdrehung pro Minute
UTR	„untranslated region“
UV	Ultraviolett
Vol.	Volumen
X-Gal	5-Bromo-4-chloro-3-indolyl-β-D-galactosid
3´-UTR	3´-nicht translatierte Region
5´-UTR	5´-nicht translatierte Region

1 Einleitung

In den letzten 20 Jahren lag einer der Schwerpunkte der humangenetischen Forschung auf der Identifizierung und Charakterisierung bisher unbekannter Gene. Die daraus gewonnenen Erkenntnisse haben zu dem Verständnis der molekularen Ursachen genetisch bedingter Krankheiten beigetragen und stellen die Voraussetzungen für eine Verbesserung der pränatalen Diagnostik und der Entwicklung therapeutischer Ansätze dar. Zur Identifizierung neuer Gene wurden hauptsächlich drei unterschiedliche Ansätze angewendet: die funktionelle Klonierung, die positionelle Klonierung sowie der Kandidatengen-Ansatz.

Die ersten krankheitsrelevanten Gene wurden mit Hilfe der Kenntnis über einen hauptsächlich biochemischen Defekt, den ein bestimmtes krankheitsauslösendes Protein besitzt, identifiziert (funktionelle Klonierung; *McKusik*, 1991). Die Kenntnis der chromosomalen Lokalisation des Gens ist bei diesem Ansatz ohne Bedeutung, da zunächst ein Enzymdefekt aufgeklärt wurde. Anschließend erfolgte die Identifizierung des Gens und des krankheitsverursachenden Gendefekts. Ein Beispiel hierfür ist die Identifizierung des Gens *PAH*, welches für das bei der Phenylketonurie defiziente hepatische Enzym Phenylalaninhydroxylase kodiert (*Lidsky et al.,* 1985; *Kwok et al.,* 1985). Einen konträren Ansatz verfolgt die Methode der positionellen Klonierung. Hierbei wird ein Gen einzig aufgrund seiner Position innerhalb einer umschriebenen chromosomalen Region, die zuvor durch Kopplungsanalysen in vielen betroffenen Familien als Kandidatenregion identifiziert worden war, kloniert. Unter Verwendung dieses Ansatzes konnte das Gen für die Chorea Huntington-Krankheit auf die humane Chromosomenregion 4p16 eingegrenzt werden (*Gusella et al*., 1983). Um anschließend mögliche Kandidatengene innerhalb dieser eingegrenzten chromosomalen Region zu identifzieren, können putative Gene u. a. anhand von CpG-Inseln identifiziert (*Bird*, 1987) oder durch markierte cDNA-Klone direkt selektiert werden (*Korn et al.,* 1992). Auch die Methode der Exonamplifikation (*Buckler et al*., 1991) stellt eine sehr effiziente Methode zur Isolierung neuer Gene dar. Zur Erleichterung der Identifizierung neuer Gene bzw. deren Transkripte auch für den Ansatz der positionellen Klonierung wurde das humane EST-Projekt initiiert (*Adams et al*., 1991). Hierbei wird aus definierten Geweben RNA isoliert und in cDNA umgeschrieben. Die durch Sequenzierung der verschiedenen cDNA-Klone erzeugten Nukleotidsequenzen repräsentieren jeweils ein ganz spezifisches Transkript, das in dem untersuchten Gewebe exprimiert wird und möglicherweise ein Kandidatengen darstellen kann. Über den Ansatz der positionellen Klonierung konnten u. a. die Gene für die Duchenne´sche Muskeldystrophie (*Monaco et al.,* 1986), das Retinoblastom (*Friend et al*., 1986), die chronische Granulomatose (*Royer-Pokora et al.,* 1986) und die Cystische Fibrose (*Rommens et al*., 1986) identifiziert werden.

Eine Weiterführung der Methode der positionellen Klonierung ist der positionsabhängige Kandidatengen-Ansatz (im Überblick: *Collins*, 1995). Hierbei handelt es sich um eine Kombination von Kopplungsanalyse und der detaillierten Untersuchung bereits bekannter und funktionell charakterisierter Gene in einer bekannten Region. Ein Beispiel für die erfolgreiche Anwendung der Strategie ist die Identifzierung des Marfan-Syndrom-Gens *FBN*. Das für die Ausbildung des Syndroms verantwortliche Gen war über klassische Kopplungsanalyse in die chromosomale Region 11q lokalisiert worden (*Kainulainen et al*., 1990), wohin fast gleichzeitig auch das Fibrillin-Gen, ein putatives Kandidatengen für die Erkrankung, kartiert wurde (*Magenis et al*., 1991). Relativ schnell konnten dann Mutationen im Fibrillin-Gen als Ursache für die Ausbildung des Marfan-Syndroms identifiziert werden (*Dietz et al*., 1991). Neben dem positionsabhängigen Kandidatengen-Ansatz können Gendefekte auch über einen positionsunabhängigen Kandidatengen-Ansatz identifziert werden. Hierbei werden Vermutungen über ein Kandidatengen für eine menschliche Erkrankung angestellt, ohne dass eine chromosomale Zuordnung des Krankheitslocus vorliegt. Dies ist besonders dann möglich, wenn das Gen zu einer bekannten Genfamilie gehört, deren Mitglieder ähnliche Pathomechanismen oder Phänotypen erzeugen. In diesen Fällen kann das gewählte Kandidatengen direkt auf Mutationen hin untersucht werden. So konnten z. B. Mutationen, die die Oberflächenladung von α-Tropomyosin verändern, mit einer verzögerten Kardiomyopathie in Verbindung gebracht werden (*Olson et al*., 2001).

Da die hier aufgeführten Strategien, besonders die positionelle Klonierung, sehr aufwendige und limitierte Methoden zur Genidentifzierung darstellen, entstand Mitte der 80er Jahre (*Palca*, 1986; *Sinsheimer*, 1989) die Idee zur Gründung eines Humangenom-Projekts, durch das die komplette genetische Information des Menschen aufgeklärt werden sollte (Übersichtsartikel: *Guyer & Collins*, 1995). Dieser Gedanke wurde im Jahre 1988 (*National Research Council*, 1988) vom US National Research Council in ein festes Konzept umgesetzt, welches die Erstellung und Integration von genetischen und physikalischen Karten bei der Aufklärung der humangenomischen Sequenz beschloss. An dem Programm beteiligten sich neben den Vereinigten Staaten (Department of Energy and the National Institutes of Health) Großbritannien (Medical Research Council, Wellcome Trust), Frankreich (Centre d´Etude du Polymorphisme Humain, French Muscular Dystrophy Association), Japan sowie die Europäische Gemeinschaft. Dieses international koordinierte Projekt wurde 1990 mit einem Förderungsumfang von ca. 3 Milliarden US-Dollar begonnen. Ziel war es, alle ca. 3 Milliarden Basenpaare des menschlichen Genoms bis zum Jahr 2005 zu entschlüsseln und alle vorhandenen Gene zu identifizieren. Seit 1995 beteiligt sich auch Deutschland mit jährlich ca. 40 Millionen DM an dieser internationalen Kooperation (siehe auch http://www.dhgp.de). 1998 verkündete die US-amerikanische Firma *Celera Genomics*, das Ziel des Humangenom-Projektes mit einer alternativen Sequenzierungsstrategie schon zum

Jahr 2001 realisieren zu wollen. Während das öffentlich geförderte Humangenom-Projekt das menschliche Genom in einem „Klon-für-Klon"-Ansatz (siehe auch 4.2) sequenziert und so jedem analysierten Klon eine exakte chromosomale Lokalisation zuweisen kann, entschied sich die Firma *Celera Genomics* dafür, den Großteil der menschlichen Sequenz durch den „whole genome shotgun"-Sequenzierungsansatz zu ermitteln (*Venter et al*., 1998). Dieser Ansatz wurde bereits erfolgreich zur Sequenzierung von Bakteriengenomen und kleineren Eukaryoten-Chromsomen angewendet (*Fleischmann et al*., 1995; *Gardner et al*., 1998) und sollte durch die Sequenzierung von ca. 36 Millionen nicht-kartierten, kleinen Subklonen des humanen Gesamtgenomes umgesetzt werden. Die Verwendung der Methode des „whole genome shotgun"-Sequenzierungsansatzes war zu diesem Zeitpunkt umstritten (*Weber & Myers*, 1997), da aufgrund der Größe des humanen Genoms und des hohen Anteils repetitiver Elemente die Erstellung einer lückenlosen Gesamtsequenz fraglich schien (*Green*, 1997, *Goodman*, 1998). Im Juni 2000 wurde die erste Rohfassung des Humangenoms veröffentlicht. *Francis Collins*, Direktor des „National Human Genome Research Institute" verkündete, dass etwa 80% der erforderlichen Kartierungsarbeiten abgeschlossen seien und dass eine vorläufige „working draft"-Sequenz, die ca. 90% des humanen Genoms repräsentiert, im Internet frei zugänglich ist (*Ferry*, 2000). Am 15. Februar 2001 veröffentlichten *Francis Collins* (HGP) und *Craig C. Venter* (*Celera Genomics*) gemeinsam die Fertigstellung der vorläufigen Genomsequenz des Menschen (*IHGSC*, 2001; *Venter et al*., 2001). Die vorläufige Gesamtsequenz des Menschen wurde vom *IHGSC* mit Hilfe physikalischer Karten erstellt, die über 96% des Euchromatin-Anteils abdecken und zusammen mit zusätzlichen Sequenzen aus den Datenbanken etwa 94% des menschlichen Genoms ausmachen. Zum Zeitpunkt der Publikation beider Sequenzen lagen über 91% der menschlichen DNA-Sequenz mit einer Genauigkeit von 99,99% vor. Allerdings bestanden sowohl in den Datenbanken des Humangenom-Projektes als auch in denen von *Celera Genomics* zu diesem Zeitpunkt noch über 100 000 Lücken, die zusammen schätzungsweise 5% der humangenomischen Sequenz ausmachten. Es kann davon ausgegangen werden, dass es sich bei den noch nicht sequenzierten Bereichen um DNA-Abschnitte handelt, deren spezifische Basenzusammensetzung eine Klonierung und Sequenzierung erschweren. Trotzdem ist vorgesehen, möglichst alle Lücken bis zum Jahr 2003 zu schließen.

Mit Hilfe einer ersten Computer-gestützten Auswertung der generierten humangenomischen Sequenzen wurde eine Abschätzung der gesamten Genzahl vorgenommen. *Celera Genomics* konnte 26 383 Gene identifizieren und schätzt die Anzahl der in der endgültig fertig gestellten humanen DNA-Sequenz vorliegenden Gene auf 39 114. Der vom *IHGSC* ermittelte Wert von 31 778 Genen befindet sich etwa in der gleichen Größenordnung. Beide Werte liegen allerdings deutlich unter den bisherigen Annahmen. Bis zum Ende des Jahre 1996 waren insgesamt 16 354 Gene identifiziert worden, die nach damaliger Beurteilung

etwa 1/5 der auf eine Gesamtzahl von 50 000 bis 100 000 geschätzten Gene darstellten (*Schuler et al*., 1996). Diese Schätzungen basierten hauptsächlich auf der Anzahl der in den EST-Datenbanken vorhandenen Einträge. Dabei ist zu beachten, dass ein einzelnes Gen in der Regel durch mehrere EST-Klone repräsentiert wird, welche unterschiedlichen Regionen oder verschiedenen alternativen Spleißformen eines Transkriptes entsprechen. Durch die Generierung von EST-Clustern in der UNIgene Datenbank konnten EST-Sequenzen des gleichen Transkripts zu Contigs zusammengefasst werden. Hierdurch werden sowohl Informationen über die möglichst vollständige Transkriptlänge erhalten als auch die Redundanz einzelner EST-Sequenzen des gleichen Transkriptes in den Datenbanken reduziert. 1999 wurde durch Computer-gestützte Analysen zur Exonidentifizierung in der von *Dunham et al*. (1999) ermittelten, etwa 33,4 Mb großen DNA-Sequenz des humanen Chromosoms 22 die Anwesenheit von 545 Genen lokalisiert. Daraufhin wurde die zu erwartende Genzahl des Menschen auf ca. 45 000 reduziert. Dass diese Zahl deutlich unter den damaligen Erwartungen lag, wurde mit noch unzureichenden Analysemethoden erklärt. Allerdings scheinen sich die 1999 ermittelten Daten zu verifizieren, da die Analysen zur Genidentifikation in den beiden vorläufigen Gesamt-DNA-Sequenzen des Menschen eine Gesamtzahl von 30 000 bis 40 000 Genen nahe legen. Somit scheint der Mensch zum Beispiel nur etwa doppelt so viele Gene wie der Wurm *Caenorhabditis elegans* (19 099 Gene, *The C. elegans Sequencing Consortium*, 1998), die Fruchtfliege *Drosophila melanogaster* (13 601 Gene; *Adams et al.*, 2000) oder *Arabidopsis thaliana* (25 498 Gene, *The Arabidopsis Genome Initiative*, 2000) zu besitzen. Trotz dieser unerwartet geringen Differenz in der Anzahl der Gene zu weit weniger komplexen Organismen weisen die menschlichen Gene deutliche Unterschiede zu denen der anderen Spezies auf. So erstrecken sich die humanen Gene über deutlich größere genomische Regionen, was z. B. die variable Konstruktion alternativ gespleißter Genprodukte begünstigt. So können 35% bis 40% der menschlichen Gene alternativ gespleißt werden. Dies betrifft vor allen Dingen solche Gene, die für Membran-Rezeptoren und Proteine des Immun- und Nervensystems kodieren (*Mironov et al.,* 1999). Auf der Basis der Daten des Humangenom-Projektes kann davon ausgegangen werden, dass jedes Gen durch alternatives Spleißen durchschnittlich für drei unterschiedliche, wenn auch ähnliche Proteine kodiert (*Venter et al*., 2001). Dieser Mechanismus könnte gewährleisten, dass im Menschen eventuell ca. fünfmal so viele primäre Proteinprodukte hergestellt werden könnten wie in *C. elegans* oder *D. melanogaster*. Auch auf Protein-Ebene sorgt zusätzlich eine hohe kombinatorische Vielfalt dafür, dass z. B. bei der Proteinarchitektur, der Kontrolle von Transkription und Translation sowie posttranskriptionaler bzw. posttranslationaler Modifikation die unerwartet geringe Genzahl kompensiert wird. Durch die gesteigerte Flexibilität bei der Zusammensetzung einzelner Proteindomänen können somit durch komplette Neukombination alter Domänen funktionell

neue Proteinstrukturen entstehen. Es wird angenommen, dass das humane „Proteom“, d. h. die Summe aller vom Menschen herstellbaren Proteine, insgesamt ca. 250 000 Proteine umfasst (*Pawson & Nash*, 2000).
Neben der möglichst vollständigen Identifizierung aller menschlichen Gene im Zuge des Humangenom-Projektes sind noch weitere interessante Informationen zu erwarten. So sind Einblicke z. B. in den Genomaufbau möglich, was wiederum Rückschlüsse auf die Genomevolution zulassen kann. Auch die Verteilung und Evolution repetitiver Elemente kann, ebenso wie das Auftreten von SNPs (single nucleotide polymorphisms), nach Erhalt der vollständigen humangenomischen Sequenz weiter untersucht werden. Insbesondere durch die Identifizierung und Auswertung von SNPs wird wesentlich dazu beigetragen, neue diagnostische und therapeutische Verfahren in der Medizin zu entwickeln (Pharmakogenetik).

In den vergangenen Jahren wurden Genomprojekte zur Aufklärung der Erbinformation diverser prokaryotischer und eukaryotischer Organismen durchgeführt. Unter den bisher sequenzierten Genomen befinden sich die zahlreicher humanpathogener Organismen, wie z. B. *Neisseria meningititis* oder auch *Mycobakterium tuberculosis* bzw. *Mycobakterium leprae*. Die beiden ersten vollständig sequenzierten Genome waren die von *Haemophilus influenza RD* (*Fleischmann et al*., 1995) und *Mycoplasma genitalium* (*Fraser et al*., 1995). Genome weiterer pathogener Bakterien, z. B. *Borrelia burgdorferi* (*Fraser et al*., 1997) oder *Mycobacterium tuberculosis* (*Cole et al.,* 1998), schlossen sich an. Inzwischen kann auf die komplett und auch teilweise sequenzierten Genomsequenzen von über 800 Organismen aus allen drei Stämmen des Lebens (Bakterien, Archaebakterien und Eukaryoten) öffentlich in der Genom-Datenbank (http://www.ncbi.nlm.nih.gov:80/Entrez/Genome) zugegriffen werden. Hierbei sind jedoch Genome eukaryotischer Organismen für die humangenetische Forschung besonders interessant, da mit Hilfe dieser Modellorganismen funktionelle Analysen interessanter menschlicher Gene möglich sind. Hierbei kommt der Hefe, dem Fadenwurm, der Fruchtfliege und der Maus zentrale Bedeutung zu.

Das Genom der Bäckerhefe, *Saccharomyces cerevisiae*, ist seit April 1996 vollständig sequenziert (*Basset jr. et al.,* 1996). Die Hefe besitzt ca. 6 000 Gene, wovon 37% Homologien zu humanen Proteinen aufweisen. Es wird vermutet, dass diese homologen Gene der minimalen eukaryotischen Genausstattung entsprechen (*Goffeau et al*, 1996). Unter den bisher identifizierten Hefe-Genen befinden sich auch Homologe zu menschlichen Genen, die in mutierter Form mit diversen Krankheiten assoziiert sind (*Bassett et al*., 1996; siehe auch Tab. 1.1). Deletionsstämme der Hefe, die von allen annotierten offenen Leserahmen hergestellt werden, sollen auf phänotypische Auswirkungen hin untersucht und mit möglicherweise korrelierenden Phänotypen des Menschen assoziiert werden (*Winzeler*

et al., 1999). Die Sequenzanalyse kompletter Genome dient jedoch nicht nur der Identifizierung von Genen, sondern ermöglicht auch Einblicke in die Architektur sowie in die Mechanismen der Gen- und Genomevolution. Bereits 1970 wurde vermutet, dass der Mechanismus der Duplikation von Chromosomen bzw. chromosomaler Abschnitte ein wichtiger Faktor bei der Evolution von Genomen während der Phylogenese ist (*Ohno*). Im Hefe-Genom konnten verschiedene Formen der genetischen Redundanz, die auf Duplikationsereignisse zurückführbar sind, entdeckt werden (*Mewes et al*., 1997). So besitzen 20% aller Gene, die auf dem Hefe-Chromosom IV lokalisiert sind, paraloge Partner, also Gene, die durch Duplikation aus einem gemeinsamen Vorläufer-Gen entstanden sind (*Jacq et al*., 1997). Die Konservierung dieser paralogen Gene könnte darauf zurückzuführen sein, dass ihre Genprodukte an grundlegenden funktionellen Mechanismen beteiligt sind.

Der Fadenwurm, *Caenorhabditis elegans*, ist der erste vielzellige Organismus, dessen Genom vollständig sequenziert wurde (*The C. elegans Sequencing Consortium*, 1998). Dieser eukaroytische Organismus stellt aufgrund diverser Eigenschaften einen idealen Modellorganismus dar. So ermöglicht z. B. die geringe Anzahl von nur 959 Zellen eine detaillierte Untersuchung einzelner Zelllinien (z. B. jeder einzelnen der ca. 300 Nervenzellen) während der Entwicklung (*White et al*., 1983). Darüber hinaus erlaubt die Transparenz des Organismus die Expression bestimmter Gene mit Hilfe des GFP-Proteins („green fluorescent protein“) *in vivo* nachzuweisen und so Aufschluss über den Ort und das zeitliche Auftreten der Expression zu geben. Die Funktionen bestimmter Gene können dann durch gezieltes Ausschalten dieser Gene mittels komplementärer RNA-Injektion (*Kuwabara & Coulson*, 2000) untersucht werden. Durch Computeranalysen wurden in dem 97 Mb großen Genom von *C. elegans* knapp 20 000 Protein-kodierende Gene vorhergesagt, von denen 36% Homologien zu humanen Proteinen zeigen.

Auch das ca. 180 Mb umfassende Genom der Fruchtfliege, *Drosophila melanogaster*, ist inzwischen von der Firma *Celera Genomics* aufgeklärt worden (*Adams et al*., 2000). Eine Computer-gestützte Analyse zur Genvorhersage ermittelte 13 601 Gene (*Celniker*, 2000), gerade etwas mehr als die Hälfte der Genzahl von *C. elegans*. Hier zeigt sich, dass die Komplexität im Bauplan oder den Verhaltensweisen eines Organismus offensichtlich nicht mit der Anzahl der Gene zu korrelieren scheint. In den letzten 90 Jahren wurde die Fruchtfliege genetisch intensiv untersucht, so dass viele Informationen über Genstrukturen, Genregulationen und Genfunktionen vorliegen. *D. melanogaster* stellt somit einen guten Modellorganismus für die Erforschung menschlicher Krankheiten dar. Mit Hilfe der Daten aus den humanen EST-Projekten konnten viele menschliche Gene identifziert werden, die Homologien zu *Drosophila*-Genen aufweisen. Diese Gene („DRES“: *Drosophila*-related expressed sequences) wurden systematisch charakterisiert und katalogisiert, was z. T. auch eine Beschreibung ihrer genomischen Lokalisation in Mensch und Maus sowie der

spezifischen Expressionsmuster beinhaltete (Übersichtsartikel: *Banfi et al.,* 1997). Vergleiche solcher DRES-Gene mit den Homologen aus *Drosophila* können wesentlich dazu beitragen, die Funktion dieser Gene in Säugern und ihre mögliche Beteiligung an der Krankheitsentstehung aufzuklären. Mit der erfolgreichen Klonierung von *Drosophila*-Homeobox-Genen (*Lewis*, 1992) wurde deutlich, dass zahlreiche Prozesse, die die Entwicklung der Metazoa steuern, auch in höheren Organismen konserviert sind. Bei dem Versuch, das Ausmaß zu bestimmen, in dem homologe Proteine unterschiedlicher menschlicher Krankheitsgene in *D. melanogaster*, *C. elegans* und *S. cerevisiae* vorhanden sind, konnten *Rubin et al*. (2000) 177 Gene *in D. melanogaster* identifizieren, die Homologie zu den 289 untersuchten menschlichen Krankheitsgenen aufwiesen. Darunter befanden sich Gene, die bei der Ausbildung von Krebs, kardiovaskulären, neurologischen und endokrinologischen Erkrankungen, angeborenen Immunschwächen sowie metabolischen und hämatologischen Krankheiten assoziiert sind (http://www.sciencemag.org/feature/data/1049664t1.shl; siehe auch Tab. 1).

Tab. 1.1: Übersicht über unterschiedliche Krankheits-assoziierte Gene des Menschen, die auch in *S. cerevisiae*, *C. elegans* und *D. melanogaster* konserviert sind. Der Grad der Sequenzkonservierung auf Aminosäure-Ebene ist durch den e-Wert angegeben.

Beteiligung an Krankheits-entstehung	humanes Gen	OMIM-Nummer	e-Werte bei blastx *D. melanogaster*	*C. elegans*	*S. cerevisiae*
Krebs	*BRCA1*	113705	2,00 e^{-06}	9,00 e^{-11}	5,00 e^{-05}
	CDK4	123829	3,00 e^{-72}	2,00 e^{-58}	8,00 e^{-55}
	P53	191728	2,00 e^{-08}	0,81	1
	WT1	194070	4,00 e^{-28}	3,00 e^{-28}	9,00 e^{-21}
Fehlbildungen	Achondroplasie–*FGFR3*	134934	1,00 e^{-129}	1,00 e^{-114}	9,00 e^{-22}
	Wardenburg-*PAX3*	193500	1,00 e^{-30}	1,00 e^{-22}	2,00 e^{-09}
	Usher-*USH2A*	276901	1,00 e^{-106}	1,00 e^{-112}	1,20 e^{-01}
neurologische Erkrankungen	Alzheimer-*PS1*	104311	1,00 e^{-96}	5,00 e^{-85}	3,20 e^{+00}
	Fragile X-*FRAXA*	309550	3,00 e^{-28}	2,00 e^{-22}	5,00 e^{-13}
	Huntington-*HD*	143100	1,00 e^{-171}	1,00 e^{-118}	8,00 e^{-10}
kardiovaskul. Erkrankungen	Long QT 2-*KCNQ2*	152427	0,00 e^{+00}	0,00 e^{+00}	3,30 e^{-01}
	Long QT 1-*KCNQ1*	192500	8,00 e^{-64}	1,00 e^{-100}	1,40 e^{+00}
endokrinolog. Erkrankung	Diabetes-*INS*	176730	3,40 e^{-02}	9,70 e^{+00}	1,00 e^{+00}

Einen weiteren, sehr wichtigen Modellorganismus für die Analyse humaner Gene stellt die Maus dar. Das Maus-Genom besitzt mit einer Größe von ca. 3 Mrd Basenpaaren einen dem Humangenom entsprechenden Umfang. Zytogenetisch ist die Maus gut charaktierisiert. Eine vergleichende genetische Kartierung von 1461 Genen in Mensch und Maus ergab eine Verteilung dieser Gene auf 181 syntäne chromosomale Abschnitte (*DeBry & Seldin*, 1996). Weiterhin wurden in den letzten 10 Jahren Techniken für das gezielte Ausschalten von Genen und für die Erzeugung transgener Mäuse entwickelt (*Capecchi*, 1989; *Smith et al*., 1995; *Zheng et al*., 1999). Hierdurch kann die biologische Funktion einzelner Gene detailliert ermittelt werden. Wegen der leichten Verfügbarkeit aller Entwicklungsstadien und Gewebe bietet sich die Maus darüber hinaus für Untersuchungen differentiell exprimierter Gene, besonders von Entwicklungsgenen, an. Aufgrund der Bedeutung der Maus als Modellorganismus wurde das Maus-EST-Projekt initiiert (*Marra et al*., 1999). Dieses Projekt ergänzt ebenso wie die Aufklärung der murinen gesamtgenomischen Sequenz auf ideale Weise das humane Genom-Projekt, da hierdurch eine Identifizierung bisher unbekannter Gene im Menschen ebenso ermöglicht wird wie eine direkte funktionelle Analyse der orthologen Gene in der Maus (*Clark*, 1999).
Da Mensch und Maus relativ nah verwandt sind (ihre Entwicklungslinien trennten sich erst vor ca. 100 Mio Jahren (*Graur/Li*, 1999)), zeigen sich somit auch beim direkten Vergleich beider genomischer Sequenzen große Übereinstimmungen. Bisherige komparative Studien zwischen Mensch und verschiedenen Nagetieren (*Koop*, 1995; *Oeltjen et al*., 1997; *Ansari-Lari et al*., 1998; *Jang et al*., 1999; *Onyango et al*., 2000; *Wu et al*., 2001) zeigten, dass die kodierenden Regionen orthologer Gene in der Regel erwartungsgemäß stärker konserviert sind als die Intron- und Intergenbereiche. So zeigte der Vergleich der mRNAs von 1196 orthologen Genen aus Mensch und Maus im Durchschnitt eine Konservierung der translatierten Bereiche von ca. 85% (*Makalowski et al*., 1996). Die bisher durchgeführten vergleichenden Sequenzanalysen zwischen Mensch und Maus zeigen jedoch auch, dass nicht-kodierende Sequenzen mit möglicherweise regulatorischer Funktion stärker konserviert sind als nicht-funktionelle Bereiche des Genoms. Somit könnten umgekehrt die durch einen direkten Sequenzvergleich von z. B. Mensch und Maus identifizierten konservierten Bereiche einen Hinweis auf die Anwesenheit von regulatorischen Elementen geben (*Hardison et al*., 1997).

Eine unter dem Aspekt der vergleichenden Sequenzanalyse besonders interessante Region stellt im Menschen die chromosomale Region 11p15 und der orthologe Bereich auf dem Maus-Chromosom 7 dar. Die humane Chromosomenregion 11p15 gilt als besonders genreich. Zahlreiche Gene, die an der Ätiologie verschiedener Krankheiten involviert sind, konnten in diese Region kartiert werden. Eines dieser Krankheitsbilder ist das Beckwith-

Wiedemann-Syndrom (BWS). Es handelt sich hierbei um ein sowohl sporadisch als auch autosomal dominant vererbt auftretendes Syndrom, das durch ein variables Krankheitsbild mit Gigantismus, Makroglossie und Exomphalos gekennzeichnet ist (*Wiedemann*, 1964). Zudem weisen Patienten mit dem Beckwith-Wiedemann-Syndrom ein deutlich erhöhtes Tumorrisiko auf (*Beckwith*, 1963; *Wiedemann*, 1964, *Wiedemann*, 1983). Unter den auftretenden Tumoren ist der Wilmstumor (Nephroblastom) am häufigsten vertreten (56%), aber auch Nebennierenrindenkarzinome, Rhabdomyosarkome, Hepatoblastome und Neuroblastome sind zu beobachten (*Wiedemann*, 1983). Eine Eingrenzung der für die Krankheitsentstehung verantwortlichen zytogenetischen Region erfolgte mit Hilfe zytogenetischer Untersuchungen von Patienten mit Duplikationen des distalen Abschnitts von Chromosom 11 (*Waziri et al*., 1983; *Turleau et al*., 1984) sowie durch balancierte Translokationen bzw. Inversionen in BWS-Patienten mit Bruchpunkten in 11p15 (*Norman et al.,* 1992; *Weksberg et al*., 1993; *Sait et al*., 1994). Aufgrund dieser Befunde wurden drei kritische Regionen in 11p15 definiert, die an der Entstehung des BWS bzw. der assoziierten Tumoren beteiligt und durch chromosomale Bruchstellen gekennzeichnet sind (*Redeker et al*., 1995; *Hoovers et al*., 1995). Die erste BWS-kritische Region (BWSCR1) wurde innerhalb des chromosomalen Bereiches 11p15.5, proximal von *IGF2*, lokalisiert. Die zweite Region, BWSCR2, ist ca. 4 MB weiter centromerwärts, proximal von *HBBC*, in der Region 11p15.4 lokalisiert. Die BWSCR3 schließlich wurde in den chromosomalen Abschnitt 11p15.3 kartiert. Innerhalb der BWS-kritischen Regionen konnten Gene lokalisiert werden, die als Kandidatengene für das BWS in Frage kommen; hierzu zählen das Onkogen *HRAS* sowie die allelspezifisch exprimierten Gene *IGF2* und *H19*. *Redeker et al*. (1995) untersuchte die beiden distalen BWS-kritischen Regionen BWSCR2 und BWSCR3 auf ihre Beziehung zu den drei dort lokalisierten, bereits bekannten Tumorsuppressor-Genen *LMO1*, *ST5* und *WEE1* hin. Das Gen *LMO1*, auch *RBTN1* oder *TTG1* genannt, liegt in unmittelbarer Nähe des chromosomalen Bruchpunktes t(11;14)(p15;q11), der in einer kindlichen T-Zell akuten lymphoblastischen Leukämie entdeckt wurde. Es wird sowohl aufgrund dieser Lokalisation (*Boehm et al*., 1988) als auch wegen der vorhandenen konservierten Cystein-reichen Region, die eine Funktion als Transkriptionsfaktor nahe legt (*Boehm et al*, 1990), als potenzielles Tumorsuppressor-Gen diskutiert. Bei dem Gen *ST5* (suppression of tumorigenicity 5) handelt es sich ebenfalls um ein Tumorsuppressor-Gen aus der Region 11p15 (*Richard et al.,* 1993), welches die Tumorigenität von HeLa-Fibroblasten-Zellhybriden (*Lichy et al*., 1992) reduziert. Das Gen *WEE1*, das über starke Homologien zum *wee1*$^+$-Gen aus *Schizosaccharomyces pombe* identifiziert wurde, kodiert für eine Tyrosin/Serin-Kinase, die den Übergang von der postsynthetischen Phase (G2-Phase) zur Mitose koordiniert und aufgrund der Mitose-inhibierenden Eigenschaft ebenfalls als Tumorsuppressor-Gen betrachtet wird (*Heald et al*., 1993). Trotz der Lokalisation innerhalb der BWS-kritischen

Region 3 und ihrer wichtigen zellulären Funktionen konnte keines dieser drei Gene direkt mit der BWS-Entstehung in Verbindung gebracht werden (*Redeker et al*., 1995). Die Daten zeigen, dass eine intensive Suche nach weiteren, in der genreichen Region 11p15.3 lokalisierten Genen einen wichtigen Beitrag zur Aufklärung des molekulargenetischen Verständnisses bei der Pathogenese des BWS sowie Tumorerkrankungen (z. B. auch Lungentumoren (*Bepler & Koehler*, 1995)) leisten kann. Die humane Chromosomenregion 11p15.3 sowie der orthologe Bereich des murinen Chromosoms 7 ist somit für eine umfassende Sequenzaufklärung und komparative Sequenzanalyse besonders attraktiv.

Zielsetzung der Arbeit

Die vorliegende Dissertation wurde im Rahmen des Deutschen Humangenom-Projektes (Förderungs-Nummer 01KW9624) in Kooperation mit dem Institut für Molekulargenetik der Johannes Gutenberg-Universität durchgeführt. Dabei sollte ein insgesamt ca. 1 Mb großer Bereich der chromosomalen Region 11p15.3 des Menschen und die orthologe Region auf dem Maus-Chromsom 7 sequenziert und anschließend direkt verglichen werden. Die in dieser Region lokalisierten und z. T. über FISH-Analyse zueinander angeordneten Gene *LMO1*, *ST5*, *CEGF1* und *WEE1* (*Seipel*, 1996) dienten dabei als Startpunkte und wurden in vier eigenständigen Dissertationen bearbeitet. Zu Beginn der Arbeiten lagen noch keine Sequenzinformationen über diese Region aus dem Humangenom-Projekt vor.

Ziel der vorliegenden Arbeit war die komparative Sequenzanalyse eines etwa 250 kb umfassenden genomischen Bereiches um das in der chromosomalen Region 11p15.3 lokalisierte humane Gen *WEE1* und der orthologen Region auf dem Maus-Chromosom 7. Die Arbeit gliederte sich in drei Abschnitte: 1. Erstellung jeweils eines Klon-Contigs, der die interessierende genomische Region abdeckte, 2. Etablierung der Sequenzierung mit Hilfe von Hochdurchsatzmethoden und Sequenzanalyse der genomischen Klone mit einer Sequenzgenauigkeit von weniger als 3 Fehlern in 10 000 bp, 3. Vergleichende Analyse der generierten humanen und murinen genomischen Sequenzen.

zu 1. Erstellung von Klon-Contigs: Zur Erstellung eines lückenlosen humanen Klon-Contigs sollten unter Verwendung von hochdichten PAC-Klonfiltern weitere genomische Klone isoliert werden, die an den von *Seipel* (1996) isolierten und über FISH-Analyse verifizierten Klon PAC-142M6 anschließen. Auch aus der orthologen Region der Maus sollte ein durchgängiger PAC-Contig erstellt und die chromosomale Herkunft überprüft werden.

zu 2. Etablierung und Durchführung der Sequenzierung im Hoch-Durchsatzmaßstab: Die isolierten Klone sollten nach der Herstellung von „shotgun“-Klon-Bibliotheken sequenziert werden. Hierbei sollte eine Strategie entwickelt werden, die im Rahmen der vorhandenen Infrastrukturen eine möglichst effiziente, den Hochdurchsatz-Methoden angepaßte Probenherstellung und –verarbeitung gewährleistete. Darüber hinaus sollten für die Generierung der beiden Konsensus-Sequenzen zur Verfügung stehende Computerprogramme hinsichtlich ihrer Effizienz verglichen werden.

zu 3. Komparative Sequenzanalyse: Nach der Analyse der generierten genomischen Sequenzen in Mensch und Maus hinsichtlich des GC-Gehaltes, des Anteils an repetitiven Elementen u. ä. sollten die Sequenzen beider Spezies mit Hilfe diverser Computerprogramme verglichen werden. Hierbei sollten putativ neue Gene sowie konservierte Inter- und Intragenbereiche mit möglicher funktioneller Bedeutung

aufgrund ihrer Konservierung identifiziert werden. Die Leistungsfähigkeit der verwendeten Computerprogramme, die zur Genvorhersage eingesetzt wurden, sollte anhand der untersuchten Sequenzen beurteilt werden.

2 Material und Methoden

2.1 Versuchsmaterial

PAC-Klone

Die humanen PAC-Klone 142M6 und 180B11 wurden von *Seipel* (1996) und *Busch* (unveröffentlicht) aus der humanen PAC-Klonbibliothek Nr. 704 (*Ioannou et al.*, 1994), die über das Ressourcenzentrum im Deutschen Humangenom-Projekt (Berlin) erhältlich ist, isoliert. Alle weiteren im Rahmen dieser Arbeit isolierten humanen PAC-Klone stammen aus derselben Klonbibliothek.

Der murine PAC-Klon 256N10 wurde aus der Klonbibliothek Nr. 711, die ebenfalls über das Ressourcenzentrum im Deutschen Humangenom-Projekt (Berlin) erhältlich ist und von *Osoegawa et al.* (2000) hergestellt wurde, isoliert.

Tab. 2.1: Auflistung der diversen Bezeichnungen der bearbeiteten PAC-Klone

Offizielle Bezeichnungen	Arbeitsbezeichnungen	Accession-Nummern
RPCI21N10256	256N10	AJ278435
RPCIP704M06142	142M6	AJ277546
RPCIP704B11180	180B11	AJ295844
RPCIP704D07151	151D7	/
RPCIP704C211100	1100C21	/
RPCIP704D221139	1139D22	/
RPCIP704N1839	39N18	/
RPCIP704E01690	690E1	/

IMAGE-Klone

Die menschlichen cDNA-Klone IMAGp998N19612, IMAGp998P151314 und IMAGp998J191724 wurden durch Datenbanksuchen (siehe 2.12.3) von sequenzierten DNA-Bereichen identifiziert und über das Ressourcenzentrum im Deutschen Humangenomprojekt bezogen. Die IMAGE-Klone stammen aus dem „WashU-Merck-EST-Projekt“ der Washington University School of Medicine, St. Louis, *USA (Hillier et al.*, 1996).

Tab. 2.2: Auflistung der diversen Bezeichnungen der sequenzierten cDNA-Klone.

Name	Arbeitsbezeichnung	Accession-Nummer	EST-Bezeichnung
IMAGp998P151314	P151314	AA083037	zn10g08.r1
		AA082948	zn10g08.s1
IMAGp998N19612	N19612	N48731	yy55c10.r1
		N56865	yy55c10.s1
IMAGp998J191724	J191724	AA279570	zs86c10.r1
IMAGE: 2090250	te51e10.x1	AI539442	te51e10.x1
DKFZ564C2163	564C2163	AL117596	/

2.2 DNA-Standardmethoden

2.2.1 Fällung

Die DNA-Fällung erfolgte unter Zugabe von 1/10 Volumen 3 M Natriumacetat, pH 4,8 und 2,5 Volumen Ethanol absolut. Die DNA wurde pelletiert (4°C, 30 min, 14000 Upm), mit 70%-igem Ethanol gewaschen, vakuumgetrocknet und in einem geeigneten Volumen 1/4xTE-Puffer oder H_2O aufgenommen.

PCR-Produkte wurden durch Zugabe von 1 Volumen 4 M Ammoniumacetat und 2 Volumen Isopropanol gefällt (RT, 30 min, 14000 Upm). Nach dem Waschen und Trocknen erfolgte die Probenaufnahme in der Regel in 10 µl Wasser.

2.2.2 Verdau von DNA durch Restriktionsendonukleasen

Der DNA-Verdau durch Restriktionsendonukleasen erfolgte nach den Angaben des jeweiligen Enzymherstellers und unter Verwendung der empfohlenen Puffer. Die Enzymmengen und die Dauer der Restriktion variierte mit der Enzymaktivität sowie der Qualität und Quantität der zu schneidenden DNA.

2.2.3 Photometrische Quantifizierung von Nukleinsäuren

Für die photometrische Konzentrationsbestimmung der Nukleinsäuren wurde ein Aliquot von 1 µl DNA mit Aqua bidest. in einer Quarzküvette auf 50 µl aufgefüllt und im Spektralphotometer die Extinktion bei 260 nm ermittelt. Der Konzentrationsberechnung lagen folgende Richtwerte zugrunde (*Sambrook et al.*, 1989):

OD=1 (260nm)	ca. 50 μg/ml	doppelsträngige DNA
	ca. 40 μg/ml	einzelsträngige DNA/RNA
	ca. 45 μg/ml	DNA/RNA-Gemische
	ca. 20 μg/ml	kurzkettige Oligonukleotide

Durch die Bestimmung des Verhältnis der Absorption bei 260 bzw. 280 nm konnte die Reinheit der Probe ermittelt werden ($OD_{260}/OD_{280} \geq 1,8$).

2.2.4 Gel-Elektrophoresen

Agarose-Gelelektrophorese:

Die Auftrennung von restringierter DNA und PCR-Produkten erfolgte auf horizontal gelagerten 0,8-2%-igen Agarosegelen mit 0,5x TE als Laufpuffer. Die aufzutrennende DNA wurde vor dem Auftrag auf das Gel mit 1/6 Vol. 6x DNA-Auftragspuffer oder OrangeDye-Auftragspuffer versetzt; als Molekulargewichtsstandards dienten Hind III-restringierte λ-DNA und eine 100 bp- bzw. 123 bp-Leiter. Je nach Gelgröße erfolgte die eindimensionale Fragmentauftrennung bei Spannungen zwischen 60-140 V.

Nach dem Lauf wurden die Gele 5-10 min im Ethidiumbromidbad (5 μg Ethidiumbromid pro ml H_20) gefärbt und für 15-20 min in Wasser entfärbt. Unter UV-Licht (λ=312 nm) wurden die Gele mit Hilfe des E.A.S.Y. Videosystems (Fa. Herolab, Wiesloch) ausgewertet und dokumentiert.

Pulsfeld-Gelelektrophorese (PFGE):

Zur Auftrennung von DNA-Fragmenten, die über 100 kb groß waren (z. B. von Not I-restringierten PAC-Klonen zur Bestimmung ihrer Integratgrößen), wurde die „Contour-clamped-homogeneous-electric-field“ (CHEF) Pulsfeld-Gelelektrophorese (*Schwartz & Cantor*, 1986) eingesetzt.

Durch den Aufbau eines jeweils um 120° alternierenden elektrischen Feldes zwischen 24 hexagonal angeordneten Elektroden können sich auch DNA-Fragmente im Megabasen-Bereich im Gel reorientieren und mit unterschiedlicher Geschwindigkeit durch die Poren eines LMP-Agarosegels hindurchbewegen. Daraus resultiert im vorher definierten Trennbereich eine lineare Auftrennung der Fragmentgrößen.

In Agaroseblöcke eingegossene DNA-Proben und Molekulargewichtstandards (λ-DNA-Concatemere und Hind III-restringierte λ-DNA; siehe 2.13.4) wurden in einem 1%-igen LMP-Agarosegel aufgetrennt. Als Laufpuffer diente 0,5x TBE-Puffer, dessen Temperatur konstant

auf 16°C gehalten wurde. Der lineare Auftrennungsbereich wurde in der Regel zwischen 20 und 200 kb eingestellt, was einer Laufzeit von ca. 28 h entsprach.
Das PFGE-Gel wurde nach dem Lauf für 10 min im Ethidiumbromid-Bad gefärbt und im Wasserbad für 10 min entfärbt.

2.2.5 DNA-Wiedergewinnung aus Gelen

Nach gelelektrophoretischer Auftrennung wurden DNA-Banden der gewünschten Größe oder DNA-Fragmente eines bestimmten Größenbereiches unter UV-Licht mit dem Skalpell aus dem LMP-Agarosegel ausgeschnitten. Die Wiedergewinnung der DNA erfolgte mit Hilfe des GeneClean-Kit™ (BIO101, Dianova) gemäß den Angaben des Herstellers. Um die Effizienz der Wiedergewinnung zu überprüfen, wurde 1/10 Vol. des Eluats auf ein 1%-iges Agarosegel aufgetragen und mit DNA bekannter Konzentration verglichen.

2.2.6 Phenol-Extraktion

Phenolextraktionen wurden durchgeführt, um DNA von Proteinen zu reinigen. Dazu wurde 1 Vol. Phenol/Chloroform/Isoamylalkohol (25:24:1) zu der DNA-Lösung gegeben, vollständig gemischt und für 10 min bei RT und 14000 Upm zentrifugiert. Die wässrige, DNA-haltige Phase wurde erneut mit 1 Vol. Chloroform extrahiert und von der organischen Phase getrennt. Die anschließende Fällung erfolgte wie unter 2.2.1 beschrieben.

2.3 Isolierung von DNA

2.3.1 Isolierung von PAC-DNA

Die Isolierung größerer Mengen von PAC-DNA, die zur Herstellung der „shot-gun"-Bibliotheken benötigt wurde, beruhte auf der Methode der alkalischen Lyse nach *Birnboim & Doly* (1979) und wurde mit Hilfe des QIAGEN® Plasmid Maxi-Kits nach Herstellerangaben durchgeführt.
Die DNA-Ausbeute konnte auf 20-100 μg gesteigert werden, indem das vom Hersteller empfohlene Protokoll zur Herstellung von „very low copy-plasmid/cosmid-DNA" verwendet wurde. Neben einem effizienteren Aufschluß der Bakterien durch die Verdopplung der Puffermengen bei der alkalischen Lyse wurde hierbei die DNA vor dem Säulenlauf über eine

Isopropanol-Fällung aufgereinigt. Eine Erwärmung des Elutionspuffers QF steigerte zusätzlich die DNA-Ausbeute.
Das Volumen der Bakterien-üN-Kulturen betrug unabhängig von der Präparations-Methode 400-500 ml; die gewaschene DNA wurde in der Regel in 50 µl H_2O oder 1/4x TE gelöst und die Konzentration photometrisch bestimmt (siehe 2.2.3). Zusätzlich wurde jeweils 1 µl der präparierten DNA gelelektrophoretisch aufgetrennt, um eventuelle Degradation bzw. RNA-Reste zu erkennen.

2.3.2 Präparation von Plasmid-DNA

Die Präparation von Plasmid-DNA, die zur Sequenzierung der Subklone im 96-Loch-Format eingesetzt wurde, erfolgte nach einem modifizierten Protokoll der alkalischen Lyse (*Birnboim & Doly*, 1979) mit Hilfe des „R.E.A.L. Prep 96 Plasmid Kit“ (Rapid Extraction Alkaline Lysis Plasmid Minipreps) der Firma Qiagen (Hilden) nach den Angaben des Herstellers.
Die DNA einzelner Plasmidklone wurde unter Verwendung des ebenfalls auf dem Prinzip der alkalischen Lyse beruhenden RPM® (Rapid Pure Miniprep)–Kits (BIO101, USA) nach Herstellerangaben isoliert.

2.4 Präparation von Gesamt-RNA aus Mausgeweben

2.4.1 RNA-Präparation

Die Isolierung von Gesamt-RNA erfolgte aus tiefgefrorenen murinen Geweben (C57/Bl10 bzw. JF1) unter Verwendung des „RNeasy Mini Kit“ der Firma QIAGEN (Hilden). Die Präparation wurde streng nach Herstellerangaben durchgeführt. Die Konzentration der isolierten RNA wurde photometrisch bestimmt (2.2.3) und die Integrität der RNA gelelektrophoretisch überprüft.

2.4.2 DNase-Verdau

Die präparierte RNA wurde einem DNase I-Verdau (*Sambrook et al.,* 1989) unterzogen, um Kontaminationen genomischer DNA zu beseitigen. Dazu wurde die RNA in einem Volumen von 50 µl (1xTE, 10 mM $MgCl_2$, 100 mM DTT, 40 U RNase Inhibitor) mit 25 U RNase-freier DNase I (Roche Diagnostics) für 2 h bei 37°C inkubiert. Die anschließende Aufreinigung von

der bei der reversen Transkription störenden DNase I erfolgte nach dem „RNeasy protocol for RNA clean-up“ (Qiagen, Hilden) nach entsprechenden Angaben des Herstellers. Die Elution erfolgte in 30 µl RNAse-freiem Wasser. Die RNA-Konzentration des Eluats wurde photometrisch bestimmt (siehe 2.2.3) und die RNA-Integrität mittels Agarosegelelektrophorese überprüft.

2.5 Polymerase-Kettenreaktion (PCR)

Die PCR dient der Vervielfältigung linearer DNA-Fragmente (*Saiki et al*., 1986*, Saiki et al.*,1988) und erfolgte in einem Reaktionsvolumen von 35 bzw. 50 µl. Der Reaktionsansatz setzte sich aus 40 mM KCl; 10 mM TrisHCl (pH 8,3); 0,1 mg/ml Gelatine; 1,5 mM $MgCl_2$; 20 pmol jedes Primers, je 200 µM der vier dNTPs und 1 U Taq-DNA-Polymerase (Gibco BRL (USA)) zusammen. Bei schwierigen Amplifikationen wurde der PCR-Ansatz durch 10% Glycerin ergänzt. Als Matrize dienten zwischen 20 und 100 ng DNA, 1 µl Bakterien-üN-Kultur oder eine ausgewählte rekombinante Bakterienkolonie, die im PCR-Ansatz geschwenkt wurde. Einer ersten Denaturierung von 3 min bei 96°C folgten in der Regel 35 Amplifikationszyklen (1 min 96°C Denaturierung, 1 min 54-60°C Primer-Anlagerung, 1 min Elongation pro kb Amplifikat bei 72°C). Abschließend erfolgte eine 7-minütige Kettenverlängerung bei 72°C.

Die PCR-Reaktionen wurden in einem PTC100™/ PTC200™ (MJ Research, USA) oder einem Gene Amp PCR System PE 9700 (PE Applied Biosystems (USA)) durchgeführt. Nach erfolgter Amplifikation wurde 1/10 Vol. des Reaktionsansatzes auf einem 1-2%-igen Agarosegel elektrophoretisch überprüft (siehe 2.2.4).

2.6 Reverse Transkriptase-Polymerasekettenreaktion (RT-PCR)

Die Reverse Transkriptase-Polymerasekettenreaktion wurde eingesetzt, um Genexpression auf RNA-Ebene zu untersuchen.

Es wurden 4 µg Gesamt-RNA in einem Gesamtvolumen von 16 µl DEPC-Wasser für 10 min bei 70°C inkubiert und danach für 1 min auf Eis abgekühlt. Nach Zugabe von 24,1 µl cDNA-Synthesemix (s. u.) erfolgte die Erststrang-Synthese in einem Gesamtvolumen von 40 µl für

90 min bei 37°C. Die reverse Transkriptase wurde durch eine Inkubation des Reaktionsansatzes bei 95°C für 10 min inaktiviert.

cDNA-Synthesemix:
- 5,0 µl Oligo(dT)$_{16}$ (50 µM)
- 3,6 µl dNTP-Gemisch (10 mM je dNTP)
- 2,0 µl RNase Inhibitor (20 U/µl; Fa. MBI Fermentas, USA)
- 8,0 µl 5x 1st strand buffer (Fa. Gibco BRL, Eggenstein)
- 4,0 µl DTT (0,1 M; Fa. Gibco BRL, Eggenstein)
- 1,5 µl MMLV-Reverse Transkriptase (200U/µl; Fa. GibcoBRL, Eggenstein)

Für die anschließende PCR wurden in der Regel 2 µl cDNA als Matrize mit genspezifischen Primern verwendet.

2.7 Automatische DNA-Sequenzierung

Die DNA-Sequenzierung erfolgte nach der von *Sanger et al.* (1977) entwickelten Kettenabbruch-Reaktion, die nach Abwandlungen durch *Prober et al*. (1987) und *Carothers et al*. (1989) von *Lee et al*. (1992) durch Verwendung von Fluoreszenz-markierten Didesoxynukleotiden modifiziert wurde.

Die Sequenzierung erfolgte zunächst mit dem „ABI PRISM™ Ready Reaction Dye Deoxy Terminator Cycle Sequencing Kit“ (Fa. Applied Biosystems) und nach Markteinführung mit dem „ABI PRISM™ Big-Dye™ Terminator Cycle Sequencing Ready Reaction Kit (Fa. Applied Biosystems), da diese Sequenzierungschemie längere Leseweiten und bessere Ergebnisse bei Direktsequenzierungen von PAC-Klonen lieferte.

Für einen Sequenzierungsansatz wurden 50-100 ng aufgereinigtes PCR-Produkt, 0,5-1 µg Plasmid-DNA bzw. 5-10 µg PAC-DNA als Matrize verwendet. Das empfohlene Reaktionsvolumen von 20 µl wurde nicht verändert, obwohl bei Sequenzierungen von PCR-Produkten nur 4 µl, bei Plasmid-DNA 6 µl und bei PAC-Direktsequenzierungen 12 µl Premix (enthält Reaktionspuffer, markierte und nicht-markierte Nukleotide sowie das Enzym) eingesetzt wurden. Weiterhin wurden dem Ansatz 10 pmol eines spezifischen Primers zugegeben. Die Sequenzierung erfolgte je nach verwendeter Matrizen-DNA mit unterschiedlichen „cycle-sequencing“2-Programmen (s. u.). Die Produkte wurden durch eine Natriumacetat/Ethanol-Fällung präzipitiert, mit 70%-igem EtOH gewaschen oder mit Hilfe des „MultiScreen®“-Assay-Systems (Millipore, Eschborn) und Sephadex G50-Pulver (Pharmacia, Freiburg) im 96-Loch-Format nach Protokoll aufgereinigt. Die gereinigten Sequenzierungs-Produkte wurden Vakuum-getrocknet und in 3 µl Auftragspuffer für Sequenzgele (siehe

2.13.1) gelöst. Die anschließende Elektrophorese auf einem Polyacrylamid-Gel, die Detektion der Fluoreszenzsignale sowie deren Auswertung erfolgte auf den automatischen 377-Sequenziergeräten nach den Angaben des Herstellers.

Verwendete Standard-Sequenzierprogramme:

für PCR-Produkte und Plasmide:

	1 min / 96°C
25-30 Zyklen à	15 sec / 96°C, 15 sec / 50°C, 4 min / 60°C

für PAC-DNA:

	4 min / 96°C (hotstart)
30 Zyklen à	30 sec / 96°C, 5 sec / 50°C, 4 min / 60°C

bei Repeat-Strukturen:

	1 min / 98°C
25 Zyklen à	5 sec / 98°C, 90 sec / 60°C, 90 sec / 50°C

2.8 Herstellung einer „shot-gun“-Klon-Bibliothek

Grundlage für die Sequenzierung der PAC-Klone war die Erstellung einer „shot-gun“-Klon-Bibliothek. Hierzu wurde vorbehandelte PAC-DNA (siehe 2.8.1.) mechanisch geschert, größenfraktioniert und in den mit Sma I-restringierten „Sequenzierungsvektor“ pUC18 kloniert.
Einen Überblick über die Vorgehensweise gibt die Abb. 2.1.

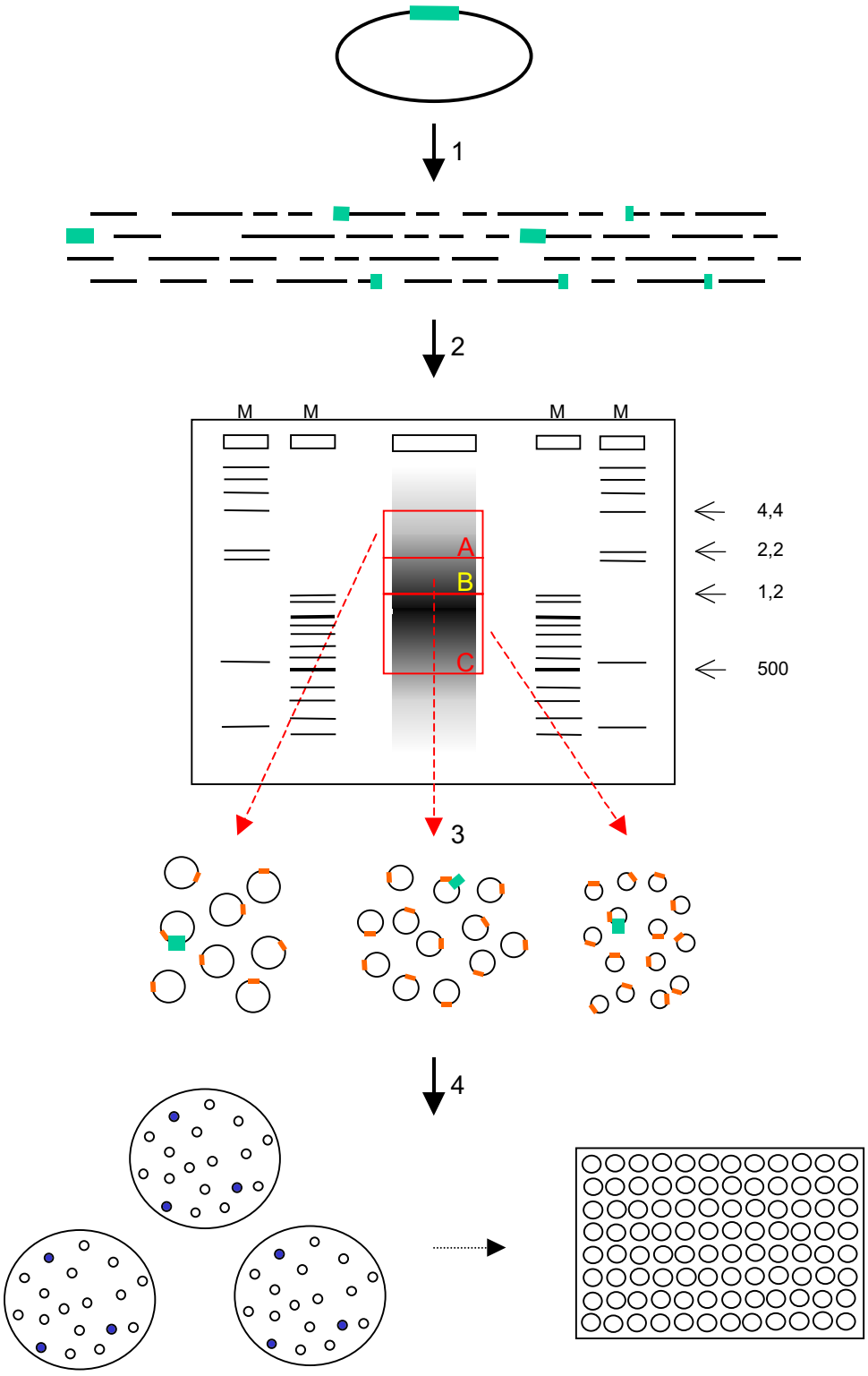


Abb. 2.1: Überblick über die Herstellung einer „shot gun"-Klon-Bibliothek. Der komplette PAC-Klon (grüner PAC-Vektoranteil mit Integrat) wird durch Scherung in unterschiedlich große Fragmente zerlegt, welche zufällig miteinander überlappen (1). Eine gelelektrophoretische Auftrennung (2) ermöglicht die Größenselektion der Fragmente vor der „blunt end"-Ligation der drei unterschiedlich großen Fragmentbereiche in den pUC18-Vektor (orange) (3). Im Anschluss an eine Transformation in elektrokompetente *E. coli*-Zellen (4), die eine Blau-Weiß-Selektion bzgl. der erfolgten Vektoraufnahme ermöglichten, wurden für 100 kb PAC-Integrat etwa 1000 verschiedene Subklone in 96-Loch-Platten angezogen und konserviert. Auch die DNA-Präparation und Sequenzierungen erfolgten anschließend im Mikrotiter-Format.

2.8.1 Plasmid-Safe™-Behandlung der PAC-DNA

Um Kontamination der zu erstellenden „shot-gun“-Klon-Bibliothek durch bakterielle DNA zu minimieren, wurde die isolierte PAC-DNA (siehe 2.3.1) einer Plasmid-Safe™-Behandlung (Epicentre Technologies Corporation (USA)) unterzogen. Hierbei handelt es sich um eine ATP-abhängige DNase-Reaktion, die selektiv lineare einzel- und doppelsträngige DNA abbaut, während zirkuläre Doppelstrang-DNA nicht angegriffen wird.
50-70 µg PAC-DNA wurden in einem Reaktionsvolumen von 150 µl (15 µl *PS* Reaction Buffer, 12 µl ATP (25mM) und 10 µl Plasmid-Safe™ DNase (10 U), mit H_20 ad 150 µl) für 4-16 h bei 37°C inkubiert. Es folgte eine Hitze-Inaktivierung des Enzyms durch eine Inkubation bei 75°C für 15 min.

2.8.2 Nebulisierung der Plasmid-Safe™-behandelten DNA

Die, wie unter 2.8.1 beschrieben, vorbehandelte PAC-DNA sollte durch mechanische Scherung (Nebulisierung) in unterschiedlich große Fragmente zerlegt werden.
Dazu wurde der Reaktionsansatz der Plasmid-Safe™-Behandlung mit 1x TE auf 2 ml aufgefüllt und im „DNA-Nebulizer“ für 1 min bei einem Druck von 1 bar durch eine sehr kleine Öffnung gepresst, wobei die DNA in Fragmente von ca. 500 bp bis ca. 4 kb Größe geschert wurde (siehe auch 2.8.4). Anschließend wurde die nebulisierte DNA gefällt, indem sie in 500 µl-Aliquots mit 40 µl 10x Dialysepuffer und 1 ml Ethanol abs. versetzt, für 1 h zu –20°C gestellt und für 30 min bei 4°C und 14 000 Upm zentrifugiert wurde. Die mit 70% Ethanol gewaschene und getrocknete DNA wurde in je 20 µl 1/4x TE ohne Resuspendierung vorsichtig gelöst und vereinigt. Durch anschließende gelelektrophoretische Auftrennung von 1/10 Volumen wurden die durch Nebulisierung entstandenen Fragmentgrößen und die Konzentration der gereinigten DNA überprüft.

2.8.3 „End-filling“

Die durch Nebulisierung entstandenen überhängenden Enden der doppelsträngigen DNA-Fragmente mußten aufgefüllt werden, um eine spätere Ligation (siehe 2.8.5) zu ermöglichen. Für diese „end-filling“-Reaktion wurde ein Gemisch aus T4-DNA-Polymerase und dem großen Fragment der DNA-Polymerase I (Klenow-Fragment) eingesetzt. Das Klenow-Fragment katalysiert ebenso wie die T4-DNA-Polymerase die DNA-Synthese in 5' → 3' Richtung, jedoch mit einer schwächeren 3' → 5'-Exonuklease-Aktivität. Da beide Enzyme keine 5' → 3' Exonuklease-Funktion aufweisen, eignen sie sich zur Entfernung von

3'-Überhängen und zum Auffüllen von 5'-Überhängen („end-filling“), was zur Herstellung von doppelsträngigen DNA-Fragmenten mit geraden Enden („blunt ends“) notwendig ist.
Die Reaktion erfolgte in einem Volumen von 60 µl, bestehend aus 30 µl nebulisierter, gereinigter DNA, 4 µl (20 U) Klenow-Enzym, 2 µl T4-DNA-Polymerase (6 U), 5 µl dNTPs (0,25 mM), 6 µl T4-DNA-Polymerase-Puffer und 0,5 µl 100x BSA. Der Ansatz wurde für 30 min bei RT inkubiert.
Die Aufreinigung des Reaktionsgemisches erfolgte gelelektrophoretisch (siehe 2.2.4).

2.8.4 Selektion der Integratgrößen

Von jedem zu sequenzierenden PAC-Klon wurden „shot-gun“-Klon-Bibliotheken mit unterschiedlichen Integratgrößen hergestellt. Die Kenntnis dieser Integratgrößen lieferte wichtige Informationen über zu schließende Lückengrößen in der „finishing“-Phase.
Der „end-filling“-Ansatz wurde auf einem 1%-igen LMP-Agarosegel aufgetrennt und gereinigt. Es folgte eine Abtrennung und Anfärbung der Markerspuren im Ethidiumbromid-Bad. Danach wurden die interessierenden Größenbereiche (in der Regel 500 bp - 1,5 kb; 1,5-2 kb und 2–3 kb) mit Hilfe der angefärbten und durchstochenen Markerspuren mit einem Skalpell aus dem Gel herausgeschnitten. Dieses Vorgehen sollte eine Schädigung der DNA durch UV-Einwirkung vermeiden. Die Wiedergewinnung der DNA aus dem Gel erfolgte mit Hilfe des „GeneClean-Kit™“ (BIO101, Dianova; siehe 2.2.5).

2.8.5 Ligation in pUC18

Die Ligation von 100-250 ng der größenselektionierten DNA-Fragmente erfolgte in 25 ng des Sma I-geschnittenen, dephosphorylierten Plasmid-Vektors pUC18 (Pharmacia, Freiburg) unter Verwendung von 1,5 µl 10x Puffer für T4-DNA-Ligase und 1 µl T4-DNA-Ligase (400 U) in einem Reaktionsvolumen von 15 µl. Die Inkubation erfolgte für 16 h bei 16°C. Der hohe Einsatz von T4-DNA-Ligase resultierte daraus, dass für die erfolgreiche Ligation von „blunt-end“-Fragmenten etwa 50% mehr Enzym benötigt wird als bei einer vergleichbaren Ligation von Fragmenten mit überhängenden Enden („sticky ends“). Der Ligationsansatz wurde mit Hilfe des „GeneClean-Kit™“ (BIO101, Dianova) nach Angaben des Herstellers gereinigt, in 20 µl H_2O eluiert und 1/10-1/20 Vol. auf einem Agarosegel überprüft.

2.8.6 Transformation und Selektion positiver Klone

Transformation

Für die Transformation wurden elektrokompetente Epicurian Coli®SURE- bzw. Epicurian Coli® XL1-Blue MRF´-Zellen (siehe 2.13.3.) verwendet. Diese eignen sich durch das limitierte Vorkommen von Rekombinations-Ereignissen und die Erhaltung des Methylierungsstatus besonders für die Transformation genomischer DNA.

Für die Transformation wurden zwischen 33 µl und 50 µl elektrokompetenter Bakterien mit 1 µl des gereinigten Ligationsansatzes auf Eis vermischt und bei einem Widerstand von 200 Ω, einer Spannung von 1,7 kV und 25 µF unter Beachtung der Zeitkonstanten, die zwischen 4,2 ms und 4,5 ms liegen sollte, elektroporiert. Der Ansatz wurde anschließend in 950 µl SOC-Medium aufgenommen und für 1 h bei 37°C im Schüttler inkubiert. Je nach Transformationseffizienz der Zellen wurden 50-500 µl des Transformationsansatzes auf LB-Agarplatten (75 µg/ml Ampicillin, 40 µl/ml X-Gal, 50 µg/ml IPTG) ausgestrichen und zwischen 16 und 20 h bei 37°C bebrütet.

Selektion

Weiße, rekombinante Subklone wurden in einer 96-Loch-Mikrotiter-Platte, die pro Loch mit je 250 µl LB-Gefriermedium (siehe 2.13.1) und Selektionsantibiotikum beschickt war, gepickt. Es wurden, abhängig von der Integratgröße des fragmentierten PAC-Klons, so viele Klone gepickt, dass bei einem reinen Sequenzneugewinn von 100 bp pro Sequenzierungsreaktion der zu sequenzierende genomische PAC-Klon abgedeckt sein sollte (ein Integrat von ca. 140 kb entsprach also mindestens 1400 rekominanten Subklonen).

Diese Dauerkulturen wurden bei –80°C gelagert; eine konsequente Benennung der Subklone (Durchnummerierung der 96-Loch-Mikrotiter-Platten kombiniert mit der Bezeichnung der dort vorgegebenen Koordinaten) machte bereits sequenzierte Klone erneuten Untersuchungen zugänglich.

2.9 Sequenzierungsstrategie

Der Subklonierung in den Sequenzierungsvektor pUC18 x Sma I folgte nach der DNA-Präparation die Sequenzierung der genomischen Fragmente. Die Nachbearbeitung der generierten Sequenzdaten und das Zusammensetzen der einzelnen Sequenz-Fragmente zu einer durchgängigen Gesamtsequenz (Konsensussequenz) wurde mit zwei verschiedenen

Computerprogrammen durchgeführt. Die Sequenzierungsstrategie war jedoch, unabhängig vom verwendeten Computerprogramm, bei allen Klonen gleich.

Durch die Scherung der PAC-DNA entstanden unterschiedlich große, zufällig miteinander überlappende Fragmente, die im Sequenzierungsvektor pUC18 x Sma I kloniert wurden. Es wurden dabei jedoch auch Fragmente generiert, die ausschließlich aus dem Vektorrückgrat der PAC-Klone (pCYPAC2 bzw. pPAC4) stammten. Von einer Selektion dieser Subklone vor der Sequenzierung wurde abgesehen, da die Information über die relative Abdeckung des Vektoranteils durch diese Vektor-tragenden Subklone auch Rückschlüsse auf die relative Abdeckung des zu sequenzierenden genomischen PAC-Integrats erlaubte.
Nach Erstellung der „shot gun“-Bibliothek (siehe 2.8) lässt sich die dann erfolgende Sequenzierung der PAC-Klone in 2 Phasen unterteilen:

1 Sequenzierungs-Phase („sequencing-“ oder „shot gun-phase“) und
2 Abschluss-Phase („finishing-phase“).

Während der Sequenzierungsphase werden alle zur Verfügung stehenden Subklone der jeweiligen Klon-Bibliotheken mit den gleichen Primern sequenziert; diese Phase dient der großangelegten Datengenerierung. In der „finishing“-Phase dagegen wird die Methode des „primer walking“ (s. u.) verwendet, um einzelne Contigs zu einer durchgehenden Konsensussequenz zusammenzufügen.

1 „shot gun“-Phase:

Die Sequenzierungen der Subklone erfolgte sowohl mit dem „M13-universal“- als auch mit dem „M13-reverse“-Primer. Die nach dem Sequenzgel-Lauf erhaltenen Sequenzdaten wurden je nach dem zur Assemblierung benutzten Computerprogramm unterschiedlich verarbeitet:

Assemblierung einer Konsensus-Sequenz für den PAC-Klon 142M6 mit Hilfe des Programmes „SEQUENCHER™3.1“:

Nach der manuellen Editierung falsch identifizierter Basen und der manuellen Entfernung vorhandener Vektoranteile wurden die Sequenzen durch das Computerprogramm „SEQUENCHER™3.1“ miteinander verglichen (siehe 2.12.1). Sequenzen, die gleiche Bereiche des PAC-Klons 142M6 beinhalteten, wurden anschließend automatisch nach definierten Parametern zu „contigs“, überlappenden Einzelsequenzen, zusammengeführt. Die Schließung letzter Lücken und die z. T. manuell vorzunehmenden notwendigen Sequenzkorrekturen wurden ebenfalls mit Hilfe des Computerprogramms „SEQUENCHER™3.1“ ausgeführt.

Assemblierung von Contigs für die Klone PAC-180B11 und PAC-256N10 mit dem Programmpaket PHREDPHRAP und CONSED:

Die nach dem Sequenzgel-Lauf generierten Sequenzdaten wurden entsprechend der verwendeten Nomenklatur umbenannt und in das von PHREDPHRAP benötigte Chromatogramm-Verzeichnis („chromat.dir“) kopiert. Die dort befindlichen Elektropherogramme („trace data“) wurden dann über ein erneutes „base-calling“ von PHRED nachbearbeitet. Hierbei erreicht das „base-calling“-Programm PHRED eine 40-50% geringere Fehlerrate als die von ABI mitgelieferte „base-calling“-Software. Zusätzlich ist das Programmpaket PHREDPHRAP in der Lage, Vektoranteile in den Sequenzen automatisch zu detektieren und zu maskieren, so dass sie die folgende Generierung einer Konsensus-Sequenz nicht stören. Dabei wird sowohl eine Erkennung des Klonierungsvektors (pPAC4 bzw. pCYPAC2) als auch des Sequenzierungsvektors (pUC18) gewährleistet.

Ebenso wie im Programm „SEQUENCHER™3.1“ erfolgt unter PHREDPHRAP nach definierten Parametern eine automatische Zusammenführung überlappender Einzelsequenzen zu „contigs“.

2 „finishing“-Phase:

Die von PHREDPHRAP zusammengefügten „contigs“ wurden mit Hilfe des Programms CONSED während der „finishing“-Phase bearbeitet. Hierbei handelt es sich um ein auf das Programmpaket PHREDPHRAP abgestimmtes graphisches Werkzeug, das der Bearbeitung der zusammengefügten Sequenzen, wie z. B. der Überprüfung des korrekten Zusammenbaus, dient.

Während dieser Phase eines Sequenzier-Projektes sollen noch vorhandene Lücken geschlossen und Sequenzbereiche mit geringer Qualität verbessert werden. Dieser Arbeitsschritt dient der Generierung einer durchgängigen Konsensus-Sequenz.

Die zwischen den Contigs noch vorhandenen Lücken wurden mit Hilfe drei verschiedener Strategien geschlossen:

1 *wiederholte Sequenzierung des Subklons mit Standard-Sequenzierungsprimern:*

Durch technische Probleme existierten von einigen Subklonen nicht beide Randsequenzierungsdaten von den verwendeten Sequenzierungsprimer.

Befand sich ein solcher Subklon an einem Contig-Rand und könnte die fehlende Randsequenz die Lücke schließen, dann wurde die ausstehende Sequenzreaktion an diesem Klon wiederholt.

2 *Vorhandensein eines großen, die Lücke umgreifenden Subklons*:

Von den Rändern der beiden die Lücke flankierenden Contigs wurden Primer generiert, mit deren Hilfe die fehlende Sequenz entweder direkt sequenziert wurde, oder erst über PCR amplifiziert und, wenn erforderlich, durch die Generierung weiterer interner „primer walking"-Primer sequenziert werden konnte.

Der die Lücke umgreifende Subklon diente auch zur Verifikation dieser erhaltenen Sequenz. Hierzu wurde die über M13-PCR oder Restriktion ermittelte Integratgröße des Subklons mit dem erhaltenen *in silico*-Sequenz-Zusammenbau verglichen. Weiterhin wurde das *in silico*-Restriktionsmuster der erhaltenen Sequenzinformation mit Hilfe der restringierten Subklon-DNA überprüft.

3 *Fehlen eines die Lücke überspannenden Subklons:*

Auch in diesem Fall wurde von jedem Contig-Rand ein Primer generiert. Durch beliebige Kombination der so ermittelten Primer konnten dann mit Hilfe der Standard-PCR-Amplifikation die zusammengehörigen Contigs ermittelt werden. Die Größen der erhaltenen PCR-Produkte dienten gleichzeitig zur Kontrolle des nach der Sequenzierung des PCR-Produktes erfolgten Zusammenbaus („alignment") der Sequenzen.

Konnte mit randständigen „primer walking"-Primern kein PCR-Produkt erzeugt werden, wurde mit den betreffenden Primern direkt am PAC-Klon sequenziert.

Nachdem alle Lücken geschlossen waren und die DNA-Sequenz der PAC-Klone als durchgängige Konsensus-Sequenz vorlag, mussten Sequenzunsicherheiten korrigiert und die Qualität der Sequenz einem Projekt-internen Standard angepasst werden: dabei sollte die Konsensussequenz in Exonbereichen an jeder Position doppelsträngig (d.h. von jeder Richtung sequenziert) bzw. in Intronbereichen mindestens durch 3-4 Sequenzen einer Richtung vorliegen.

2.10 Hybridisierungstechniken

2.10.1 Radioaktive Markierung von DNA-Sonden

Die radioaktive Markierung doppelsträngiger DNA erfolgte mit Hilfe der „Random-Primed–Oligo-Labelling"-Methode (*Feinberg & Vogelstein*, 1983). Der Markierungsansatz setzte sich zusammen aus 30 µCi α-^{32}P-dCTP (Amersham, Braunschweig), 6 µl Oligolabelling-Puffer, je 10 pM der spezifischen Primer, 60 µg BSA und 3 U Klenow DNA-Polymerase (2 U/µl; Boehringer, Mannheim). Die Markierungsreaktion wurde für 3 h bei 37°C

inkubiert und nachfolgend zur Entfernung nicht eingebauter Nukleotide über eine Sephadex-G50-Säule (Nick-Columns, Pharmacia) nach Angaben des Herstellers aufgereinigt. Ein Aliquot (1/100 Vol.) der eluierten Sonde wurde im Szintillationszähler gemessen, um die spezifische Aktivität der markierten Sonde zu ermitteln.

2.10.2 PAC-Filter-Hybridisierung

Zur Isolierung putativer Anschlußklone an die sequenzierten PAC-Klone wurden die über das Ressourcenzentrum im Deutschen Humangenom-Projekt in Berlin verfügbaren hochdichten Klonbibliotheken # 704 und # 711 (siehe auch 2.1) durchsucht.

Tab. 2.3: Informationen über die verwendeten hochdichten Klon-Banken. Diese Bibliotheken sind über das Ressourcenzentrum im Deutschen Humangenom-Projekt in Berlin verfügbar.

RZPD-Nummer	# 704	# 711
Name	RPCI1,3-5 Human PAC	RPCI21 Mouse PAC
Hersteller	*P. Ioannou*, *P. de Jong*; Roswell Park Cancer Institut	*K. Osoegawa*, *P. de Jong*; Roswell Park Cancer Institute
Quelle	Blutzellen eines Mannes	Milz einer weiblichen Maus (129/SvevTACfBr)
Vektor	pCYPAC2 (PAC)	pPAC 4 (PAC)
Ø Integratgröße	140 kb	146 kb
Redundanz	16x	13x
Klonanzahl	461 0184 Klone	258 048 Klone
Filteranzahl	15	10

Die PAC-Filter wurden zunächst für 2 h bei 65°C in Church-Hybridisierungspuffer (*Church & Gilbert*, 1984) präinkubiert. Die radioaktiv markierte Sonde (siehe 2.10.1) wurde durch 10-minütiges Kochen denaturiert und nach Abkühlung auf Eis dem Church-Puffer zugegeben. Es wurden zwischen 2x 10^5 und 1x 10^6 cpm pro ml Hybridisierungspuffer eingesetzt. Nach einer Hybridisierungsdauer von 16-20 Stunden bei 65°C wurden die Filter mit zunehmender Stringenz in Na_2HPO_4-Puffer mit 1% SDS für jeweils ca. 20 min so oft gewaschen, bis die auf einem Müller-Geiger-Zähler gemessene Aktivität der Filter zwischen 20 und 50 counts/sec lag. Die Filter wurden auf einer Plastikunterlage mit Frischhaltefolie luftdicht fixiert und ein Röntgenfilm (Hyperfilm-MP RPN8, Amersham) aufgelegt. Die Exposition der Autoradiographie erfolgte in einer mit Verstärkerfolie ausgekleideten Kassette (Quanta III/ Cronex Kassette, DuPont) bei –70°C in der Regel für 1-5 Tage.

2.10.3 RNA-DNA-Hybridisierung (Northern-Hybridisierung)

Der verwendete „Mouse Multiple Tissue Northern Blot“ der Firma CLONTECH enthielt elektrophoretisch aufgetrennte und auf Nylonmembran fixierte RNA aus unterschiedlichen murinen adulten Geweben. Pro Spur waren ca. 2 µg polyA$^+$-RNA aufgetragen.
Der Filter wurde mit einem radioaktiv markierten PCR-Produkt hybridisiert (siehe 2.10.1); dabei erfolgte die Präinkubation, Hybridisierung und das Waschen genau nach den Angaben und mit den empfohlenen Lösungen des Herstellers.
Nach dem Waschen wurde der feuchte Filter in Frischhaltefolie gewickelt und in eine mit Verstärkerfolie ausgekleidete Kassette (Quanta III/Cronex-Kassette, Fa. DuPont, USA) zwischen 2 h und 8 Tagen exponiert.

2.11 Fluoreszenz-*in situ*-Hybridisierung

2.11.1 Markierung der Sonden-DNA für die CISS-Hybridisierung

Die nicht-radioaktive Markierung von Sonden-DNA für die Fluoreszenz-*in situ*-Hybridisierung (siehe 2.11.2) wurde nach der Methode der *nick*-Translation (*MacGregor & Mizuno*, 1976) durchgeführt.
Hierzu wurden 0,5-2 µg Pst I-restringierte, Phenol-Chloroform-aufgereinigte PAC-DNA mit Hilfe des „Nick-Translations-Kit“ (Gibco BRL, USA) gemäß den Herstellerangaben mit Digoxigenin-11-dUTP bzw. Biotin-16-dUTP (Boehringer, Mannheim) markiert. Nach einer 2- bis 4-stündigen Inkubation bei 15°C wurde die Reaktion durch Inkubation auf Eis bzw. Einfrieren bei –20°C abgestoppt. Durch Gelelektrophorese wurde 1/10 Vol. des Markierungsansatzes auf eine erfolgreiche Nick-Translation überprüft. Die Mehrheit der entstandenen DNA-Fragmente sollten hierbei zwischen 200 und 500 bp groß sein.

2.11.2 Fluoreszenz-*in situ*-Hybridisierung (FISH) auf Metaphase-Chromosomen

Zur Überprüfung der chromosomalen Herkunft eines PAC-Klones wurde Pst I-restringierte PAC-DNA über *nick*-Translation markiert (siehe 2.11.1) und auf gespreitete Metaphasechromosomen aus humanen Lymphozyten bzw. einer murinen Fibroblastenzelllinie (*Seipel*, 1996) hybridisiert (*Lemieux et al.,* 1992). Die markierte DNA wurde in Anwesenheit von Cot-1 DNA (Gibco BRL, USA) und einzelsträngiger Lachssperma-DNA (Stratagene, USA) gefällt und in 50% Formamid, 2x SSC, 10% Dextransulfat gelöst.

Die Sonde wurde für 10 min bei 75°C denaturiert und anschließend für 30 min bei 37°C inkubiert, um repetitive Sequenzen abzusättigen. Die nach *Klever et al.* (1991) denaturierten Chromosomenpräparate wurden mit 40-50 ng pro Objektträger der wie oben beschrieben vorbereiteten Sonde beschickt, mit einem Deckglas abgedeckt und in einer feuchten Kammer inkubiert (16 h, 38°C). Die Objektträger wurden dann 3 x 5 min bei 42°C in 50% Formamid/1x SSC und 3 x 5 min in 0,3x SSC gewaschen und für 15 min in 4x SSC/ 0,1% Tween 20/ BSA inkubiert. Die Detektion der Sonde erfolgte nach der Methode von *Lichter et al.* (1988) mit monoklonalen Maus-anti-Dig IgG1- bzw. Avidin-FITC-Antikörpern; TRITC-markierten Hase-anti-Maus-IgG- bzw. anti-Avidin-Biotin-Antikörpern und TRITC-markierten Ziege-anti-Hase-IgG- bzw. Avidin-FITC-Antikörpern (Sigma, Deisenhofen). Die Chromosomen wurden mit DAPI (Boehringer, Mannheim) gegengefärbt und in „Vectashield Mounting Medium“ (Vector Laboratories, USA) eingebettet. Die Bildanalyse erfolgte mit einem DNRBE-Mikroskop der Firma Leica (Bensheim), einem 100x/1.30 PC Fluotar Ölimmersions-Objektiv und durch eine CCD-Kamera mit dem zugehörigen Computer-Software-Paket (Applied Imaging, GB).

2.11.3 *In situ*-Hybridisierung einer α-Satelliten-Sonde

Als Kontrolle für die *in situ*-Hybridisierung wurde biotinylierte bzw. Digoxigenin-markierte Chromosom 11-spezifische α-Satelliten-DNA (Oncor, USA) in einer Konzentration von 10 ng/μl verwendet. Pro Objektträger wurden jeweils 1 μl α-Satelliten-Sonde mit 2 μl 50% Dextransulfat, 5 μl deionisiertem Formamid, 1 μl 20x SSC und 1μl A. bidest gemischt, 10 min bei 75°C denaturiert und mit der *nick*-translatierten Hybridisierungsprobe auf den Objektträger aufgetragen. Die Hybridisierungsbedingungen, die Waschschritte und der Nachweis des Fluoreszenz-markierten α-Satelliten erfolgte wie unter 2.11.2 beschrieben.

2.12 Computerauswertung von Nukleotidsequenzen

Der Hauptteil der computergestützten Sequenzauswertung erfolgte im „world wide web“ (www). Als Arbeitsoberfläche für alle Arbeiten im Internet wurde das Programm NETSCAPE NAVIGATOR™ (Netscape Communications Corporation, USA) verwendet.
Die dabei am häufigsten verwendeten Adressen sind in der Tab. 2.4 zusammengestellt.

Tab. 2.4: Auflistung der am häufigsten verwendeten Internet-Adressen zur Analyse der generierten Nukleotidsequenzen:

BlastN/X	http://www.ncbi.nlm.nih.gov/blast http://www.ncbi.nlm.nih.gov/genome/seq/page.cgi?F=HsBlast.html&&ORG=Hs http://www.ncbi.nlm.nih.gov/gorf/bl2.html
Chromosomen-Karten	http://www-shgc.stanford.edu/Mapping/rh/MapsV2 http://www.ncbi.nlm.nih.gov/genemap/ http://www.ncbi.nlm.nih.gov:80/cgi-bin/Entrez/maps.cgi?
CONSED	http://bozeman.mbt.washington.edu/consed/consed.html
Homologievergleiche	http://www.genebee.msu.su/genebee.html http://genius.embnet.dkfz-heidelberg.de/menu/cgi-bin/w2h/w2h.start (bestfit-option) http://www.ncbi.nlm.nih.gov/blast http://www3.ncbi.nlm.nih.gov/Homology/human11.html http://www.ncbi.nlm.nih.gov/UniGene
Informationen über Chromosom 11 („genome view“)	http://www.ncbi.nlm.nih.gov:80/cgi-bin/Entrez/maps.cgi
PHRAP	http://bozeman.mbt.washington.edu/phrap.docs/phrap.html
PHRED	http://bozeman.mbt.washington.edu/phrap.docs/phred.html
PIP-Maker	http://bio.cse.psu.edu/pipmaker/
Primer-Design	http://www.genome.wi.mit.edu//cgi-bin/primer/primer3_www.cgi
Promotervorhersage (TESS, TSSW, TSSG, NNPP-eukaryotic)	http://searchlauncher.bcm.tmc.edu:9331/seq-search/gene-search.html
Proteinvorhersage	http://expasy.ch
RepeatMasker	http://ftp.genome.washington.edu/cgi-bin/RepeatMasker http://www.mgu.har.mrc.ac.uk/repeat/RepeatMasker

2.12.1 Zusammenbau der Sequenzen

Bei der Sequenzierung des PAC-Klons 142M6 wurden alle vom automatischen DNA-Sequenziergerät generierten Sequenzdaten manuell unter Verwendung des Programmes „ABI PRISM Sequencing Analysis 3.3“ auf einem Apple Macintosh Computer G3 editiert.

Diese editierten Sequenzen wurden mit Hilfe des Programmes SEQUENCHER™ (Version 3.1.1.; Gene Code Corporation, USA) zu einer Gesamtsequenz zusammengesetzt (Parameter: „clean data“; 90% Sequenzübereinstimmung über einen Bereich von mindestens 20 bp) und editiert. Das Programm SEQUENCHER™ stellt erhaltene Überlappungen von Sequenzen graphisch dar und generiert von den überlappenden Bereichen eine durchgängige Basenabfolge. Weiterhin diente es der *in silico*-Erstellung von Restriktionskarten.

Zur Generierung der Gesamtsequenz der PAC-Klone 180B11 und 256N10 wurden die vom Sequenziergerät generierten Daten direkt in das Programmpaket PHREDPHRAP, v.1.1 (*Ewing et al*., 1998; *Ewing & Green*, 1998), geladen. Dieses Programm nahm automatisch ein erneutes „basecalling“ vor und eliminierte mögliche Vektoranteile vor dem Zusammenbau der Einzelsequenzen.
Zum „Finishing“ dieser Projekte wurde das Programm CONSED (v.7.0; Department of Molecular Biotechnology, University of Washington (*Gordon et al*., 1998)), verwendet. Hierbei handelt es sich um ein graphisches Werkzeug, das die Überprüfung und Verifikation der Sequenzqualität erleichtert.

2.12.2 Paarweise Sequenzvergleiche

Zur komparativen Sequenzanalyse wurden „Dotplot“- und „percent-identitiy plot“-Analysen durchgeführt.
Der Vergleich der genomischen Sequenzen von Mensch und Maus wurde mit Hilfe der „dotplot“-Option des Programms MEGALIGN aus dem Programmpaket DNA-STAR™-LASERGENE (DNASTAR Inc., Madison, USA) durchgeführt. Als Parameter wurde dabei standardisiert eine 65%-ige Übereinstimmung in einem mindestens 50 bp großen Fenster gewählt.

2.12.3 Datenbankanalysen

Homologievergleiche einzelner Sequenzabschnitte mit z. T. schon maskierten repetitiven Anteilen (siehe 2.12.4) mit bekannten Nukleotidsequenzen, die in öffentlichen Datenbanken zugänglich sind, wurden mit den Programmen BLASTN (Vergleich von Nukleotidsequenzen mit Einträgen aus Nukleotiddatenbanken) und BLASTX (Übertragung einer Nukleotidsequenz in die sechs möglichen Aminosäureabfolgen und anschließender Vergleich mit Einträgen in Protein-Datenbanken) untersucht (*Pearson & Lipman*, 1988; *Altschul et al.,* 1990; *Altschul et al.,* 1997). In der Regel wurden dazu die Datenbanken „nr“,

est, htgs (high throughput genomic sequence) und sts benutzt, die über das Internet frei zugänglich sind (http://www.ncbi.nlm.nih.gov/blast).

Die Datenbanksuchen wurden am 20. März 2001 abgeschlossen.

2.12.4 Identifikation repetitiver Elemente

Die Identifikation und Maskierung vorhandener repetitiver Elemente in genomischen Sequenzen erfolgte mit dem Programm REPEATMASKER *(Smit & Green*, unveröffentlicht), das u. a. über das Programmpaket RUMMAGE-DP (siehe 2.12.5) zur Verfügung stand. Hierbei konnten Sequenzen bis zu einer Größe von 4999 bp online über http://www.mgu.har.mrc.ac.uk/repeat/RepeatMasker bzw. Sequenzen über 5 kb unter http://ftp.genome.washington.edu/cgi-bin/RepeatMasker maskiert werden.

2.12.5 RUMMAGE-DP-Programm-Paket

Das Programm-Paket RUMMAGE-DP (*Taudien et al*., 2000) wurde für die Analyse der nach Sequenzierung erhaltenen Nukleotidsequenzen benutzt. Es beinhaltet Programme, die der Identifizierung funktionell relevanter Strukturen, wie z. B. Exon-Intron-Struktur, Promotoren, polyA-Stellen u. ä. dienen. Zu den darin verwendeten Exon-Vorhersage-Programmen zählen z. B. GRAIL 2.0 (*Uberbacher*, 1991), MZEF (*Zhang*, 1997), GENSCAN (*Burge*, 1997) und COMPILE (*Schattevoy*, unveröffentlicht), die zur Auswertung der erstellten Nukleotidsequenzen benutzt wurden. Weiterhin sind Programme implementiert, die u. a. den GC-Gehalt bestimmen und somit CpG-Inseln identifizieren können (siehe auch 3.4.4).

2.13 Reagenzien und Materialien

2.13.1 Puffer, Lösungen und Kulturmedien

Agarosegel (x%-ig)	x g Agarose 100 ml 0,5xTBE
Agar-Platten	15 g Agar-Agar ad 1000 ml LB-Medium
Antibiotika	Ampicillin Stammlösung: 50 mg/ml Arbeitskonzentration: 50-75 μg/ml Kanamycin Stammlösung: 25 mg/ml Arbeitskonzentration: 25 μg/ml
Auftragspuffer für Sequenzgele	5 Vol. deionis. Formamid 1 Vol. 25 mM EDTA, pH 8,0 50 mg/ml Blue Dextran
BlueMarker (DNA-Auftragspuffer)	0,25% Bromphenolblau 0,25% Xylencyanol 40% Sucrose
Church-Hybridisierungs-Puffer	1 mM EDTA 0,5 M Na_2HPO_4, pH 7,2 7% SDS
Dialysepuffer (1x)	25 mM Tris 300 mM NaCl 10 mM Na_2EDTA
DNA-Auftragspuffer (6x)	0,25% Bromphenolblau 0,25% Xylencyanol 40% Sucrose
DTM	Je 100 μM dATP, dGTP, dTTP in 250 mM TrisCL 25 mM $MgCl_2$ 50 mM β-Mercaptoethanol pH 7,0
Ethidiumbromid-Färbelösung	5 µg/ml H_2O
FM-Medium (2x)	65% Glycerin 100 mM MgSO4 25 mM Tris/HCl, pH 8,0

LB-Gefrier-Medium	10 g Trypton 5 g Hefeextrakt 10 g NaCl ad 1000 ml A. bidest pH 7,5 7% Glycerin
LB-Medium	10 g Trypton 5 g Hefeextrakt 10 g NaCl ad 1000 ml A. bidest pH 7,5
MOPS (10x)	200 mM 3(N-morpholin) Propan Sulfonsäure 50M mM Na-Acetat 10 mM EDTA pH 7,0
OL	90 OD U 5´-pd(N6)/ml TE
OLB	1 M HEPES/DTM/OL (25/25/7)
OrangeDye-Auftragspuffer (6x)	15 g Sucrose 0,175 g Orange G (Sigma, Steinheim) ad 50ml A. bidest
Puffer 1 (Qiagen) (Resuspendierungspuffer)	50 mM Tris/HCl 10 mM EDTA 100 µg/ml RNase A pH 8,0
Puffer 2 (Qiagen) (Lysepuffer)	200 mM NaOH 1% SDS
Puffer 3 (Qiagen) (Neutralisierungspuffer)	2,55 M Kaliumacetat pH 4,8
Puffer QBT (Qiagen)	750 mM NaCl 50 mM MOPS 15% Ethanol 0,15% Triton X-100 pH 7,0
Puffer QC (Qiagen)	1 M NaCl 50 mM MOPS 15% Ethanol pH 7,0
Puffer QF (Qiagen)	1,25 M NaCl 50 mM MOPS 15% Ethanol pH 7,0

SOB-Medium	20 g Trypton 5 g Hefeextrakt 0,5 g NaCl ad 1 l H20 pH 7,0 10 mM MgCl2 10 mM MgSO4
SOC-Medium	SOB-Medium 20 mM Glucose
SSC-Puffer (10x)	1,5 M NaCl 150 mM Natriumcitrat pH 7,0
TBE-Puffer (1x)	45 mM Tris 45 mM Borsäure 0,625 mM Na2EDTA
TE-Puffer (1x)	10 mM Tris/HCl 1 mM Na2EDTA
Waschpuffer für FISH	4xSSC 0,1% Triton
Waschpuffer nach Church (1994)	0,4 M Na3PO4, pH 7,2 0,1% (w/v) SDS

2.13.2 Enzyme, Radioisotope und Markierungssysteme:

ABI PRISM™ Big-Dye™ Terminator Cycle Sequencing Ready Reaction Kit	Applied Biosystems
ABI PRISM™ Ready Reaction Dye Deoxy Terminator Cycle Sequencing Kit	Applied Biosystems
Biotin-16-dUTP	Boehringer (Mannheim)
Cot 1-DNA	Gibco BRL (USA)
Digoxigenin-11-dUTP	Boehringer (Mannheim)
DNase I	Roche Diagnostics (Mannheim)
DNTPs	Boehringer (Mannheim)
Klenow-DNA-Polymerase	New England Biolabs (USA)
Nick-Translationssystem	Gibco BRL (USA)
Plasmid-Safe™ ATP-dependent DNase	Epicentre Technologies Corporation (USA)
Restriktionsenzyme	New England Biolabs (USA)

Reverse Transkriptase	Perkin Elmer (USA)
RNase A	Boehringer (Mannheim) oder Qiagen (Hilden)
RNase Inhibitor	MBI Fermentas GmbH (St. Leon-Rot)
RNeasy™ Total RNA Kit	Qiagen (Hilden)
Salmon Sperm DNA	Pharmacia (Schweden)
T4-DNA-Ligase	New England Biolabs (USA)
Taq-DNA-Polymerase	Gibco BRL (USA)
α-^{32}P-dCTP	Amersham (Braunschweig)
α-Satellit Chromosom 11 (Mensch)	Oncor (USA)
α-Satellit Chromosom 7 (Maus)	Oncor (USA)

2.13.3 Bakterienstämme und Vektoren

Epicurian Coli ® SURE	Stratagene (USA)
Epicurian Coli ® XL1-Blue MRF´	Stratagene (USA)
pUC18 x Sma I	Pharmacia (Freiburg)

2.13.4 Molekulargewichtsstandards

λ x Hind III	Boehringer Mannheim
λ x Hind III–Concatemere (für PFGE)	Pharmacia (Freiburg)
100 bp-Leiter	Gibco BRL (USA)
123 bp-Leiter	Gibco BRL (USA)
1 kb-Marker	Gibco BRL (USA)

2.13.5 Bezugsquellen

Kunststoff-Verbrauchsmaterialien wurden von den Firmen AB (Abgene, UK), Biozym (Oldendorf), Greiner (Frickenhausen), Nunc (Dänemark) und Roth (Karlsruhe) bezogen.
Alle verwendeten Chemikalien besaßen den Reinheitsgrad „reinst“ oder „pA“.

Acrylamid	Fa. Roth (Karlsruhe)
Agar-Agar	Fa. Difco (USA)
Agarose	Fa. Eurogentec (Belgien)
Ampicililn	Fa. Ratiopharm (Ulm)
Bacto-Hefeextrakt	Fa. Difco (USA)
Bacto-Trypton	Fa. Difco (USA)
EDTA	Fa. Boehringer (Mannheim)
Ethanol	Fa. Riedel-de-Haen (Seelze)
Ethidiumbromid	Fa. Oncor (USA)
Harnstoff	Fa. Roth (Karlsruhe)
IPTG	Fa. Boehringer (Mannheim)
LMP-„SeaPlaque®“-Agarose	Fa. BMA (USA)
Kanamycin	Fa. Ratiopharm (Ulm)
Phenol	Fa. Roth (Karlsruhe)
Phenol/Chloroform	Fa. Roth (Karlsruhe)
SDS	Fa. ICN (USA)
Tris	Fa. Roth (Karlsruhe)
Tris-HCl	Fa. Gerbu (Gaiberg)
Vectashield Mounting Medium	Fa. Vector Laboratories, USA
X-Gal	Fa. Eurogentec (Belgien)

Alle hier nicht aufgeführten Chemikalien wurden von folgenden Firmen bezogen :

Fa. Merck (Darmstadt)

Fa. Roth (Karlsruhe)

Fa. Serva Feinbiochemika (Heidelberg)

Fa. Sigma (Deisenhofen)

2.13.6 Materialien

„DNA-Nebulizer“	Fa. GATC (Konstanz)
Elektroporationsküvetten	Fa. BioRad (München)
Mouse Multiple Tissue Northern Blot	Fa. CLONTECH Laboratories GmbH (Heidelberg)
MultiScreen® Assay-System-Filtrations-Platten	Fa. Millipore (Eschborn)
PAC-Filter	Deutsches Ressourcenzentrum (Berlin)
Qiagen tip-500-Säulen	Fa. Qiagen (Hilden)
Röntgenfilm Hyperfilm-MP RPN 8	Fa. Amersham Life Sciences (Braunschweig)
Sephadex-G50-Pulver	Fa. Pharmacia (Freiburg)
Sephadex-G50-Säulen (Nick-Columns)	Fa. Pharmacia (Freiburg)

2.13.7 Geräte

Agarose-Gel-Elektrophorese	Easy-Cast electrophoresis system (versch. Grössen) Owl Scientific (USA)
Computer	Apple Macintosh G3 Apple Power Macintosh 7600/132 Sun workstation; UltraSparc 1, 200 MHz; Betriebssystem Solaris 2.6 PC 100 MHz, 16 Mb RAM; Betriebssystem Windows 95
Bakterien-Inkubatoren	Typ B5942 Heraeus (Hanau) Schüttelinkubator „Lab-Therm“ Kühner (Schweiz)
DNA-Sequenziergerät	ABI PRISM™ 377 Applied Biosystems (Weiterstadt)
Elektroporationsgerät	Gene Pulser BioRad (München)
Gel-Dokumentations-Gerät	E.A.S.Y. System Herolab (Wiesloch)
Hybridisierungsofen	Hybridiser HB-2 Techne (GB)

PCR-Geräte	PTC100™ / PTC200™ MJ Research (USA) Gene Amp PCR System PE 9700 PE Applied Biosystems (USA)
Pulsfeld-Gelelektrophorese-Gerät	CHEF Mapper™ BioRad (München)
Photometer	Ultrospec III Pharmacia (Freiburg)
Spannungsgeräte	LKB-GPS 200/400 Pharmacia (Freiburg)
Sterilbank	Antair BSK
Szintillationszähler	Liquid Scintillations Counter WALLACE-1410 Pharmacia (Freiburg)
Vakuumtrockner	Univapo 100H, Refrigerated Aspirator Uni-Equip (Martinsried)
Zentrifugen	Sigma 3K12 Zentrifuge Sigma (Deisenhofen) Sorvall RT6000D Du Pont (USA) Sorvall RC5C Du Pont (USA) Zentrifuge 5415C Eppendorf (Hamburg)
Wasserbäder	Thermomix-1420 Braun AG (Melsungen) F10 Julabo (USA)

2.13.8 Primer

Die Primer wurden hauptsächlich über die Firmen Life Technologies (jetzt Invitrogen, Karlsruhe) und Metabion (München) bezogen und sind auf der beigefügten CD aufgeführt (siehe 7.4).

3 Ergebnisse

In der vorliegenden Dissertation sollte eine Sequenzierung und Analyse eines ca. 250 kb und ca. 200 kb großen Bereiches im menschlichen und murinen Genom durchgeführt werden. Diese Arbeit stellt einen Teil eines größeren Projektes dar, das im Rahmen des Deutschen Humangenom-Projektes (Förderungs-Nummer 01KW9624) durchgeführt wurde und die komparative Sequenzanalyse eines 1 Mb großen Bereiches auf dem kurzen Arm des Chromosoms 11 des Menschen (11p15.3) sowie des homologen Bereiches auf Chromosom 7 der Maus (*Blake et al.,* 2000) zum Ziel hatte. Anhand spezifischer Sequenzkonservierungen zwischen den untersuchten Spezies sollten bereits bekannte Gene innerhalb der analysierten Regionen lokalisiert sowie neue Gene und deren regulatorische Sequenzen identifiziert werden. Durch den Vergleich der humanen mit der murinen Sequenz wird ein idealer Zugang zu funktionellen Analysen im Maussystem ermöglicht.

3.1 Isolierung und Auswahl der zu sequenzierenden Klone

Ausgangspunkte der im Rahmen des Deutschen Humangenom-Projektes durchgeführten Analyse waren die vier Gene *LMO1*, *ST5*, *CEGF1* und *WEE1*, die innerhalb des zu untersuchenden Bereiches in der Chromosomenregion 11p15.3 lokalisiert sind (siehe auch Abb. 3.1). Dieser Bereich wird beim Menschen proximal von *WEE1* und distal von *LMO1* begrenzt. *WEE1* spielt eine wichtige Rolle bei der Regulation des Zellzyklus (*Heald et al.*, 1993), während das *LMO1*-Genprodukt an der Ausbildung des Zentralnervensystems beteiligt ist und mit der Entstehung einer bestimmten Form der lymphoblastischen Leukämie in Verbindung gebracht wird (*Boehm et al.*, 1988). Beiden Genen und dem innerhalb der Region lokalisierten Gen *ST5* wird Tumorsuppressor-Aktivität (*Richard et al.*, 1993; *Lichy et al*., 1992) zugeschrieben. Über FISH-Analyse wurden die vier Gene relativ zueinander sowohl in die Region 11p15.3 des Menschen als auch auf Chromosom 7 der Maus kartiert (*Seipel*, 1996). Ziel der vorliegenden Dissertation war es, die genomische Region um das Gen *WEE1/Wee1* (*Igarashi et al.*, 1991) sowohl im Menschen als auch in der Maus zu sequenzieren.

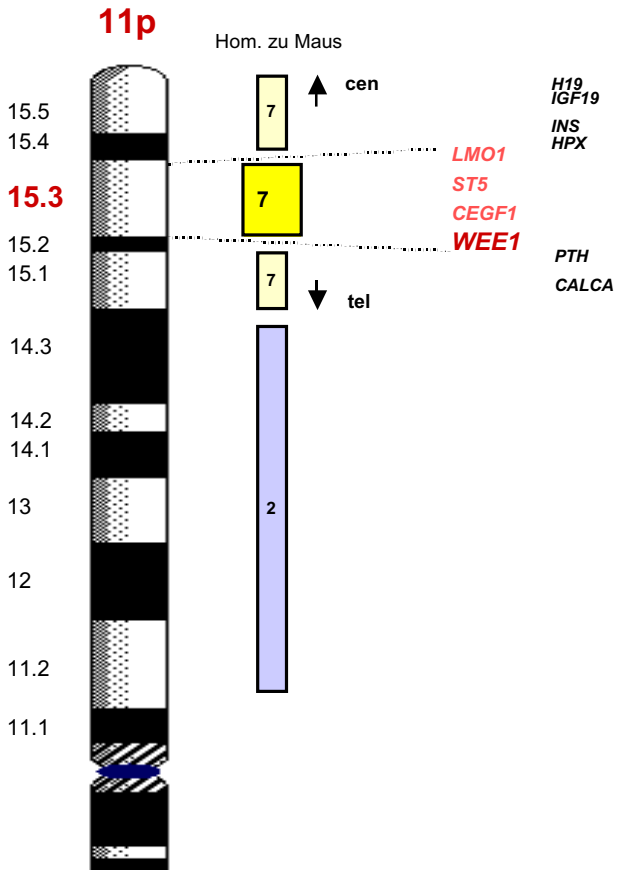


Abb. 3.1: Darstellung der humanen Chromosomenregion 11p15.3. Es sind Syntäniegruppen des murinen Chromosoms 7 (hellblau) bzw. 2 (gelb) gezeigt, die bestimmten chromosomalen Regionen aus 11p entsprechen. Weiterhin wurde die relative Lage der bei der Sequenzierung als Startpunkte dienenden Gene *LMO1*, *ST5*, *CEGF1* und *WEE1* dargestellt.

3.1.1 Klone aus der Region um das *WEE1*-Gen des Menschen

Für die Sequenzanalyse der genomischen Region um das humane *WEE1*-Gen auf dem kurzen Arm von Chromosom 11 (11p15.3) diente der von *Seipel* (1996) isolierte und über FISH-Analyse verifizierte PAC-Klon 142M6 als Startpunkt. Der PAC-Klon 142M6 besitzt eine Integratgröße von ca. 132 kb und wurde aus der humanen PAC-Klonbibliothek Nr. 704 (*Ioannou et al.,* 1994) isoliert. Mit Hilfe spezifischer Primer aus dem Randbereich des Klons PAC-142M6 wurde ein PCR-Produkt generiert, das als Sonde bei der Suche nach Anschlussklonen eingesetzt wurde. Durch die Hybridisierung der radioaktiv markierten Sonde auf hochdichte PAC-Klon-Filter wurde der Klon PAC-180B11 isoliert, der mit dem Klon PAC-142M6 überlappt und über Pulsfeld-Gelelektrophorese auf eine Größe von ca. 120 kb geschätzt wurde. Die Klone PAC-142M6 und PAC-180B11 wurden im Rahmen der vorliegenden Arbeit vollständig sequenziert (siehe 3.2). Zusätzlich wurden weitere PAC-Klone isoliert, um den genomischen Bereich zwischen den Startpunkten *WEE1* und *CEGF1* komplett durch einen PAC-Contig abzudecken. Diese Klone wurden mittels PCR-Technik auf

eine Überlappung zu sequenzierten Klonen hin überprüft und die Integratränder wurden ansequenziert. Der erstellte Kloncontig ist in Abb. 3.2 zusammen mit allen im Rahmen des Gesamtprojektes isolierten und sequenzierten Klonen dargestellt.

3.1.2 Klone aus der orthologen Region der Maus (Chromosom 7)

Vorversuche haben gezeigt, dass die vier als Startpunkte dienenden humanen Gene *WEE1*, *CEGF1*, *ST5* und *LMO1* aus der Chromosomenregion 11p15.3 auch in ihrer relativen Lage zueinander auf Chromosom 7 der Maus konserviert vorliegen (*Seipel*, 1996). Deshalb konnte davon ausgegangen werden, dass es sich bei der in Mensch und Maus zu untersuchenden Chromosomenregion um einen orthologen Bereich handelt. Allerdings ist dieser Chromosomenbereich in der Maus relativ zum Menschen invertiert (siehe http://www3.ncbi.nlm.nih.gov/Homology/human11.html).

Um einen genomischen PAC-Klon aus der entsprechenden Region der Maus auf Chromosom 7 zu isolieren, wurden hochdichte murine Klonfilter der PAC-Bibliothek # 711 mit einer für das humane Zinkfingergen *ZNF143* spezifischen Sonde hybridisiert (siehe 2.10.1 und 2.10.2). Das von *Tommerup & Vissing* 1995 beschriebene humane Gen *ZNF143* konnte innerhalb des humanen Klons PAC-142M6 lokalisiert werden. Ein für das Exon 15 des *ZNF143*-Gens spezifisches PCR-Produkt diente als Sonde, da bei der Hybridisierung mit einer Maus-*Wee1*-spezifischen Sonde viele falsch-positive Klone erhalten wurden (*Seipel*, 1996). Der auf diese Weise isolierte PAC-Klon 256N10 wurde mit Hilfe spezifischer Primer sowohl aus dem 5´- als auch aus dem 3´-Bereich des murinen *Wee1*-Gens mittels PCR verifiziert. Die Integratgröße des Klons PAC-256N10 wurde nach Not I-Restriktion über Pulsfeld-Gelelektrophorese (siehe 2.2.4) auf ca. 200 kb geschätzt. Die Lage des murinen PAC-Klons 256N10 im Verhältnis zu den sequenzierten humanen PAC-Klonen 142M6 bzw. 180B11 relativ zu *WEE1* und den weiteren, im Rahmen des Gesamtprojektes sequenzierten Klonen, ist in Abb. 3.2 dargestellt.

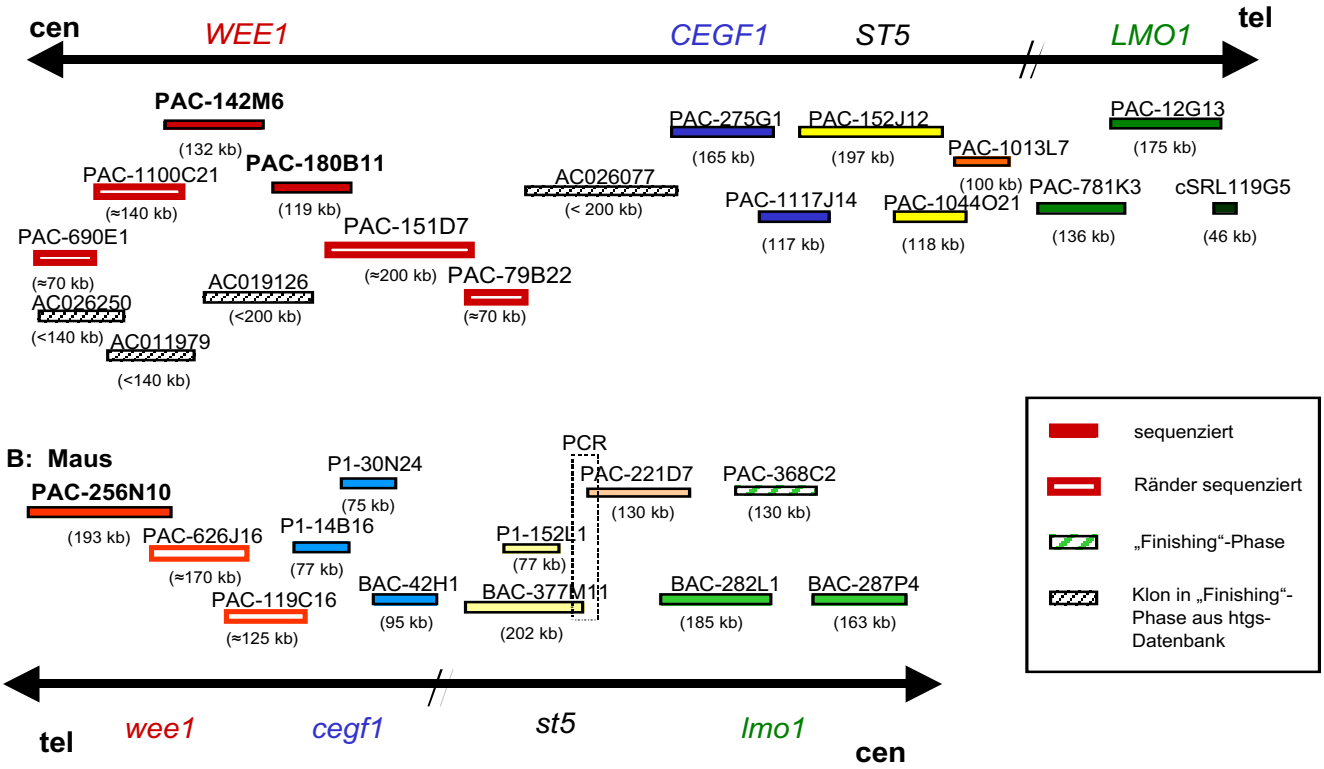


Abb. 3.2: Anordnung der im Rahmen des Projektes sequenzierten Klone relativ zu den als Startpunkten dienenden Genen. Die im Rahmen der vorliegenden Dissertation sequenzierten PAC-Klone sind rot unterlegt, die isolierten Klone rot eingerahmt. Die restlichen Klone sind den anderen Startpunkten zugeordnet und wurden jeweils im Rahmen einer eigenständigen Dissertation untersucht. Der gepunktete Kasten mit dem Text „PCR" gibt einen über PCR erfolgten Lückenschluss an. *LMO1/Lmo1*: T. Brückmann (und Diplomarbeit Silke Schlaubitz: cSRL119G5); Lücke zwischen *ST5* und *LMO1*: A. Mujica; *ST5/St5*: C. Amid; *CEGF1/cegf1*: A. Bahr; *WEE1*/*Wee1*: A. Cichutek

Zusätzlich wurde der PAC-Klon 256N10 über eine Fluoreszenz-*in situ*-Hybridisierung (FISH) chromosomal lokalisiert. Zur Identifizierung des murinen Chromosoms 7 wurde eine simultane Hybridisierung des PAC-Klons 256N10 und des murinen BAC-Klons 287P4 durchgeführt, welcher von *Kleyn et al.* (1996) auf dem Chromosom 7 der Maus lokalisiert wurde und das Gen *LMO1* enthält. Beide Sonden zeigten klare Signale auf demselben murinen Chromosom (Abb. 3.3), so dass durch die Kenntnis der Herkunft des BAC-Klons 287P4 von Chromosom 7 auch die Lokalisation des Klons PAC-256N10 auf dem murinen Chromosom 7 bestätigt werden konnte. Weiterhin zeigte sich eine Inversion der genomischen Region zwischen *Wee1* und *Lmo1* in der Maus im Verhältnis zum Menschen, wo *WEE1* den untersuchten genomischen Bereich proximal und *LMO1* distal begrenzt. Darüber hinaus konnte auf diese Weise gezeigt werden, dass es sich bei dem PAC-Klon 256N10 nicht um einen chimären Klon handelt.

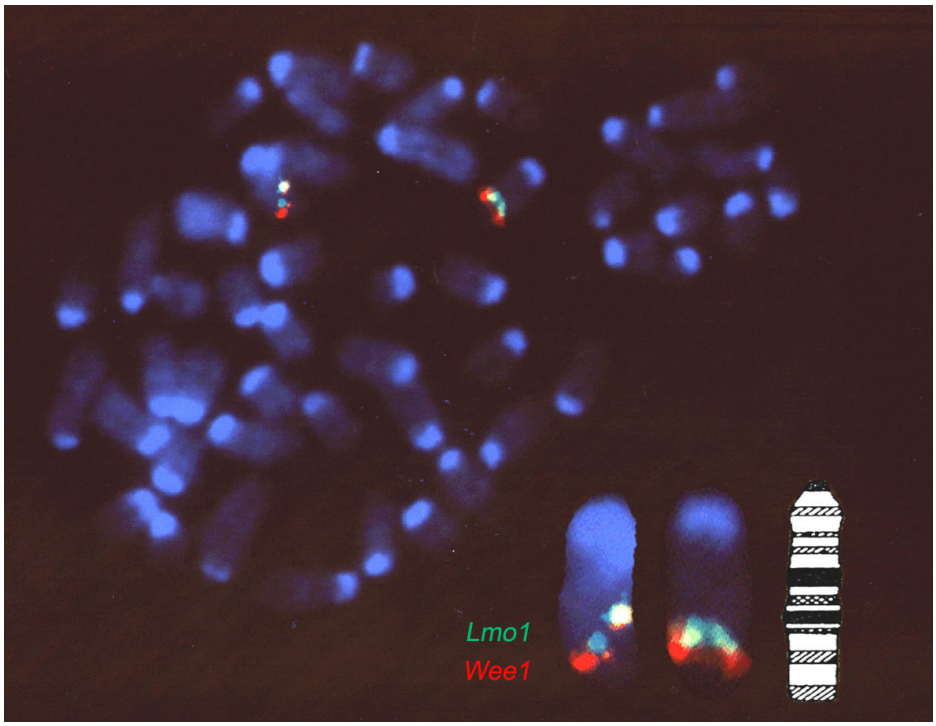


Abb. 3.3: „2-Farben-FISH"-Analyse auf Maus-Metaphase-Chromosomen. Der murine PAC-Klon 256N10, welcher das Gen *Wee1* vollständig beinhaltet, wurde biotinyliert und simultan mit dem das Gen *Lmo1* enthaltenden, Digoxigenin-markierten Klon BAC-287P4 hybridisiert. Da der BAC-Klon 287P4 von *Kleyn et al.* (1996) auf das Maus-Chromosom 7 kartiert wurde, konnte somit auch die Lokalisation des *Wee1*-Gens auf Maus-Chromosom 7 bestätigt werden. Weiterhin zeigte sich, daß die Orientierung der genomischen Region zwischen *Wee1* und *Lmo1* in der Maus im Verhältnis zur humanen Anordnung auf Chromosom 11 invertiert vorliegt.

3.2 Sequenzierung der humanen Klone PAC-142M6 und PAC-180B11 sowie des murinen Klons PAC-256N10

3.2.1 Sequenzierungsstrategie

Wie unter 2.8 detailliert beschrieben, wurden von den zu sequenzierenden PAC-Klonen „shot-gun“-Bibliotheken mit Subklonen unterschiedlicher Integratgröße angelegt. Es sollte gewährleistet sein, dass bei einem durchschnittlichen Sequenzneugewinn von nur 100 bp pro Sequenzreaktion durch einen der beiden standardmäßig verwendeten Primer die Integratlänge des PAC-Klons abgedeckt ist. Deshalb wurden für 100 kb PAC-Integrat etwa 1000 Subklone mit unterschiedlichen Integratgrößen (siehe Tab. 3.1) isoliert. Die verschieden großen Subfragmente der PAC-Klone wurden mit Hilfe des Primers M13-universal ansequenziert (siehe 2.9) und die erhaltenen Sequenzdaten mit Hilfe von Computerprogrammen auf identische Sequenzabschnitte in unterschiedlichen Subklonen untersucht, die sich somit zu überlappenden Einzelsequenzen („contigs“) aneinanderfügen ließen.

Ein Teil der Subklone wurde zusätzlich unter Verwendung des Primers M13-revers sequenziert, um noch vorhandene Lücken zwischen den Contigs zu schließen. Nach der Assemblierung dieser zusätzlichen Sequenzen in das Gesamtprojekt wurden die noch bestehende Lücken mittels „primer walking“ geschlossen. Hierbei wurde mit Hilfe spezifischer Primer innerhalb von Lücken-überspannenden Subklonen oder direkt am PAC-Klon sequenziert. Weiterhin wurden Lücken zwischen Contigs, die von keinem Subklon überspannt wurden, mit Lücken-flankierenden Primern über PCR (siehe 2.5) amplifiziert und anschließend sequenziert. Nach Erhalt einer durchgängigen Konsensus-Sequenz wurde die Qualität dieser Sequenz überprüft und, wenn nötig, durch erneute Sequenzreaktionen verbessert. Hierbei wurden alle potenziellen Exonbereiche doppelsträngig, Intronbereiche entweder doppelsträngig oder durch mindestens drei Sequenzen in einer Richtung sequenziert.

3.2.2 Statistik zur Sequenzierung der humanen Klone PAC-142M6 und PAC-180B11

Die Sequenzlängen der PAC-Klone 142M6 und 180B11 betragen 131 998 bp bzw. 119 794 bp und überlappen in einem Bereich von ca. 7,8 kb. Somit decken die Integrate beider Klone zusammen einen genomischen Bereich von 243 966 bp der humanen chromosomalen Region 11p15.3 ab. Zur Generierung der durchgängigen humanen Konsensus-Sequenz wurden insgesamt 4140 Sequenzreaktionen durchgeführt.

Für den Klon PAC-142M6 lässt sich eine detaillierte Sequenzierungsstatistik aufstellen, da zu dessen Assemblierung das Programm SEQUENCHER™3.1 verwendet wurde. Wie unter 2.9 aufgeführt, müssen bei Verwendung dieses Programms die Sequenzdaten manuell nachbearbeitet werden, so dass Aussagen über die Unterschiede zwischen der vom Sequenziergerät erzeugten und der editierten Sequenz gemacht werden können. Bei dem mit Hilfe des Programmes PHREDPHRAP zusammengesetzten Klon PAC-256N10 lassen sich solche Unterschiede nicht erkennen, weil die vom Sequenziergerät erzeugten Sequenzen automatisch auf Vektorkontamination und Qualität hin untersucht und korrigiert werden, bevor sie in „contigs“ zusammengesetzt werden. Somit ist hierbei kein Vergleich möglich.

Für den mit einer Integrat-Größe von knapp 120 kb kleineren Klon PAC-180B11 wurden deutlich mehr Sequenzreaktionen (2606) durchgeführt als für den fast 132 kb großen PAC-Klon 142M6. Grund dafür waren Probleme beim Zusammensetzen der Konsensus-Sequenz des Klons PAC-180B11, die durch einen größeren Anteil repetitiver Elemente (siehe auch 3.4.5) verursacht wurden. Eine Übersicht über die abhängig vom Assemblierungs-Programm verfügbaren Daten der Sequenzierungsstatistik beider humaner PAC-Klone gibt die Tab. 3.1.

Tab. 3.1: Sequenzierungsstatistik der beiden humanen Klone PAC-142M6 und PAC-180B11. Die durchschnittliche Leseweite ergibt sich aus dem Quotienten von Rohbasen und der Anzahl der durchgeführten Sequenzreaktionen; die Redundanz ist über den Quotienten von ermittelten Rohbasen und tatsächlicher Integratgröße definiert. n. f.: nicht feststellbar (s. o.).

Klon **Accession-Nummer**	**PAC-142M6** **AJ 277546**	**PAC-180B11** **AJ295844**	**Mensch gesamt**
verwendetes Programm	SEQUENCHER™3.1	PHREDPHRAP / CONSED	
geschätzte Größe nach PFGE (bp)	142 000	125 000	
Tatsächliche Größe (bp)	131 998	119 794	243 966
Größenbereiche der subklonierten Fragmente	500 bp – 1,5 kb 1,5 – 2,5 kb	500 bp – 1,3 kb 1,3 – 2 kb 2 - ≤ 3 kb	
Anzahl isolierter Subklone	1452	1344	
auswertbare Sequenzen	1534	2606	4140
Rohbasen (bp)	926 770	n. f.	n. f.
Ø Leseweite (bp)	604	n. f.	n. f.
Redundanz	7,02	n. f.	n. f.
Primeranzahl	202	128	330

Die generierten humanen Sequenzen werden aus Platzgründen nicht in der vorliegenden Arbeit aufgeführt, sondern sind auf der beigefügten CD-ROM einsehbar (siehe 7.4). Diese Sequenzen entsprechen der Orientierung der PAC-Klone im Gesamt-Contig (von *WEE1* bis *LMO1;* siehe Abb. 3.2). Die generierten Sequenzen sind außerdem in der GenBank-Datenbank unter den Accession-Nummern AJ277546 (PAC-142M6) und AJ295844 (PAC-180B11) in revers-komplementärer Orientierung eingetragen.

Beide humanen PAC-Klone wurden entsprechend der oben aufgeführten Parameter sequenziert. Somit liegen 99,54% der Gesamtsequenz doppelsträngig vor, nur in den Abschnitten nt 44 044 - nt 44 388, nt 66 951 - nt 67 323, nt 77 912 - nt 78 312 sowie im Abschnitt nt 180 116 - nt 180 125 konnte wahrscheinlich aufgrund problematischer Sekundärstrukturen die DNA-Sequenz nur auf einem Strang ermittelt werden.

Darüber hinaus weist die sequenzierte humane DNA-Sequenz insgesamt 5 Positionen auf, an denen sich das hier befindliche Nukleotid nicht eindeutig bestimmen lässt. Solche Basen wurden nach dem allgemein gültigen Code für ambige Nukleotide bezeichnet. So unterscheiden sich die Klone PAC-142M6 und PAC-180B11 an zwei Positionen innerhalb der überlappenden Region von Position nt 124 173 - nt 131 998 (Überlappung: 7826 bp). Während der Klon PAC-180B11 an der Position nt 125 288 eindeutig ein Cytosin zeigt, befindet sich an entsprechender Position des PAC-Klons 142M6 ein Adenosin. Deshalb wurde diese Position in der Konsensus-Sequenz mit einem M belegt. Auch an Position nt 127 701 findet sich ein Sequenzunterschied. Während der PAC-180B11 hier deutlich ein Cytosin aufweist, zeigt sich bei dem Klon PAC-142M6 ein Thymidin. Diese Position wurde mit Y gekennzeichnet. Desweiteren weist der Klon PAC-142M6 drei Positionen auf, an denen trotz wiederholter doppelsträngiger Sequenzierung und manueller Editierung nicht eindeutig festgelegt werden konnte, um welche Base es sich hierbei handelt. An Position nt 35 616 konnte nicht entschieden werden, ob es sich um ein Thymidin oder ein Guanosin handelt; folglich wurde diese Position durch K belegt. Weitherhin konnte nicht geklärt werden, ob die Basen an den Positionen nt 91 823 und nt 104 609 ein Adenosin oder Guanosin darstellen und wurden deshalb durch ein R gekennzeichnet. Alle Sequenzunstimmigkeiten lagen außerhalb kodierender Bereiche und wurden deshalb nicht weiter untersucht.

3.2.3 Statistik zur Sequenzierung des murinen Klons PAC-256N10

Die komplette Sequenzierung des murinen Klons PAC-256N10 ergab eine Integratgröße von 192 519 bp. Für die Erstellung der PAC-256N10-Konsensus-Sequenz wurden 2382 Einzelsequenzierungen und 75 „primer walking“-Oligonukleotide benötigt. Da der Zusammenbau einzelner Sequenzen zu einer Konsensus-Sequenz mit Hilfe des

Programmes PHREDPHRAP durchgeführt wurde, konnte nur eine eingeschränkte Sequenzierungsstatistik erstellt werden (siehe Tabelle 3.2).

Tab. 3.2: Sequenzierungsstatistik des murinen Klons PAC-256N10. Die durchschnittliche Leseweite, Anzahl der Rohbasen und die Redundanz konnte nicht ermittelt werden, da die Nachbearbeitung der Sequenzen mit Hilfe des Programmes PHREDPHRAP erfolgte und somit keine Angaben über Rohbasen und durchschnittliche Leseweite möglich waren. n. f.: nicht feststellbar

Klon **Accession-Nummer**	**PAC-256N10** **AJ278435**
verwendetes Programm	PHREDPHRAP / CONSED
geschätzte Größe nach PFGE (bp)	200 000
Tatsächliche Insertgröße (bp)	192 519
Größenbereiche der subklonierten Fragmente	500 bp - 1,4 kb 1,4 - 2 kb 2 - 3 kb
Anzahl isol. Subklone	1824
auswertbare Sequenzen	2382
Rohbasen (bp)	n.f.
Ø Leseweite (bp)	n.f.
Redundanz	n.f.
Primeranzahl	75

Die Konsensus-Sequenz des murinen Klons PAC-256N10 ist in der innerhalb des Gesamtcontigs vorliegenden Orientierung auf der beigefügten CD-ROM gespeichert; der hierzu revers-komplementäre, unter der Accession-Nummer AJ278435 identifizierbare Eintrag in der Datenbank GenBank ist dort ebenfalls aufgeführt.

Die DNA-Sequenz des Klons PAC-256N10 liegt nicht durchgängig vor, sondern weist eine Lücke auf, die auch unter Verwendung unterschiedlicher Sequenzierungsprotokolle nicht geschlossen werden konnte. Grund für den Abbruch der Sequenzreaktionen war vermutlich eine Basenabfolge von komplementären Dinukleotidwiederholungen ((GT)n bzw. (CA)n) beiderseits der Lücke, die eine Sekundärstruktur ausbilden, welche auch nicht durch verlängerte Denaturierungszeiten und Zugabe von DMSO bzw. Glycerol zum Sequenzierungsansatz zerstört werden konnte. Die Größe dieser Lücke konnte anhand eines überspannenden Subklons (m10F4) aus der „shot-gun“-Bibliothek genau abgeschätzt werden. Durch die Bestimmung von dessen Integratgröße über Restriktion (da die

Amplifikation des Inserts aufgrund der vorliegenden schwierigen Sekundärstruktur nicht möglich war) und der davon über Sequenzierungen abgedeckten Bereiche konnte die Lückengröße auf maximal 170 bp eingeschränkt werden und erstreckt sich von Position nt 9884 bis Position nt 9983 (durch eine Basenabfolge von 100 N´s markiert). Diese konstruierte Konsensus-Sequenz diente bei allen durchgeführten Untersuchungen als Grundlage. Der murine Klon PAC-256N10 wurde mit Ausnahme der vorhandenen Lücke komplett doppelsträngig sequenziert. An Position nt 103 874 wies er eine Ambiguität zwischen Thymidin und Cytosin auf, die entsprechend des Standard-Codes durch Y gekennzeichnet wurde.

3.3 Auswertung der Sequenzen

Zur Identifizierung von Genen wurden die humanen und murinen Sequenzen unter drei verschiedenen Gesichtspunkten analysiert (siehe Abb. 3.4):

i) *Suche nach Genen über Exonvorhersageprogramme*. Die generierten Sequenzdaten wurden hauptsächlich mit Hilfe des Programmpakets „RUMMAGE-DP“ (siehe 2.12.5) analysiert. Da es sich hierbei um eine Kombination verschiedenster Programme handelt, können somit mehrere Analysen, wie z. B. Homologievergleiche mit diversen Datenbanken oder die Maskierung repetitiver Bereiche durch unterschiedliche Programme, in einem Arbeitsgang durchgeführt werden. Darüber hinaus sollten durch die Verwendung von Exon-Vorhersage-Programmen bisher unbekannte Gene, die nicht durch EST-Klone repräsentiert sind, identifiziert werden. Die parallele Verwendung diverser Exon-Vorhersage-Programme erleichterte hierbei die Beurteilung der Vorhersagesicherheit einzelner Programme. Da in den analysierten Sequenzen auch schon bekannte Gene lokalisiert waren, konnte die Vorhersagegenauigkeit der unterschiedlichen Exon-Vorhersage-Programme relativ gut eingeschätzt werden.

ii) *Datenbanksuchen*. Parallel zur Analyse über das Programm „RUMMAGE-DP“ wurden repetitive Bereiche der humanen und murinen Sequenz getrennt mit Hilfe des Programmes REPEATMASKER maskiert. Die sich anschließenden Datenbanksuchen (i.d.R. nr / htgs) der maskierten Sequenzen über den Algorithmus blastn zeigten meist Homologien zu bekannten Sequenzen. Auf diese Weise ließen sich sowohl bekannte Gene als auch solche Gene identifizieren, die in anderen Spezies schon beschrieben waren. Zeigten sich keine oder nur schwache Homologien zu bekannten Genen, wurden Datenbanksuchen in der EST-Datenbank durchgeführt um festzustellen, ob sich mit Hilfe dieser Daten eine putative cDNA-Sequenz erstellen läßt. Parallel dazu wurden experimentelle Untersuchungen wie RT-PCRs bzw. Northern-Analysen vorgenommen, um die „*in silico*“-erhaltenen Daten zu verifizieren und zu ergänzen.

iii) *Komparative Sequenzanalyse*. Die Sequenzvergleiche zwischen Mensch und Maus wurden unter Verwendung von Dotplot–Analysen durchgeführt. Liegen Konservierungen einer definierten Größe vor, wird innerhalb eines Koordinatensystems ein Punkt („dot“) gesetzt, der die Position dieses konservierten Bereiches innerhalb der zu vergleichenden Sequenzen markiert. Das Ergebnis dieses Sequenzvergleiches wurde mit Hilfe zweier Programme mit unterschiedlichen Analyseschwerpunkten graphisch dargestellt (siehe 2.12.2 und 3.4).

Die Abb. 3.4 zeigt schematisch die verfolgte Strategie zur Sequenzauswertung im Überblick.

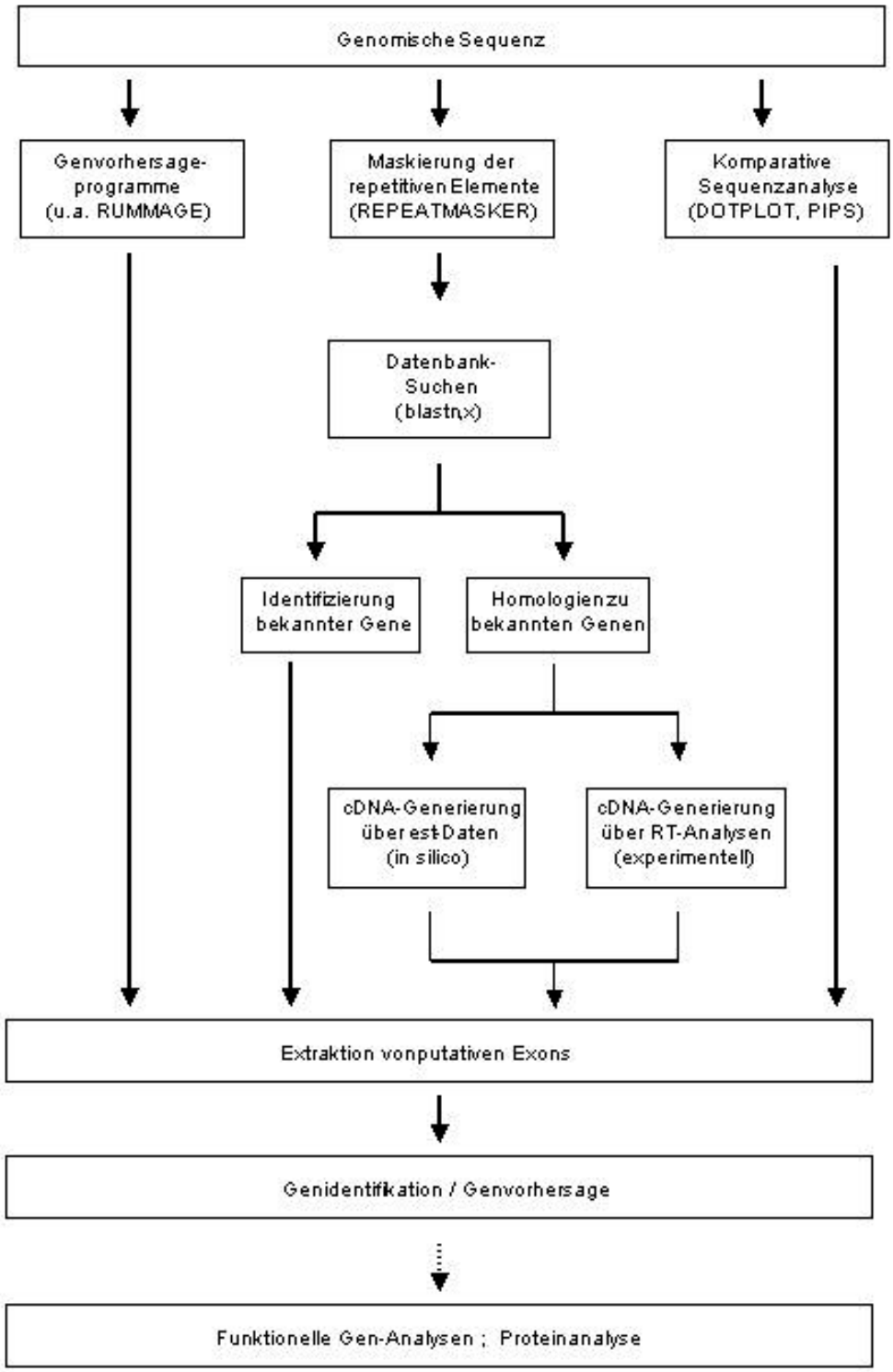


Abb. 3.4: Überblick über die verfolgten Strategien zur Genidentifizierung innerhalb der erstellten genomischen Sequenzen. Hierbei stellten die Analysen mit Hilfe von Exonvorhersage-Programmen, Homologievergleichen zu Datenbankeinträgen sowie die komparative Sequenzanalyse die hauptsächlichen Untersuchungsmethoden dar. Details siehe Text.

Diese drei aufgeführten Analysewege bieten bei Kombination aller Ergebnisse eine sehr gute Möglichkeit, unbekannte Gene innerhalb einer großen genomischen Region zu identifizieren. Mit dieser Strategie konnten insgesamt drei bekannte Gene sowie ein neues Pseudogen in dem untersuchten humangenomischen Bereich identifiziert werden. Da von den identifizierten Genen jeweils schon die cDNA-Sequenz veröffentlicht war, erfolgte eine genaue Untersuchung der genomischen Struktur über Homologievergleiche mit Hilfe der cDNA-Sequenz und eine exakte Lokalisation dieser Gene innerhalb des untersuchten genomischen Bereiches.

Eines der drei in der vorliegenden humanen Sequenz lokalisierten Gene konnte im Rahmen dieser Arbeit über komparative Sequenz- und Homologievergleiche erstmals auch in der Maus beschrieben werden. Die Abb. 3.5 zeigt eine Übersicht über die Lokalisation und Orientierung der Gene innerhalb der genomischen Sequenzen, auf die im Folgenden im Detail eingegangen wird.

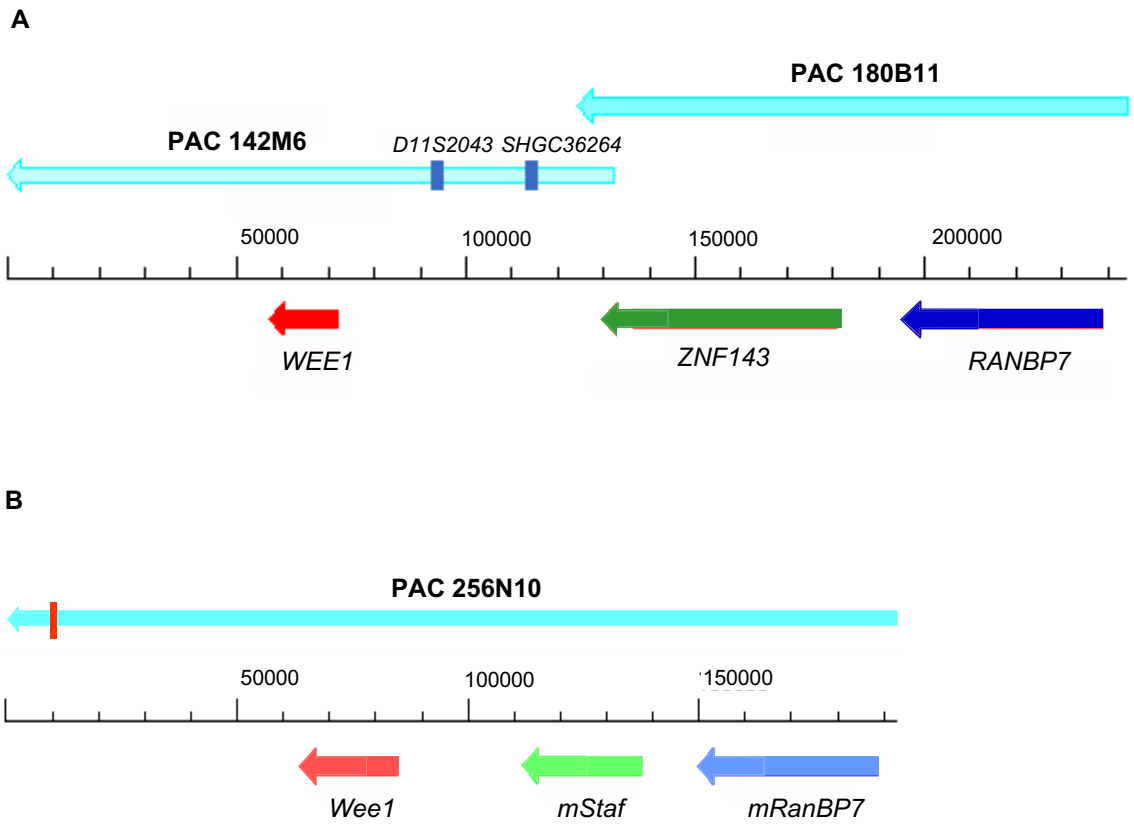


Abb. 3.5: Darstellung der genomischen Anordnung und Ausdehnung der identifizierten Gene innerhalb der sequenzierten humanen (A) bzw. murinen (B) Sequenz. Die Lage und Orientierung der PAC-Klone ist unter Berücksichtigung des Gesamt-Contigs dargestellt; farbige Pfeile kennzeichnen die genomische Ausdehnung einzelner Gene. Der rote senkrechte Balken innerhalb der murinen Sequenz markiert die Lage des nicht-sequenzierten Bereiches von max. 170 bp.

3.3.1 Identifikation bekannter Gene

3.3.1.1 *WEE1*/*Wee1*

Das *WEE1*-Gen kodiert für eine Tyrosin/Serin-Kinase, die den Übergang von der postsynthetischen Phase (G2-Phase) zur Mitose koordiniert und aufgrund der Mitose-inhibierenden Eigenschaften als Tumorsuppressor-Gen gilt (*Heald et al*., 1993).

WEE1

Das humane Gen *WEE1*, dessen mRNA-Länge von *Watanabe et al.* (1995; Acc.-Nr. U10564) mit einer Größe von 2194 bp beschrieben wurde, liegt, da es als Startpunkt für die Sequenzierung der genomischen Umgebung diente, wie erwartet vollständig in der sequenzierten Region (Abb. 3.5). Das Gen setzt sich aus 11 Exons zusammen, die sich über einen genomischen Bereich von knapp 15 kb erstrecken (siehe Abb. 3.6).

Beim Vergleich der von *Watanabe et al.* (1995) veröffentlichten mRNA-Sequenz des humanen *WEE1*-Gens mit den entsprechenden kodierenden Bereichen des PAC-Klons 142M6 ergibt sich an einer Position ein Unterschied in der Nukleotidsequenz. An Position 505 der veröffentlichten mRNA-Sequenz befindet sich ein Cytosin, während sich an der entsprechenden Stelle (Position 243 966) in der hier ermittelten humangenomischen Sequenz ein Guanosin befindet. Da sich dieses Nukleotid jedoch an der dritten Position („wobble"-Position) des entsprechenden Codons befindet, ergibt sich keine Änderung der publizierten Aminosäuresequenz des WEE1-Proteins. Es handelt sich somit vermutlich um einen Polymorphismus.

Wee1

Das murine *Wee1*-Gen, das 1995 von *Honda et al.* beschrieben wurde (Acc.-Nr. D30743), liegt ebenso wie das humane *WEE1*-Gen vollständig in der sequenzierten Region. Das Gen, dessen mRNA eine Größe von 2270 bp aufweist, besteht aus ebenfalls 11 Exons und erstreckt sich über einen genomischen Bereich von ca. 20 kb.

Einen Vergleich der genomischen Organisation des humanen und murinen *WEE1/Wee1*-Gens sowie eine Gegenüberstellung der jeweiligen Exongrößen zeigt Abbildung 3.6 und Tabelle 3.3.

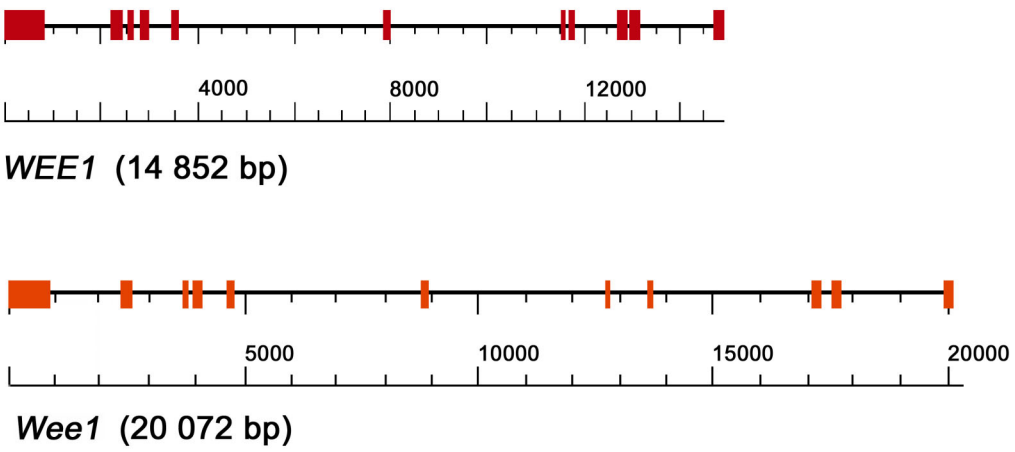


Abb. 3.6: Gegenüberstellung der genomischen Ausdehnung des humanen und murinen *WEE1*- bzw. *Wee1*-Gens. Das humane *WEE1*-Gen, das eine mRNA-Länge von 2194 bp (Acc.-Nr. U10564) aufweist, erstreckt sich über einen genomischen Bereich von 14 853 bp. Die Exons des murinen *Wee1*-Gens, das mit einer mRNA-Länge von 2270 bp (Acc.-Nr. NM009516) größer als das homologe humane Gen ist, sind in einer genomische Region von 20 072 bp lokalisiert.

Tab. 3.3: Gegenüberstellung der Exon- bzw. Introngrößen des humanen und murinen *WEE1*- bzw. *Wee1*-Gens. Die Positionen der Exons innerhalb der publizierten mRNAs sind aufgeführt. Das Exon, in dem das Start-Codon lokalisiert ist, wurde grün markiert. Die Größe des letzten Exons wurde bis zum Stop-Codon (rot) und bis zum polyA- bzw. putativen polyA-Signal (?) ermittelt. Als putative polyA-Signale wurden polyA-Konsensus-Sequenzen bezeichnet, die aufgrund fehlender EST-Daten nicht über EST-Cluster verifiziert werden konnten.

Exon	Intron	Größe (bp) Mensch	HSU10564 (bp)	Größe (bp) Maus	D30743 (bp)
1		829	1 – 829	863	6 – 868
	1	1377		1530	
2		206	830 – 1035	206	869 – 1074
	2	136		1145	
3		64	1036 – 1099	61	1075 – 1135
	3	193		128	
4		173	1100 – 1272	173	1136 – 1308
	4	482		532	
5		122	1273 – 1394	122	1309 – 1430
	5	4264		4007	
6		147	1395 – 1541	147	1431 – 1577
	6	3516		3780	
7		96	1542 – 1637	96	1578 – 1673
	7	88		787	
8		86	1638 – 1723	86	1674 – 1759
	8	921		3425	
9		171	1724 – 1894	171	1760 – 1930
	9	91		242	
10		146	1895 – 2040	146	1931 – 2076
	10	1591		2248	
11		154 452 ?	2041 - 2194	157 755	2077 - 2253

Da die Exongrößen des *WEE1*-/*Wee1*-Gens in Mensch und Maus konserviert sind (Ausnahme: Exon 3 ist beim Menschen um 1 Codon verlängert), resultiert die größere genomische Ausdehnung des murinen *Wee1*-Gens aus den im Vergleich zum Menschen längeren Intronbereichen (Ausnahme: Intron 3 und 5). Besonders in der zweiten Hälfte des Transkripts (ab einschließlich Intron 7) sind die Intronlängen in der Maus deutlich vergrößert.

Ein Vergleich der aus den Sequenzdaten des PAC-Klons 256N10 erstellten murinen *Wee1*-Sequenz mit der von *Honda et al.* (1995) veröffentlichten mRNA-Sequenz weist im Protein-kodierenden Bereich an sieben Positionen Unterschiede auf. Zwei Sequenzunterschiede führen dabei nicht zu einer Änderung der Aminosäure-Sequenz. An fünf Positionen befinden sich allerdings Nukleotidunterschiede, die bei der Translation für eine andere Aminosäure kodieren und somit zu einem veränderten Protein führen. Das partielle Alignment der veröffentlichten (Acc.-Nr. D30743) und der aus der vorliegenden genomischen Maus-Sequenz (Acc.-Nr. AJ278435; revers-komplementär) abgeleiteten Nukleotid- bzw. Aminosäuresequenzen ist im direkten Vergleich dargestellt. Dabei sind veränderte Codons, die nicht zu einem Aminosäureaustausch führen, grün und solche, die für eine andere Aminosäure kodieren, rot hervorgehoben.

```
D30743            500   gccgaggcgcagcgccgccgtcgctcgcccggcgcggagccc      541
                          A  E  A  Q  R  R  R  R  S  P  G  A  E  P

                          A  E  A  E  R  R  R  R  S  P  G  A  E  P
AJ278435rk     84145  gccgaggccgagcgccgccgtcgctcgcccggcgcggagcc     84105


D30743            998   aaccccttactccggatcctgtgctgctc      1027
                          N  P  F  T  P  D  P  V  L  L

                          N  P  F  T  P  D  P  V  L  L
AJ278435rk     82088  aaccccttactccggatcctgtactgctc      82117


D30743            1634  gaatatgactggatatccaataaagttatgttta  1667
                          E  Y  D  W  I  S  N  K  V  M  F

                          E  D  D  W  I  S  N  K  V  M  F
AJ278435rk     71891  gaagatgactggatatccaataaagttatgttta  71858


D30743            1682  gggcatgatacaagaatctctagtcctcaacttgaa      1717
                          G  H  D  T  R  I  S  S  P  Q  L  E

                          G  H  V  T  R  I  S  S  P  Q  V  E
AJ278435rk     71056  gggcatgtaacaagaatctctagtcctcaagttgaa      71021


D30743            1841  agaaatggagagcactggcac   1861
                          R  N  G  E  H  W  H

                          R  N  G  E  Q  W  H
AJ278435rk     67472  agaaatggagagcagtggcac   67446
```

Die Sequenzen wurden einem Homologievergleich mit Daten aus der Maus-EST-Datenbank und dem Maus-UniGene-Cluster unterzogen. Dabei zeigte sich, dass mehrere ESTs vorlagen, welche die hier gefundenen Basenaustausche ebenfalls enthielten. Es handelt sich hierbei somit vermutlich um Polymorphismen.

3.3.1.2 *ZNF143*/*mStaf*

Sowohl das humane Zinkfingergen *ZNF143* (*Tommerup & Vissing*, 1995) als auch das homologe murine *mStaf*-Gen (*Adachi et al*., 1998) konnten in der untersuchten humanen bzw. murinen Sequenz lokalisiert werden. Das *ZNF143*-Gen kodiert für einen Transkriptionsfaktor, der Mitglied der „Krüppel"-Familie ist.

ZNF143

Das humane Zinkfingergen *ZNF143* wurde von *Tommerup & Vissing* (1995) über FISH-Analyse in die Chromosomenregion 11p15.4 lokalisiert. Die publizierte mRNA-Größe beträgt 3908 bp (Acc.-Nr. U09850). Der gesamte Protein-kodierende Bereich des Zinkfingergens *ZNF143* konnte in der sequenzierten Region identifiziert werden. Der Bereich von nt 7 bis nt 2291 der veröffentlichten mRNA-Sequenz von *ZNF143* ist in der vorliegenden genomischen Region enthalten. Die Nukleotide 1 – 7 der von *Tommerup & Vissing* (1995) veröffentlichten mRNA-Sequenz konnten weder über die vorliegenden Sequenzdaten noch über EST-Daten bestätigt werden und scheinen somit Sequenzierungsfehler oder Klonierungsartefakte darzustellen.

Der sich in der publizierten cDNA-Sequenz (U09850) von Position nt 2287 bis Position nt 3908 erstreckende 3´-untranslatierte Bereich findet sich in der hier untersuchten genomischen Region nicht. Ein Homologievergleich dieser Sequenz mit der humanen EST-Datenbank zeigte Ähnlichkeiten zu zahlreichen Klonen. Dabei war auffällig, dass die Klone entweder zu dem Bereich vor der Position nt 2280 oder nach der Position nt 2309 der publizierten mRNA U09850 Homologien aufwiesen und somit kein Klon diesen Bereich überspannte. Blast-Analysen gegen die humangenomische Datenbank des NCBI (siehe Tab. 2.4) zeigten ab dem Bereich nt 2309 (U09850) 99,7%-ige Sequenzidentität mit dem Klon CTD-2317F5 aus der chromosomalen Region 14q14.3, der auch in dem Contig Hs14_10297 (NT_010140.2) angeordnet werden konnte. Der Bereich vor nt 2287 (U09850) zeigte 100%-Sequenzübereinstimmung mit dem Contig Hs11_24352 (NT_024206), der auf dem kurzen Arm von Chromosom 11 kartiert werden kann. Da sich nach Position nt 2286 der publizierten *ZNF143*-mRNA (U09850) eine polyA-Abfolge von insgesamt 17 aufeinanderfolgenden Adenosinen befindet, scheint es sich hier um das tatsächliche Ende der mRNA handeln. Unterstützt wird diese Vermutung dadurch, dass sich in der mRNA-Sequenz von nt 2265 bis nt 2270 (entspricht nt 114 587 bis nt 114 582 der vorliegenden humangenomischen

Sequenz) das Polyadenylierungssignal AATAAA befindet (siehe Abb. 3.7). Bei der publizierten Sequenz könnte es sich somit um eine künstlich zusammengesetzte Sequenz, also um ein Klonierungsartefakt, handeln.

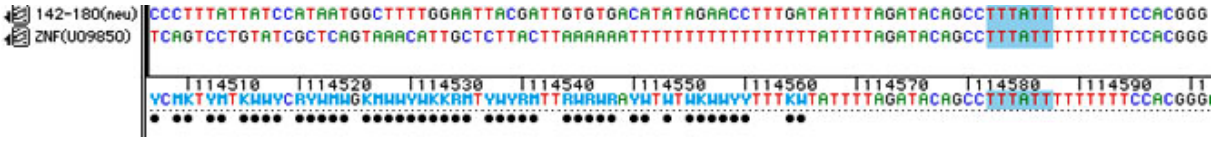


Abb. 3.7: Sequenzvergleich der ermittelten humangenomischen Sequenz mit der revers komplementär dargestellten *ZNF143*-mRNA-Sequenz (U09850). Das putative Polyadenylierungssignal in beiden Sequenzen sowie der Konsensus-Sequenz ist grau unterlegt. Die publizierte mRNA-Sequenz weist 16 bp nach dem polyA-Signal einen polyA-Bereich (aufgrund der revers komplementären Darstellungsweise hier ein polyT-Bereich) auf, der wahrscheinlich das echte Transkriptende darstellt. Im Anschluss an diesen putativen polyA-Bereich finden sich zwischen der *ZNF143*-mRNA-Sequenz und der genomischen Sequenz keine Übereinstimmungen mehr.

Da *Adachi et al.* (2000) bei dem zum Zinkfingergen *ZNF143* homologen Maus-Gen *mStaf* bei der Untersuchung dessen 5´-Bereiches ein weiteres Exon identifizieren konnten (s. u.), das einen Teil des 5´-untranslatierten Bereiches enthält und eine Größe von 40 bp aufweist, wurde auch die 5´-Region des *ZNF143*-Gens auf das Vorhandensein eines weiteren, bisher noch nicht beschriebenen Exons überprüft. Hierzu wurde über eine komparative Sequenzanalyse zwischen Mensch und Maus mittels Dotplot (siehe 3.4.1) auf eine Sequenzkonservierung im 5´-Bereich des *ZNF143*-Gens geachtet. Der interessierende Bereich des sequenzierten Maus-Klons PAC-256N10 (nt 144 000 - nt 146 000) zeigte in vier Regionen starke Ähnlichkeit mit der humanen Sequenz. Die Ergebnisse der blastn-Untersuchungen dieser Bereiche sind in Tab. 3.4 zusammengefasst. Es zeigte sich, dass zwei der vier identifizierten konservierten Bereiche (Region 3 und 4) starke Ähnlichkeit (bis zu 96%) mit der publizierten 5´-Sequenz des murinen *mStaf*-Gens aufweisen.

Tab. 3.4: Konservierte Regionen zwischen der humanen und der murinen genomischen Sequenz aus der 5´-Region von *ZNF143* und *mStaf*. Neben der prozentualen Übereinstimmung („Ähnlichkeit“) zwischen den interessierenden Bereichen sind die ermittelten Homologien der humanen Sequenz nach blastn-Analyse aufgeführt.
AF145372: „Mus musculus selenocysteine tRNA gene transcription activating factor (Staf) gene, 5' flanking region and partial mRNA“; *Adachi et al*., 2000;
X84996: „X.laevis mRNA for selenocysteine tRNA acting factor (Staf)“; *Schuster et al*., 1995.

konserv. Region	Bereich der Sequenzkonservierung (Maus)	Bereich der Sequenzkonservierung (Mensch)	Ähnlichkeit zwischen Mensch und Maus	blastn-Ergebnis des humanen Sequenzbereiches
1	144 149 – 144 235	179 510 – 179 596	65 %	keine signifikanten Homologien
2	144 685 – 144 780	180 157 – 180 252	68,8 %	keine signifikanten Homologien
3	144 771 – 145 252	180 245 – 180 726	79,9%	134/146 bp (91%): AF145372: nt 1966 – nt1821 41/48 bp (85%): AF145372: nt 1710 – nt 1663 48/50 bp (96%): X84996: nt 115 – nt 66
4	145 232 – 145 470	180 707 – 180 945	74,9 %	99/111 bp (89%): AF145372, nt 1609 – nt 1499

Nachdem im Menschen zwei konservierte Regionen (3 und 4) identifiziert wurden, die Homologien zu dem 5´-Bereich von *mStaf* besaßen, sollte nachgewiesen werden, ob sich in diesen konservierten Bereichen ein weiteres, bisher unbekanntes Exon von *ZNF143* befindet. Mit Hilfe von Datenbanksuchen konnten fünf EST-Klone ermittelt werden, welche die mRNA von *ZNF143* um 77 bp in den 5´-Bereich hinein verlängern. Diese Klone stammen aus einem EST-Projekt, das die Aufklärung von Volllängen-cDNA-Sequenzen zum Ziel hat (*Ota et al.,* 2000; unveröffentlicht; *Gu et al*., 2000, unveröffentlicht). Einen Überblick über die relative Lage der identifizierten EST-Klone zur bekannten *ZNF143*-mRNA-Sequenz gibt die Abbildung 3.8.

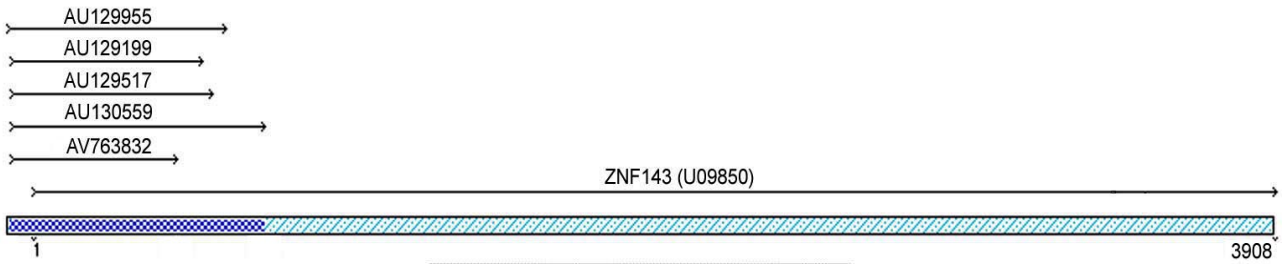


Abb. 3.8: Überblick über EST-Klone, welche die publizierte mRNA-Sequenz von *ZNF143* (U09850) im 5´-Bereich verlängern.

Die hierdurch neu gewonnene Sequenzinformation konnte ebenfalls in der hier erstellten humangenomischen Nukleotidsequenz lokalisiert werden (nt 180 453 – nt 180 389 und nt 170171 – nt 170 160; entspricht nt 11 – nt 75 bzw. nt 76 - nt 87 von AU129955). Sie weist eine 100%-ige Sequenzübereinstimmung auf und konnte der konservierten Region 3 (siehe Tab. 3.4) zugeordnet werden. Die Konservierung im Bereich 4 beruht auf Sequenzähnlichkeiten mit dem 5´-untranslatierten Bereich von *mStaf*. Da die Homologie zwischen der über EST-Klone verifizierten Basenabfolge und der genomischen Sequenz nicht kontinuierlich in die Nukleotidabfolge des bekannten ersten *ZNF143*-Exons übergeht und sich direkt im Anschluss an die Sequenzübereinstimmungen von EST- und genomischer Sequenz die Nukleotide Guanosin und Thymidin befinden, weist dies auf die Anwesenheit einer Spleiß-Donorstelle hin. Unterstützt wird diese Annahme dadurch, dass sich auch vor Beginn des Exons 2 (erstes Exon in U09850) ein Spleiß-Akzeptor, also Adenin gefolgt von Guanosin, befindet. Somit scheint das humane *ZNF143*-Gen ebenso wie das homologe murine *mStaf*-Gen ein weiteres, bisher noch nicht beschriebenes Exon stromaufwärts der bisher bekannten kodierenden Sequenz zu besitzen. Aussagen über die exakte Größe des ersten Exons können nicht gemacht werden, da bisher keine Informationen über den Transkriptionsstart vorliegen. Die *in silico* erweiterte, vermutlich vollständige cDNA-Sequenz der *ZNF143*-mRNA ist auf der beigefügten CD-ROM enthalten.
Ebenso wie sein murines homologes Gen *mStaf* setzt sich somit das humane Zinkfingergen *ZNF143* aus 16 Exons zusammen und erstreckt sich über einen genomischen Bereich von knapp 66 kb. Die genomische Organisation ist sowohl in der Abb. 3.9 als auch detailliert in der Tab. 3.5 dargestellt.

mStaf

Das murine Gen *mStaf* (mouse selenocysteine tRNA gene transcription-activating factor) wurde 1998 von *Adachi et al.* kloniert und charakterisiert und mit einer cDNA-Größe von 2962 bp beschrieben (Acc.-Nr. AF011758). Weiterführende Arbeiten konnten die genomische Organisation und die genaue chromosomale Lage des Gens *mStaf* in der Region E3-F1 auf Chromosom 7 der Maus aufklären (*Adachi et al,* 2000; Acc.-Nr. 145372). Das *mStaf*-Gen befindet sich komplett in der sequenzierten genomischen Region der Maus und besteht aus 16 Exons, die sich über einen Bereich von knapp 35 kb erstrecken. Einen Überblick über die genomische Ausdehnung und somit die relativen Exon-/Introngrößen des murinen *mStaf*-Gens im Vergleich zum orthologen humanen *ZNF143*-Gen gibt die Abb. 3.9 sowie die Tabelle 3.5.

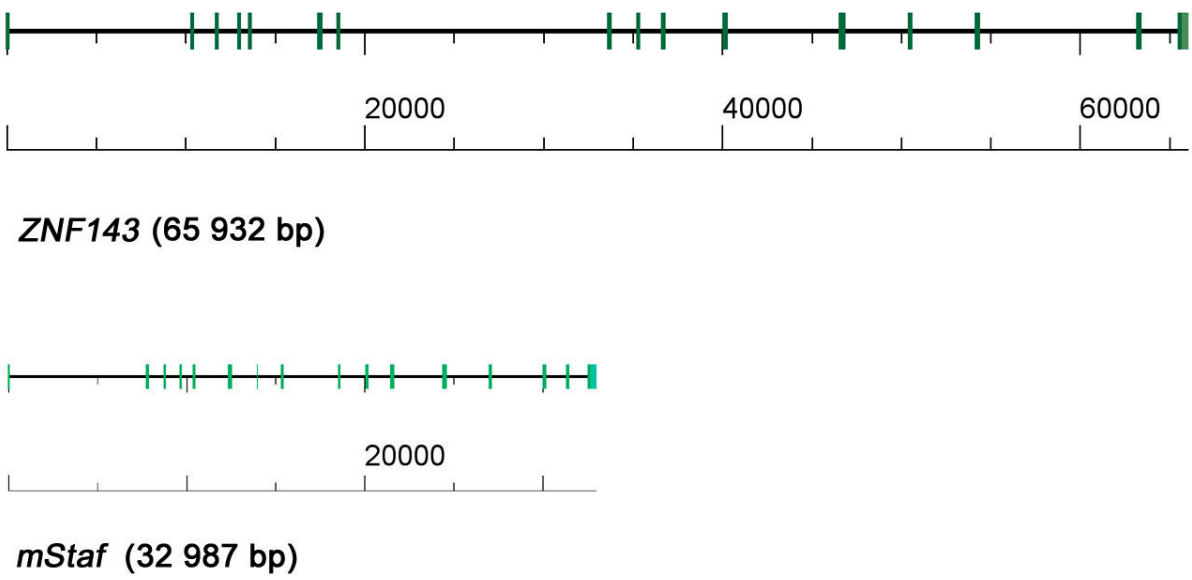


Abb. 3.9: Gegenüberstellung der genomischen Ausdehnung der humanen und murinen orthologen Gene *ZNF143* bzw. *mStaf*. Das humane *ZNF143*-Gen, dessen publizierte mRNA eine Länge von 3908 bp (Acc.-Nr. U09850) aufweist, erstreckt sich über einen genomischen Bereich von knapp 66 kb. Die 16 Exons des murinen *mStaf*-Gens, das mit einer mRNA-Länge von mindestens 2962 bp (Acc.-Nr. AF011758) kleiner ist als die humane mRNA, sind innerhalb einer genomischen Region von nur knapp 35 kb lokalisiert.

Die deutlich größere genomische Ausdehnung des humanen *ZNF143*-Gens im Vergleich zum orthologen murinen *mStaf*-Gen beruht vor allem auf den größeren Intronbereichen (Ausnahmen sind die Introns 6, 8 und 9). Besonders auffällig ist dies beim Intron 7, das beim Menschen eine Größe von über 15 kb aufweist, während es bei der Maus nur 1341 bp groß ist. Die Exongrößen dagegen sind bis auf die den Transkriptionsstart und –stop enthaltenden Exons 2 und 16 konserviert (siehe Tab. 3.5).

Tab. 3.5: Gegenüberstellung der Exon- bzw. Introngrößen des humanen *ZNF143*- und murinen *mStaf*-Gens. Die Lage der jeweiligen Exons innerhalb der publizierten mRNAs sind aufgeführt. Das Exon, in dem das Start-Codon lokalisiert ist, ist grün markiert. Die Größe des letzten Exons wurde bis zum Stop-Codon (rot) und bis zum polyA-Signal ermittelt. Die Größe des ersten humanen Exons konnte aufgrund fehlender Informationen über den Transkriptionsstart experimentell nicht ermittelt werden und ist daher mit einem Fragezeichen versehen.

Exon	Intron	Größe (bp) Mensch	U09850 (bp)	Größe (bp) Maus	AF011758 (bp) AF145372
1		66 ?		40	1930 - 1969
	1	10219		7691	
2		119	7 - 113	130	10 - 139
	2	1255		875	
3		93	114 - 206	93	140 - 232
	3	1173		801	
4		84	207 - 290	84	233 - 316
	4	525		663	
5		84	291 - 374	84	317 - 400
	5	3752		1865	
6		197	375 - 571	197	401 - 597
	6	884		1376	
7		75	572 - 646	75	598 - 672
	7	15047		1341	
8		120	647 - 766	120	673 - 792
	8	1537		3091	
9		76	767 - 842	76	793 - 868
	9	1296		1448	
10		126	843 - 968	126	869 - 994
	10	3289		1273	
11		180	969 - 1148	180	995 - 1174
	11	6330		2742	
12		228	1149 - 1376	228	1175 - 1402
	12	3600		2389	
13		149	1377 - 1525	149	1403 - 1551
	13	3604		2888	
14		162	1526 - 1687	162	1552 - 1713
	14	8878		1143	
15		147	1688 - 1834	147	1714 - 1860
	15	2144		1074	
16		84 493	1835-2291	84 436	1861-2106

Der Vergleich der unter der „accession“-Nummer AF011758 publizierten mRNA-Sequenz des *mStaf*-Gens mit den vorliegenden Sequenzdaten weist an drei Positionen im 5´-UTR-Bereich (am Anfang des Exons 2) Unterschiede in der Nukleotidsequenz auf.

```
mStaf (AF011758); nt 1 - 75

          gttggaaaattttctcggaggtagaaaatgttgttagcacaaataaatcgggattctcaggaatgacagaattt
                  ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                  ttttcttcaaggtagaaaatgttgttagcacaaataaatcgggattctcaggaatgacagaattt
                                                                      M  T  E  F

AJ278435; nt 137225 – 137160 (rk)
```

Es konnten jedoch keine Maus-EST-Klone identifiziert werden, die diesen Bereich der *mStaf*-mRNA abdecken, so dass die Sequenzunterschiede nicht verifiziert werden konnte.
Die Enden der publizierten mRNA-Sequenz des *mStaf*-Gens wurden ursprünglich von *Adachi et al*. (1998) über 5´- und 3´-RACE-Experimente ermittelt. Da jedoch keine polyA-Abfolge in der Sequenz vorhanden ist, wurde im Rahmen der vorliegenden Arbeit nach Maus-EST-Klonen gesucht, welche die bekannte mRNA-Sequenz in den 3´-Bereich hinein verlängern.
Es konnten so mehrere Maus-EST-Klone (AV234351, AW109739, BB215632, BB225496, BB237303, BB344361) isoliert werden, welche die *mStaf*-mRNA um 167 bp in den 3´-UTR-Bereich hinein erweitern (siehe Abb. 3.10). Sie weisen jedoch weder einen polyA-Schwanz, noch eine potentielle Polyadenylierungsstelle auf (trotz eines leicht erhöhten AT-Gehaltes von 65,9%). Die über EST-Sequenzen ermittelte Sequenz konnte innerhalb der vorliegenden murinen genomischen Sequenz in dem Bereich von nt 111 469 bis nt 111 303 identifiziert werden.

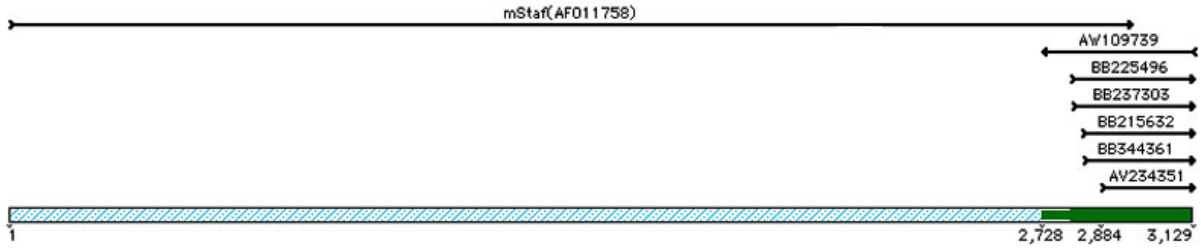


Abb. 3.10: Verlängerung der publizierten *mStaf*-mRNA (Acc.-Nr. AF011758) durch murine EST-Klone. Die bekannte mRNA des *mStaf*-Gens konnte um 167 bp in den 3´-UTR hinein verlängert werden. Es konnte allerdings kein Klon isoliert werden, der den polyA-Bereich enthält.

Vermutlich befindet sich das mRNA-Ende noch weiter stromabwärts und wird momentan nicht von EST-Klonen abgedeckt. Dafür spricht die Sequenzkonservierung im Bereich nt 112 602 bis nt 111 236 in der analysierten murinen Sequenz, die 66%-ige Identität zur humanen Sequenz aus dem Bereich nt 115 209 bis nt 113 949 aufweist und welcher das Exon 16 mit dem putativen Polyadenylierungssignal (nt 114 587 – 114 582) enthält. Unterstützt wird diese Vermutung zudem dadurch, dass sich das am ehesten in räumlicher Nähe lokalisierte putative *mStaf*-Polyadenylierungssignal von nt 111 920 – nt 111 915 befindet.

3.3.1.3 *RanBP7*/*„mRanBP7“*

Ein weiteres bekanntes Gen, das innerhalb der untersuchten humangenomischen Sequenz lokalisiert werden konnte, ist das humane *RanBP7*-Gen. Das Genprodukt RanBP7 (Ran-binding protein 7) bindet an die GTPase Ran, die eine zentrale Rolle beim Transport von Proteinen und RNAs zwischen Nukleus und Zytoplasma spielt (*Görlich et al*., 1997, 1998).

RanBP7

Die von *Jäkel & Görlich* (1998) publizierte mRNA-Sequenz des humanen *RanBP7*-Gens (Acc.-Nr. AF098799) besitzt eine Größe von 3557 bp und liegt mit der Protein-kodierenden Sequenz vollständig in dem im Rahmen der vorliegenden Dissertation analysierten Bereich aus der humanen Chromosomenregion 11p15.3.

Das *RanBP7*-Gen besteht aus 25 Exons, von denen sich 24 in der untersuchten Region wiederfinden. Diese 24 Exons erstrecken sich über einen genomischen Bereich von mehr als 51 kb (siehe Abb. 3.14), wobei das Exon 1 mittels PCR in den Anschlussklonen PAC-151D7 und PAC-79B22 (siehe Abb. 3.2) nachgewiesen werden konnte. Eine genaue Lokalisation dieses Exons innerhalb beider überlappender PAC-Klone war aufgrund fehlender interner Sequenzinformation nicht möglich. Zur Abschätzung der Introngröße zwischen Exon 1 und 2 wurde die Länge der PAC-Sequenz bis zum Exon 2 ermittelt und die aus der Randsequenzierung des Anschlussklones PAC-79B22 gewonnene Sequenzinformation bis zum Ende von Exon 1 des *RanBP7*-Gens addiert. Da nicht ermittelt werden konnte, wie groß die Überlappung der PAC-Klone ist, konnte die minimale Introngröße zwischen Exon 1 und 2 des humanen *RanBP7*-Gens mit mindestens 6381 bp angegeben werden. Mit Hilfe von Expand-PCR-Analysen wurde die Introngröße zwischen Exon 1 und Exon 2 auf ca. 23 kb geschätzt (Daten nicht gezeigt). Die aus den vorliegenden genomischen Sequenzdaten abgeleitete *RanBP7*-mRNA-Sequenz ab Exon 2 wurde mit der publizierten mRNA-Sequenz (Acc.-Nr. AF098799) verglichen und zeigte 100% Übereinstimmung.

Datenbanksuchen des 3´-Bereiches der publizierten *RanBP7*-mRNA-Sequenz mit der nicht-redundanden Datenbank des NCBI ergaben starke Homolgien zu dem cDNA-Klon DKFZ564C2163 (Acc.-Nr. AL117596). Dieser Klon wurde im Rahmen der vorliegenden Arbeit vollständig sequenziert und überlappt mit der *RanBP7*-mRNA (nt 3476–nt 3557: AL117596) mit 100%-iger Sequenzübereinstimmung. Beim Vergleich der aus der vorliegenden humangenomischen Sequenz (nt 196025 – nt 193627) abgeleiteten, kodierenden *RanBP*-Sequenz mit der publizierten mRNA-Sequenz des Gens zeigten sich drei Nukleotid-Unterschiede. Somit erweiterte der Klon DKFZ564C2163 die bisher bekannte mRNA-Sequenz um 2371 bp in den 3´-UTR-Bereich hinein. Am 3´-Ende der

DKFZ564C2163-Sequenz befindet sich 21 bp vor dem polyA-Ende eine konservierte Polyadenylierungssequenz (AATAAA; von nt 193648 – nt 193643). Somit repräsentiert dieser cDNA-Klon den 3´-untranslatierten Bereich des humanen *RanBP7*-Gens und stellt keine genomische Kontamination dar. Eine Bestätigung der auf diese Weise ermittelten mRNA-Größe des *RanBP7*-Transkripts erfolgte über Northern-Blot (siehe 2.10.3, Daten nicht gezeigt). Die Lage des DKFZ-Klons 564C2163 in Relation zur bekannten mRNA-Sequenz von *RanBP7* ist in Abb. 3.11 dargestellt. Die Nukleotidsequenz des verlängerten *RanBP7*-Gens ist aus Platzgründen nicht im Folgenden abgedruckt, sondern auf der beigefügten CD-Rom aufgeführt.

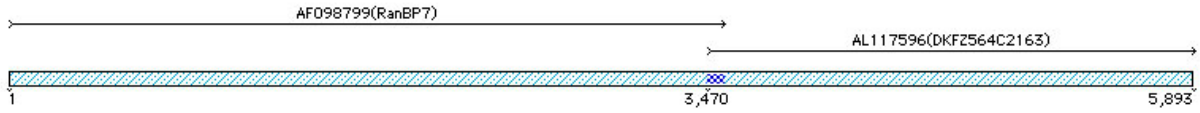


Abb. 3.11: Darstellung der durch den DKFZ-Klon 564C2163 verlängerten mRNA-Sequenz des humanen *RanBP7*-Gens.

mRanBP7

Erste blastn-Analysen des murinen Klons PAC-256N10 zeigten starke Homologien zu dem humanen *RanBP7*-Gen. Da Orthologe des humanen *RanBP7*-Gens in ihrer vollständigen kodierenden Länge bisher nur beim afrikanischen Krallenfrosch *Xenopus laevis* und der Fruchtfliege *Drosophila melanogaster* bekannt waren, sollte die komplette mRNA-Sequenz des murinen *mRanBP7*-Gens beschrieben werden.

mRNA-Sequenz des murinen *mRanBP7*-Gens

Da sich der murine Klon PAC-256N10 zum Zeitpunkt der ersten Untersuchungen in der Sequenzierungsphase befand und deshalb keine durchgängige Sequenzinformation vorlag, wurde versucht, die murine cDNA durch RT-PCR-Analysen mit RNA aus verschiedenen fetalen und adulten murinen Geweben zu generieren. Dabei wurde das humane *RanBP7*-Gen und die vorliegenden murinen Sequenzdaten mit starken Homologien zum humanen *RanBP7*-Gen zur Primergenerierung benutzt. Die Primer für die Amplifikation eines cDNA-Abschnittes wurden dabei so gewählt, dass bei entsprechend optimierten PCR-Bedingungen kein Produkt auf genomischer DNA gebildet werden konnte. Es wurde ein Maus-spezifischer Primer aus dem 5´-Bereich mit einem Primer aus dem mittleren Bereich bzw. aus dem mittleren und dem 3´-Bereich der humanen cDNA für eine Amplifikation eingesetzt (siehe Abb. 3.12) und die PCR-Produkte anschließend direkt sequenziert. Bei den Untersuchungen zeigte sich, dass *mRanBP7* in Gehirn und Milz von fetalen und adulten Mäusen exprimiert

wird. Die sequenzierten RT-PCR-Produkte sind in ihrer relativen Lage zur humanen *RanBP7*-mRNA-Sequenz in der Abb. 3.12 dargestellt.

Da Versuche, den mittleren Teil der *mRanBP7*-RNA zu amplifizieren, fehlschlugen, wurden murine EST-Datenbanken nach Klonen mit Homologien zu diesem Teil der humanen *RanBP7*-RNA durchsucht. Die ermittelten EST-Sequenzen konnten zueinander angeordnet werden und lieferten auf diese Weise die fehlende Sequenzinformation. In der Abb. 3.12 sind die Maus-EST-Klone abgebildet, die zusammen mit den erhaltenen RT-PCR-Produkten eine möglicherweise vollständige *mRanBP7*-RNA bilden.

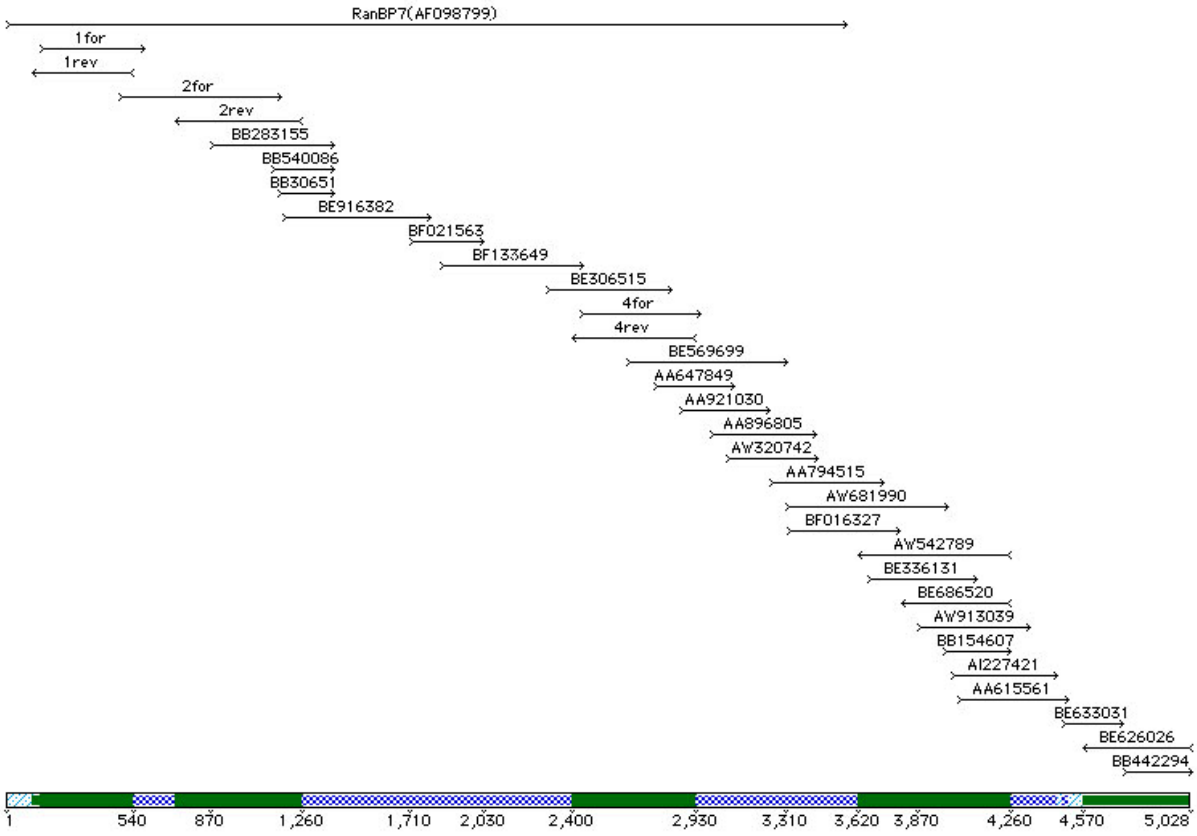


Abb. 3.12: Übersicht über die Lage der generierten RT-PCR-Produkte und murinen EST-Klone zur Aufklärung der mRNA-Sequenz des murinen *mRanBP7*-Gens, dargestellt im Verhältnis zur humanen *RanBP7*-mRNA-Sequenz (oben). Die Anordnung der murinen cDNA-Fragmente gegen die humane Sequenz erfolgte bei einer Überlappung von mindestens 18 bp, die zu 84% identisch sein sollten. Die Fragmente aus den RT-PCR-Analysen wurden mit folgenden Primern generiert und sequenziert: 1: mRan5´for und mRanBP4; 2: mRanmitterev und mRan3; 4: mRan3´rev und mRanBP1 (siehe auch 2.13.8). Die Bezeichnungen der murinen EST-Klone entsprechen den GenBank-Accession-Nummern.

Innerhalb der konstruierten cDNA-Sequenz des murinen *mRanBP7*-Gens ist ein Polyadenylierungssignal (AATAAA; nt 149 836-149 831) lokalisiert. Da die Homologien zu Maus-EST-Klonen 29 bp nach dem PolyA-Signal (bis Position nt 149 808) abbrechen, kann

davon ausgegangen werden, dass das komplette 3´-Ende bestimmt wurde. Die Länge der neu beschriebenen murinen *mRanBP7*-RNA entspricht mit 5028 bp somit ungefähr der Länge des humanen *RanBP7*-Transkriptes, das eine Länge von 5893 bp aufweist.

Bestimmung der Transkriptgröße von *mRanBP7* über Northern-Analyse

Um die ermittelte mRNA-Sequenz-Länge des murinen *mRanBP7*-Gens zu überprüfen, wurde eine Northern Blot-Analyse (siehe 2.10.3) durchgeführt. Hierzu wurde eine Sonde mittels RT-PCR auf Maus-Nieren-mRNA generiert, die aus dem vom EST-Klon BE916382 stammenden Sequenzabschnitt stammt und die Sequenz der Exons 12 bis 14 enthält. Diese Sonde wurde radioaktiv markiert und mit einem kommerziell erhältlichen murine „Multiple Tissue-Northern-Blot“ (CLONTECH) hybridisiert. Es zeigten sich bei allen aufgetragenen RNAs deutliche Signale bei einer Transkriptgröße von ca. 5 kb, was die aus RT-Analysen und mit Hilfe von EST-Klonen bestimmte RNA-Länge bestätigt (siehe Abb. 3.13).

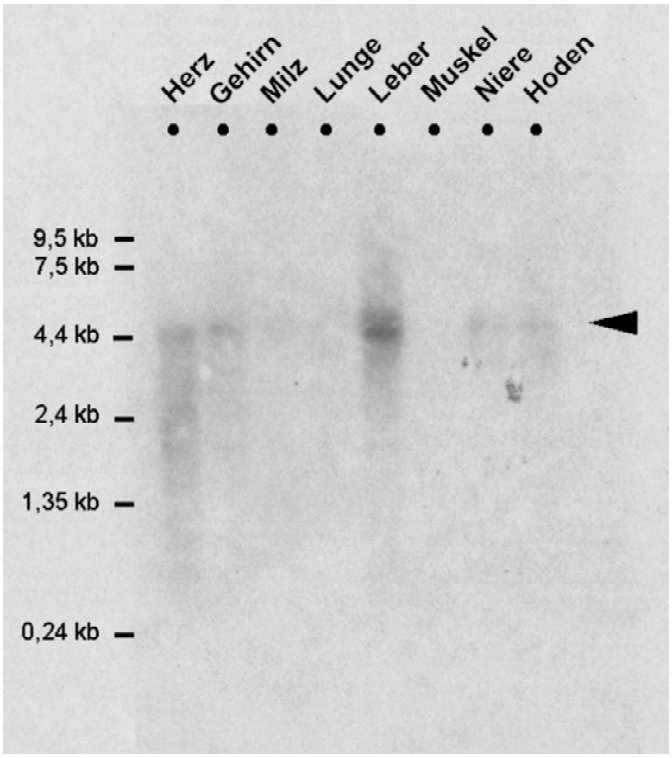


Abb. 3.13: Northern Blot-Analyse zur Bestimmung der Transkriptgröße des murinen *mRanBP7*-Gens. Ein muriner „Multiple Tissue Northern Blot“ (CLONTECH) mit polyA$^+$-RNAs aus den oben aufgeführten adulten Geweben wurde mit einer für das *mRanBP7*-Gen-spezifischen Sonde hybridisiert. Es zeigte sich in allen untersuchten Geweben außer im Muskel eine deutliche Bande bei ca. 5 kb (siehe Pfeilspitze). Dies korreliert mit der über RT-Analyse und mit Hilfe von EST-Klonen ermittelten Transkriptgröße. Die Positionen der Banden des RNA-Größenmarkers sind markiert.

Sowohl das bekannte humane als auch das hier erstmals beschriebene murine *mRanBP7*-Gen setzt sich aus 25 Exons zusammen. Die genomische Ausdehnung des humanen *RanBP7*-Gens ist mit über 45 kb (inkl. Exon und Intron 1 wahrscheinlich über 67 kb) deutlich größer als die Ausdehnung des murinen Gens. Grund dafür sind die beim Menschen in der Regel größeren Intronbereiche. Besonders das Intron 1, das auf eine Größe von ca. 23 kb geschätzt werden konnte, ist deutlich größer als im orthologen murinen Gen. Eine Übersicht über die genomische Organisation des humanen und murinen *RanBP7*-Gens ist in Abb. 3.14 dargestellt.

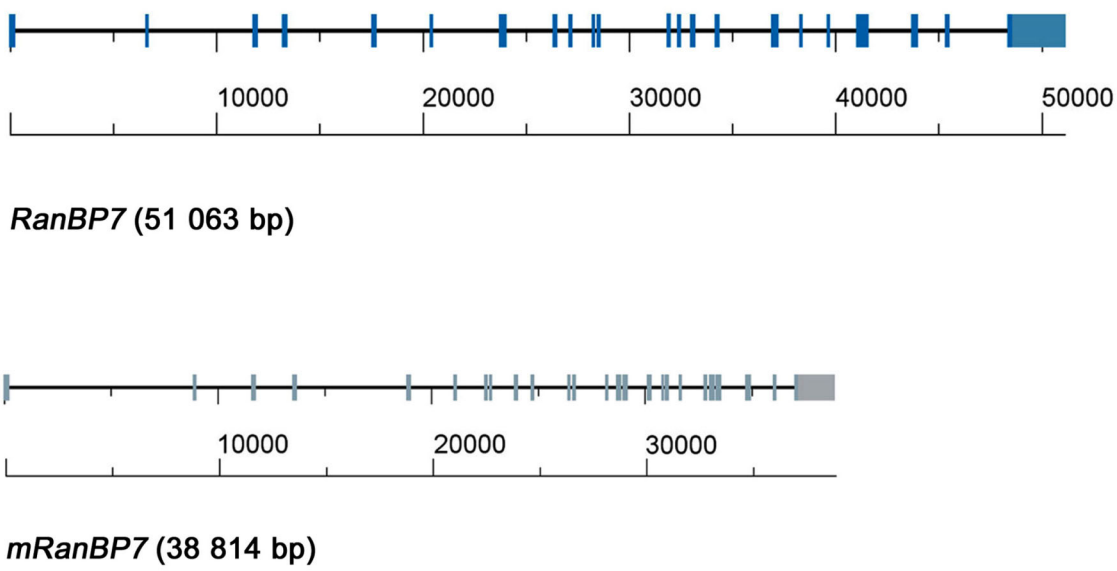


Abb. 3.14: Genomische Organisation der Protein-kodierenden Exons des humanen *RanBP7*-Gens (Acc.-Nr. AF098799) und des murinen *mRanBP7*-Gens. Das Exon 1 ist nicht in der untersuchten humangenomischen Sequenz enthalten, sondern befindet sich auf den Klonen PAC-151D7 und PAC-79B22 (vergl. Abb. 3.2), durch die der Anschluss an die restlichen, den Gesamt-Contig abdeckenden Klone erfolgt. Die Größe des Introns 1 konnte mit mindestens 6 kb angegeben und über Expand-PCR auf etwa 23 kb geschätzt werden.

Die einzelnen Größen der Exons und Introns des humanen *RanBP7*-Gens sind in der Tabelle 3.6 im Vergleich zum orthologen Maus *mRanBP7*-Gen detailliert aufgeführt.

Tab. 3.6: Gegenüberstellung der Exon- bzw. Introngrößen des humanen *RanBP7*- und des murinen orthologen *mRanBP7*-Gens. Die Größe des murinen Introns 1, das über die sequenzierte humangenomische Sequenz hinausgeht, konnte über Expand-PCR auf ca. 23 kb geschätzt werden. Das Exon, welches das Start-Codon enthält, ist grün markiert. Die Größe des letzten Exons wurde bis zum Stop-Codon (rot) und bis zum polyA- bzw. putativen polyA-Signal („?“) ermittelt. Als putative polyA-Signale wurden polyA-Konsensus-Sequenzen bezeichnet, die aufgrund fehlender EST-Daten nicht über EST-Cluster verifiziert werden konnten.

Exon	Intron	Größe (bp) Mensch	AF098799	Größe (bp) Maus	
1				≈189	1-191
	1	≥ 6381		8660	
2		86	192-277	86	192-277
	2	5113		2634	
3		154	278-427	154	278-427
	3	1287		1762	
4		159	428-586	159	428-586
	4	4168		5177	
5		157	587-743	157	587-743
	5	2652		2063	
6		90	744-833	90	744-833
	6	3263		1352	
7		95	834-928	95	834-928
	7	108		112	
8		85	929-1013	85	929-1013
	8	2307		1083	
9		135	1014-1148	135	1014-1148
	9	636		670	
10		100	1149-1248	100	1149-1248
	10	1037		1623	
11		77	1249-1325	77	1249-1325
	11	155		132	
12		117	1326-1442	117	1326-1442
	12	3278		1427	
13		90	1443-1532	90	1443-1532
	13	401		426	
14		166	1533-1698	166	1533-1698
	14	477		143	
15		161	1699-1859	161	1699-1859
	15	1040		981	
16		129	1860-1988	129	1860-1988
	16	2577		536	
17		67	1989-2055	67	1989-2055
	17	84		84	
18		126	2056-2181	126	2056-2181
	18	1085		525	
19		98	2182-2279	98	2182-2279
	19	1235		1086	
20		96	2280-2375	96	2280-2375
	20	1334		137	
21		221	2376-2596	221	2376-2596
	21	90		88	
22		206	2597-2802	206	2597-2802
	22	2173		1183	
23		207	2803-3009	207	2803-3009
	23	1419		1082	
24		117	3010-3126	117	3010
	24	2906		900	
25		98 2732	3127-3557	98 1830 ?	3127-3557

3.3.2 Identifizierung bisher unbekannter Gene

3.3.2.1 Putatives Pseudogen L23a

Innerhalb der untersuchten humangenomischen Sequenz (nt 39 022– nt 39 568) konnte ein Nukleotidbereich identifiziert werden, der über eine Länge von 549 bp eine 92%-ige Sequenzidentität zu dem humanen ribosomalen Protein L23a (*Fan et al*., 1996; 1997; Acc.-Nr. U37230) aufweist. Allerdings finden sich an 19 Positionen Unterschiede in der Nukleotidsequenz beider Sequenzen. Davon führen 8 Unterschiede innerhalb der kodierenden Region zu einem Austausch in der Aminosäuresequenz, wobei ein Nukleotidaustausch innerhalb des offenen Leserahmens zu einem vorzeitigen Stop bei der Translation führt. Es kann somit davon ausgegangen werden, dass es sich bei dem in dem humanen Klon PAC-142M6 lokalisierten Gen um ein Pseudogen handelt. Dies wird durch die typischen 13 bp langen direkten Sequenzwiederholungen unterstützt, die Pseudogene typischerweise flankieren (*Vanin et al*., 1985) und auch im vorliegenden Fall vorhanden sind. Die Sequenzwiederholungen erstrecken sich von Position nt 39 022 bis nt 39 010 und von Position 39 580 bis nt 39 568, wo sie direkt auf den polyA-Bereich folgen.

Die Abbildung 3.15 zeigt die Nukleotid- und Aminosäuresequenz des putativen, prozessierten Pseudogens L23a in der im Rahmen der vorliegenden Arbeit ermittelten humangenomischen Nukleotidsequenz.

In der sequenzierten Maus-Sequenz lässt sich keine dem putativen prozessierten Pseudogen L23a homologe Sequenz finden.

```
39680 gtaatagagagagaacccagatgtcaaggggttgaagaccctggcgccagtctacagggg

      V  I  E  R  E  P  R  C  Q  G  V  E  D  P  G  A  S  L  Q  G

39620 aatggagctttccctacactgctcagcactacgtgatgtcaagaagatggaaggcctttt
                                                          g
      N  G  A  F  P  T  L  L  S  T  T  -  C  Q  E  D  G  R  P  F

39560 cacaagatggcaccgaaagcgaagaaggaagctcctgcccctcctaaagctgaagccaaa
        a     g
      H  K  M  A  P  K  A  K  K  E  A  P  A  P  P  K  A  E  A  K

39500 gcgaaggctttaaaggccaagaaggcagtgttgaacggtgtccacagccacaaaaaaaag
                                        a                   ^g^ag
      A  K  A  L  K  A  K  K  A  V  L  N  G  V  H  S  H  K  K  K

39440 atccacacgtcacccaccttccggcggcccaagacactgcgactccggagacagcccgaa
                              g                             a
      I  H  T  S  P  T  F  R  R  P  K  T  L  R  L  R  R  Q  P  E

39380 tatcctcggaagagcactcccaggaaaaacaagcttgaccactatgctatcatcaagttt
            a  g        g                              t
      Y  P  R  K  S  T  P  R  K  N  K  L  D  H  Y  A  I  I  K  F

39320 ccgctgaccactgagtctgccatgaagaagatagaagacaacaacacacttgtgttcatt

      P  L  T  T  E  S  A  M  K  K  I  E  D  N  N  T  L  V  F  I

39260 gtggttgttaaagccaacaagcaccagattaaacaggctgtgaagtagctctatgacatt
         a                          gt          a    g
      V  V  V  K  A  N  K  H  Q  I  K  Q  A  V  K  -  L  Y  D  I

39200 gatgtggccaaggtcaacaccctgattcggcctgatggagagaagaaggcatatgttcga

      D  V  A  K  V  N  T  L  I  R  P  D  G  E  K  K  A  Y  V  R

39140 ctggctcctgattacgatgctttggatgttgtcaacaaaattgggatcatctaaactgag
                                    c
      L  A  P  D  Y  D  A  L  D  V  V  N  K  I  G  I  I  -  T  E

39080 tccagatgcctaattctaaatatatatatatatatatctttcaccataaaaaaaaaaaa
         c          g
      S  R  C  L  I  L  N  I  Y  I  Y  I  S  F  H  H  K  K  K  K

39020 gaagatggaagtttaaggggacacagatgtaatgagataaggaaggggagctgtaagaa

      E  D  G  S  L  R  G  H  R  C  N  E  I  R  K  G  E  L  -  E

38960 caggtgcctatgtttggataatcggattttgtcagagtattacttttaggt

      Q  V  P  M  F  G  -  S  D  F  V  R  V  L  L  L  G
```

Abb. 3.15: Darstellung der Nukleotid- und Aminosäuresequenz des putativen prozessierten Pseudogens L23a in der ermittelten humangenomischen Sequenz. Nukleotidunterschiede zum ribosomalen Protein L23a sind an entsprechender Position aufgeführt. Die Transversion an Position nt 39 215 führt zur Entstehung eines Stop-Codons und somit zum verfrühten Ende des offenen Leserahmens. Die 13 bp lange direkte Sequenzwiederholung beiderseits des Pseudogens ist fett und unterstrichen dargestellt und schließt im 3´-Bereich des Pseudogens an den vorhandenen polyA-Bereich an. ^ kennzeichnet die Stelle, wo im ribosomalen Protein L23a ein oder mehrere zusätzliche Nukleotide vorkommen. Grau dargestellte Sequenzanteile sind nicht mehr Bestandteil des putativen Pseudogens.

3.4 Komparativer Sequenzvergleich Mensch – Maus

Die komparative Analyse der im Rahmen der vorliegenden Dissertation generierten humanen und murinen Sequenzdaten erfolgte mit Hilfe der Programme DOTPLOT und PIP-Maker, die unterschiedliche Dotplots generieren.

3.4.1 Dotplot (MegAlign)

Die humane und murine Sequenz wurden unter Verwendung der Option DOTPLOT aus dem Programm MEGALIGN (Programmpaket DNA-STAR) miteinander verglichen. Bei dieser Analyse werden die zu untersuchenden Sequenzen auf einem 2-achsigen Koordinatensystem aufgetragen und vorher definierte Sequenzabschnitte (im vorliegenden Fall wurde ein zu vergleichender Basenbereich auf 50 Nukleotide festgesetzt) paarweise miteinander verglichen. Bei einer Sequenzkonservierung oberhalb eines vorher definierten Wertes (hier ab 65%) setzt das Programm die Position der Konservierung innerhalb beider Spezies graphisch um. Beim Vergleich zweier identischer Sequenzen entstünde somit eine durchgängige Diagonale (siehe 2.12.2 und 4.3.1). Die folgende Abbildung 3.16 zeigt das Resultat der Dotplot-Analyse der humanen und murinen genomischen Sequenz.

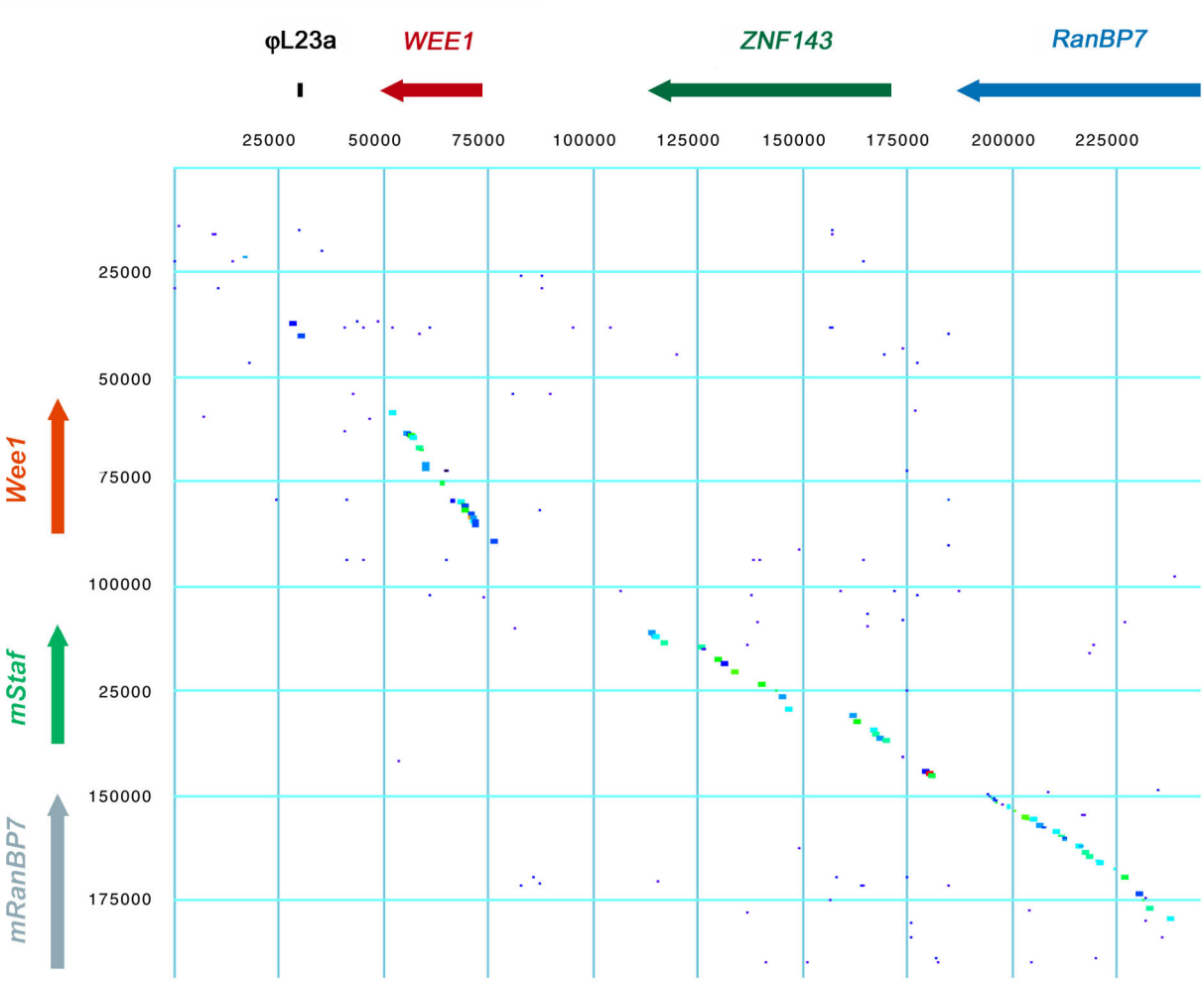


Abb. 3.16: Dotplot-Ergebnis der komparativen Sequenzanalyse der untersuchten genomischen Regionen um *WEE1*/*Wee1* in Mensch und Maus. Hierbei ist die 243 966 bp lange humane Sequenz auf der x-Achse und die 192 519 bp lange murine Sequenz auf der y-Achse entsprechend ihrer Orientierung im Gesamt-Contig aufgetragen (Parameter: minimale Übereinstimmung der beiden Sequenzen über eine Länge von 50 bp: 65%, Fenstergröße: 50 bp). Die Lokalisiation, Orientierung und genomische Ausdehnung der in den untersuchten Bereichen identifizierten Gene ist mit Pfeilen angegeben.

Der Dotplot zeigt eine durchgängige, fast diagonale Linie, was für die Syntänie der in beiden Spezies untersuchten Regionen spricht. Die Anzahl, Reihenfolge und Orientierung der innerhalb der Sequenzbereiche lokalisierten Gene ist, mit Ausnahme des Pseudogens L23a im Menschen, zwischen den Spezies konserviert. Es ist ein Bereich erkennbar, an denen die Diagonale leicht unterbrochen wird (Bereich bei ca. nt 113 699 im Menschen und nt 73 000 der Maus). Hier hat in einem der beiden Genome eine Insertion oder Deletion stattgefunden. Eine weitere leichte Unterbrechung der Diagonalen tritt ungefähr von Position nt 146 800 bis nt 162 005 in der humanen Sequenz auf. Hierbei handelt es sich um den Bereich zwischen Exon 7 und 8 des *ZNF143*-Gens. Das Intron 7 ist im Menschen mit 15 047 bp deutlich größer als der entsprechende Bereich des murinen *mStaf*-Gens (1341 bp). Grund hierfür ist das

Vorkommen zahlreicher repetitiver Elemente an dieser Stelle in der humanen Sequenz. Während die murine Sequenz im Intron 7 zwei SINE-Elemente aufweist, befinden sich in der ensprechenden humanen Sequenz 30 SINE-Elemente (siehe auch 4.3.3).
Da der Dotplot in der vorgestellten Form aufgrund der graphischen Aufarbeitung nur schwer detaillierte Analysen zulässt, wurde eine zusätzliche komparative Sequenzanalyse mit Hilfe des Programmes PIP-Maker durchgeführt.

3.4.2 PIP-MAKER

Ebenso wie beim klassischen Dotplot werden bei den PIPs (percent-identity plot; *Schwartz et al*., 2000) die zu untersuchenden Sequenzen aus beiden Spezies miteinander auf Sequenzübereinstimmungen untersucht. Hierbei wird die menschliche Sequenz auf der horizontalen Achse aufgetragen und unter festgesetzten Parametern mit der murinen Sequenz verglichen (siehe Abb. 3.17) und umgekehrt (siehe Abb. 3.18). Ebenso wie beim Dotplot werden konservierte Sequenzabschnitte durch einen Punkt markiert. Allerdings wird statt der Lokalisation innerhalb der Maus-Sequenz zusätzlich angegeben, zu wieviel Prozent die humane Sequenz an betreffender Stelle mit der Maus-Sequenz übereinstimmt. Somit generiert der PIP-Maker ein viel detaillierteres graphisches Resultat des Sequenzvergleiches als der Dotplot. Die Lokalisationen vorhandener Exons können hierbei ebenso wie unterschiedliche repetitive Elemente oberhalb des Plots dargestellt werden und ermöglichen so eine schnelle, übersichtliche Auswertung.
Die vorliegenden grafischen PIP-Ergebnisse (siehe Abb. 3.17 und Abb. 3.18) zeigen, dass die Exons auf Nukleotid-Ebene in beiden Spezies stark konserviert (i. d. R. über 75%) sind. Die Konservierung der 5´- und 3´-untranslatierten Bereiche ist in der Regel niedriger, aber noch immer deutlich vorhanden.

Abbildung siehe Anhang

Abbildung siehe Anhang

Durch die Auswertung der PIP-Analyse wird deutlich, dass auch zahlreiche vermutlich nicht-kodierende Bereiche deutlich konserviert sind (z. B. weist der Bereich von nt 52 000 – nt 53 000 im Menschen ca. 65%-ige Konservierung mit dem murinen Bereich von nt 58 000 bis nt 59 500 auf). Solche Bereiche könnten prinzipiell Hinweise auf noch nicht bekannte, aber in beiden Spezies vorhandene Gene oder auf konservierte regulatorische Elemente sein. Die Analyse der über die PIP-Grafik ermittelten konservierten nicht-kodierenden Bereiche wird unter 3.4.3 behandelt.
Die Daten zeigen, dass alle hier vorliegenden Gene allein aufgrund der Sequenzähnlichkeit vollständig identifizierbar waren. Somit scheint die komparative Sequenzanalyse eine sehr effiziente Methode Genidentifikation darzustellen.

3.4.3 Konservierte Bereiche zwischen Mensch und Maus

Die genomischen humanen und murinen Sequenzen wurden über das Programm PIP-Maker miteinander verglichen. Dabei zeigten sich erwartungsgemäß starke Sequenzkonservierungen in Genbereichen, aber auch an vier Stellen, an denen kein bekanntes Gen identifiziert werden konnte. Es stellte sich die Frage, ob hier noch nicht bekannte Gene oder konservierte regulatorische Bereiche lokalisiert sind. Um dies zu klären, wurden die Sequenzen Homologievergleichen unterzogen. Innerhalb der interessierenden Sequenzbereiche wurden evtl. vorhandene repetitive Elemente mit Hilfe des Programms REPEATMASKER maskiert und die so bearbeitete Gesamtsequenz anschließend erneut analysiert. Ein erster Vergleich erfolgte mit der nr-Datenbank, um mögliche Ähnlichkeiten zu bekannten Genen anderer Spezies zu identifizieren. Dann wurde ein Homologievergleich mit der humanen bzw. murinen EST-Datenbank vorgenommen, um Hinweise auf eventuell transkribierte Bereiche zu erhalten. Zusätzlich wurde mit Hilfe des im HUSAR-Programmpaket verfügbaren ESTCLUSTER-Programmes versucht, identifzierte EST-Sequenzen weitmöglichst zu verlängern. Um sicherzustellen, dass auch regulatorische Bereiche innerhalb der konservierten Regionen erkannt werden, wurde der GC-Gehalt der interessierenden Abschnitte analysiert und Promotor-Vorhersagen mit verschiedenen Programmen durchgeführt. Die Ergebnisse aus diesen Untersuchungen sind im Folgenden dargestellt, wobei die Tabelle 3.7 zunächst einen allgemeinen Überblick gibt.

Tab. 3.7: Übersicht über die zwischen Mensch und Maus konservierten Bereiche außerhalb der bekannten Gene. Die über den PIP-Maker identifizierten konservierten Bereiche wurden durch blastn-Untersuchungen und Exon- sowie Promotor-Vorhersageprogramme analysiert (siehe Text).

	Konservierter humaner Bereich (Position)	Entsprechender konservierter muriner Bereich (Position)	PIP-Ergebnis	
1	16 500 – 17 500	21 000 – 22 000	Prozentuale Übereinstimmung d. Bereichs	53%
			Anzahl d. Übereinstimmungen	329
			Anzahl d. Unterschiede	115
			Gesamtlänge aller Lücken	172
			Bereich innerhalb der humanen Sequenz	16910 – 17483
			Bereich innerhalb der murinen Sequenz	21462 - 21947
2	30 000 – 33 000	29 600 – 43 000	Prozentuale Übereinstimmung d. Bereichs	53%
			Anzahl d. Übereinstimmungen	627
			Anzahl d. Unterschiede	227
			Gesamtlänge aller Lücken	323
			Bereich innerhalb der humanen Sequenz	29605 – 30687
			Bereich innerhalb der murinen Sequenz	39600 – 40547
			Prozentuale Übereinstimmung d. Bereichs	48%
			Anzahl d. Übereinstimmungen	1189
			Anzahl d. Unterschiede	629
			Gesamtlänge aller Lücken	611
			Bereich innerhalb der humanen Sequenz	30705 – 32943
			Bereich innerhalb der murinen Sequenz	41037 - 43044
3	43 800 – 45 000	53 000 – 54 000	Prozentuale Übereinstimmung d. Bereichs	54%
			Anzahl d. Übereinstimmungen	565
			Anzahl d. Unterschiede	272
			Gesamtlänge aller Lücken	192
			Bereich innerhalb der humanen Sequenz	43621 - 44634
			Bereich innerhalb der murinen Sequenz	53163 – 54014
4	52 000 – 53 000	58 000 – 59 500	Prozentuale Übereinstimmung d. Bereichs	65%
			Anzahl d. Übereinstimmungen	427
			Anzahl d. Unterschiede	147
			Gesamtlänge aller Lücken	73
			Bereich innerhalb der humanen Sequenz	52200 – 52815
			Bereich innerhalb der murinen Sequenz	58763 - 59367

Bereich 1

Über Homologievergleiche mit der nicht-redundanten sowie der EST-Datenbank zeigte sich weder in der humanen noch in der murinen Sequenz eine signifikante Homologie zu bekannten Genen oder EST-Klonen. Auch die verwendeten Exonvorhersageprogramme konnten keine Exons in diesen konservierten Bereichen identifizieren.

Die Analyse der interessierenden Sequenzabschnitte durch Promotor-Vorhersage-Programme detektierte ebenfalls keine potenziellen Promotorbereiche. Damit übereinstimmend konnte in den entsprechenden Regionen kein erhöhter GC-Gehalt entdeckt

werden. Der GC-Gehalt innerhalb des humanen Bereiches 1 liegt bei 35,19%, während er bei dem homologen murinen Bereich 41,36% beträgt.

Bereich 2

Der Homolgievergleich mit der nicht-redundanden Datenbank erbrachte keine signifikante Ähnlichkeit zu einer beschriebenen Sequenz, allerdings zeigte die Suche in der humanen EST-Datenbank eine starke Homolgie zu dem Klon te51e10.x1 (Acc.-Nr.: AI539442). Der EST-Klon te51e10.x1 weist am 3´-Ende einen polyA-Bereich auf, es konnte in der Nähe dieses Bereiches jedoch kein putatives Polyadenylierungs-Signal detektiert werden. Auch durch die Suche nach einem EST-Cluster konnte die putativ transkribierte mRNA des EST-Klons nicht erweitert werden. Der Klon te51e10.x1 besitzt von nt 295 – nt 110 (entspricht nt 31 238 bis nt 31 053 der untersuchten humangenomischen Sequenz) einen offenen Leserahmen. Die blastx-Analyse dieses Bereiches zeigt 38%-ige Ähnlichkeit zu dem hypothetischen Protein SPBC16C6.04 (AL021767) aus *Schizosaccharomyces pombe*. Bei der Suche nach transkribierten Sequenzen aus dem entsprechenden konservierten murinen Bereich konnte jedoch weder in der nicht-redundanden noch in der murinen EST-Datenbank ein Eintrag gefunden werden, der deutliche Sequenzübereinstimmungen aufweist.

Sowohl das PIP-Ergebnis als auch die RUMMAGE-Analyse zeigte, dass es sich bei diesem Bereich um eine Region mit erhöhtem GC-Gehalt handelt, der in der humangenomischen Sequenz in der Region von nt 31 171 bis nt 32 266 lokalisiert ist (GC-Gehalt: 61,3%; siehe auch 3.4.4). Die Promotor-Vorhersage-Programme TESS, TSSG, TSSW und NNPP-eukaryotic (http://searchlauncher.bcm.tmc.edu:9331/seq-search/nucleic_acid-search.html; Baylor College of Medicine, Houston, USA) zeigten die Präsenz je eines Promotors und / oder von Bindungsstellen für Transkriptionsfaktoren in diesem Bereich an. Die Promotoren wurden von den unterschiedlichen Programmen jedoch an deutlich verschiedenen Stellen identifiziert, so dass nicht mit Sicherheit auf die Anwesenheit eines Promotors in der untersuchten Region geschlossen werden kann. Die Untersuchung des korrespondierenden murinen Bereiches (nt 41 513 – nt 42 324) mit Hilfe der oben aufgeführten Promotor-Vorhersage-Programme konnte keine potentiellen Promotoren in der interessierenden Region identifizieren. Somit kann nicht sicher gesagt werden, dass es sich hier um einen Promotorbereich handelt. Eine regulatorische Funktion dieses Bereiches kann allerdings nicht ausgeschlossen werden.

Bereich 3

Die Region von ca. nt 43 800 bis nt 45 000 der analysierten Humansequenz zeigte in einigen Bereichen zwischen 75%- und 85%-ige Sequenzidentität zur murinen Sequenz des Bereiches von etwa nt 53 000 bis nt 54 000. Homologievergleiche mit nicht-redundanden

bzw. EST-Datenbanken zeigten weder in der humanen noch in der murinen Sequenz deutliche Homologien zu vorhandenen Einträgen. Auch die Exonvorhersageprogramme konnten keine putativen Exonbereiche innerhalb beider Sequenzbereiche entdecken. Bei der Untersuchung des Bereiches 3 konnte weder ein erhöhter GC-Gehalt noch potenzielle Promotor-Bindungsstellen detektiert werden. Somit lässt sich keine Aussage über eine mögliche Funktion des konservierten Bereiches 3 in Mensch und Maus machen.

Bereich 4

Bei der Untersuchung des konservierten Bereiches 4 (im Menschen von ca. nt 52 000 bis nt 53 000; in der Maus von nt 58 700 bis nt 59 500) auf Homologien zu veröffentlichten Sequenzen in den EST-Datenbanken zeigten sich sowohl in der humanen als auch in der murinen Sequenz deutliche (beim Menschen 85%-ige, bei der Maus über 90%-ige) Sequenzübereinstimmungen zu zwei Maus-EST-Klonen, AV154313 und BB427901.

Die sequenzierte mRNA beider EST-Klone wurde aus dem Hippocampus einer männlichen adulten Maus isoliert und im Rahmen des RIKEN-Maus-EST-Projektes sequenziert (*Carninci et al*., 1999; unveröffentlicht). Das Alignment der murinen genomischen Sequenz aus dem Bereich 4 mit der publizierten Sequenz des EST-Klons AV154313 (der EST-Klon BB4327901 liegt fast vollständig im EST-Klon AV154313) ist im Folgenden aufgeführt.

```
>dbj|AV154313.1|AV154313 AV154313 Mus musculus hippocampus C57BL/6J adult Mus
                                  musculus cDNA clone 2900060N11.
 Length = 296
 Score =  436 bits (220), Expect = e-120
 Identities = 274/296 (92%)
 Strand = Plus / Minus


Query: 50  ctttgaaatatttattagctacatacaactaaaccacaggaattctgtttcttcaaggta 109
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 296 ctttgaaatatttattagctacatacaactaaaccacaggaattctgtttcttcaaggta 237


Query: 110 attctctttgaaggctggaatggttgaggtctcccagactcttagcttaatgaagttttg 169
           ||||||||||||||| |||||||||||||||||| | ||||||||||||||||||||
Sbjct: 236 attctctttgaagggtggaatggttgaggtctcccaaaatcttagcttaatgaagttttg 177


Query: 170 acagaacatgtcacctttttcctggtaaatcttcctagagcatacacagctatttctcctt 229
           | |||||||||||||||||||||||||||||||| | ||||||||||||||||||||
Sbjct: 176 aaagaacatgtcacctttttcctggtaaatcttcctaaaacatacacagctatttctcctt 117


Query: 230 actccacagacctttgaactgcattgaaaattaggaatatctgccacacnnnnnnnnnnn 289
           |||||||||||||||||||||||||||||||||||||||||||||||| ||
Sbjct: 116 actccacagacctttgaactgcattgaaaattaggaatatctgccatacaaaaaaaaaaa 57


Query: 290 ngaagaagttcaaagatgctcaggtgagtgcttctccaaagctgttgatgctaaat 345
            |||||| |||||| |||||||||||||||||||||||||||||| |||||||
Sbjct: 56  agaagaaattcaaaaatgctcaggtgagtgcttctccaaagctgttgaagctaaat 1
```

Exon-Vorhersageprogramme haben weder in der humanen noch in der murinen Sequenz Exongrenzen innerhalb der Homologien zu AV154313 aufzeigenden Regionen detektiert. Eine Suche nach putativen mRNA-verlängernden EST-Klonen über das Programm ESTCLUSTER brachte keine weiteren Sequenzinformationen. Darüber hinaus weist der konservierte Bereich 4 keinen erhöhten GC-Gehalt auf (GC-Gehalt: 35,03%). Es konnten weder beim Menschen noch in der Maus über entsprechende Vorhersageprogramme Promotoren identifiziert werden. Es könnte möglich sein, dass es sich hier um ein Exon eines neuen Gens handelt. Um diese Vermutung allerdings zu bestätigen, müssten weiterführende Expressionsanalysen durchgeführt werden.

3.4.4 GC-Gehalt und CpG-Inseln

Der durchschnittliche GC-Gehalt der untersuchten humangenomischen Sequenz beträgt 42,73% und entspricht dem GC-Gehalt, der in der murinen Sequenz vorliegt (42,76%). Der erstellte GC-Plot zeigt sowohl die Verteilung der GC-haltigen Bereiche als auch die deutlichen Schwankungen im GC-Gehalt in den analysierten genomischen Regionen von Mensch und Maus an (siehe Abb. 3.19). Zusätzlich wurde mit Hilfe des Programmpakets RUMMAGE die genomische Mensch- bzw. Maus-Sequenz auf das Vorhandensein von CpG-Inseln untersucht, die sich häufig im 5´-Bereich von Genen befinden (*Bird*, 1986) und aus diesem Grund eine Hilfe bei der Identifizierung bisher unbekannter Gene sein können.
Im GC-Plot der menschlichen und der murinen Sequenz zeigen sich zwei bzw. drei Bereiche mit deutlich erhöhtem GC-Gehalt, die mit dem 5´-Bereich von Genen assoziiert sind. Allerdings können in der humangenomischen Sequenz auch zwei Bereiche mit leicht erhöhtem GC-Gehalt identifiziert werden, die scheinbar keine Korrelation zu 5´-Bereichen von Genen aufweisen (siehe Abb. 3.19 und Tab. 3.8), aber dennoch Hinweise auf neue Gene geben können.

Abbildung siehe Anhang

Abb. 3.19: GC-Plot der orthologen genomischen Regionen um das *WEE1*/*Wee1*-Gen in Mensch (A) und Maus (B). Der GC-Gehalt, die Lokalisation und Orientierung der Gene innerhalb der untersuchten Regionen sind dargestellt. Es ist deutlich, dass die 5´-Regionen von Genen mit einem erhöhten GC-Gehalt (> 60%) korreliert sind.

In den erstellten Sequenzen aus Mensch und Maus wurden Bereiche identifiziert, die sowohl einen erhöhten GC-Gehalt (>55%) als auch große von über 300 bp Ausdehnung aufwiesen. Es konnten vier Bereiche so in der humanen Sequenz identifiziert werden, die diese Anforderungen erfüllten (siehe Tabelle 3.8).

Tab. 3.8: Übersicht über die in der humanen Sequenz identifizierten GC-reichen Regionen. Es sind solche Bereiche aufgeführt, die sowohl einen GC-Gehalt über 55% als auch eine größere Erstreckung (>300 bp) aufwiesen.

	Lokalisation in humaner Sequenz	Länge (bp)	CpG-Anteil	Homologie
A	31 171 – 32 266	1096	61,3%	keine signifikante Homologie
B	70 650 – 72 838	2189	70,6%	5´-UTR von WEE1 incl. Exon 1
C	179 860 - 180 902	1043	64,9%	5´-UTR von ZNF143 incl. Exon 1
D	236 564 - 236871	308	56,2%	Intron 2 zwischen Exon 2 und 3 von RanBP7

Die GC-reichen Bereiche B und C sind mit dem 5´-UTR der Gene *ZNF143* bzw. *WEE1* assoziiert, der GC-Reichtum erstreckt sich hierbei bis ins erste Exon hinein. Die GC-reiche Region D befindet sich im Intron 2 des RanBP7-Gens, das die Exons 2 und 3 voneinander trennt.

Auch der Bereich A (nt 31 171 bis nt 32 266) der vorliegenden humangenomischen Sequenz zeichnet sich durch einen deutlich erhöhten GC-Anteil von 61,3% aus. Dieser Abschnitt zeigte schon bei der PIP-Analyse (siehe 3.4.2; Bereich 2) deutliche Sequenzkonservierung zur murinen Sequenz. Homologievergleiche zur EST-Datenbank ergaben Ähnlichkeiten mit dem humanen EST-Klon te51e10.x1 (Acc.-Nr.: AI539442).

Auch in der murinen genomischen Sequenz wurden drei Bereiche mit hohem und ausgedehntem GC-Reichtum analysiert.

Tab. 3.9: Übersicht über GC-reiche Regionen in der murinen Sequenz. Es sind solche Bereiche aufgeführt, die sowohl einen GC-Gehalt über 55% als auch eine größere Erstreckung (>300 bp) aufwiesen.

	Lokalisation innerhalb muriner Sequenz	Länge	CpG-Anteil	Homologien
1	83 654 – 85 143	1490	70%	5´-UTR von Wee1 incl. Exon 1
2	144 427 – 145 769	1343	63,4%	5´-UTR von mStaf incl. Exon 1
3	187 799 – 188 507	709	68,8%	5´-UTR von mRanBP7, incl. Exon1

Ebenso wie bei der humanen Sequenz befinden sich die detektierten GC-reichen Regionen in unmittelbarer Umgebung des 5´-UTRs der in der Region enthaltenen Gene.

Die Länge der einzelnen GC-reichen Regionen vor den einzelnen Genen variiert deutlich zwischen Mensch und Maus. So ist der 5´-UTR des murinen *Wee1*-Gens mit 1490 bp größer als der entsprechende Bereich im Menschen (1096 bp). Dagegen ist der humane 5´-UTR des *ZNF143*-Gens mit 2189 bp deutlich größer als in der Maus (1343 bp). Über die Verhältnisse beim Gen *RanBP7* ist keine Aussage möglich, da das erste Exon des humanen *RanBP7*-Gens nicht in der sequenzierten Region enthalten ist.

3.4.5 Repetitive Sequenzen

Die ermittelten genomischen Sequenzen in Mensch und Maus wurden mit Hilfe des Programmes REPEATMASKER auf das Vorhandensein von repetitiven Bereichen untersucht. Die Ergebnisse dieser Analysen, die sowohl für die beiden humanen Einzelsequenzen getrennt als auch für die murine und humane Gesamtsequenz zusammen durchgeführt wurden, sind in den Tabellen 3.10 und 3.11 zusammengefasst.
Insgesamt beläuft sich der Anteil repetitiver Elemente im Klon PAC-180B11 auf 51,28% und im Klon PAC-142M6 auf 58,44%. Die beiden sequenzierten humanen PAC-Klone unterscheiden sich im Vorkommen bestimmter repetitiver Elemente sehr deutlich. Während beim Klon PAC-180B11 39,8% der Gesamtsequenz aus SINEs besteht, weist der PAC-142M6 diese Art repetitiver Elemente nur zu 28,51% auf. Dagegen besitzt der Klon PAC-180B11 mit 8,31% einen deutlich geringeren Anteil an LINEs als der Klon PAC-142M6 (20,15%). Auch das Vorkommen von LTR- und MER-Elementen ist im Klon PAC-180B11 niedriger als im Klon PAC-142M6 (0,98% bzw. 1,11% gegenüber 3,49% und 4,06% bei PAC142M6).

Tab. 3.10: Gegenüberstellung der repetitiven Anteile in den beiden sequenzierten humanen PAC-Klonen. Die verwendeten Abkürzungen stehen für folgende Bezeichnungen: SINE: „short interspersed nuclear element"; LINE: „long interspersed nuclear element"; LTR: „long terminal repeat"; MIR: „mammalian-wide interspersed repeat"; MER: „medium reiteration frequency interspersed repeat"; MaLR: „mammalian LTR-retrotransposon"; ERV: „endogenous retrovirus"

PAC-180B11 (119 794 bp)

Rep. Elemente	Zahl d. Elemente	Länge (bp)	% der Sequenz
Σ		61 435	51,28
SINE	184	47 683	39,80
ALU´s	170	45 821	38,25
MIR	14	1 862	1,55
LINE	25	9 950	8,31
LINE1	18	8 119	6,78
LINE2	6	1 766	1,47
L3/CR1	1	65	0,05
LTR-Elemente	2	1 171	0,98
MaLR	1	124	0,1
ERVL	/	/	/
ERLV-cl.I	/	/	/
ERLV-cl.II	1	1 047	0,87
DNA-Elemente	6	1 326	1,11
MER1	5	1 005	0,84
MER2	1	321	0,27
„small RNA"	/	/	/
„simple repeats"	16	733	0,61
„low complex."	19	584	0,49

PAC-142M6 (131 998 bp)

Rep. Elemente	Zahl d. Elemente	Länge (bp)	% der Sequenz
Σ		77 146	58,44
SINE	150	38 954	29,51
ALU´s	144	37 933	28,74
MIR	6	1 021	0,77
LINE	43	26 601	20,15
LINE1	25	19 374	14,68
LINE2	17	7 139	5,41
L3/CR1	1	88	0,07
LTR-Elemente	11	4 608	3,49
MaLR	5	1530	1,1
ERVL	/	/	/
ERLV-cl.I	5	3 043	2,3
ERLV-cl.II	1	36	0,03
DNA-Elemente	13	5 365	4,06
MER1	7	2 006	1,52
MER2	3	1 019	0,77
„small RNA"	2	435	0,33
„simple repeats"	12	363	0,28
„low complex."	24	900	0,68

Der Vergleich der murinen mit der humanen Gesamtsequenz zeigt ebenfalls deutliche Unterschiede im Vorkommen repetitiver Elemente (siehe Tab. 3.11).

Der Anteil repetitiver Elemente ist in der untersuchten humanen Sequenz mit 55,26% höher als in der untersuchten murinen Sequenz (41,87%). Der direkte Vergleich des prozentualen Vorkommens bestimmter repetitiver Elemente in Mensch und Maus zeigt ebenfalls eine deutlich unterschiedliche Verteilung.

Tab. 3.11: Gegenüberstellung der repetitiven Anteile in der sequenzierten humanen und murinen Region. Die verwendeten Abkürzungen stehen für folgende Bezeichnungen: SINE: „short Interspersed nuclear element"; LINE: „long interspersed nuclear element"; LTR: „long terminal repeat"; MIR: „mammalian-wide interspersed repeat"; MER: „medium reiteration frequency interspersed repeat"; MaLR: „mammalian LTR-retrotransposon"; ERV: „endogenous retrovirus"

Humane Sequenz (243 966 bp)

Rep. Elemente	Zahl d. Elemente	Länge (bp)	% der Sequenz
Σ		134817	55,26
SINE	324	83 867	34,38
ALU´s	305	81 137	33,26
MIR	19	2 730	1,12
LINE	66	35 573	14,58
LINE1	40	26 448	10,84
LINE2	24	8 972	3,68
L3/CR1	2	153	0,06
LTR-Elemente	13	5779	2,37
MaLR	6	1 654	0,68
ERVL	/	/	/
ERLV-cl.I	5	6 042	1,25
ERLV-cl.II	2	1 087	0,44
DNA-Elemente	19	6 691	2,74
MER1	12	3 011	1,23
MER2	4	1 340	0,55
„small RNA"	2	435	0,18
„simple repeats"	27	1 070	0,44
„low complex."	43	1 484	0,61

Murine Sequenz (192 519 bp)

Rep. Elemente	Zahl d. Elemente	Länge (bp)	% der Sequenz
Σ		80 609	41,87
SINE	410	52 699	27,37
B´s	393	51 462	26,73
MIR	3	279	0,14
ID´s	14	958	0,50
LINE	19	4 556	2,37
LINE1	12	3 758	1,95
LINE2	5	667	0,35
L3/CR1	2	131	0,07
LTR-Elemente	40	16 164	8,40
MaLR	20	6 930	3,60
ERVL	2	1 598	0,83
ERLV-cl.I	/	/	/
ERLV-cl.II	1	2 658	1,38
DNA-Elemente	8	1 797	0,93
MER1	7	1 264	0,66
MER2	/	/	/
„small RNA"	1	80	0,04
„simple repeats"	85	3 791	1,97
„low complex."	41	1 533	0,80

Sowohl SINE- als auch LINE-Elemente sind in der humanen Gesamtsequenz häufiger zu finden als im analysierten Maus-Klon. SINEs machen in der humanen Sequenz ca. 34% der Gesamtsequenz aus, während es in der Maus nur ca. 27% sind. Noch gravierender ist der Unterschied beim Vorkommen von LINE-Elementen. Während nur ca. 2% der murinen Sequenz aus LINEs besteht, weist diese Familie repetitiver Elemente in der untersuchten humangenomischen Sequenz einen Anteil von fast 15% auf.

4 Diskussion

Über die komparative Sequenzanalyse orthologer chromosomaler Abschnitte können Gene oder konservierte regulatorische Bereiche zuverlässig erkannt werden. In der vorliegenden Arbeit wurde im Rahmen des Deutschen Humangenom-Projektes eine vergleichende Sequenzanalyse eines Bereiches aus der humanen Chromosomenregion 11p15.3 und der orthologen Region der Maus auf Chromosom 7 durchgeführt. Die analysierte Region ist im Menschen besonders interessant, weil hier Gene lokalisiert wurden, die an der Ausprägung des Beckwith-Wiedemann-Syndroms (*Beckwith*, 1963; *Wiedemann*, 1964) und assoziierter Tumoren beteiligt sind. Auch konnten die Tumorsuppressor-Gene *WEE1* (*Igarashi et al*., 1991), *ST5* (*Lichy et al*., 1992) und *LMO1* (*Boehm et al.,* 1990) in diese Region kartiert werden. Untersuchungen von *Bepler & Koehler* (1995) legten aufgrund eines häufig beobachteten Allelverlusts in Lungenkrebs-Zelllinien das Vorhandensein potenzieller Tumorsuppressor-Gene in der Region 11p15.3 nahe. Kartierungsversuche zeigten, dass die Tumorsuppressor-Gene *WEE1*, *ST5* und *LMO1* im Menschen in einer syntänen Region angesiedelt sind (*Higgins et al*., 1994; *Redeker et al*., 1995); die exakte Reihenfolge und Orientierung der einzelnen Gene war jedoch unklar. Im Rahmen des Deutschen Humangenom-Projektes sollte die genomische Region zwischen *WEE1* und *LMO1* in Mensch und Maus durchgängig von überlappenden PAC-Klonen abgedeckt und die Bereiche um diese bekannten Tumorsuppressorgene jeweils im Rahmen von eigenständigen Dissertationen sequenziert werden.

In der vorliegenden Dissertation wurde ein ca. 250 kb großer Bereich um das *WEE1*-Gen herum aus der humanen Chromosomenregion 11p15.3 und der orthologe, ca. 190 kb große Abschnitt auf dem Maus-Chromosom 7 sequenziert. Nach Aufklärung beider Nukleotidsequenzen wurde die Struktur und Organisation der darin identifizierten Gene sowie das Vorkommen repetitiver Elemente in Mensch und Maus untersucht. Aufgrund des komparativen Ansatzes konnten auch Konservierungen außerhalb kodierender Bereiche identifiziert und mit Sequenz-spezifischen Eigenschaften, wie z. B. dem GC-Gehalt, korreliert werden.

4.1 Chromosomale Lokalisation der orthologen Gengruppen

Die vier aus der humanen Chromosomenregion 11p15.3 stammenden Gene *LMO1*, *ST5*, *CEGF1* und *WEE1*, die als Startpunkte für die Sequenzierung dienen sollten, wurden beim Menschen als auch teilweise in der Maus zueinander kartiert (*Taviaux et al*., 1993; *Higgins et al*., 1994; *Redeker et al*., 1995). Obwohl die etwa 14 Mb umfassende Chromosomenbande

11p15 (NCBI-Genome: „map viewer“; siehe Tab. 2.4) die am intensivsten untersuchte Region von Chromosom 11 darstellt, lag zum Zeitpunkt des Projektstartes keine gesicherte Aussage über die relative Lage der vier Gene zueinander vor; existierende Anordnungsversuche gaben widersprüchliche Informationen (*van Heyningen & Little*, 1995; *Redeker et al.*, 1995; *Higgins et al.,* 1994). Die zu diesem Zeitpunkt aktuellen Karten unterschiedlicher Autoren sind in der Abbildung 4.1 dargestellt.

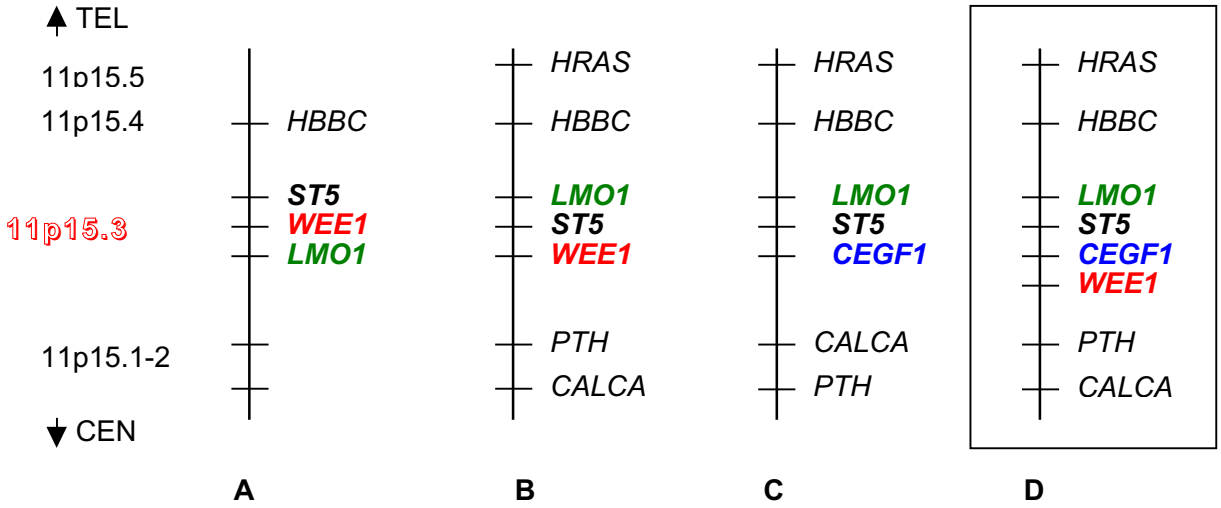


Abb. 4.1: Anordnungen einiger Gene in der humanen Chromosomenregion 11p15 nach verschiedenen Autoren (nach *Seipel*, 1996). Die umrahmte Anordnung (D) war Basis des vorliegenden Sequenzierprojektes (*Seipel*, 1996). A: nach *Redeker et al.* (1995); B: nach *van Heyningen & Little* (1995), C: nach *Higgins et al.* (1994); D: Gen-Lokalisationen über 2-Farben-FISH nach *Seipel* (1996)

Im Rahmen einer Diplomarbeit konnte *Seipel* (1996) über FISH-Analyse *LMO1* als distale und *WEE1* als proximale Begrenzung der humanen Gengruppe ermitteln. Das Gen *ST5* wurde zwischen *LMO1* und *WEE1*, jedoch distal von *CEGF1* angeordnet (siehe Abb. 4.1). In den letzten Jahren konnte gezeigt werden, dass ein Teil der auf dem kurzen Arm des humanen Chromosoms 11p lokalisierten Gene auf dem murinen Chromosom 7 in drei voneinander getrennten Syntäniegruppen über eine Distanz von fast 50 cM verstreut liegen (*Rinchik et al.*, 1992; *Holdener et al.*, 1993; *Brilliant et al.*, 1994; *Stubbs et al.*, 1994). Auch die am „National Center for Biotechnology Information“ (NCBI) zugängliche Gegenüberstellung (http://www3.ncbi.nlm.nih.gov/Homology/human11.html) zeigt, dass weite Bereiche aus 11p15 ortholog zu Maus-Chromosom 7 sind. Eine Ausnahme bildet nur die Region um das in 11p15.5 gelegene Gen *CTSD*, dessen orthologes Maus-Gen sich auf dem murinen Chromosom 4 befindet. Ebenso wie auf dem humanen Chromosom 11 war auch auf

dem murinen Chromosom 7 versucht worden, die interessierenden Gene *Lmo1*, *St5* und *Wee1* zueinander anzuordnen. Nachdem die bisher angenommene Lokalisation der Gene *Lmo1* und *St5* auf dem murinen Chromosom 7 durch CISS-Hybridisierung überprüft worden war (*Foroni et al*., 1992; *Angel et al*., 1993), konnte *Seipel* (1996) mit Hilfe der 2-Farben-FISH-Analyse die murinen Gene *St5* und *Lmo1* zueinander auf dem Maus-Chromosom 7 anordnen. Beide Gene hybridisieren weit distal von *Art1* (*Koch-Nolte et al*., 1996), wobei *Lmo1* als proximaler und *St5* als distaler Marker angeordnet werden konnte. *Seipel* lokalisierte aufgrund vorhandener Daten (*Brilliant et al*., 1994; *Shows et al*., 1996) das Gen *Wee1* distal von *St5*; experimentelle Bestätigung konnte jedoch aufgrund des Fehlens eines genomischen Klons, der das murine *Wee1*-Gen enthält, zu diesem Zeitpunkt noch nicht erbracht werden. Aus diesem Grunde wurde in der vorliegenden Dissertation eine 2-Farben-FISH-Analyse des murinen Klons PAC-256N10, der das Gen *Wee1* komplett beinhaltet, durchgeführt. Gleichzeitig mit dem Klon PAC-256N10 wurde der murine Klon BAC-287P4 hybridisiert. Dieser Klon, der das Gen *tubby* (*tub*) teilweise umfasst (*Kleyn et al*., 1996), ist auch Bestandteil eines Contigs, der ausgehend vom *Lmo1*-Gen etabliert wurde (*Cichutek et al*., 2001; *Brückmann*, unveröffentlicht). Das Ergebnis dieser Doppelhybridisierung bestätigte die postulierte Lokalisation des murinen *Wee1*-Gens auf dem Maus-Chromosom 7 (siehe 3.1.2 und Abb. 3.3). Eine weitere Feinkartierung wurde nicht vorgenommen, da die vorliegenden genomischen Sequenzen die relative Lage der Startgene zueinander aufklären konnten (*Amid et al*., 2001; *Cichutek et al*., 2001). Beim Vergleich der Anordnung der als Startpunkte dienenden Gene zwischen Mensch und Maus wurde deutlich, dass die Abfolge der Gene konserviert ist; allerdings liegt der entsprechende chromosomale Bereich im Mausgenom invertiert vor (siehe 3.1 und Abb. 3.2). So ist in der Maus das Gen *Wee1* das distal begrenzende Gen, während *Lmo1* die proximale Grenze der Gengruppe darstellt.

Durch das Vorliegen der vorläufigen humangenomischen Nukleotidsequenz (*IHGSC*, 2001; *Venter et al*., 2001) lassen sich experimentell ermittelte Genanordnungen konkret überprüfen. Zum Zeitpunkt der Veröffentlichung wurde die ca. 3,2 Mrd Basenpaare umfassende humangenomische Sequenz durch etwa 23 Mrd Rohbasen abgedeckt (was einer Redundanz von 7,6 entspricht; siehe auch 4.2) (*IHGSC*, 2001). Von diesen 23 Mrd Rohbasen entsprachen allerdings nur 9 Mrd Basen (39%) der geforderten hohen Qualität („finished“) vor. Der Großteil, also knapp 13,4 Mrd Basen, lag als sogenannte „draft“-Sequenz. Hierunter wird die vorläufige Anordnung bereits sequenzierter, aber noch nicht der geforderten Qualität entsprechender Klone zu einem übergeordneten Contig verstanden. Durch solche Contigs sind weite chromosomale Bereiche abgedeckt und somit weiteren Analysen zugänglich. Aus diesem Grund wurde die in der vorliegenden Arbeit sequenzierte humangenomische Region des Chromosoms 11p15.3 im Bereich des *WEE1*-Gens mit physikalischen Gen-Karten verglichen. Hierbei zeigt sich, dass die über das NCBI

zugängliche, vorläufige humangenomische Sequenz (http://www.ncbi.nlm.nih.gov:80/cgi-bin/Entrez/hum_srch?chr=hum_chr.inf&query) die vorgenommene Anordnung der Gene *LMO1, ST5* und *WEE1* bestätigt. Auch die Lokalisation der Gene *ZNF143* und *RanBP7* (siehe 3.3.1), die sich schon vor der Veröffentlichung der vorläufigen humangenomischen Sequenz im Rahmen der vorliegenden Dissertation aus der Sequenzaufklärung des hierbei untersuchten Bereiches ergab, konnte mit Hilfe der humangenomischen Sequenz im nachhinein überprüft werden. Hierbei zeigten sich Übereinstimmungen in der Anordnung, aber Unterschiede in der postulierten Orientierung der Gene. Während sich in der Sequenz, die im Rahmen der vorliegenden Dissertation angefertigt wurde, die drei Gene *WEE1*, *ZNF143* und *RanBP7* in der gleichen Orientierung befinden (siehe Abb. 3.5), weist das Gen *RanBP7* in der Sequenz des *IHGSC* (2001) eine gegenläufige Orientierung auf. Die drei untersuchten Gene *WEE1*, *ZNF143* und *RanBP7* liegen hier innerhalb des 649 461 bp umfassenden Contigs NT_024206.2, der aus 6 Einzelklonen besteht. Darunter befindet sich auch der Klon RP11-555O6 (Acc. Nr. AC019126), der zur Assemblierung des Klons PAC-180B11 zu Hilfe genommen wurde (siehe auch Abb. 3.2). Da die Nukleotidsequenzen der einzelnen Klone, die zur Erstellung des genomischen Contigs NT_024206.2 verwendet wurden, noch nicht vollständig aufgeklärt waren, scheinen bei der Assemblierung sowohl der einzelnen Klone als auch des Contigs NT_024206.2 fehlerhafte Anordnungen aufgetreten zu sein. So zeigt eine detaillierte Analyse dieses Contigs, dass die angenommene genomische Ausdehnung des Klons AC053531 innerhalb des Contigs NT_024206.2 seine ermittelte Größe deutlich übersteigt. Auch der Bereich um *CEGF1* wird von Klonen, die im Rahmen des *IHGSC* sequenziert werden, abgedeckt. Der 199 203 bp große Klon RP11-299K14 (Acc. Nr. AC026077, siehe auch Abb. 3.2) überlappt um knapp 34 kb mit dem von *Bahr* sequenzierten Klon PAC-275G1 und verlängert den Contig in Richtung *WEE1*. Allerdings konnte vom *IHGSC* der genomische Bereich zwischen den Klonen RP11-555O6 und RP11-299K14 nicht durch überlappende Klone abgedeckt werden. In der vorliegenden Arbeit konnte der Bereich zwischen *WEE1* und *CEGF1* durch einen durchgängigen PAC-Contig abgedeckt (siehe Abb. 3.2) und auf eine Grösse von mindestens 580 kb geschätzt werden. Diese Entfernung von über einer halben Megabase ist deutlich größer als von *Seipel* (1996) angenommen, die die Entfernung zwischen *WEE1* und *CEGF1* über Interphase-2-Farben-FISH-Analyse auf nur ca. 250 kb schätzte.

Somit unterstützen die Daten, die im Rahmen des Deutschen Humangenom-Projektes mit Hilfe der komparativen Sequenzanalyse eines Bereiches aus der humanen Chromosomenregion 11p15.3 und der orthologen Region der Maus auf Chromosom 7 resultieren, die von *Seipel* (1996) ermittelte Anordnung der Gene *WEE1*, *ST5* und *LMO1* in Mensch und Maus. Die öffentlich zugängliche, vorläufige Sequenz des Contigs NT_024206.2 konnte nicht bestätigt werden; die dort postulierte Genanordnung scheint auf fehlerhaften

Assemblierungen sowohl innerhalb der einzelnen Klone als auch von Klonfragmenten innerhalb der vorläufigen Contig zu beruhen.

4.2 Sequenzanalyse großer chromosomaler Bereiche

Zur Zeit werden zwei unterschiedliche Ansätze zur Sequenzierung großer Genome verfolgt. Bei der 1997 noch kontrovers diskutierten Methode der „whole genome-shotgun"-Sequenzierung (*Weber & Myers*, 1997; *Green*, 1997) wird das komplette Genom in Fragmente zerlegt und erst nach der Sequenzierung mit Hilfe von genetischen Markern, deren Anordnung relativ zueinander bestimmt wurde, zusammengesetzt. Computer-gestützte Analysen sollen hierbei fehlerhaftes Zusammensetzen verhindern. Dieser Ansatz wurde bis vor einigen Jahren zur Aufklärung der Erbinformation von „repeat"-armen Virus-, Bakterien- und Fliegengenomen verwendet. Die Firma *Celera Genomics* hat die „whole genome-shotgun"-Methode verwendet, um das humane Genom zu sequenzieren (*Venter et al.*, 2001). Die erfolgreiche Assemblierung von Einzelsequenzen zur vorliegenden Rohfassung der menschlichen Nukleotidsequenz war jedoch nur dadurch möglich, dass Daten des öffentlich geförderten Humangenom-Projektes zu Hilfe genommen wurden. Für das Humangenom-Projekt wurde der Ansatz der hierarchischen oder auch Karten-gestützten „shotgun"-Sequenzierung verwendet. Hierbei werden zunächst Klone mit großen Integraten hergestellt, die das komplette Genom redundant abdecken und eindeutig bestimmten chromosomalen Regionen zugeordnet werden können. Es ergeben sich hierbei Contigs zusammenhängender, überlappender Klone. Die Nukleotidsequenz der auf diese Weise generierten PAC- oder BAC-Klone wird anschließend mittels „shotgun"-Sequenzierung ermittelt. Hierbei werden Klone (zumeist BAC- oder PAC-Klone, da sie im Gegensatz zu YAC-Klonen bis zu 250 kb große Integrate stabil propagieren können), die definierte Bereiche des humanen Genoms tragen, durch Scherkräfte in zufällige, unterschiedlich große Fragmente zerlegt und sequenziert. Jeder Bereich des genomischen Integrats wird somit durch mehrere Fragmente abgedeckt. Diese Redundanz ist ein wichtiges Qualitätsmerkmal bei der Rekonstruktion der durchgängigen Gesamtsequenz des fragmentierten Ausgangsklons (s. u.). Weil die so generierte Sequenzinformation begrenzt ist, kann die Gefahr eines großflächigen falschen Zusammenbaus eliminiert und eine fehlerhafte Rekonstruktion aufgrund repetitiver Sequenzen innerhalb des Klones stark reduziert werden. Ein Nachteil der hierarchischen „shotgun"-Sequenzierung besteht darin, dass einige Klone mit großen Integraten (wie z. B. YAC-Klone) Rearrangement-Ereignisse durchlaufen und somit chimäre Klone repräsentieren können. Durch geeignete Kontrollen (wie z. B. FISH-

Analyse oder die Erstellung Klon-spezifischer Restriktionsmuster („fingerprinting“)) kann dieses Risiko allerdings minimiert werden.

Unabhängig von der verwendeten Sequenzierungsmethode ist ein hoher Automatisierungsgrad sowohl bei der praktischen Probengenerierung als auch bei der computergestützen Sequenzauswertung notwendig. Nur auf diese Weise kann unter ständiger Einbeziehung der neuesten technischen Entwicklungen und Optimierung von automatisierten Vorgängen der benötigte hohe Probendurchsatz zur schnellstmöglichen Sequenzgenerierung gewährleistet werden. Ein grundlegender Aspekt ist dabei neben der Hochdurchsatz-Sequenzierung eine möglichst effiziente Qualitätskontrolle der Sequenzen. Bei der Sequenzierung des humanen Klons PAC-142M6 erfolgte die Assemblierung der Sequenzen mit Hilfe des Programmes SEQUENCHER™3.1. Die Editierung der Daten erfolgte manuell: eventuell vorhandene Vektoranteile wurden dabei entfernt und die generierte Nukleotidabfolge wurde anhand des vom Sequenzierautomaten über die Geräte-eigene Software erstellten Elektropherogrammes überprüft und wenn erforderlich korrigiert. Neben des starken manuellen Eingriffs beim Editieren bestand ein weiterer Nachteil der auf diese Weise überarbeiteten Daten darin, dass Sequenzbereiche mit geringer Qualität entfernt werden mussten, da das Programm SEQUENCHER™3.1 alle Sequenzbereiche zusammenbaut, welche die definierten Parameter bzgl. der Überlappung und der prozentualen Übereinstimmung erfüllen. Um somit zu verhindern, dass die qualitativ schlechten, meist randständigen Sequenzbereiche zu einem falschen Zusammenbau führen, mussten diese entfernt werden. Allerdings bedeutete dies auch einen irreversiblen Informationsverlust. Dies waren wichtige Gründe, für die Sequenzeditierung und -assemblierung des humanen Klons PAC-180B11 und des murinen Klons PAC-256N10 das Programmpaket PHREDPHRAP (*Ewing et al*., 1998; *Ewing & Green*, 1998) zu verwenden. Bei der automatischen Sequenzanalyse mit dem ABI377-Sequenziergerät wird das generierte Gelbild durch eine Computeranalyse in eine Basensequenz umgewandelt. Hierfür wird, nachdem der Verlauf jeder einzelnen Probe auf dem Sequenzgel kontrolliert wurde („lane-tracking“), ein Profil („trace“) erstellt. Dieses Profil zeigt von jedem der vier Basen-spezifisch verwendeten Fluorochrome eine getrennte Darstellung der Emissionsintensität (Signale) im Verlauf der Elektrophorese. In einem weiteren Rechenschritt werden störende Faktoren, wie z. B. der Hintergrund, reduziert und Effekte einzelner Farbstoffe auf das Laufverhalten der Fragmente oder das während des Elektrophoreselaufes sinkende Auflösungspotenzial des Gels ausgeglichen. Erst danach wird das prozessierte Profil in eine Abfolge von Basen (Chromatogramm) übersetzt („base calling“). Das Programm PHRED führt unabhängig von der über das Sequenziergerät hergestellten Basensequenz ein erneutes „base calling“ durch, das im Durchschnitt 40-50% weniger Fehler macht als die dem ABI-Sequenziergerät angepasste ABI-Software (*ABI*, 1996). *Ewing et al*. (1998) konnte

zeigten, dass die Fehlerrate für Substitutionen bei beiden Programmen ungefähr gleich ist. Die Indel-Rate bei PHRED-generierten Sequenzen ist jedoch niedriger als bei den durch ABI-Software generierten. Da unbedingt zwischen durch Substitionen und Insertionen/Deletionen (Indels) einzelner Nukleotide hervorgerufenen Sequenzunsicherheiten unterschieden werden muss (Indels haben in der Regel gravierendere Auswirkungen bei der Sequenzinterpretation als Substitutionen, da sie potenzielle Leserahmen verschieben können), ist neben der erniedrigten Fehlerrate auch die weniger starke Auswirkung auf den „Sinngehalt“ der sequenzierten DNA bei fehlerhaftem „base-calling“ ein deutlicher Vorteil von PHREDPHRAP. Zusätzlich erstellt PHRED um etwa 5% längere auswertbare Sequenzen. Ein weiterer, grundlegender Vorzug des Programmes PHRED liegt darin, dass jeder mittels verbessertem „base calling“ verifizierten Base eine spezifische Fehlerwahrscheinlichkeit zugeordnet wird (*Ewing & Green*, 1998). Somit erfolgt kein Informationsverlust durch Eliminierung qualitativ schlechter Sequenzbereiche, sondern eine Art Markierung von Nukleotiden entsprechend ihrer wahrscheinlichen Zuverlässigkeit. Diese Fehlerwahrscheinlichkeiten werden bei der verbesserten Sequenzzusammenführung im Assemblierungsprogramm PHRAP und dem Programm CONSED, mit dessen Hilfe die korrekte, den Standards (Fehlergenauigkeit unter 1 Fehler in 10000 bp) entsprechende Konsensus-Sequenz generiert wird, berücksichtigt. Zusätzlich führt PHRAP eine automatische Maskierung vorhandener Vektoranteile durch. Zusammenfassend kann gesagt werden, dass die im Verlaufe der Arbeit etablierte Verwendung der Programme PHREDPHRAP und CONSED die Generierung der Konsensus-Sequenzen wesentlich erleichtert sowie qualitativ verbessert hat.

Um die Effizienz bei der Erstellung der Konsensus-Sequenzen unter Berücksichtigung der beiden benutzten Programme zu untersuchen, müssen mehrere Faktoren detailliert analysiert werden. Dazu zählt der Zeitpunkt des Überganges von der „shotgun"-Sequenzierungsphase zur „finishing“-Phase, die durchschnittliche Leselänge und die generierte Rohbasenanzahl ebenso wie die Redundanz. Ein Vergleich der aus den unterschiedlichen Computerauswertungen resultierenden Daten und deren Einordnung in generelle Richtwerte aus anderen vergleichbaren Projekten erweist sich in der vorliegenden Arbeit als schwierig. Das Programm PHREDPHRAP gibt keine Informationen über die Summe aller im Contig befindlichen Nukleotide (Rohbasen). Auch die Aussage über die durchschnittliche Leselänge der integrierten Sequenzen ist nicht direkt möglich, da die Sequenzen in ihrer kompletten Leselänge (wenn auch mit Informationen über die Fehlerwahrscheinlichkeit versehen) integriert werden. Somit lässt sich die durchschnittliche Leselänge nur schätzen. Demzufolge kann die aus diesen Werten berechnete Redundanz nur in grober Näherung angegeben werden. Bei der hierarchischen „shotgun“-Sequenzierung sollte jeder DNA-Abschnitt etwa sechs- bis elfmal sequenziert werden. Diese

Redundanz ist ein wichtiges Qualitätsmerkmal bei der Generierung der gewünschten Konsensus-Sequenz. Der mit Hilfe des Programmes SEQUENCHER™3.1 zusammengesetzte PAC-142M6 wurde mit einer Redundanz von 7 sequenziert. Diese recht niedrige Redundanz korreliert mit der hohen Anzahl an benötigten Primern (202) durch die relativ früh eingeleitete „primer walking"-Phase. Bei einer angenommenen durchschnittlichen Leselänge von ebenfalls 600 bp ist der murine PAC-Klon 256N10 ebenfalls mit einer Redundanz von 7 sequenziert worden, allerdings wurden nur 75 Primer zur Generierung einer durchgängigen Konsensus-Sequenz aus 22 Contigs benötigt. Demgegenüber ist die geschätzte Redundanz 13, mit welcher der Klon PAC-180B11 sequenziert wurde, relativ hoch. Dies ist auf den hohen Anteil an repetitiven LINEs (20,15%, siehe auch 4.3.6) zurückzuführen, der sehr viele Sequenzreaktionen zum korrekten Zusammenbau der Konsensus-Sequenz erforderte. Auch die im Gegensatz zum entsprechenden murinen Klon PAC-256N10 benötigte hohe Primerzahl von 128 resultiert aus den beschriebenen Schwierigkeiten. Zusammenfassend kann aber gesagt werden, dass die ermittelten Redundanzen bei der Sequenzierung der vorliegenden Klone dem internationalen Standard entsprechen.

Eine gesicherte Aussage über die vorliegende Sequenziergenauigkeit im Vergleich zum internationalen Standard (angestrebte Fehlerrate von 0,01%) lässt sich zum jetzigen Zeitpunkt nicht machen, da die entsprechenden Bereiche in den öffentlichen Datenbanken zum Zeitpunkt der Auswertung lediglich als vorläufige „draft"-Sequenzen verzeichnet waren. Die humanen Klone PAC-142M6 und PAC-180B11 überlappen über einen Bereich von knapp 8 kb. Innerhalb dieses Bereiches unterscheiden sich die Klone an zwei Positionen (nt 125 288: M; nt 127 701: Y). Vergleiche dieser ambigen Sequenzstellen mit der htgs-Datenbank zeigten 99%-ige Homologien zu dem Klon RP11-555O6 (Acc. Nr. AC019126), der vom Genome Sequencing Center in Washington (USA) sequenziert wurde und sich zum Zeitpunkt der Auswertung in der „finishing"-Phase befand. Dessen fragmentarische Nukleotidsequenz unterstützt an Position nt 125 288 die Sequenz des Klons PAC-142M6, während sie an Position nt 127 701 die Sequenz des Klons PAC-180B11 verifiziert (siehe 3.2.2). Aus diesen beiden Sequenzunterschieden im Überlappungsbereich der humanen Klone kann allerdings nicht zwingend auf eine daraus resultierende Fehlerwahrscheinlichkeit der generierten Nukleotidsequenz von 0,025% geschlossen werden. Die PAC-Bibliothek, aus der die verwendeten Klone stammen (*Iannou et al.*, 1994), wurde von DNA hergestellt, die aus dem Blut männlicher Spender isoliert wurde. Dabei handelte es sich um mehrere Individuen, deren Genome sich durch Sequenzvariationen, SNPs („single nucleotide polymorphisms"), unterscheiden. Die *INTERNATIONAL SNP MAP WORKING GROUP* (2001) konnte 1,42 Millionen SNPs in der humangenomischen Sequenz detektieren, was einem durchschnittlichen Vorkommen eines SNPs pro 1,9 kb entspricht. Die Daten von

Venter et al. (2001) legen sogar ein Auftreten von 2,1 Millionen humanen SNPs nahe, was einer statistischen Verteilung eines Polymorphismus pro 1,25 kb entspricht. Da die Bereiche der beiden Sequenzunterschiede in der überlappenden Region der sequenzierten PAC-Klone 142M6 und 180B11 jeweils hohe Sequenzqualitäten (auch in unmittelbarer Nachbarschaft) aufweisen, scheinen sie durch SNPs erklärbar zu sein, da jeder einzelne Haplotyp der Spender in jedem einzelnen Klon repräsentiert sein kann.
Eine abschließende Schätzung der Fehlerwahrscheinlichkeit ist daher erst möglich, wenn die humangenomische Gesamtsequenz in ihrer endgültigen Form vorliegt, in der die detektierten Polymorphismen eindeutig identifiziert sein sollten.

4.3 Sequenzauswertung und komparative Sequenzanalyse

4.3.1 Methoden zur Genidentifizierung und komparativen Sequenzanalyse

Gene bzw. die kodierenden Bereiche von Genen stellen zwar nur einen kleinen Anteil der humanen DNA dar, dennoch ist die Identifizierung und nachfolgende Untersuchung dieser Gene das Ziel der meisten großen Sequenzierprojekte. Dementsprechend war auch in der vorliegenden Dissertation die Identifizierung von Genen ein zentrales Thema. Bei der Auswertung der hier ermittelten Sequenzdaten wurde eine Kombination von drei unterschiedlichen Untersuchungsmethoden zur Identifizierung möglichst aller in den untersuchten Genombereichen lokalisierter Gene eingesetzt.

(1) Datenbanksuchen
Mit Hilfe bestehender Datenbanken können sequenzierte Genomabschnitte auf Homologien zu bereits bekannten Genen bzw. ESTs untersucht werden. Hierbei werden Datenbanksuchen mit Hilfe der nr- und der dbEST-Datenbank durchgeführt.
Während mit Hilfe der nr-Datenbank bekannte Gene bzw. Homologien zu bekannten Genen identifiziert werden können, ist es über Homologievergleiche mittels der EST-Datenbanken möglich, Teilbereiche, in der Regel ein oder zwei Exons, zumeist unbekannter Gene zu identifizieren. Erstmals wurde 1991 von *Adams* und Mitarbeitern ein EST-Projekt durchgeführt. Generell wird bei diesen Projekten aus einem bestimmten Gewebe RNA isoliert und daraus eine cDNA-Bibliothek erstellt. Nach der Isolierung einer möglichst großen Anzahl zufällig ausgewählter cDNA-Klone werden diese ansequenziert. Hierbei können durchschnittlich 300 bp an Sequenzinformation erhalten werden, welche ein bestimmtes Transkript als sogenanntes EST („expressed sequence tag“) eindeutig charakterisieren.

Inzwischen existieren diverse EST-Projekte, die Teile von Transkripten aus unterschiedlichen Spezies wie Mensch, Maus, Ratte und anderen Organismen identifizieren (z. B. *Adams et al*,. 1993; *Liew et al.*, 1994; *Adjaye et al*., 1997; *Marra et al*., 1999). Im Rahmen des UniGene-Programms wurde eine überlappende Anordnung der verschiedenen ESTs vorgenommen und Daten aus dem Jahr 2000 ließen die Vermutung zu, dass die bis dahin bekannten EST-Klone ungefähr 86 000 verschiedene Gene repräsentieren, wovon nur etwa 11% vollständig sequenzierte cDNAs darstellen (http://www.ncbi.nlm.nih.gov/UniGene). Dabei ist zu beachten, dass viele stark exprimierte Gene von zahlreichen ESTs abgedeckt werden (*Schuler et al*., 1996). So ist z. B. das humane Serumalbumin durch mehr als 1300 Sequenzen in der EST-Datenbank vertreten. Allerdings wird versucht, diese Redundanz der ESTs durch die Herstellung normalisierter cDNA-Banken zu umgehen, wodurch der Anteil schwach exprimierter Gene relativ erhöht wird (*Bonaldo et al.,* 1996). Als deutlicher Schwachpunkt neben der oben beschriebenen Redundanz häufig exprimierter Transkripte muss die oft schlechte Qualität der EST-Sequenzen angefügt werden. Da die in der EST-Datenbank publizierten Klone in der Regel nur einzelsträngig sequenziert und nicht editiert werden, kann von einer Lesegenauigkeit von etwa 97% ausgegangen werden (*Hillier et al.,* 1996). Auch die auf ca. 10% geschätzten genomischen Kontaminationen unter den EST-Sequenzen (*Glöckner et al.,* 1998) stellen ein Problem dar. Da für die Synthese der verwendeten cDNA-Bibliotheken zumeist ein oligo-dT-priming verwendet wurde, repräsentieren etwas 65% aller veröffentlichten EST-Sequenzen das 3´-Ende und somit größtenteils den untranslatierten Bereich der cDNAs. Nur etwa 26% repräsentieren das 5´-Ende von Genen (*Aaronson et al*., 1996). Daher wird sowohl in öffentlichen als auch privaten Organisationen versucht, durch die Etablierung neuer Methoden zur cDNA-Bibliothek-Synthese Volllängen-cDNAs zu isolieren und in die bestehenden EST-Projekte zu integrieren (*Neto et al*., 2000; *Ota et al*., 2000 und *Gu et al*., 2000; beide unveröffentlicht, *Das et al*., 2001).

Zur Analyse der hier vorliegenden Sequenzbereiche in Mensch und Maus wurden, wie oben angeführt, Homologievergleiche mit EST-Datenbanken verwendet. Auf diese Weise konnte die schon publizierte mRNA-Sequenz des humanen Zinkfingergens *ZNF143* (Acc.-Nr. U09850) in den 5´-Bereich hinein verlängert, bzw. die mRNA-Sequenz des orthologen murinen *mStaf*-Gens (Acc.-Nr. AF011758) in den 3´-Bereich erweitert werden (siehe 3.3.1.2.; Abb. 3.8 und Abb. 3.10). Weiterhin fand sich über Homologievergleiche mit der humanen EST-Bank ein Klon (DKFZ564C2163), der die mRNA-Sequenz des humanen *RanBP7*-Gens im 3´-Bereich um ca. 2,4 kb verlängert (siehe 3.3.1.3. und Abb. 3.11), was über Northern-Analyse bestätigt wurde.

Die Maus-EST-Klone, die in der vorliegenden Arbeit über Homologievergleiche der genomischen Maus-Sequenz mit der Maus-EST-Datenbank identifiziert wurden, konnten zur

Überprüfung und Vervollständigung der bisher unbekannten murinen *mRanBP7*-RNA herangezogen werden. Auf diese Weise konnten Teile des Transkriptes aufgeklärt werden, die einer Identifizierung über RT-PCR nicht zugänglich waren (siehe 3.3.1.3. und Abb. 3.12).

(2) Exonvorhersageprogramme

In den letzten Jahren sind zahlreiche Computerprogramme entwickelt worden, die eine Computer-gestützte Genidentifikation erlauben. Für die Entwicklung dieser Programme wurden zwei grundlegende Strategien verfolgt (*Fickett*, 1997): die Suche nach Sequenz-Ähnlichkeiten („lookup"-Methode) und die Suche nach Signalen (Matrizen-Methode). Bei der Suche nach Sequenz-Ähnlichkeiten werden Informationen über publizierte und annotierte DNA- und Protein-Sequenzen sowie EST-Klone benutzt. Daten, die aufgrund der Aufklärung der Sequenz des menschlichen Chromosoms 22 erhoben wurden (*Dunham et al*., 1999), zeigten allerdings, dass nur 50% der hier kodierten Proteine Ähnlichkeiten zu bereits bekannten Proteinen aufweisen, so dass vermutlich nur etwa 50% der bisher noch unbekannten Vertebraten-Gene aufgrund von Sequenz-Ähnlichkeits-Suchen zwischen verschiedenen Phyla detektiert werden können. Bei der wesentlich effektiveren Matrizen-Methode werden Statistiken über kodierende Bereiche mit der Detektion von bestimmten Signalen (Folgen von Nukleotiden) kombiniert. Statistiken über das Auftreten solcher Nukleotidfolgen zeigen Unterschiede in kodierenden und nicht-kodierenden Regionen und sind somit Maße, die Hinweise auf mögliche Protein-kodierende Funktionen geben können. Die Auswertung einer Vielzahl solcher Experimente zeigte, dass die Messung der Häufigkeit des Auftretens von Oligonukleotiden einer Länge von sechs Nukleotiden in einem bestimmten Leserahmen der effektivste Ansatz zur Identifizierung einer „typischen" Exonstruktur ist (*Fickett & Tung* (1992). Diese Methode ist z. B. auch Grundlage des von *Burge* (1997) entwickelten und u. a. in der vorliegenden Arbeit verwendeten Programms GENSCAN.

Bei der Computer-gestützten Identifizierung von kodierenden Bereichen innerhalb einer genomischen Sequenz ist zu beachten, dass diese Programme nicht in der Lage sind, aufgrund alternativer Spleißmechanismen unterschiedliche mRNA-Sequenzen eines einzelnen Gens korrekt zu identifizieren. Dies betrifft jedoch vermutlich mindestens 35% aller menschlichen Gene (*Mironov et al.,* 1999), so dass sich hierbei eine Vielzahl interessanter Genprodukte der *in silico*-Gen-Identifizierung entziehen könnte. Auch ist es diesen Programmen meist nicht möglich, unterschiedliche Gene zu identifizieren, die innerhalb der gleichen genomischen Region lokalisiert sind (*Schulz & Butler*, 1989; *Dunham et al*., 1999*; Cooper et al*., 1998). Darüber hinaus kann die Anwesenheit von Pseudogenen, die in zahlreichen Kopien im ganzen Genom verteilt sind, die Identifikation tatsächlicher Protein-kodierender Regionen mit Hilfe von Exon-Vorhersageprogrammen erschweren. Während vor

wenigen Jahren nur ca. 50% der Exons mit Hilfe von Exonvorhersageprogrammen korrekt identifiziert werden konnten (*Burset & Guigo*, 1996), zeigen die Programme der neuen Generation eine größere Vorhersagegenauigkeit (*Rogic et al*., 2001), die inzwischen bei etwa 80% der Exons liegt (Rogic et al., 2001; *Solovyev*, unveröffentlicht).

(3) Komparative Sequenzanalyse

Die komparative Sequenzanalyse ist eine weitere Methode zur Identifizierung kodierender Bereiche, wenn orthologe genomische Sequenzen vorliegen. Da Exonsequenzen einem hohen Selektionsdruck unterliegen, sind Exons orthologer Gene in verschiedenen nahverwandten Spezies stärker konserviert. Diese konservierten Bereiche lassen sich durch die vergleichende Analyse mit Hilfe des Dotplots oder PIP detektieren (siehe 3.4). Die konservierten transkribierten Bereiche können anschließend über weitere Methoden (*in silico* oder mittels RT-PCR) bestätigt werden.

Zum direkten Vergleich zweier langer genomischer Sequenzen werden in der Regel zwei unterschiedliche Darstellungsweisen gewählt. Bei der von *Gibbs & McIntyre* (1970) eingeführten „Dot-Matrix-Methode“ (auch „Dot-Plot“ genannt) werden die miteinander zu vergleichenden Sequenzen in einer zweidimensionalen Matrix aufgetragen und definierte Gruppen von Nukleotiden der einen Sequenz mit definierten Gruppen der anderen Sequenz überlappend und kontinuierlich miteinander verglichen. Übersteigt die prozentuale Übereinstimmung zwischen beiden Sequenzabschnitten einen bestimmten Schwellenwert, wird die Lokalisation der betreffenden Nukleotidgruppe in dem Koordinatensystem durch einen Punkt angezeigt. Beim Vergleich zweier identischer Nukleotidsequenzen entsteht auf diese Weise eine aus Einzelpunkten bestehende lückenlose Diagonale. Beim Vergleich orthologer Sequenzen aus z. B. verschiedenen Spezies sind konservierte Abschnitte durch auftretende Diagonalen leicht erkennbar. Ein Nachteil dieser Methode bzw. der resultierenden Darstellungsweise besteht jedoch darin, dass Aussagen über den Grad der Sequenzidentität innerhalb der konservierten Bereichen schwierig sind. Bei der Verwendung der Option „DotPlot“ aus dem Programm MegAlign (DNASTAR) wird dieses Problem durch die Verwendung eines Farbcodes gelöst, mit dem die Punkte entsprechend der prozentualen Sequenzidentität markiert werden (siehe auch 3.4.1 und Abb. 3.16).

Eine sehr viel detailliertere Darstellungsweise eines Sequenzvergleiches zweier langer genomischer Sequenzen als der klassische Dot-Plot bietet der „PipMaker“ (*Schwartz et al.,* 2000). Ebenso wie beim Dot-Plot handelt es sich hierbei um ein Computerprogramm, das zwei lange Sequenzen miteinander vergleicht, um konservierte Segmente zu identifizieren. Die graphische Darstellung des Sequenzvergleiches ist ein sogenannter „PIP“ („percentage identity plot“; *Hardison et al*., 1997). Auch hierbei handelt es sich um eine zweidimensionale Matrix, deren X-Achse eine der beiden zu vergleichenden Sequenzen repräsentiert. Die

Sequenzbereiche, die Konservierung mit der Referenz-Sequenz zeigen, werden auf der Y-Achse entsprechend der prozentualen Konservierung markiert. Zusätzlich können auf der X-Achse weitere Informationen, wie die Lokalisation verschiedener Exons eines Gens, repetitive Elemente sowie Bereiche mit hohem GC-Gehalt integriert werden (siehe 3.4.2 und Abb. 3.17 bzw. Abb. 3.18). Hierdurch werden nicht nur erwartete Konservierungen in Exonbereichen, sondern auch in Intron- bzw. Intergen-Bereichen deutlich. Gerade Konservierungen außerhalb kodierender Sequenzen können Anhaltspunkte für die Präsenz z. B. regulatorisch wichtiger Regionen geben. Damit deletierte bzw. inserierte genomische Bereiche in einer der beiden Spezies anhand des PIP erkannt werden können, muss bei jeder PIP-Analyse eine zweifache Darstellung generiert werden. Hier dient jede der zwei untersuchten genomischen Sequenzen einmal als Referenz-Sequenz.
Somit ermöglicht der PipMaker eine sehr viel detailliertere Darstellung des Sequenzvergleiches als der DotPlot und erlaubt dem Benutzer auf diese Weise eine effiziente und schnelle Auswertung der Daten. Dies äußert sich in der ausschließlichen Verwendung dieser Analysemethode in aktuellen Veröffentlichungen (*Hardison et al*., 1997; *Oeltjen et al.,* 1997; *Onyango et al*., 2000; *Wu et al*., 2001).

4.3.2 Identifizierte Transkriptionseinheiten

Neben der Identifizierung bekannter Gene oder EST-Sequenzen über Homologievergleiche wurden in der vorliegenden Arbeit die Genvorhersage-Programme GRAIL 2.0 (*Uberbacher,* 1991); MZEF *(Zhang,* 1997), GENSCAN (*Burge*, 1997) und COMPILE (*Schattevoy*, unveröffentlicht) zur Identifizierung kodierender Bereiche verwendet. Diese Programme sind Bestandteile des Programmpakets RUMMAGE (*Glöckner et al.,* 1998; *Taudien et al*., 2000). Das Programm GRAIL 2.0 konnte in der humanen Sequenz 97 Exons, in der murinen Sequenz 93 Exons identifizieren, diese Anzahl entspricht etwa der mit Hilfe des Programmes MZEF vorhergesagten Exonzahl (Mensch: 100 Exons, Maus: 90 Exons). Die Programme GENSCAN und COMPILE konnten in beiden Spezies deutlich weniger Exons identifzieren: Während GENSCAN im Menschen 71 Exons und in der Maus 51 Exons bestimmte, konnte das Programm COMPILE in der humanen Sequenz 38 Exons bzw. nur 30 Exons in der murinen Sequenz identifizieren. Aufgrund der bekannten Gene konnten die *in silico*-bestimmten Exonbereiche direkt auf ihre Richtigkeit überprüft und somit zur Beurteilung der Vorhersagegenauigkeit der benutzten Programme herangezogen werden. Dabei zeigte sich, dass sich zwischen 89% der bestimmten humanen Exons (COMPILE) und 87% der murinen Exons (COMPILE) bzw. nur 34% der humanen Exons (GRAIL) und 33% der murinen Exons (GRAIL) den bereits bekannten Genpaaren *WEE1*/*Wee1* und *ZNF143*/*mStaf* bzw. *RanBP7* zuordnen ließen. Einen Überblick über die Ergebnisse der Computer-gestützten

Exonvorhersagen für die Genpaare *WEE1/Wee1*, *ZNF143/mStaf* und *RanBP7/mRanBP7* zeigt die Tab. 4.1. Die Exons 1 und 16 (hierbei handelt es sich um das letzte Exon) des humanen und murinen *ZNF143*- bzw. *mStaf*-Gens und das humane Exon 9 sowie das murine Exon 25 (auch hierbei handelt es sich um das letzte Exon des Gens) des *RanBP7/mRanBP7*-Genpaares konnte von keinem der vier verwendeten Vorhersageprogramme identifiziert werden. Prinizipiell ist zu beachten, dass das erste und letzte Exons eines Genes häufig nicht durch Vorhersageprogramme detektiert wird. Da Computerprogramme, die zur Exonvorhersage entwickelt wurden, u. a. nach den typischen konservierten Konsensus-Spleiß-Sequenzen am 5´- und 3´-Ende eines Exons (tcgca**ag**CTGCGA an der Spleißakzeptorstelle bzw. GACAG**gt**cagcaat an der Spleißdonorstelle) suchen, ist die Vorhersage des jeweils ersten bzw. letzten Exons aufgrund der Abwesenheit typischer 5´- und 3´-Konsensus-Spleiß-Stellen schwierig. Bei den vorliegenden Genen wurde zumindest im Fall des humanen und murinen *WEE1*- bzw. *Wee1*-Gens das erste Exon teilweise (Spleissdonor-Stelle) und das letzte Exon des humanen WEE1-Gens vom Programm GENSCAN korrekt erkannt. Auffallend ist, dass das humane Exon 9 des *RanBP7*-Gens von keinem der verwendeten Computerprogramme identifiziert wurde. Dies überrascht, da die Exon-Introngrenzen dieses Exons nicht von der Spleiß-Konsensus-Sequenz abweichen. Das zweite Exon des humanen und murinen *RanBP7*-Gens konnte nur an der Spleißakzeptor-Stelle erkannt werden. Grund dafür ist eine von der typischen Spleißdonorsequenz abweichende Nukleotidabfolge von **gc** statt **gt** in beiden Spezies.

Bei dem Vergleich der Genauigkeit der unterschiedlichen Computerprogramme zur Exonvorhersage wurde zum einen untersucht, wieviel Prozent der anhand der bekannten Gene nachweislich in der Sequenz vorhandenen Exons korrekt, teilweise richtig bzw. nicht vorhergesagt wurden. Zusätzlich wurde ermittelt, welcher Anteil der insgesamt vom Computerprogramm vorhergesagten Exons komplett bzw. teilweise tatsächlich vorhandenen Exons entspricht. Auf diese Weise konnte analysiert werden, wieviel Prozent der insgesamt ermittelten Exons falsch positive Vorhersagen darstellen. Die Daten sind der Tabelle 4.1 dargestellt.

Tab. 4.1: Überblick über die Resultate verschiedener Computerprogramme zur Exon-Identifikation der Gene *WEE1*/*Wee1*, *ZNF143*/*mStaf* und *RanBP7*/*mRanBP7*.

+	Exon korrekt vorhergesagt	(+	5´-Speißstelle nicht korrekt vorhergesagt
-	Exon nicht vorhergesagt	+)	3´-Speißstelle nicht korrekt vorhergesagt

WEE1/Wee1:

Exon # Mensch	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
GRAIL 2.0	-	(+	+	+)	+	+	-	-	+)	-	+)
MZEF	-	+	+	+	+	+)	+	+	+)	+	-
GENSCAN	(+	+	+	-	-	-	-	-	+)	+	+
COMPILE	-	-	+	-	-	-	-	-	+)	+	-

Exon # Maus	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
GRAIL 2.0	(+	-	+	-	+	+)	-	-	-	+)	+)
MZEF	-	+	+	+	+	+	+	+	+	+	+)
GENSCAN	(+	+	+	+	+	+	+	+	+	+	+)
COMPILE	(+	-	+	+	+	+)	-	-	+	+)	+)

ZNF143/mStaf

Exon # Mensch	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
GRAIL 2.0	-	(+	+	+	+	+	+	-	-	-	-	+)	+	(+	(+	-
MZEF	-	(+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
GENSCAN	-	(+	-	+	-	+	+	+	+	+	+	+	+	+	-	-
COMPILE	-	(+	-	+	-	+	+	-	-	-	-	+)	+	(+	+)	-

Exon # Maus	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
GRAIL 2.0	-	-	+	+	-	+	+	+)	-	-	-	+)	(+	+)	-	-
MZEF	-	(+	+	+	(+	+	+	+	+	+	+	+	+	+)	+	-
GENSCAN	-	(+	+	+	-	+	+	+	+	+	+	+	+	+)	+	-
COMPILE	-	-	+	+	-	+	+	+)	-	-	-	+)	(+	+)	(+	-

RanBP7/mRanBP7

Exon # Mensch	1	2 gc	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
GRAIL 2.0	?	-	(+	+	(+	-	-	+	-	+	-	+)	+	-	+	+	-	+	-	+	+	+	+	+	-
MZEF	?	+)	+	+	(+	+	+	+	-	+	+	+	+	+	+	+	-	+	-	+	+	+	+	+	-
GENSCAN	?	+)	+	+	+	-	+	+	-	-	-	+	-	+	+	+	+	+	+	+	+	+	+	+	+)
COMPILE	?	-	+	+	(+	-	+	+	-	-	-	+)	-	+)	+	+	-	+	-	+	+	+	(+	(+	-
XPOUND																									
Exon # Maus	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
GRAIL 2.0	+)	-	(+	+	+	+	-	+	-	+	+	+	+	+	+	+)	-	+	+)	+	+	+)	+	+	-
MZEF	-	+)	(+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	-
GENSCAN	-	-	(+	+	+	+	+	+	-	+	+)	+	-	+	+	+	+	+	+	+	+	+	+	+	-
COMPILE	-	-	(+	+	+	+	-	+	-	+	+)	+	-	+	+	+)	-	+	-	+	+	+)	(+	+	-
XPOUND																									

\+ Exon korrekt vorhergesagt
(+ 5´-Speißstelle nicht korrekt vorhergesagt
+) 3´-Speißstelle nicht korrekt vorhergesagt
\- Exon nicht vorhergesagt

Insgesamt lag der Anteil an korrekt vorhergesagten Exons, bezogen auf die tatsächlich vorhandenen Exons, zwischen 31,37% (COMPILE; *Schattevoy*, unveröffentlicht) und 74,51% (MZEF; *Zhang*, 1997) in der humanen und zwischen 38,46% (COMPILE) und 75% (MZEF) in der murinen Sequenz. Werden die teilweise korrekt vorhergesagten Exons zu dieser Fraktion addiert, bewegt sich der Bereich der richtig vorhergesagten Exons zwischen 50,98% (COMPILE) und 75,31% (MZEF) in der menschlichen Sequenz. In der Maus-Sequenz liegt der Anteil an korrekt und teilweise korrekt vorhergesagten Exons mit Werten zwischen 65,38% (COMPILE und GRAIL 2.0) und 86,54% (MZEF) deutlich höher. Diese Werte entsprechen ungefähr der von *Solovyev* (unveröffentlicht) postulierten Vorhersage-Genauigkeit einzelner Programme von bis zu 80%. Der Anteil an falsch positiv vorhergesagten Exons stellt ein wichtiges Qualitätskriterium der untersuchten Exonvorhersageprogramme dar. Die ersten Computerprogramme, mit deren Hilfe in genomischen Sequenzen mögliche neue Exonbereiche identifiziert werden sollten, waren mit sehr stringenten Parametern ausgestattet, die eine hohe Spezifität, dafür aber niedrige Sensitivität aufwiesen. Die neuere bzw. weiterentwickelte Software dagegen ermöglicht eine sehr viel sensitivere Exonidentifizierung, woraus aber eine deutliche Erhöhung der falsch positiv vorhergesagten Exons resultiert (*Rogic et al*., 2001). Bezogen auf die Summe der in den Genen *WEE1*, *ZNF143* und *RanBP7* vorhandenen 51 Exons im Menschen variieren die Anteile an falsch-positiv bestimmten Exons in der humanen Sequenz zwischen 13,33% (COMPILE) und 66,67% (GRAIL 2.0). In der Maus-Sequenz schwankt die Rate an falsch-positiv vorhergesagten Exons zwischen 10,53% (COMPILE) UND 74,14% (MZEF). Dementsprechend liegen die Raten für falsch-positiv bestimmte Exons bei dem Programm COMPILE erwartungsgemäß am niedrigsten, während das Programm MZEF den höchsten Anteil an falsch-positiv bestimmten Exons ermittelt. Die Verwendung von mehreren Computerprogrammen zur Exonvorhersage ermöglicht allerdings eine wesentlich zuverlässigere Identifizierung tatsächlich vorhandener Exons. In der von *Glöckner et al. (*1998) durchgeführten Analyse von 644 kb genomischer DNA wurden nur solche Exons bewertet, die von mindestens drei der fünf verwendeten Computerprogramme identifiziert wurden. Insgesamt werden bei leicht modifizierter Anwendung der von *Glöckner et al*. verwendeten Parameter (mindestens zwei der vier hier verwendeten Computerprogramme bestimmen die gleiche potenzielle Exonposition richtig) durch die Kombination von vier verschiedenen Computerprogrammen zur Exonvorhersage in der vorliegenden humanen Sequenz 56,86% der vorliegenden Exons korrekt bzw. 73,08% teilweise korrekt (eine der beiden Exongrenzen korrekt identifiziert) erkannt. Entsprechend können in der murinen Sequenz durch diese Methode 68,63% korrekt identifiziert bzw. 84,62% teilweise richtig

identifiziert werden. Diese Werte liegen bei beiden untersuchten Spezies deutlich über der prozentualen Vorhersagegenauigkeit der in dieser Analyse als am wenigsten zuverlässig eingestuften Computerprogramme und reicht an die Vorhersagegenauigkeit der Programme heran, die den höchsten Prozentsatz an korrekt vorhergesagten Exons aufweisen, jedoch auch so stringent arbeiten, dass sie den höchsten Prozentwert an nicht identifizierten Exons besitzen (siehe Tab. 4.2).

Tab. 4.2: Übersicht über die prozentuale Genauigkeit der verwendeten Exonvorhersageprogramme gemessen an dem Verhältnis von tatsächlich vorhandenen zu *in silico*-bestimmten Exons. Die prozentualen Anteile der vorhergesagten Exons in der untersuchten Sequenz sind in schwarz angegeben. Die Anteile der entsprechenden *in silico*-bestimmten Exonfraktionen an den insgesamt ermittelten Exons der einzelnen Programme sind in grau dargestellt.

Exon-Vorhersage-programm Mensch	Summe vorherge-sagter Exons (von 51)	korrekt vorherge-sagte Exons	teilweise korrekt vorherge-sagte Exons	nicht vorherge-sagte Exons	falsch vorherge-sagte Exons
GRAIL 2.0	93	20 (39,22%) (21,51%)	11 (21,57%) (11,83%)	20 (39,22%)	62 (66,67%)
MZEF	90	38 (74,51%) (42,22%)	5 (9,80%) (5,55%)	8 (15,69%)	47 (52,22%)
GENSCAN	51	31 (60,78%) (60,78%)	5 (9,80%) (9,80%)	15 (29,41%)	15 (29,41%)
COMPILE	30	16 (31,37%) (53,33%)	10 (19,61%) (33,33%)	25 (49,02%)	4 (13,33%)

Exon-Vorhersage-programm Maus	Summe vorherge-sagter Exons (von 52)	korrekt vorherge-sagte Exons	teilweise korrekt vorherge-sagte Exons	nicht vorherge-sagte Exons	falsch vorherge-sagte Exons
GRAIL 2.0	97	21 (40,38%) (21,65%)	13 (25%) (13,40%)	18 (34,62%)	63 (64,95%)
MZEF	174	39 (75%) (22,41%)	6 (11,54%) (3,45%)	7 (13,46%)	129 (74,14%)
GENSCAN	74	38 (73,08%) (51,35%)	6 (11,54%) (8,11%)	8 (15,38%)	30 (40,54%)
COMPILE	38	20 (38,46%) (52,63%)	14 (26,92%) (36,84%)	18 (34,62%)	4 (10,53%)

Neben der Genvorhersage über Computerprogramme wurde versucht, Gene aufgrund ihrer evolutionären Konservierung zu identifzieren. Bei der direkten komparativen Analyse der vorliegenden Sequenzen aus Mensch und Maus mit Hilfe des Dotplots konnten insgesamt

282 Bereiche identifiziert werden, die über eine Länge von mindestens 50 Basen eine Identität von mindestens 65% aufwiesen. Alle in den untersuchten Sequenzen vorliegenden bekannten Exons (Mensch: 51 Exons, Maus: 52 Exons) wurden auf diese Weise identifiziert. Innerhalb der konservierten Bereiche jedoch konnten die Exons nicht immer in ihren exakten Exon-Intron-Grenzen bestimmt werden. Grund dafür ist eine Ausdehnung der Sequenzkonservierung in die Intronbereiche hinein. Zusätzlich weisen diese Intronbereiche weitere Positionen auf, die der ag/gt-Regel für potenzielle Spleißstellen genügen würden. Dies ist anhand des Exons 2 des humanen *WEE1*-Gens in Abbildung 4.2 exemplarisch dargestellt.

Abb. 4.2: Beispiel für einen über Dotplot identifizierten konservierten Bereich (*WEE1-/Wee1*-Gen, Exon 2), der sich bis in den Intronbereich hinein erstreckt. Es können mehrere potenzielle Spleiß-Konsensus-Stellen identifziert werden (revers komplementäre Darstellungsweise). Der Exonbereich ist unterstrichen; echte Spleißstellen sind fett in grün, weitere Motive, die der ag/gt-Regel folgen, sind fett in rot hervorgehoben. In dem abgebildeten Sequenzabschnitt wurde die Sequenz des interessierenden Exons 2 des humanen und murinen *WEE1*-Gens in 3´→ 5´-Richtung abgebildet, da diese Orientierung der Genlage innerhalb des übergeordneten Contigs entspricht (siehe Abb. 3.2).

Neben der Identifizierung bisher unbekannter Gene können durch den komparativen Ansatz auch konservierte, nicht-kodierende Sequenzabschnitte identifiziert werden, die möglicherweise regulatorische Funktion haben. Der Nachteil des direkten Sequenzvergleiches besteht jedoch darin, dass Gene, die nur in einer der beiden untersuchten Spezies vorhanden sind, nicht durch diese Methode erkannt werden können.

Die in der vorliegenden Arbeit durchgeführte Kombination von drei Analysemethoden (Homologievergleich mit Datenbankeinträgen, Computer-gestützte Exonvorhersage und komparative Sequenzanalyse) verbessert die Genvorhersagegenauigkeit deutlich, da die aus der komparativen Analyse gewonnenen Daten als Filter bei der Auswertung der über die beiden anderen Methoden erzeugten Ergebnisse dienen und somit die Zahl der falsch vorhergesagten Exons auf ein Minimum reduzieren können. So können Exons aus humanen und/oder murinen ESTs, deren Sequenzen zwischen Mensch und Maus konserviert sind, mit sehr hoher Sicherheit als wahre Exons angesehen werden. Artefakte unter den EST-Sequenzen, wie genomische Kontaminationen, *in vitro*-Ligate (siehe 3.3.1.2) oder unvollständig gespleißte cDNA-Klone können ebenfalls durch einen einfachen Vergleich zu den vorliegenden genomischen Sequenzen detektiert werden. Auch Exons mit einer von der konservierten Spleiß-Konsensus-Sequenz abweichenden Spleißstelle, die sich einer *in silico*-Identifzierung entziehen können, sind aufgrund des vergleichenden Ansatzes zwischen den Spezies eindeutig identifizierbar. So konnte auch das Exon 2 des murinen *mRanBP7*-Gen durch einen Mensch-Maus-Sequenzvergleich sicher erkannt werden.
Es ist abschließend festzuhalten, dass jedes Exon der bekannten Gene *WEE1*, *ZNF143* und *RanBP7* in der humanen Sequenz und der bekannten orthologen Gene in der murinen Sequenz über die hier angewandte kombinierte komparative Untersuchungsmethode sicher bestimmt wurde. Auch die Exons des bisher bei der Maus noch nicht beschriebenen Gens *mRanBP7* konnten auf diese Weise identifiziert werden.

4.3.3 Sequenz- und Organisationsvergleiche der bekannten Gene

Bei der Sequenzanalyse der genomischen Bereiche um das Gen *WEE1*/*Wee1* herum konnten zwei weitere bekannte Gene identifiziert werden. Es handelt sich hierbei um das orthologe Genpaar *ZNF143*/*mStaf* sowie das Gen *RanBP7*, dessen murines Homologes, *mRanBP7*, im Rahmen der vorliegenden Dissertation erstmals ermittelt wurde.

Das humane *WEE1*-Gen konnte 1991 von *Igarashi et al.* als humanes Homologes des *wee1*$^{+}$-Gens aus *Schizosaccharomyces pombe* erstmals beschrieben werden. Das *WEE1*-Gen kodiert für eine 94 kDa große Tyrosin-Kinase, die antagonistisch zu der CDC25-Phosphatase den Übergang eukaryotischer Zellen von der G2-Phase zur Mitose koordiniert. Dies geschieht, indem die nukleäre WEE1-Tyrosin-Kinase den Zellkern durch Phosphorylierung vor dem Einfluss der zytoplasmatisch aktivierten CDC2-Kinase schützt (*Parker et al*., 1992; *Heald et al*., 1993). Es wird vermutet, daß *WEE1* eine Kontrollfunktion in bestimmten Stadien des Zellzyklus und während der Mitose spielen könnte (*Baldin & Ducommun,* 1995). Dementsprechend stellen *WEE1* und sein murines orthologes *Wee1*, das von *Honda et al*. (1995) beschrieben wurde, mögliche Kandidaten-Tumorsuppressor-Gene dar. Die chromosomale Lokalisation des humanen *WEE1*-Gens wurde von *Taviaux & Demaille* (1993) mit Hilfe der FISH-Analyse auf die chromosomale Region 11p15.1-15.3 eingeengt, welche eine enge Kopplung mit Krankheits-assoziierten Genen aufweist.

Das Gen für den Zinkfinger-Transkriptionsfaktor *Staf* („selenocysteine tRNA gene transcription-activating factor“) konnte 1995 von *Schuster et al*. aus *Xenopus* laevis isoliert werden. Transkriptionsfaktoren nehmen Schlüsselpositionen in der Regulation der Proliferation und Differenzierung sowohl normaler als auch krankhafter Zellen ein. Eine verbreitete Klasse dieser Transkriptionsfaktoren gleicht in ihrer DNA-Bindungsdomäne dem Krüppel-Genprodukt von *Drosophila melanogaster*, welches durch die Anwesenheit wiederholter Cys2-His2-Zinkfinger-Domänen charakterisiert ist. Diese Cys2-His2-Zinkfinger-Domänen sind unter Ausbildung der namensgebenden fingerartigen Struktur verkettet. Das *Staf*-Genprodukt bindet an den Promotor des Selenocystein-tRNA-Gens und aktiviert so den RNA-Polymerase III-Promotor des Selenocystein-tRNA-Gens. Das Zinkfingergen *Staf* enthält 7 Zinkfinger und eine saure Aktivierungsdomäne. Durch die Suche mit einem degenerierten Oligonukleotid aus der H/C-Linker-Region in einer humanen Insulinoma-cDNA-Bank konnten *Tommerup et al*. (1993) diverse Zinkfinger-Protein-kodierende Gene, unter anderem das Gen *ZNF143,* identifizieren (*Tommerup & Vissing*, 1995). Die *ZNF143*-cDNA kodiert für ein Protein, das Mitglied der „Krüppel“-Familie ist und weist starke Homologie zu der *Staf*-cDNA von *Xenopus laevis* auf. Unabhängig davon konnten *Rincon et al*. (1998) ebenfalls die *ZNF143*-cDNA isolieren, die das Gen jedoch *SBF* („SPH-binding factor“) nannte. Die Autoren konnten zeigen, dass Antikörper gegen SBF den Zusammenbau nativer SBF-DNA-Komplexe verhindern. *In vitro* stimuliert ZNF143/SBF die Transkription an einem SPH-Element enthaltenden U6-snRNA-Gen-Promotor durch die RNA-Polymerase III. Auch das orthologe murine *ZNF143*-Gen, *mStaf*, wurde kloniert und charakterisiert (*Adachi et al.,* 1998). Das *mStaf*-Gen wurde im Rahmen dieser Arbeit auf dem Maus-Chromosom 7 lokalisiert, *Adachi et al.* (2000) klärten die genomische Organisation auf und führten Promotor-Analysen des

murinen *mStaf*-Gens („mouse selenocysteine tRNA gene transcription-activating factor"; *Adachi et al*. 2000) durch.

Die GTPase RAN („Ras-related nuclear protein") spielt bei dem Energie-verbrauchenden Transport von Proteinen, die ein nukleäres Lokalisations-Signal (NLS) besitzen, durch die nukleären Porenkomplexe (NPCs) eine Schlüsselrolle (*Izaurralde et al.,* 1997; *Stochaj & Rother*, 1999; *Jäkel et al*., 1999). Die RanBP7-Proteine von Mensch und Frosch sind zu ca. 95% identisch und gehören der Superfamilie Ran-bindender Proteine an. Mitglieder dieser Familie weisen, ebenso wie Importin–β, ein charakteristisches N-terminales Sequenzmotiv auf, das für die RanGTP-Bindung verantwortlich ist. Sowohl das RanBP7-Protein von *X.laevis* als auch das des Menschen (*Görlich et al*., 1997) binden über Importin-β an Importin-α. Anhand von Oocyten-Injektionen konnte gezeigt werden, dass es sich bei RanBP7 aus *Xenopus* hauptsächlich um ein zytoplasmatisches Protein handelt, das zwischen Zytoplasma und Nukleus pendelt (*Görlich et al*.; 1997). Versuche mit Fluoreszenz-markiertem RanBP7-Protein zeigten, dass RanBP7 mit Importin-β oder Transportin um Bindungsstellen am NPC kompetiert und zusammen mit diesen unter anderem den nukleären Import von ribosomalen Proteinen und Histon H1 in Säugerzellen vermittelt (*Izaurralde et al.,* 1997; *Jäkel & Görlich*, 1998, *Jäkel et al.,* 1999). In neueren Untersuchungen konnte nachgewiesen werden, dass die *RanBP7*-Transkription in humanen Colorektal-Karzinomen signifikant erhöht ist (*Li et al*., 2000). Dieser Bezug zu einer möglichen Beteiligungen an einer Tumorausbildung sowie die Lokalisation in der Krankheits-assozierten, insbesondere Tumor-assoziierten (*Bepler & Köhler*, 1995) chromosomalen Region 11p15.3 macht das Gen *RanBP7* für weitere Untersuchungen attraktiv. Das murine Orthologe des *RanBP7*-Gens, *mRanBP7*, konnte in dieser Arbeit erstmals beschrieben werden.

In die vorliegende vergleichende Analyse der orthologen Gene wurden die zur Verfügung stehenden cDNA-Daten, aber auch Informationen über die genomische Architektur der identifizierten Gene einbezogen. Die kodierenden Sequenzen der bekannten Gene wurden durch einen Vergleich der genomischen Sequenzen mit den publizierten cDNA-Sequenzen bestimmt. Nach der Überprüfung von konservierten Exon-Intron-Grenzen wurde durch „*in silico*"-Spleißen eine cDNA-Sequenz abgeleitet. Die so ermittelten Daten konnten daraufhin mit Ergebnissen anderer komparativer Untersuchungen orthologer Gene von Mensch und Maus verglichen werden (*Koop* (1995); *Lamerdin et al.* (1996); *Makalowski et. al.* (1996); *Oeltjen et al.* (1997); *Batzoglou et al.* (2000)).

Die identifizierten orthologen cDNA-Sequenzen der drei Genpaare zeigten auf Nukleotid-Ebene eine Übereinstimmung zwischen 85,25 und 92,34% und zwischen 90,1 und 99,19% auf Aminosäure-Ebene (Tab. 4.3). Damit ist der Grad der Konservierung dieser drei orthologen Genpaare höher als der von *Makalowski et al.* (1996) anhand einer komparativen

Analyse von 1196 orthologen Genpaaren aus Mensch und Maus ermittelte Wert, der bei einer durchschnittlichen Nukleotid-Sequenzidentität von ca. 85% lag. Auch die Ähnlichkeiten der entsprechenden Aminosäuresequenzen liegen mit Werten zwischen 90,9 und 99,51% hoch.

Tab. 4.3: Prozentuale Übereinstimmungen der lokalisierten orthologen Gene auf Nukleotid- und Aminosäure-Ebene. *WEE1*, *ZNF143*, *RanBP7*: Mensch; *Wee1*, *mStaf*, *mRanBP7*: Maus

Gen	Identität (nt) %	Identität (AS) %	Ähnlichkeit (AS) %
WEE1 *Wee1*	85,25	90,1	90,9
ZNF143 *mStaf*	88,60	96,97	98,24
RanBP7 *mRanBP7*	92,34	99,19	99,51

Die bisher bekannten cDNA-Sequenzen von *RanBP7*, *ZNF143*, *mStaf* und *WEE1* konnten durch die im Rahmen der vorliegenden Dissertation erstellten Sequenzdaten verifiziert werden. Allerdings zeigten sich an sieben Positionen der aus den vorliegenden Daten ermittelten „*Wee1*“-cDNA-Sequenz Unterschiede zu der von *Honda et al*. (1995) publizierten murinen *Wee1*-cDNA-Sequenz (siehe auch 3.3.1.1). Da fünf dieser Nukleoidaustausche nicht-synonym sind, resultiert die Translation dieser mRNA in einer veränderten Aminosäuresequenz. Eine Übersicht über die prozentualen Konservierungen auf Nukleotid- und Aminosäure-Ebene des humanen *WEE1*-Gens und der von *Honda et al*. (1995) publizierten *Wee1*-Sequenzen sowie des aus der vorliegenden genomischen Maus-Sequenz ermittelten „*Wee1*“-Gens gibt die Tabelle 4.4.

Tab. 4.4: Prozentuale Übereinstimmungen der orthologen WEE1-Gene auf Nukleotid- und Aminosäure-Ebene. *Wee1* bezeichnet die von *Honda et al*. (1995) publizierte Sequenz, „*Wee1*“ bezeichnet die aus der vorliegenden genomischen Sequenz abgeleiteten kodierenden Sequenz für das murine Orthologe des humanen *WEE1*-Gens.

Genpaar	Identität (nt) %	Identität (AS) %	Ähnlichkeit (AS) %
WEE1/*Wee1*	85,25	90,1	90,9
WEE1/„*Wee1*“	85,53	90,84	91,77
Wee1/„*Wee1*“	99,38	99,23	99,38

Da die Konservierung zwischen der von *Honda et al.* (1995) publizierten murinen *Wee1*-cDNA und der aus der vorliegenden Maus-Sequenz abgeleiteten „*Wee1*“-Sequenz auf Nukleotid-Ebene über 99% beträgt, kann ausgeschlossen werden, dass es sich bei dem „*Wee1*“-Gen um ein bisher unbekanntes Homologes des publizierten *Wee1*-Gens handelt. Vielmehr scheint es sich bei den Unterschieden in der Nukleotidsequenz möglicherweise um Stamm-spezifische Polymorphismen der Maus zu handeln. Dafür spricht auch die sehr hohe Konservierung der prozentualen Ähnlichkeit bzw. Identität auf Aminosäure-Ebene zwischen der publizierten und abgeleiteten *Wee1*-Sequenz. Tatsächlich wurde die von *Honda et al.* (1995) publizierte murine *Wee1*-mRNA aus einer Zelllinie des Mausstamms C57BL/6 isoliert, während die Milz einer weiblichen Maus des Stammes 129S6/SvEvTac als DNA-Spender zur Erstellung der verwendeten PAC-Bibliothek #711 (siehe 2.10.2) diente (*Osoegawa et al.,* 2000). Beim Vergleich der „*Wee1“*-cDNA-Sequenz mit Einträgen in der murinen EST-Datenbank konnten vier der in der abgeleiteten cDNA-Sequenz manifesten Sequenzunterschiede durch EST-Sequenzen bestätigt werden. Es zeigte sich jedoch auch, dass 75% dieser identifizierten EST-Klone aus mRNA von Mäusen des Stammes C57BL/6 erstellt worden waren. Somit kann nicht ausgeschlossen werden, dass es sich bei den detektierten Sequenzunterschieden in den cDNA-Sequenzen nicht um Polymorphismen, sondern wahrscheinlich um Fehler in der von *Honda et al.* (1995) publizierten Sequenz handelt.

Die vergleichende Untersuchung der genomischen Sequenzen mit den publizierten cDNAs zeigte bei den drei analysierten murinen und humanen Genpaaren *WEE1*/*Wee1*, *ZNF143*/*mStaf* und *RanBP7*/*mRanBP7* die identische Anzahl und Anordnung von Exons. Dies entspricht den Ergebnissen von *Batzoglou et al*. (2000), der bei der Untersuchung von 117 orthologen Genpaaren aus Mensch und Maus bei 95% dieser Gene eine völlige Konservierung der Exonanzahl feststellen konnte; 73% der untersuchten Genpaare wiesen zudem eine identische Exonlänge auf. Auch die vorliegenden orthologen Genpaare zeigen, unter Ausschluss des jeweils ersten und letzten Exons aufgrund der dort lokalisierten variablen 5´- und 3´-UTRs, eine 100%-ige Konservierung der Exonlängen. Einzige Ausnahme ist das Exon 3 des humanen *WEE1*-Gens, welches eine um 3 bp verlängerte Nukleotidsequenz aufweist. Es kommt somit lediglich zur Insertion einer Aminosäure und nicht zu einer Veränderung des Leserahmens. Die Längen des jeweils ersten und letzten Exons unterscheiden sich jedoch bei allen analysierten Genen häufiger, da die 5´- und 3´-UTRs im Vergleich zum kodierenden Anteil einer mRNA aufgrund des geringeren Selektionsdrucks weniger stark konserviert sind (*Makalowski et al*., 1997; *Graur & Li*, 1999). Sowohl die Genpaare *WEE1*/*Wee1* als auch *ZNF143*/*mStaf* zeigen Größenunterschiede des Exons, in dem sich das Startcodon befindet (*WEE1*/*Wee1*: Exon 1; *ZNF143*/*mStaf*: Exon 2,

siehe Tab.3.3 uns Tab. 3.5); auch die Länge des jeweils letzten Exons ist variabel. Allerdings ist die Länge der Exons bis zum Erreichen des Stop-Codons konserviert. Einzige Ausnahme stellt das *Wee1*-Gen dar, das in der kodierenden Region des letzten Exons der Maus drei Nukleotide, also eine zusätzliche Aminosäure, mehr als beim *WEE1*-Gen des Menschen aufweist.

Obwohl, wie oben aufgeführt, die cDNAs der untersuchten Gene in Mensch und Maus aus Exons aufgebaut sind, deren Länge bis auf wenige Ausnahmen identisch sind, erstrecken sich die jeweiligen humanen und murinen Gene über einen sehr unterschiedlich großen genomischen Bereich. Diese Unterschiede kommen hauptsächlich durch verschieden lange Intronbereiche zustande. Während *Oeltjen et al.* (1997) und *Lamerdin et al.* (1996) bei der Untersuchung von fünf Genen aus der BTK-Region (*BTK*, *FCI-12*, *FTP-3*, *GLA* und *L44L*) bzw. der *ERCC2*-Genregion im Menschen kleinere Introns als in der genomischen Maus-Sequenz entdecken konnten, zeigen die Daten von *Bahr* (1999) die umgekehrte Situation. Bei den hier vorliegenden Daten läßt sich keine einheitliche Aussage über generell größere Intronlängen bei einer der beiden Spezies machen. Die genomische Ausdehnung des humanen *WEE1*-Gens ist mit knapp 15 kb nur um ca. 5 kb kleiner als die des murinen orthologen *Wee1*-Gens (20 kb). Dagegen erstrecken sich die für *ZNF143* und *RanBP7* kodierenden Regionen (66 kb bzw. 51 kb) über einen deutlich größeren genomischen Bereich als die orthologen Maus-Gene *mStaf* und *mRanBP* (33 kb bzw. 39 kb). Die größte Differenz in der Intronlänge beider Spezies findet sich in Intron 7 des *ZNF143*- bzw. *mStaf*-Gens. Sie beträgt im Menschen knapp 14 kb und resultiert aus der stark erhöhten Präsenz von SINEs. So weist das humane Intron 7 insgesamt 30 SINEs auf, während sich im entsprechenden Intron der Maus nur 2 SINEs befinden. Abschließend kann festgestellt werden, dass die unterschiedlichen Intronlängen hauptsächlich aus der Integration von repetitiven Elementen nach der Aufspaltung der Entwicklungslinien von Mensch und Maus vor ca. 80 Millionen Jahren (*Graur & Li*, 1999) zu resultieren scheinen (siehe auch 4.3.6).

4.3.4 Sonstige konservierte Sequenzbereiche

Bei der vergleichenden Sequenzanalyse der vorliegenden humanen und murinen Nukleotidsequenz zeigen sich Konservierungen, die sich nicht eindeutig Gen-Bereichen zuordnen lassen. Bisherige vergleichende Studien zwischen Mensch und Maus haben unterschiedliche Ergebnisse hinsichtlich der Konservierung von kodierenden und nicht-kodierenden Sequenzen ergeben (*Lamerdin et al*., 1996; *Oeltjen et al*., 1997; *Ansari-Lari et al*., 1998; *Onyango et al*., 2000; *Jang et al*., 2000; *Wu et al*., 2001). Während beim β-*Globin*-Cluster und der *ERCC2*-Region so gut wie keine konservierten Abschnitte außerhalb der kodierenden Bereiche vorhanden sind, zeigen der *T-Zell-Rezeptor*-Locus, der *BTRK*-Locus

und eine genreiche Region von Chromosom 12p13 Konservierung in Intergen- und Intronbereichen. Beim *Immunglobulin*-Locus (*J-Cμ-Cδ*) sind konservierte und nicht-konservierte Regionen in den Intergenbereichen sogar direkt benachbart. Diese Daten legen die Vermutung nahe, dass verschiedene Abschnitte des Genoms mit unterschiedlichen Geschwindigkeiten evolvieren (*Koop*, 1995; *Hardison et al.*, 1997).

Der hier durchgeführte Vergleich der ermittelten Sequenzen mittels des Programms DOTPLOT zeigt neben den erwartungsgemäß hoch-konservierten Exonbereichen der besprochenen Gene noch weitere konservierte Abschnitte in den Intergen-Bereichen (siehe 3.4.3). Funktionelle Untersuchungen dieser Regionen liegen nicht vor. Es könnte sich dabei möglicherweise um nicht-Protein-kodierende Gene handeln, die mit den derzeitigen Analysemethoden nicht detektiert werden können. Auch eine mögliche Funktion solcher konservierter Sequenzbereiche bei der Aufrechterhaltung der chromosomalen Struktur kann nicht ausgeschlossen werden (*Koop*, 1995). Allerdings konnte keine der konservierten Regionen über Computer-gestützte Analysen als eine potenzielle „matrix attachment region“ (MAR) eingestuft werden. Vermutungen, dass es sich bei den beobachteten konservierten Bereichen um Promotor-Regionen handeln könnte, konnten durch Computeranalysen (siehe 3.4.3) nicht bestätigt werden. Da bisherige Erkenntnisse eine Sequenzkonservierung generell mit Regionen in Verbindung bringen, denen eine funktionelle, z. B. regulatorische Aufgabe zugeordnet wird (*Hardison et al.*, 1997), könnten Funktionsanalysen im Mausmodell, bei denen die konservierten Sequenzen gezielt ausgeschaltet werden, Aufschluss über die Bedeutung dieser konservierten Bereiche geben.

Vermutungen, dass im konservierten Bereich 3 des Menschen (siehe 3.4.3) ein noch nicht bekanntes Gen lokalisiert sein könnte, konnten nicht eindeutig bestätigt werden. Dieser Bereich zeigt starke Homologien zu dem humanen EST-Klon te51e10.x1 (Acc.-Nr.: AI539442), welcher einen offenen Leserahmen einer Länge von 186 bp als auch einen erhöhten GC-Gehalt aufweist (siehe 3.4.4). Dies scheint auf die Anwesenheit eines 5´-Genbereiches hinzudeuten. Allerdings konnte von den verwendeten Promotor-Vorhersage-Programmen kein potenzieller Promotor in der betreffenden Region überzeugend lokalisiert werden. Weiterhin wurde kein muriner EST-Klon identifiziert, der zu dem EST-Klon te51e10.x1 homolog wäre. Dies könnte darauf hinweisen, dass es sich bei dem EST-Klon te51e10.x1 um eine genomische Kontamination handeln könnte. Da sich jedoch die durch vergleichende Sequenzanalyse detektierbaren Sequenzkonservierungen als zuverlässiger Anzeiger bei der Genidentifizierung erwiesen haben (siehe 4.3.1), müssten zusätzliche Expressionsanalysen durchgeführt werden, um die Lokalisation eines Gens in diesem Bereich eindeutig bestätigen zu können.

4.3.5 Analyse des GC-Gehaltes und der CpG-Inseln

GC-reiche bzw. GC-arme Regionen in Säuger-Genomen wurden ursprünglich durch Experimente wie z. B. die Dichtegradientenzentrifugation dargestellt. Hierbei wurde ersichtlich, dass große DNA-Fragmente deutliche Abweichungen vom durchschnittlichen GC-Gehalt aufweisen können. Weiterführende Studien zeigten, dass diesen GC-armen bzw. –reichen Regionen unterschiedliche biologische Eigenschaften zugeordnet werden können, wie z. B. die Gendichte, der Anteil an repetitiven Sequenzen, die Zuordnung zu zytogenetischen Banden sowie die Rekombinationsrate (*Saccone et al*., 1992; *Saccone et al*., 1993; *Duret et al*., 1995; *Gardiner*, 1996; *Bernardi et al.*, 1985; *Bernardi*, 2000). Es wird angenommen, dass Vertebratengenome in der Regel einen durchschnittlichen GC-Gehalt von ca. 40% aufweisen, der allerdings nicht gleichmäßig über das Genom verteilt ist. Daher wird postuliert, dass sich die Vertebratengenome aus einem Mosaik sogenannter Isochoren (*Bernardi*, 1985) zusammensetzen. Als Isochoren werden längere Bereiche (200-1300 kb) bezeichnet, die sich im Hinblick auf ihre Basenzusammensetzung homogen verhalten und eine spezifische Gendichte aufweisen (*Bernardi*, 2000). Isochoren lassen sich in eine kleine Anzahl von Gruppen unterteilen, die einen spezifischen GC-Anteil aufweisen. Im humanen Genom werden fünf verschiedene Isochoren-Klassen unterschieden, wobei hier zwischen den GC-armen Klassen L1 und L2 und den GC-reichen Klassen H1, H2 und H3 differenziert wird (*Bernardi*, 1995; *Gardiner*, 1996), die auch deutliche Unterschiede bezüglich ihrer Gendichte zeigen (siehe Tab. 4.5).

Tab. 4.5: Überblick über die einzelnen Isochorenklassen von Säugern bezüglich des GC-Gehaltes, des Genomanteils und der Gendichte (nach *Gardiner*, 1996 und *IHGSC*, 2001). Die Werte für die Gendichte beziehen sich auf eine geschätzte Zahl von insgesamt 75 000 Genen.

Isochor	L1	L2	H1	H2	H3
GC-Gehalt (%)	<38	38-42	42-47	47-52	>52
Genomanteil (%)	30	32	21	10	3
Gendichte (pro kb)	1/75	1/75	1/37,5	1/37,5	1/4,8

Während die Gendichte (bei einer angenommenen Gendichte von 75 000 Genen *(Gardiner*, 1996)) in den GC-armen Isochoren L1 und L2 mit etwa einem Gen pro 75 kb sehr niedrig ist, steigt sie in den GC-reichen Isochoren entsprechend zum sich erhöhenden GC-Gehalt von

einem Gen pro 37,5 kb bei H1 und H2 auf knapp 5 kb im Isochor H3 an. Im Gegensatz zu anderen Säugergenomen weist das murine Genom eine zum humanen Genom deutlich unterschiedliche Isochoren–Struktur auf. So finden sich in der Maus aufgrund einer homogeneren GC-Verteilung weniger Isochoren-Klassen, woduch ein Äquivalent für die humane Klasse H3 völlig fehlt und die Unterschiede in der Klasse L1 und H2 deutlich schwächer ausgeprägt sind (*Bernardi*, 1995).

Die in der vorliegenden Arbeit untersuchten Genomabschnitte weisen beim Menschen einen durchschnittlichen GC-Gehalt von 42,73% bzw. von 42,76% in der Maus auf. Der GC-Gehalt der humanen Sequenz ist im Vergleich zum durchschnittlichen GC-Gehalt des Menschen nur leicht erhöht und lässt sich, ebenso wie der GC-Gehalt der Maus, der Isochoren-Klasse H1 zuordnen. Auf der Grundlage von drei identifizierten Genen liegt die Gendichte des untersuchten humanen Genomabschnitts bei einem Gen pro 81 kb, welcher eine identifizierte Gendichte von einem Gen pro 61 kb in der murinen Sequenz gegenübersteht. Die Gendichte in den analysierten Sequenzen entspricht somit eher der GC-armen Isochoren-Klasse L2. Aktuelle Untersuchungen der vorläufigen gesamtgenomischen Sequenz des Menschen (*IHGSC*, 2001; *Venter et al*., 2001) konnten die Präsenz der postulierten, in sich homogenen Isochoren-Klassen allerdings nicht uneingeschränkt bestätigen. Hierzu wurde die vorliegende Sequenz in 300 kb-große Abschnitte („Fenster“) unterteilt, welche wiederum in 20 kb große Unterabschnitte („Unterfenster“) aufgegliedert wurden. Der durchschnittliche GC-Gehalt jedes Fensters und Unterfensters wurde berechnet und daraufhin untersucht, inwieweit die Varianz des GC-Gehaltes eines Unterfensters duch den durchschnittlichen GC-Gehalt jedes einzelnen Fensters erklärt werden kann. Dabei zeigte sich, dass ca. 3/4 der durchschnittlichen genomischen Varianz im GC-Gehalt der kleinen 20 kb-Fenster durch statistische Abweichungen im GC-Gehalt der großen 300 kb-Fenster erklärt werden kann. Die verbleibende Varianz zwischen den kleinen Fenstern allerdings war zu groß, um die Hypothese der homogenen Verteilung von GC-Anteilen in den postulierten Isochoren-Klassen über das Genom hinweg zu unterstützen (*IHGSC*, 2001). Ähnliche Ergebnisse wurden auch bei der Verwendung von anderen Fenster- und Unterfenstergrößen erzielt. Es ist somit fraglich, ob die bisher recht strikte Vorstellung, dass es sich bei Isochoren um streng homogene Bereiche des gleichen GC-Gehaltes handelt, durch die aktuellen Daten uneingeschränkt aufrecht erhalten werden kann. Es ist darüberhinaus vorstellbar, dass die experimentell beobachtete und postulierte Heterogenität des GC-Gehaltes auch durch Insertionen transposabler Elemente verursacht sein könnte (*IHGSC*, 2001), da diese Elemente typischerweise einen höheren GC-Gehalt als die umgebende genomische Sequenz aufweisen. Andererseits ist deutlich, dass das Genom große Regionen mit distinktem GC-Gehalt aufweist, so dass auch über eine Modifikation der

bisherigen Isochoren-Definition nachgedacht werden könnte, mit der sich das Genom in definierte Bereiche einteilen ließe.

Ein in Verbindung mit dem GC-Gehalt häufig diskutierter Aspekt der Architektur genomischer Bereiche ist das Vorkommen von CpG-Inseln. Untersuchungen von *Cooper et al*. (1983) haben gezeigt, dass in Vertebraten der Großteil der CpG-Dinukleotide des Genoms methyliert vorliegt. Nicht-methylierte Sequenzen, die nur ca. 1% des Genoms ausmachen, sind meist auf kurze CpG-Inseln reduziert. Das Dinukleotid CpG tritt im menschlichen Genom nicht in der erwarteten Häufigkeit von ca. 4% auf (Multiplikation der typischen Fraktionen an Cytosinen und Guanosinen, also 0,21 x 0,21), sondern ist mit einem Auftreten von nur ca. 1/5 der erwarteten Häufigkeit deutlich unterrepräsentiert. Dieses Defizit resultiert daraus, dass die meisten CpG-Dinukleotide, welche das methylierte Nukleotid Cytosin aufweisen, spontanen Desaminierungen unterworfen sind. Diese führen zu einer Umwandlung des Cytosins zu Thymidin, so dass methylierte CpG-Dinukleotide häufiger zu TpG-Dinukleotiden mutieren (*Holliday & Pugh*, 1975). CpG-Inseln sind bei der Analyse von genomischen Nukleotidsequenzen von besonderem Interesse, da sie zumeist mit dem 5´-Bereich von Genen assoziiert sind (*Bird*, 1985; *Larsen et al*., 1992) und sich hier häufig vor der transkribierten Region befinden (*Gardiner-Garden & Frommer*, 1987). Sie können aber auch im ersten oder zweiten Exon eines Gens gefunden werden (*Antequera & Bird*, 1993). Zusätzlich gibt es experimentelle Hinweise darauf, dass die Methylierung dieser CpG-Inseln mit einer Gen-Inaktivierung korreliert ist (*Cross & Bird*, 1995) und bei der gewebsspezifischen Expression (*Grunau et al*., 2000) sowie bei der allelspezifischen Expression (genomisches Imprinting) wichtig zu sein scheint (*Onyango et al*., 2000; *Grunau et al*., 2000). Die Anzahl der CpG-Inseln im menschlichen Genom wurde auf ca. 30 000 bis 45 000 geschätzt (*Bird*, 1987; *Antequera & Bird*, 1993). *Larsen et al*. (1992) sowie *Gardiner-Garden & Frommer* (1987) benutzten eine Computer-gestützte Analyse, um CpG-Inseln zu identifizieren. Hierzu wurden die CpG-Inseln als mindestens 200 bp lange Sequenzbereiche definiert, die einen GC-Gehalt von über 50% aufweisen, wobei das Verhältnis der beobachteten zur erwarteten Häufigkeit des CpG-Dinukleotids größer gleich 0,6 sein sollte. Es zeigte sich, dass ein direkter Vergleich von experimentellen und Computer-gestützten Bestimmungen von CpG-Inseln sehr schwierig ist. Das liegt zum einen daran, dass Computer-gestützte Analysen den Methylierungsstatus des Cytosins nicht beachten können, und zum anderen daran, dass experimentelle Ansätze nicht automatisch genomische Regionen mit hohem GC-Gehalt für die Untersuchungen heranziehen können.

Die Zahl der CpG-Inseln wurde im murinen Genom bei gleicher Genanzahl auf nur ca. 37 000 angenommen (*Antequera & Bird*, 1993). Somit weist die Maus etwa 18% weniger CpG-Inseln auf als der Mensch. Nach einer Hypothese von *Antequera & Bird* (1993) ist dieser Unterschied auf den Verlust von CpG-Inseln im murinen Genom durch eine verstärkte

Methylierung in der Keimbahn und eine nachfolgende Mutation von CpG zur TpG durch Desaminierung (s. o.) zurückzuführen, wobei nur gewebsspezifisch exprimierte Gene von diesem Verlust betroffen zu sein scheinen. *Cross & Bird* (1995) fanden auch im humanen Genom Anhaltspunkte für den Verlust von CpG-Inseln. Allerdings scheint es sich hierbei um einen langsameren Prozess zu handeln als im Genom der Maus. Die Analyse der in der vorliegenden Arbeit erstellten humanen und murinen genomischen Sequenz zeigt allerdings keinen deutlichen Verlust an CpG-Inseln in der murinen Sequenz, da sowohl in der humanen als auch in der murinen Sequenz drei CpG-Inseln über Computer-Analyse identifiziert werden konnten (siehe auch Tab. 4.6 und 3.4.4).

Von den jeweils in beiden Spezies identifizierten drei CpG-Inseln sind je zwei CpG-Inseln mit dem 5´-Bereich der Gene *WEE1*/*Wee1* bzw. *ZNF143*/*mStaf* assoziiert. Die dritte CpG-Insel in der murinen Sequenz ist im 5´-Bereich des Gens *mRanBP7* lokalisiert, welcher in der untersuchten humanen Sequenz nicht enthalten ist, da der Klon PAC-180B11 das erste Exon des *RanBP7*-Gens nicht enthält. Allerdings weist der Bereich des humanen Introns 2 von *RanBP7* auch einen erhöhten CpG-Gehalt auf (56,2 % CpG über eine Länge von 308 bp), der aufgrund der kurzen Ausdehnung die Kriterien für eine CpG-Insel nicht optimal erfüllt. Die dritte CpG-Insel innerhalb der hier analysierten humangenomischen Sequenz findet sich in dem zwischen Mensch und Maus konservierten Bereich 2 (siehe 3.4.3), in dem auch ein humaner EST-Klon identifiziert werden konnte (siehe 4.3.4) und in dem möglicherweise ein bisher unbekanntes Gen lokalisiert sein könnte.

Tab. 4.6: Überblick über die durchschnittliche Länge und den prozentualen GC-Gehalt der CpG-Inseln, die in den ermittelten Sequenzen identifiziert wurden.

	Mensch	**Maus**
Anzahl der CpG-Inseln	3	3
durchschnittl. Länge (bp)	1443	1181
durchschnittl. GC-Gehalt (%)	65,6	67,4

Es zeigt sich, dass die durchschnittliche Länge der in der humanen Sequenz detektierten CpG-Inseln mit 1443 bp deutlich länger ist als in der murinen Sequenz (1181 bp). Die Längen der CpG-Inseln liegen jedoch im Rahmen der vom *IHGSC* (2001) erhobenen Daten, nach denen mehr als 95% der CpG-Inseln kürzer als 1800 bp sind. Der durchschnittliche GC-Gehalt innerhalb der identifizierten CpG-Inseln liegt mit 67,4% in der Maus höher als beim Menschen (65,6%). Dies korreliert auch mit dem in der Maus detektierten leicht erhöhten GC-Gehalt der untersuchten genomischen Region. Die in der humangenomischen

Sequenz identifizierten CpG-Inseln weisen einen GC-Gehalt auf, der dem kürzlich ermittelten Durchschnittswert von 60-70% entspricht (*IHGSC*, 2001). Allerdings widerspricht die Differenz im GC-Gehalt der CpG-Inseln in Mensch und Maus Ergebnissen aus früheren Untersuchungen, wonach der Mensch einen ca. 3% höheren CG-Gehalt in den CpG-Inseln aufweisen sollte als die Maus (*Matsuo* et al., 1993; *Antequera & Bird*, 1993). Auch die Daten von *Cross et al*. (1997), die zeigten, dass die CpG-Inseln des Menschen um ca. 30% länger als die der Maus sind, konnten durch die vorliegenden Daten nicht bestätigt werden. Allerdings sollte hierbei berücksichtigt werden, dass die Anzahl von drei identifizierten CpG-Inseln in den beiden analysierten genomischen Sequenzen keinen repräsentativen Durchschnitt darstellen.

4.3.6 Repetitive Elemente

Auf der Basis der aus dem Humangenom-Projekt stammenden Daten wird der Anteil repetitiver Sequenzen am menschlichen Genom auf 36% bis 50% geschätzt (*Venter et al*., 2001; *IHGSC*, 2001). Allerdings wird der von *Venter et al.* mit 36% angegebene Anteil repetitiver Sequenzen am Gesamtgenom aufgrund der Assemblierungsstrategie, in der repetitive Sequenzen unterrepräsentiert waren, möglicherweise als zu niedrig eingeschätzt. Der Großteil der humanen repetitiven Anteile (45%) stammt von transposablen Elementen ab (Smit, 1999; *IHGSC*, 2001). Bei Säugern lassen sich fast alle transposablen Elemente bzw. repetitiven Sequenzen, die aus solchen entstanden sind, einer der folgenden vier Klassen zuordnen: LINE-, SINE- und LTR-Elemente sowie DNA-Transposons.

LINEs

Bei den entwicklungsgeschichtlich sehr alten LINEs („long interspersed nuclear element"), die mit einem Anteil von 21% im humanen Genom von allen repetitiven Elementen am stärksten vertreten sind, handelt es sich um ca. 6-8 kb lange Transposons, die einen internen Polymerase II-Promotor besitzen und zwei offene Leserahmen aufweisen. Die Insertionsstellen der LINEs werden von einer „target site duplication" von 7-20 bp flankiert. L1-Elemente weisen einen erhöhten AT-Gehalt auf und sind nicht gleichmäßig über das humane Genom verteilt, sondern befinden sich gehäuft in AT-reichen Regionen, in welchen einige niedrig exprimierte Gene lokalisiert sind. Das gehäufte Vorkommen von LINEs in AT-reichen Regionen ist auch im Maus-Genom zu beobachten (*Cross et al*., 1997; *Smit* 1999).

SINEs

Bei den SINEs („short interspersed nuclear element") handelt es sich um ca. 100-400 bp kurze transposable Elemente, die einen internen Polymerase III-Promotor besitzen. Sowohl

humane als auch murine SINEs zeichnen sich durch einen erhöhten GC-Gehalt aus und finden sich verstärkt in Gen-dichten GC-reichen Regionen des Genoms. Es wird vermutet, dass diese nicht zur autonomen Transposition fähigen mobilen Elemente die Transpositions-Maschinerie der LINEs zur Verbreitung innerhalb des Genoms verwenden. Tatsächlich zeigt sich, dass die meisten SINEs das 3´-Ende mit einem immobilen LINE-Element teilen (*Okada et al.*, 1997). Die Promotorregionen von allen bekannten humanen SINEs stammt von tRNA-Sequenzen ab. Nur eine einzige monophyletische SINE-Familie, die Alu-Elemente, hat ihren Ursprung in der 7SL-RNA und teilt sich ihr 3´-Ende auch nicht mit einem LINE-Element. Die Alu-Elemente machen 13% des humanen Genoms aus. Den Alu-Elementen des humanen Genoms entsprechen im murinen Genom die B1-Elemente.

LTR-Elemente

Die LTR-Retroposons sind durch flankierende, terminale direkte Sequenzwiederholungen („long terminal repeat“) gekennzeichnet und enthalten z. T. die für die Transkription und Transposition notwendigen regulatorischen Elemente. Entgegen Schätzungen von *Smit* (1996), der den Anteil von LTR-Elementen am menschlichen Genom auf ca. 4,6% schätzte, legen neue Untersuchungen des *IHGSC* (2001) einen LTR-Anteil von ca. 8% nahe.
Die autonomen LTR-Elemente (Retrotransposons), zu denen z. B. die Retroviren zählen, enthalten die Gene *gag* und *pol*. Die Transposition erfolgt durch einen retroviralen, durch eine tRNA initiierten Mechanismus der reversen Transkription. Obwohl eine Vielzahl von LTR-Retrotransposons existieren, scheinen nur die Vertebraten-spezifischen endogenen Retroviren (ERVs), die sich in drei Klassen (ERV I – III; siehe auch 3.4.5.) unterteilen lassen, im Säuger-Genom aktiv gewesen zu sein. Etwa 85% der von LTR-Retroposons abstammenden Überreste bestehen nur noch aus einem isolierten LTR; die interne Sequenz wurde durch homologe Rekombination zwischen flankierenden LTRs verloren. Zu den nicht-autonomen LTR-Elementen, deren Länge sich mit 1,5 bis 3 kb deutlich von den ca. 6-11 kb großen autonomen LTR-Elementen unterscheidet, zählen die MaLR- („mammalian LTR-retrotransposon“) sowie die MER4-Elemente („medium reiteration frequency interspersed repeat“).

DNA-Transposons

Die Gruppe der DNA-Transposons ähnelt bakteriellen Transposons und weist charakteristische terminale invertierte Sequenzwiederholungen auf. Die autonomen DNA-Transposons kodieren für eine Transposase, die in der Nähe der invertierten Wiederholungen bindet und die Transposition durch ein Ausschneiden und Integrieren der betroffenen Sequenz an einer anderen Stelle im Genom bewerkstelligt. Das menschliche Genom enthält mindestens sieben Hauptgruppen an DNA-Transposons, die in verschiedene

Unterfamilien mit unterschiedlicher evolutionärer Abstammung unterteilt werden können (*Smit*, 1996). Typische Vertreter der DNA-Transposons, die nach neuen Schätzungen ca. 2,8% des humanen Genoms ausmachen (*IHGSC*, 2001; *Venter et al*., 2001), sind die Mariner-ähnlichen Elemente sowie die MER1- und MER2-Elemente.

4.3.6.1 Verteilung repetitiver Elemente in den untersuchten Sequenzbereichen von Mensch und Maus

Die im Rahmen der vorliegenden Dissertation ermittelten Sequenzen aus der humanen Chromosomenregion 11p15.3 sowie des orthologen murinen Bereiches von Chromosom 7 wurden mit Hilfe des Programmes REPEATMASKER auf die Anwesenheit und prozentuale Verteilung von repetitiven Elementen untersucht. Der Gesamtanteil an repetitiven Elementen liegt in der untersuchten humanen Sequenz bei 55,26% und entspricht somit dem durchschnittlichen Vorkommen repetitiver Elemente im Gesamtgenom von wahrscheinlich über 50%. Die aus der humanen Sequenz von 11p15.3 ermittelten Ergebnisse unterscheiden sich trotzdem stark von den Daten, die aus der vorläufigen Gesamtsequenz des Menschen erhoben wurden (*Venter et al.,* 2001; *IHGSC*, 2001; siehe Tab. 4.5). Ebenso wie bei den Untersuchungen von *Martindale et al.* (2000), *Onyango et al.* (2000) und *Bahr* (1999) lag der Anteil repetitiver Elemente in dem hier untersuchten murinen Genomabschnitt unter den im Menschen ermittelten Werten. Hierbei zeigten sich Differenzen am Anteil repetitiver Elemente zwischen der murinen und humanen Sequenz von 7% (*Bahr*, 1999; *Ansari-Lari et al*., 1998), über 16% (*Oeltjen et al.,* 1997) und bis zu 18% (*Martindale et al*., 2000).

Tab. 4.7: Überblick über das prozentuale Vorkommen repetitiver Elemente in der vorläufigen menschlichen Genomsequenz sowie in den hier untersuchten humanen und murinen Genombereichen (*IHGSC*, 2001; *Venter et al*., 2001). Details siehe Text.

Repetitive Elemente	LINEs (%)	SINEs (%)	LTR-Elemente (%)	DNA-Elemente (%)	andere (%)
vorl humangenom. Sequenz	20,42	13,14	8,29	2,84	0,14
Mensch (243966 bp)	14,58	34,38	2,37	2,74	1,23
Maus (192519 bp)	2,37	27,37	8,40	0,93	2,81

Die vorliegende Analyse ergab, dass ca. 34% der untersuchten humangenomischen Sequenz aus SINEs besteht. Dieser Wert liegt deutlich über dem erwarteten Durchschnittswert von ca. 13%. Hierbei ist besonders auffällig, dass der Anteil an SINEs im

Klon PAC-180B11 mit 39,8% noch deutlich höher ist als im Klon PAC-142M6 (29,51%). Dieser hohe Anteil an SINEs könnte auch der Grund für den deutlich schwierigeren Zusammenbau der Einzelsequenzen von Klon PAC-180B11 nach der „shot gun"-Sequenzierung im Vergleich zum Klon PAC-142M6 sein. Bisher vorliegende Daten haben nur in einem Fall (*Martindale et al*., 2000) einen höheren Anteil (45,4%) an SINEs aufgewiesen. Während *Ansari-Lari et al.* (1998) für die Region 12p13 einen SINE-Anteil von 21,77% ermitteln konnte, der ebenfalls deutlich über dem angenommenen Durchschnittswert von 13% liegt, konnten *Onyango et al*. (2000) bei der Analyse eines knapp 1 Mb großen Bereiches der Region 11p15 jedoch nur 5,25% der untersuchten humanen Sequenz als SINEs identifizieren. Entsprechend dem geringeren Auftreten repetitiver Elemente liegt der Anteil von SINEs in der untersuchten murinen Sequenz mit 27,37% deutlich unter dem Anteil des im Menschen analysierten Sequenzabschnitt. Entsprechend den Ergebnissen im Menschen findet sich auch bei den von *Onyango et al.* (2000) publizierten Daten ein auffällig niedriger Anteil an SINEs in der orthologen Region der Maus.

Der durchschnittliche Anteil von LINEs an der vorläufigen humangenomischen Gesamtsequenz beträgt ca. 21% (*IHGSC*, 2001). Während der Anteil an LINEs im PAC-Klon 142M6 mit 20,15% diesem Durchschnittswert entspricht, konnte im Klon PAC-180B11 ein deutlich niedrigeres Vorkommen (8,31%) ermittelt werden. Auch Daten aus vergleichbaren komparativen Sequenzuntersuchungen zeigen ein signifikant geringeres Auftreten von LINEs. So wiesen *Martindale et al.* (2000) und *Ansari-Lari et al.* (1998) einen LINE-Anteil von 2,91% bzw. 2,15% an der untersuchten Sequenz nach. Entsprechend verhält es sich in der murinen Sequenz. Hier pendeln die prozentualen Anteile von LINEs an der jeweils untersuchten Sequenz zwischen 0,85% (*Ansari-Lari et al.,* 1998) und 8,08% (*Onyango et al.,* 2000). Die LINEs stellen eine große Gruppe an interspergierten repetitiven Elementen im Säugergenom dar. LINE-Elemente sind im Gegensatz zu SINE-Elementen zur autonomen Transposition fähig, so dass sich SINEs oft in räumlicher Nähe zu LINE-Elementen befinden, um von deren Transpositionsmaschinerie zu profitieren. Ebenso wie die SINEs haben vermutlich 50% aller L1-Elemente schon vor der Aufspaltung der Säugetiere das Genom besiedelt (*Smit et al*., 1995). Einige L1-Elemente sind immer noch aktiv und transponieren sowohl im humanen als auch im murinen Genom (*Kazazian & Moran*, 1998; *DeBerardinis* et al., 1998). Von den beiden untersuchten Sequenzabschnitten aus der humanen Region 11p15.3 und der orthologen Region auf dem Maus-Chromosom 7 stellen die LINEs 14,58% bzw. 2,37% der Gesamtsequenz dar. Dass der Anteil an LINE-Elementen in humanen Sequenzen deutlich höher ist, konnte auch von *Bahr* (1999) und *Ansari-Lari et al.* (1998) bestätigt werden.

Auch die Anteile der LTRs und DNA-Transposons in der humanen Sequenz zeigen deutliche Abweichungen von den Durchschnittswerten. Während die hier untersuchte Region mit

einem 2,37%-igem Anteil an LTR-Elementen deutlich unter dem Durchschnittswert (8%) liegt, machen die LTR-Anteile vergleichbarer Publikationen zwischen 6,37% (*Onyango et al.,* 2000) und nur 0,9% (*Ansari-Lari et al*., 2000) der untersuchten Sequenzen aus. MER-Elemente, die in der vorliegenden humanen Sequenz 2,74% ausmachen und somit dem durchschnittlichen Wert von 3% entsprechen, lagen in anderen untersuchten humangenomischen Bereichen zwischen Werten von 0,6% und 1,92% (*Ansari-Lari et al*., 1998*; Onyango et al*., 2000). LTR-Elemente sind in der untersuchten murinen Sequenz deutlich stärker vertreten als in der humanen Sequenz, was die Ergebnisse von *Bahr* (1999) bestätigt. Während LTR-Elemente 8,4% der untersuchten Sequenz ausmachen, stellen die MaLR-Elemente 3,6% des analysierten murinen Bereiches dar (siehe auch Tab. 3.11). Die Amplifikation von LTR- und MaLR-Elementen im Genom findet seit der Aufspaltung der Säugetiere statt (*Smit*, 1996). Dabei scheint die Expansion der MaLR-Elemente im murinen Genom weiter voranzuschreiten, während die Anzahl dieser repetitiven Elemente im humanen Genom zu stagnieren scheint (*Smit*, 1993, *Smit*, 1996). Diese Beobachtung lässt sich durch die vorliegenden Daten bekräftigen.

Die Ausbreitung und Integration der Alu- bzw. B-Elemente innerhalb der jeweiligen Genome begann hauptsächlich nach der Aufspaltung der Säugetiere (*Britten et al*., 1988). Die in der vorliegenden Arbeit sequenzierten und analysierten genomischen Bereiche in Mensch und Maus weisen einen unterschiedlichen Anteil an Alu- bzw. B-Elementen auf (Mensch: 33,26%; Maus: 26,73%). Bei diesen Angaben ist zu berücksichtigen, dass sich die humanen Alu-Elemente und die murinen B1-, B2- und B4-Elemente deutlich in ihrer Länge unterscheiden. Somit wird der in der humanen Sequenz ermittelte prozentuale Alu-Anteil durch die Anwesenheit von 305 Alu-Elementen verursacht, während die 393 kürzeren B-Elemente in der Maus-Sequenz nur 26,73% der untersuchten Gesamtsequenz ausmachen. Eine weitere Gruppe an SINE-Elementen wird durch MIR1-Elemente repräsentiert. Diese MIR-Elemente („mammalian-wide interspersed repeats“) sind in allen Säugergenomen vertreten und befinden sich häufig sogar an orthologen Positionen (*Smit & Riggs*, 1995). Ihre hohe Divergenz und die Präsenz an orthologen Stellen in verschiedenen Genomen indizieren, dass MIRs, zumindest teilweise, noch vor der Verbreitung der Säuger amplifiziert wurden (*Jurka et al*., 1995). Die Arbeiten von *Gilbert & Labuda* (1999) zeigen, dass eine 65 bp lange, sogenannte „core“-Sequenz das zentrale Segment dieser MIR-Elemente darstellt. Bei der Analyse der in der vorliegenden Arbeit untersuchten genomischen Bereiche aus Mensch und Maus fällt auf, dass der Anteil an MIR-Elementen deutlich variiert. Während in der gesamten untersuchten humanen Sequenz 19 MIR-Elemente (die einen prozentualen Anteil von 1,12% ausmachen) identifizierbar sind, können in der murinen Sequenz nur drei MIR-Elemente (entspricht 0,14%) lokalisiert werden. Diese Beobachtung wird sowohl durch die Ergebnisse von *Bahr* (1999) als auch von *Ansari-Lari* et al. (1998) bestätigt, die ebenfalls

in den jeweiligen untersuchten humanen Sequenzbereichen deutlich mehr MIR-Elemente detektieren konnten als in den murinen. Möglicherweise hat die in den Nagern vorhandene höhere Mutationsrate dazu geführt, dass sich ein Großteil der vorhandenen MIR-Elemente stark verändert hat und dass diese somit von Computer-gestützten Analyseprogrammen nicht mehr als solche identifiziert werden können.
Auch unter Berücksichtigung der übrigen repetitiven Elemente (SINEs, LINEs, LTRs, DNA-Elemente) wird anhand der vorliegenden Daten deutlich, dass der Anteil der interspergierten repetitiven Sequenzen (LINEs und SINEs) im Mausgenom geringer ist als im Menschen (siehe Tab. 4.7). Als mögliche Erklärung könnte dafür die unterschiedliche Mutationsrate zwischen Mensch und Maus herangezogen werden (*Graur & Li*, 1999), die mit der deutlich kürzeren Generationszeit (*Laird et. al*., 1996) der Nager korreliert. Somit mutieren repetitive Elemente nach ihrer Integration ins Mausgenom wesentlich schneller, was eine Detektion über ein Programm wie z. B. REPEATMASKER erschwert. Ein weiterer Grund für die höhere Mutationsrate im Nagergenom ist auch eine gerinere Effizienz der DNA-Reparatursysteme (*Britten*, 1986).

Abschließend muß die Bedeutung der komparativen Sequenzanalyse unterstrichen werden, die im Zuge der Aufklärung des murinen Genoms, das bis zum Jahre 2003 komplett entschlüsselt sein soll (Battey et al., 1999), weiter zunehmen wird. Hierbei wird, ebenso wie in der vorliegenden Arbeit, nicht nur die Identifizierung von Genen, sondern auch die Identifizierung von regulatorischen bzw. strukturell relevanten Bereichen im Vordergrund stehen. Die hieraus resultierenden Daten stellen, ebenso wie in der vorliegenden Arbeit, die Grundlage für sich anschließende Funktionsanalysen der konservierten Bereiche dar. Die Kenntnis der murinen Sequenz erleichtert hierbei nicht nur die Genidentifizierung im komparativen Ansatz, sondern ermöglicht darüber hinaus die Durchführung wichtiger Funktionsanalysen im Modellorganismus. Besonders durch die Möglichkeiten des gezielten Ausschaltens von Genen, der Erzeugung transgener Tiere sowie der evolutiven Nähe zum Menschen bietet sich die Maus besonders als Modellorganismus an (*Capecchi*, 1989; *Smit et al*., 1995; *Zheng et al.,* 1999).

5 Zusammenfassung

In der vorliegenden Dissertation wurde im Rahmen des Deutschen Humangenomprojektes ein 243 966 bp grosser genomischer Bereich um das humane Gen *WEE1* in der Chromsomenregion 11p15.3 und der 192 519 bp lange orthologe Bereich auf dem murinen Chromosom 7 anhand von PAC-Klonen sequenziert. Der Sequenzierung ging die Erstellung von PAC-Klon-Contigs voraus, welche die zu untersuchenden genomischen Regionen in Mensch und Maus lückenlos abdecken. Nach der Etablierung von Hochdurchsatzmethoden zur Probenherstellung und –verarbeitung wurden die Konsensussequenzen in Mensch und Maus ermittelt. Zur Identifizierung aller Gene wurde die Sequenz einer Kombination von Datenbanksuchen, computergestützten Exonvorhersageprogrammen und der komparativen Sequenzanalyse mit Hilfe von Dotplot- und PIP-Darstellungen unterzogen. In den untersuchten genomischen Regionen der beiden Spezies konnten insgesamt drei orthologe Genpaare (*WEE1*, *ZNF143* und *RanBP7*) und ein humanes Pseudogen (Pseudogen L23a) lokalisiert werden.

Das am Zellzyklus beteiligte *WEE1*-Gen, das auch als Ausgangspunkt für die Isolierung der PAC-Klone zur Erstellung der genomischen Contigs diente, ist sowohl in der humanen als auch in der murinen Sequenz vollständig enthalten. Hierbei konnte die publizierte mRNA-Sequenz des murinen *Wee1*-Gens, unterstützt von EST-Daten, korrigiert werden. Sowohl das *ZNF143*-Gen als auch sein murines Orthologes, *mStaf*, sind in den genomischen Sequenzen vollständig enthalten. Somit muss die in 11p15.4 publizierte Lokalisation des *ZNF143*-Gens in die Region 11p15.3 berichtigt werden. Weiterhin wurde die cDNA-Sequenz des humanen *ZNF143*-Gens um ein bisher noch nicht beschriebenes Exon im 5´-Bereich und die des murinen *mStaf*-Gens um knapp 170 bp im 3´-Bereich verlängert. Der in der *ZNF143*-mRNA-Sequenz publizierte 3´-UTR konnte in der vorliegenden genomischen Sequenz nicht lokalisiert werden. Es scheint sich hierbei um ein von Chromosom 14 stammendes Klonierungsartefakt zu handeln. Das im Menschen beschriebene *RanBP7*-Gen wurde mit Ausnahme des Exons 1 vollständig in der untersuchten genomischen Sequenz lokalisiert. Über Datenbank-Suchen konnte ein EST-Klon identifiziert werden, der die bisher bekannte *RanBP7*-mRNA um knapp 2,4 kb in den 3´-Bereich hinein verlängert. Eine Bestätigung der Transkriptlänge erfolgte über Northern Blot-Analyse. Das bisher unbekannte murine Orthologe, *mRanBP7*, konnte aufgrund komparativer Sequenzanalyse und Datenbanksuchen in der vorliegenden genomischen Maus-Sequenz ermittelt werden, wobei die Sequenz über RT-PCR-Experimente generiert und die Transkriptlänge durch Northern Blot-Analyse bestätigt werden konnte. Neben den drei bekannten Genen konnte in der humanen Sequenz darüber hinaus ein Pseudogen (Pseudogen L23a) identifiziert werden, welches über einen Bereich von 549 bp eine 92%-ige Sequenzidentität zu dem humanen

ribosomalen Protein L23a aufweist und die typischen, 13 bp langen direkten Sequenzwiederholungen besitzt. Acht der insgesamt 10 Nukleotidaustausche führen im Vergleich zu L23a zu einem Aminosäureaustausch, wodurch u. a. ein vorzeitiger Translations-Stop bedingt ist.

Die komparative Sequenzanalyse deckte neben den konservierten Gen-Bereichen zwischen Mensch und Maus insgesamt vier konservierte Bereiche auf. Bei der Analyse dieser Regionen mit Hilfe von EST-Daten bzw. Exonvorhersageprogrammen konnte jedoch keiner dieser vier konservierten Regionen eine eindeutige kodierende Funktion nachgewiesen werden. Es könnte sich hierbei somit um funktionell bedeutsame regulatorische Regionen handeln.

Die Analysen der ermittelten genomischen Sequenzen zeigten, dass der Anteil an repetitiven Elementen mit 55,26% in der untersuchten humanengenomischen Region gegenüber der murinen Sequenz (41,87%) deutlich erhöht ist.

Durch die vergleichende Sequenzanalyse können Artefakte in den EST-analysiert und somit die Zuverlässigkeit der verwendeten Exonvorhersage-Programme optimiert werden.

Die Ergebnisse der vorliegenden Arbeit zeigen, dass die Kombination von komparativer Sequenzanalyse, Datenbank-Suchen und Exonvorhersageprogrammen die Sicherheit bei der Identifikation von kodierenden Sequenzen stark verbessert.

6 Literaturverzeichnis

AARONSON, J. S.; ECKMAN, B.; BLEVINS, R. A.; BORKOWSKI, J. A.; MYERSON, J.; IMRAN, S.; ELLISTON, K. O. (1996): Toward the development of a gene index to the human genome: an assessment of the nature of high-throughput EST sequence date. *Genome Res*. **6**: 829-845

ABI (1996): ABI PRISM, DNA sequencing analysis software, user´s manual. *PE Applied Biosystem*s, Foster City, CA, USA

ADACHI, K.; SAITO, H.; TANAKA, T.; OKA, T. (1998): Molecular cloning and characterization of the murine Staf cDNA encoding a transcription activating factor for the selenocysteine tRNA gene in mouse mammary gland. *J. Biol. Chem*. **273**: 8598-8606

ADACHI, K.; KATSUYAMA, M.; SONG, S.; OKA, T. (2000): Genomic organization, chromosomal mapping and promoter analysis of the mouse selenocysteine tRNA gene transcription-activating factor (mStaf) gene. *Biochem. J*. **346**: 45-51

ADAMS, M. D.; KELLEY, J. M.; COCAYNE, J. D.; DUBNICK, M.; POLYMEROPOULOS, M. H.; XIAO, H.; MERRIL, C. R.; WU, A.; OLDE, B.; MOERNO, R. F.; KERLAVAGE, A. R.; MCCOMIE, W. R.; VENTER, C. J. (1991): Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**: 1651-1656

ADAMS, M. D.; KERLAVAGE, A. R.; FIELDS, C.; VENTER, J. C. (1993): 3400 new expressed sequence tags identify diversity of transcripts in human brain. *Nature Genet*. **4**: 256-267

ADAMS, M. A.; ELNIKER, S. E.; HOLT, R. A.; EVANS, C. A.; GOCAYNE, J. D.; AMANATIDS, P. G.; SCHERER, S. E.; LI, P. W.; HISKINS, R. A.; GALLE, R. F. et al. (2000): The genome sequence of Drosophila melanogaster. *Science* **287**: 2185-2195

ADJAYE, J.; DANIELS, R.; BOLTON, V.; MONK, M. (1997): cDNA libraries from single human preimplantation embryos. *Genomics* **46**: 337-344

ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. (1990): Basic local alignment search tool. *J. Mol. Biol*. **215**: 403-410

ALTSCHUL, S. F.; MADDEN, T. L.; SCHÄFFER, 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. *Nucl. Acids Res.* **25**: 3389-3402

AMID, C.; BAHR, A.; SAMPSON, N.; BIKAR, S.-E.; WINTERPACHT, A.; ZABEL, B. U.; HANKELN, T.; SCHMIDT, E. R. (2001): Comparative genomic sequencing reveals a strikingly similar architecture of a conserved syntenic region on human chromosome 11p15.3 (including gene ST5) and mouse chromosome 7. *Cytogen. Cell Genet*. **93**: 284-290

ANGEL, J. M.; MOORE, J. L.; PELPHREY, A.; RICHIE, R. (1993): The mouse homolog of the rhombotin (Ttg-1) gene maps on chromosome 7 distal to the β-globin (Hbb) locus. *Mamm. Genome* **4**: 281-282

ANSARI-LARI, M. A.; OELTJEN, J. C.; SCHWARTZ, S.; ZHANG, Z.; MUZNY, D. M.; LU, J.; GORRELL, J. H.; CHINAULT, A. C.; BELMONT, J. W.; MILLER, W.; GIBBS, R. A. (1998): Comparative sequence analysis of a gene-rich cluster at human chromosome 12p13 and its syntenic region in mouse chromosome 6. *Genome Res*. **8**: 29-40

ANTEQUERA, F.; BIRD, A. (1993): Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci. USA* **90**: 11995-11999

AVERY, O. T.; MCLEOD, C. M.; MCCARTY, M. (1944): Studies on the chemical nature of the substance inducing transformation in pneumococcal types. *J. Exp. Med*. **79**: 137-158

BAHR, A. (1999): Die molekulare Struktur der Chromosomenregion 11p15.3 des Menschen und des homologen Abschnitts der Maus: Nukleotidsequenz, neue Gene und Interspeziesvergleich. Dissertation im Fachbereich Biologie der Johannes Gutenberg-Universität, Mainz

BAILEY, J. A.; CARREL, L.; CHAKRAVARTI, A.; EICHLER, E. E. (2000): Molecular evidence for a relation between LINE-1 elements and X-chromosome inactivation: The Lyon repeat hypothesis. *Proc. Natl. Acad. Sci. USA* **97**: 6634-6639

BALDIN, V.; DUCOMMUN, B. (1995): Subcellular localization of human wee1 kinase is regulated during the cell cycle. *J. Cell Sci*. **108**: 2425-2432

BANFI, S.; BORSANI, G.; BULFONE, A.; BALLABIO, A. (1997): Drosophila-related expressed sequences. *Hum. Mol. Genet*. 6: 1745-1753

BASSETT JR., D. E.; BASRAI, M. A.; CONNELLY, C.; HYLAND, K. M.; KITAGAWA, K.; MAYER, M. L.; MORROW, D. M.; PAGE, A. M.; RESTO, V. A.; SKIBENS, R. V.; HIETER, P. (1996): Exploiting the complete yeast genome sequence. *Curr. Opin. Genet. & Dev*. **6**: 763-766

BATTEY, J.; JORDAN, E.; COX, D.; DOVE, W. (1999): An action plan on mouse genomics. *Nature Genet.* **21**: 73-75

BATZOGLOU, S.; PACHTER, L.; MESIROV, J. P.; BERGER, B.; LANDER, E. S. (2000): Human and mouse gene structure: Comparative analysis and application to exon prediction. *Genome Res.* **10**: 950-958

BEPLER, G.; KOEHLER, A. (1995): Multiple chromosomal aberrations and 11p allelotyping in lung cancer cell lines. *Cancer Genet. Cytogenet*. **84**: 39-45

BERNARDI, G.; OLOFSSON, B.; FILIPSKI, J.; ZERIAL, M.; SALINAS, J.; CUNY, G.; MEUNIER-ROTIVAL, M.; RODIER, F. (1985): The mosaic genome of warm-blooded vertebrates. *Science* **228**: 953-957

BERNARDI, G. (1995): The human genome: organization and evolutionary history. *Ann. Rev. Genet.* **29**: 445-476

BERNARDI, G. (2000): Isochores and the evolutionary genomics of vertebrates. *Gene* **241**: 3-17

BIRD, A.; TAGGART, M.; FROMMER, M.; MILLER, O. J.; MACLEOD, D. (1985): A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* **40**: 91-99

BIRD, A. P. (1986): CpG-rich island and the function of DNA methylation. *Nature* **321**: 209-213

BIRNBOIM, H. C.; DOLY, J. (1979): A rapid alkaline extraction procedure for screening recombinanat plasmid DNA. *Nucl. Acids Res.* **7**: 1513-1523

BLAKE, J. A.; EPPIG, J. T.; RICHARDSON, J. E.; DAVISSON, M. T.; THE MOUSE GENOME DATABASE GROUP (2000): The mouse genome database (MGD): expanding genetic and genomic resources for the laboratory mouse. *Nucl. Acids Res.* **28**: 1108-1111

BOEHM, T.; BAER, R.; LAVENIR, I.; FORSTER, A.; WATERS, J. J.; NACHEVA, E.; RABBITTS, T. H. (1988): The mechanism of chromosomal translocation t(11;14) involving the T-cell receptor C delta locus on human chromosome 14q11 and a transcribed region of chromosome 11p15. *EMBO J*. **7**: 385-394

BOEHM, T.; FORONI, L.; KANEKO, Y.; PERUTZ, M. F.; RABBITTS, T. H. (1991): The rhombotin family of cysteine-rich LIM-domain oncogenes: distinct members are involved in T-cell translocations of human chromosomes 11p15 and 11p13. *Proc. Natl. Acad. Sci. USA* **88**: 4367-4371

BOGUSKI , M. S.; SCHULER, G. D. (1995): ESTablishing a human transcript map. *Nature Genet.* **10**: 369-371

BONALDO, M. F.; LENNON, G.; SOARES, M. B. (1996): Normalization and substraction: two approaches to facilitate gene discovery. *Genome Res.* **6**: 791-806

BOUCK, J.; MILLER, W.; GORRELL, J. H.; MUZNY, D.; GIBBS, R. A. (1998): Analysis of the quality and utility of random shotgun sequencing at low redundancies. *Genome Res.* **8**: 362-376

BRILLIANT, M. H.; WILLIAMS, R. W.; CONTI, C. J.; ANGEL, J. M.; OAKEY, R. J.; HOLDENER, B. C. (1994): Mouse chromosome 7. *Mamm. Genome* **7**: 104-123

BRINKMANN, B.; KLINTSCHAR, M.; NEUHUBER, F.; HUHNE, J.; ROLF, B. (1998): Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am. J. Hum. Genet*. **61**: 1408-1415

BRITTEN, R. J.; KOHNE, D. E. (1968): Repeated sequences in DNA. *Science* **161**: 529-540

BRITTEN, R. J. (1986): Rates of DNA sequence evolution differ between taxonomic groups. *Science* **231**: 1393-1398

BRITTEN, R. J.; BARON, W. F.; STOUT, D. B.; DAVIDSON, E. H. (1988): Sources and evolution of human Alu repeated sequences. *Proc. Natl. Acad. Sci. USA* **85**: 4770-4774

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

BURGE, C. (1997): Identification of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**: 78-94

BURSET, M.; GUIGÓ, R. (1996): Evaluation of gene structure prediction programs. *Genomics* **34**: 353-367

CAPECCHI, M. R. (1989): Altering the genome by homologous recombination. *Science* **244**: 1288-1292

CAROTHERS, A. M.; URLAUB, G.; MUCHA, J.; GRUNBERGER, D.; CHASIN, L. A. (1989): Point mutation analysis in a mammalian gene: rapid preparation of total RNA, PCR amplification of cDNA, and *Taq sequencing* by a novel method. *Biotechniques* **7**: 494-496; 498-499

CELNIKER, S. E. (2000): The Drosophila genome. *Curr. Opin. Genet. & Dev*. **10**: 612-616

CHAMBAUD, I.; HEILIG, R.; FERRIS, S.; BARBE, V.; SAMSON, D.; GALISSON, F.; MOSZER, I.; DYBVIG, K.; WROBLEWSKI, H.; VIARI, A.; ROCHA, E. P.; BLANCHARD, A. (2001): The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis. *Nucl. Acids Res*. **29**: 2145-2153

CHISSOE, S. L.; MARRA, M. A.; HILLIER, L.; BRINKMAN, R.; WILSON, R. K.; WATERSTON, R. H. (1997): Representation of cloned genomic sequences in two sequencing vectors: correlation of DNA sequence and subclone distribution. *Nucl. Acids Res.* **25**: 2960-2966

CHURCH, G. M.; GILBERT, W. (1984): Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **81**: 1991-1995

CICHUTEK, A.; BRÜCKMANN, T.; SEIPEL, B.; HAUSER, H.; SCHLAUBITZ, S.; PRAWITT, D.; HANKELN, T.; SCHMIDT, E. R.; WINTERPACHT, A.; ZABEL, B. U. (2001): Comparative architectural aspects of regions of conserved synteny on human chromosome 11p15.3 and mouse chromosome 7 (including genes WEE1 and LMO1). *Cytogen. Cell Genet*. **93**: 277-283

CLARK, M. S. (1999): Comparative genomics: the key to understand the Human Genome Project. *BioEssays* **21**: 121-130

COLE, S. T.; BROSCH, R.; PARKHILL, J.; GARNIER, T.; CHURCHER, C.; HARRIS, D.; GORDON, S. V.; EIGLMEIER, K.; GAS, S.; BARRY III, C. E.; TEKAIA, F.; BADCOCK, K.; BASHAM, D.; BROWN, D.; CHILLINGHWORTH, T.; CONNOR, R.; DAVIES, R.; DEVLIN, K.; FELTWELL, T.; GENTLES, S.; HAMLIN, N.; HOLROYD, S. et al. (1998): Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393**: 537-544

COLLINS, F. (1995): Positional cloning moves from perditional to traditional. *Nature Genet.* **9**: 347-350

COLLINS, F. S.; PATRINOS, A.; JORDAN, E.; CHAKRAVARTI, A.; GESTELAND, R.; WALTERS, L. (1998): New goals for the U.S. Human Genome Project: 1998-2003. *Science* **282**: 682-689

CONNELL, C. S.; FUNG, S.; HEINER, C.; BRIDGHAM, J.; CHAKERIAN, V.; HERON, E.; JONES, B.; MENCHEN, S.; MORDAN, W.; RAFF, M. et al. (1987): Automated DNA sequence analysis. *Biotechniques* **5**: 342-348

COOPER, D. N.; TAGGART, M. H.; BIRD, A. P. (1983): Unmethylated domains in vertebrate DNA. *Nucl. Acids Res.* **11**: 647-658

COOPER, P. R.; SMILINICH, N. J.; DAY, C. D.; NOWAK, N. J.; REID, L. H.; PEARSALL, R. S.; REECE, M.; PRAWITT, D.; LANDERS, J.; HOUSMAN, D. E.; WINTERPACHT, A.; ZABEL, B. U.; PELLETIER, J.; WEISSMAN, B. E.; SHOWS, T. B.; HIGGINS, M. J. (1998): Divergently transcribed overlapping genes expressed in liver and kidney and located in the 11p15.5 imprinted domain. *Genomics* **49**: 38-51

COST, G. J.; BOEKE, J. D. (1998): Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. *Biochem.* **37**: 18081-18093

CROSS, S. H.; BIRD, A. P. (1995): CpG islands and genes. *Curr. Opin. Genet. Devel*. **5**: 309-314

CROSS, S. H.; LEE, M.; CLARK, V. H.; CRAIG, J. M.; BIRD, A, P.; BICKMORE, W. A. (1997): The chromosomal distribution of CpG islands in the mouse: evidence for genome scrambling in the rodent lineage. *Genomics* **40**: 454-461

DAS, M.; HARVEY, I.; CHU, L. L.; SINHA, M.; PELLETIER, J. (2001): Full-length cDNAs: more than just reaching the ends. *Physiol. Genomic*s **6**: 57-80

DEBERARDINIS, R. J.; GOODIER, J. L.; OSTERTAG, E. M.; KAZAZIAN JR., H. H. (1998): Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. *Nature Gene.* **20**: 288-290

DEBRY, R. W.; SELDIN, M. F. (1996): Human/mouse homology relationships. *Genomics* **33**: 337-351

DIETZ, H. C.; CUTTING, G. R.; PYERITZ, R. E.; MASLEN, C. L.; SAKAI, L. Y.; CORSON, G. M.; PUFFENBERGER, E. G.; HAMOSH, A.; NANTHAKUMAR, E. J.; CURRISTIN, S. M. et al. (1991): Marfan syndrome caused by recurrent de novo missense mutation in the fibrillin gene (see comments). *Nature* **352**: 337-339

DUNHAM, I.; SHIMIZU, N.; ROE, B. A.; CHISSOE, S. et al., (1999): The DNA sequence of human chromosome 22. *Nature* **402**: 489-495

EWING, B.; GREEN, P. (1998): Base-Calling of automated sequencer traces using Phred. II. Error Probabilities. *Genome Res.* **8**: 186-194

EWING, B.; HILLIER, L.; WENDL, M. C.; GREEN, P. (1998): Base-Calling of automated sequencer traces using Phred. I. Accuracy Assessment. *Genome Res.* **8**:175-185

FAN, W.; CAI, W.; PARIMOO, S.; LENNON, G. G.; WEISSMAN, S. M. (1996): Identification of seven new human MHC class I region genes around the HLA-F locus. *Immunogenetics* **44**: 97-103

FAN, W.; CHRISTENSEN, M.; EICHLER, E.; ZHANG, X.; LENNON, G. (1997): Cloning, sequencing, gene organization, and localization of the human ribosomal protein RPL23A gene. *Genomics* **46**: 234-239

FEINBERG, A. P.; VOGELSTEIN, B. (1983): A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem*. **132**: 6-13

FENG, Q.; MORAN, J. V.; KAZAZIAN JR., H. H.; BOEKE, J. D. (1996): Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**: 905-916

FERRY, J. (2000): "Working draft" of human genome available by June. *Lancet* **355**: 1337

FICKETT, J. W.; TUNG, C.-S. (1992): Assessment of protein coding measures. *Nucl*. *Acids Res*. **20**: 6441-6450

FICKETT, J. W.; HATZIGEORGIOU, A. G. (1997): Eukaryotic promoter recognition. *Genome Res.* **7**: 861-878

FIELDS, C.; ADAMS, M. D.; WHITE, O.; VENTER, J. C. (1994): How many genes in the human genome? *Nature Genet*. **7**: 345-346

FLEISCHMANN, R. D.; ADAMS, M. D.; WHITE, O.; CLAYTON, R. A.; KIRKNESS, E. F.; KERVALAGE, A. R.; BULT, C. J.; TOMB, J.-F.; DOUGHERTY, B. A.; MERRICK, J. M.; MCKENNEY, K.; SUTTON, G.; FITZHUGH, W.; FIELDS, C.; GOCAYNE, J. D. et al. (1995): Whole genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* **269**: 496-512

FORONI, L.; BOEHM, T.; WHITE, L.; FORSTER, A.; SHERRINGTON, P.; LIAO, X. B.; BRANNAN, C. I.; JENKINS, N. A.; COPELAND, N. G.; RABBITTS, T. H. (1992): The rhombotin gene family encode related LIM-domain proteins whose differing expression suggests multiple roles in mouse development. *J. Mol. Biol.* **226**: 747-761

FRASER, C. M.; CASJENS, S.; HUANG, W. M.; SUTTON, G.; CLAYTON, R. A.; LATHIGRA, R.; WHITE, O.; KETCHUM, K. A.; DODSON, R.; HICKEY, E. K.; GWINN, M.; DOUGHERTY, B. A.; TOMB, J.-F.; FLEISCHMANN, R. D.; RICHARDSON, D.; PETERSON, J.; KERLAVAGE, A. R.; QUACKENBUSH, J.; SALZBERG, S., et al. (1997): Genomic sequence of an Lyme disease spirochaeate Borrelia burgdorferi. *Nature* **390**: 580-586

FRASER, C. M.; GOCAYNE, J. D.; WHITE, O.; ADAMS, M. D.; CLAYTON, R. A.; FLEISCHMANN, R. D.; BULT, C. J.; KERLAVAGE, A. R.; SUTTON, G.; KELLEY, J. M.; FRITCHMAN, J. L.; WEIDMAN, J. F.; SMALL, K. V.; SANDUSKY, M. et al. (1995): The minimal gene complement of Mycoplasma genitalium. *Science* **270**: 397-403

FRIEND, S. H.; BERNARDS, R.; ROGELJ, S.; WEINBERG, R. A.; RAPAPORT, J. M.; ALBERT, D. M.; DRYJA, T. P. (1986): A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* **323**: 643-646

GARDINER-GARDEN, K.; FROMMER, M. (1987): CpG islands in vertebrate genomes. *J. Mol. Biol*. **196**: 261-282

GARDINER, K. (1996): Base composition and gene distribution: critical patterns in mammalian genome organization. *Trends in Genet.* **12**: 519-524

GARDNER, M. J.; TETTELIN, H.; CARUCCI, D. J.; CUMMINGS, L. M.; ARAVIND, L.; KOONIN, E. V.; SHALLOM, S.; MASON, T.; YU, K.; FUJII, C.; PEDERSON, J.; SHEN, K.; JING, J.; ASTON, C.; LAI, Z.; SCHWARTZ, D. C.; PERTEA, M. et al. (1998): Chromosome 2 sequence of the human Malaria parasite: Plasmodium falciparum. *Science* **282**: 1126-1132

GIBBS, A. J.; MCINTYRE, G. A. (1970): The diagram, a method for comparing sequences. Its use with amino acid and nucleotide sequences. *Eur. J. Biochem*. **16**: 1-11

GILBERT, N.; LABUDA, D. (1999): CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs. *Proc. Natl. Acad. Sci. U*SA **96**: 2869-2874

GLÖCKNER, G.; SCHERER, S.; SCHATTEVOY, R.; BORIGHT, A.; WEBER, J.; TSUI, L.-C.; ROSENTHAL, A. (1998): Large-scale sequencing of two regions in human chromosome 7q22: analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci reveals 17 genes. *Genome Res*. **8**: 1060-1073

GOFFEAU, A.; BARRELL, B. G.; BUSSEY, H.; DAVIS, R. W.; DUJON, B.; FELDMANN, H.; GALIBERT, F.; HOHEISEL, J. D.; JACQ, C.; JOHNSTON, M.; LOUIS, E. J.; MEWES, H. W.; MURAKAMI, Y.; PHILIPPSEN, P.; TETTELIN, H.; OLIVER, S. G. (1996): Living with 6000 genes. *Science* **274**: 546-567

GOODMAN, L. (1998): Random shotgun fire. *Genome Res.* **8**: 567-568

GORDON, D.; ABAJIAN, C.; GREEN, P. (1998): A graphical tool for sequence finishing. *Genome Res.* **8**: 195-202

GÖRLICH, D.; DABROWSKI, M.; BISCHOFF, F. R.; KUTAY, U.; BORK, P.; HARTMANN, E.; PREHN, S.; IZAURRALDE, E. (1997): A novel class of RanGTP binding proteins. *J. Cell Biol*. **138**: 65-80

GRAUR, D.; LI, W.-H. (2000): Fundamentals of molecular evolution. 2nd edition. *Sinauer Associates, Inc*.; Sunderland, Massachusetts

GREEN, P. (1997): Against a whole-genome shotgun. *Genome Res.* **409**: 410-417

GRUNAU, C.; HINDERMANN, W.; ROSENTHAL, A. (2000): Large-scale methylation analysis of human genomic DNA reveals tissue-specific differences between the methylation profiles of genes and pseudogenes. *Hum. Mol. Genet* **9**: 2651-2663

GU, J.; ZHAO, M.; HUANG, Q.; XU, X.; LI, Y.; PENG, Y.; SONG, H.; XIAO, H.; GU, Y.; LI, N.; QIAN, B.; LIU, F.; QU, J.; GAO, X.; CHENG, Z.; XU, Z.; ZENG, L.; XU, S.; GU, W.; TU, Y.; JIA, J.; FU, G.; REN, S.; ZHONG, M.; LU, G.; YANG, Y.; GAO, G.; ZHANG, Q.; CHEN, S.; HAN, Z.; CHEN, Z. (2000): Homo sapiens cDNA MDS clones. *Unveröffentlicht*

GUSELLA, J. F.; WEXLER, N. S.; CONNEALLY, P. M.; NAYLOR, S. L.; ANDERSON, M. A.; TANZI, R. E.; WATKINS, P. C.; OTTINA, K.; WALLACE, M. R.; SAKAGUCHI, A. Y.; YOUNG, A. B.; SHOULSON, I.; BONILLA, E.; MARTIN, J. B. (1993): A polymorphic DNA marker genetically linked to Huntington´s disease. *Nature* **306**: 2234-2238

GUYER, M. S.; COLLINS, F. S. (1995): How is the Human Genome Project doing, and what have we learned so far? *Proc. Natl. Acad. Sci. USA* **92**: 10841-10848

HARDISON, R. C.; OELTJEN, J.; MILLER, W. (1997): Long human-mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome. *Genome Res.* **7**: 959-966

HEALD, R.; MCLOUGHLIN, M.; MCKEON, F. (1993): Human *WEE1* maintains mitotic timing by protecting the nucleus from cytoplasmatically activated cdc2 kinase. *Cell* **74**: 463-474

HEINER, C. R.; HUNKAPILLER, K. L.; CHEN, S.-M.; GLASS, J. I.; CHEN, E. Y. (1998): Sequencing multimegabase-template DNA with BigDye terminator chemistry. *Genome Res.* **8**: 557-561

HIGGINS, M. J.; SMILINICH, N. J.; SAIT, S.; KOENIG, A.; PONGRATH, J.; GESSLER, M.; RICHARD III, C. W.; JAMES, M. R.; SANFORD, J. P.; KIM, B.-W.; CATTELANE, J.; NOWAK, N. J.; WINTERPACHT, A.; ZABEL, B. U.; MUNROE, D. J.; ERIC, E.; HOUSMAN, D. E.; JONES, C.; NAKAMURA, Y.; GERHARD, D. S.; SHOWS, T. B. (1994): An ordered Not I- fragment map of human chromosome band 11p15. *Genomics* **23**: 211-222

HILLIER, L.; LENNON, G.; BECKER, M.; BONALDO, M. F.; CHIAPELLI, B.; CHISSOE, S.; DIETRICH, N.; DUBUQUE, T.; FAVELLO, A.; GISH, W.; HAWKINS, M.; HULTMAN, M.; KUCABA, T.; LACY, M.; LE, M.; SCHELLENBERG, K.; SOARES, M. B.; TAN, F.; THIERRY-MEG, J.; TREVASKIS, E.; UNDERWOOD, K.; WOHLDMAN, P.; WATERSTON, R.; WILSON, R.; MARRA, M. (1996): Generation and analysis of 280 000 human expressed sequence tags. *Genome Res.* **6**: 807-828

HOLDENER, B. C.; BROWN, S. D. M.; ANGEL, J. M.; NICHOLLS, R. D.; KELSEY, G.; MAGNUSON, T. (1993): Mouse chromosome 7. *Mamm. Genome* **4**: 110-120

HOLLIDAY, R.; PUGH, J. E. (1975): DNA modification mechanisms and gene activity during development. *Science* **187**: 1095-1107

HONDA, R.; TANAKA, H.; OHBA, Y.; YASUDA, H. (1995): Mouse p87wee1 kinase is regulated by M-phase specific phosphorylation. *Chromosome Res*. **3**: 300-308

IGARASHI, M.; NAGATA, A.; JINNO, S.; SUTO, K.; OKAYAMA, H. (1991): Wee1(+)-like gene in human cells. *Nature* **353**: 80-83

ILYAS, M.; STRAUB, J.; TOMLINSON, I. P.; BODMER, W. F. (1999): Genetic pathways in colorectal and other cancers. *Eur. J. Cancer* **35**: 335-351

IOANNOU, P. A.; AMAMIYA, C. T.M; GARNES, J.; KROISEL, P. M.; SHIZUYA, H.; CHEN, C.; BATZER, M. A.; DE JONG, P. J. (1994): A new bacteriophage P1-derived vector for the propagation of large human DNA-fragments. *Nature Genet.* **6**: 84-89

ITO, T.; SELDIN, M. F.; TAKETO, M. M.; KUBO, T.; NATORI, S. (2000): Gene structure and chromosome mapping of mouse transcription elongation factor S-II (Tcea1). *Gene* **244**: 55-63

IZAURRALDE, E.; KUTAY, U.; VON KOBBE, C.; MATTAJ, I. W.; GÖRLICH, D. (1997): The asymmetric distribution of the constituents of the RAN system is essential for transport into and out of the nucleus. *EMBO J*. **16**: 6535-6547

JACQ, C.; ALT-MÖRBE, J.; ANDRÉ, B.; ARNOLD, W.; BAHR, A.; BALLESTA, J. P. G.; BARGUES, M.; BARON, L.; BECKER, A.; BITEAU, N.; BLÖCKER, H.; BLUGEON, C.; BOSKOVIC, J.; BRANDT, P.; BRÜCKNER, M.; BUITRAGO, M. J. et al. (1997): The nucleotide sequence of Saccharomyces cerevisiae chromosome IV. *Nature* **387**: 75-87

JÄKEL, S.; GÖRLICH, D. (1998): Importin β, transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells. *EMBO J*. **17**: 4491-4502

JÄKEL, S.; ALBIG, W.; KUTAY, U.; BISCHOFF, F. R.; SCHWAMBORN, K.; DOENECKE, D.; GÖRLICH, D. (1999): The importin β/ importin 7 heterodimer is a functional nuclear import receptor for histone H1. *EMBO J*. **18**: 2411-2423

JANG, W.; HUA, A.; SPILSON, S. V.; MILLER, W.; ROE, B. A.; MEISLER, M. H. (1999): Comparative sequence of human and mouse BAC clones from the *mnd2* region of chromosome 2p13. *Genome Res.* **9**: 53-61

JURKA, J.; ZIETKIEWICZ, E.; LABUDA, D. (1995): Ubiquitous mammalian-wide interspersed repeats (MIRs) are molecular fossils from the mesozoic era. *Nucl. Acids Res*. **23**: 170-175

KAINULAINEN, K.; PULKKINEN, L; SAVOLAINEN, A.; KAITILA, I.; PELTONEN, L. (1990): Location on chromosome 15 of the gene defect causing Marfan syndrome. *N. Engl. J. Med*. **323**: 935-939

KARNIK, P.; PARIS, M.; WILLIAMS, B. R. G.; CASEY, G.; CROWE, J.; CHEN, P. (1998): Two distinct tumor suppressor loci within chromosome 11p15 implicated in breast cancer progression and metastasis. *Hum. Mol. Genet*. **7**: 895-903

KAZAZIAN JR., H. H.; MORAN, J. V. (1998): The impact of L1 retrotransposons on the human genome. *Nature Genet.* **19**: 19-24

KLEVER, M.; GROND-GINSBACH, C.; SCHERTHAN, H.; SCHROEDER-KURTH, T. M. (1991): Chromosomal in situ suppression of hybridization after Giemsa banding. *Hum. Genet*. 86: 484-486

KLEYN, P. W.; FAN, W.; KOVATS, S. G.; LEE, J. L.; PULIDO, J. C.; WU, Y.; BERKEMEIER, L. R.; MISUMI, D. J.; HOLMGREN, L.; CHARLAT, O.; WOOLF, E. A.; TAYBER, O.; BRODY, T.; SHU, P.; HAWKINS, F.; KENNEDY, B.; BALDINI, L.; EBELING, C.; ALPERIN, G. D.; DEEDS, J.; LAKEY, N.; CULPEPPER, J.; CHEN, H.; GLÜCKSMANN-KULIS, M. A.; CARLSON, G. A.; DUYK, G. M.; MOORE, K. J. (1996): Identification and characterization of the mouse obesity gene tubby: a member of a novel gene family. *Cell* **85**: 281-290

KOCH-NOLTE, F.; KÜHL, M.; HAAG, F.; CETKOVICH-CVRLJE, M.; LEITER, E. H.; THIELE, H.-G. (1996): Assignment of the human and mouse genes for muscle ecto mono (ADPribosyl) transferase to a conserved linkage group on human chromosome 11p15 and mouse chromosome 7. *Genomics* **36**: 215-216

KOOP, B. F. (1995): Human and rodent DNA sequence comparisons: a mosaic model of genomic evolution. *Trends in Genet.* **11**: 367-371

KORN, B.; SEDLACEK, Z.; MANCA, A.; KIOSCHIS, P.; KONECKI, D.; LEHRACH, H.; POUSTKA, A. (1992): A strategy for the selection of transcribed sequences in the Xq28 region. *Hum. Mol. Genet*. **1**: 235-242

KOZAK, M. (1996): Interpreting cDNA sequences: some insights from studies on translation. *Mammal. Genome* **7**: 563-574

KUWABARA, P. E.; COULSON, A. (2000): RNAi--prospects for a general technique for determining gene function. *Parasitol. Toda*y **16**: 347-349

KWOK, S. C.; LEDLEY, F. D.; DILELLA, A. G.; ROBSON, K. J.; WOO, S. L. (1985): Nucleotide sequence of a full-length complementary DNA clone and amino acid sequence of human phenylalanine hydroxylase. *Biochem.* **24**: 556-561

LAIRD, C. D.; MCCONAUGHY, B. L.; MCCARTHY, B. J. (1969): Rate of fixation of nucleotide substitutions in evolution. *Nature* **224**: 249-254

LAMERDIN, J. E.; STILWAGEN, S. A.; RAMIREZ, M. H.; STUBS, L.; CARRANO, A. V. (1996): Sequence analysis of the ERCC2 gene regions in human, mouse, and hamster reveals three linked genes. *Genomics* **34**: 399-409

LANDER, E. S. (1996): The new genomics: global views of biology. *Science* **274**: 536-539

LARSEN, F.; GUNDERSEN, G.; LOPEZ, R.; PRYDZ, H. (1992): CpG islands as gene markers in the human genome. *Genomics* **13**: 1095-1107

LAWRENCE, J. B.; VILLNAVE, C. A.; SINGER, R. H. (1988): Sensitive, high-resolution chromatin and chromosome mapping in situ: presence and orientation of two closely integrated copies of EBV in a Lymphoma line. *Cell* **52**: 51-61

LAWRENCE, J. B.; SINGER, R. H.; MCNEIL, J. A. (1990): Interphase and metaphase resolution of different distances within the human dystrophin gene. *Science* **249**: 928-932

LEBO, R. V.; LYNCH, E. D.; BIRD, T. D.; GOLBUS, M. S.; BARKER, D. F.; O`CONNELL, P.; CHANCE, P. F. (1992): Multicolor in situ hybridization and linkage analysis order Charot-Marie-Tooth Type I (CMTIA) gene-region marker. *Am. J. Hum. Genet*. **50**: 42-55

LEE, L. G.; CONNELL, C. R.; WOO, S. L.; CHENG, R. D.; MCARDLE, B. F.; FULLER, C. W.; HALLORAN, N. D.; WILSON, R. K. (1992): DNA sequencing with dye-labeled terminators and T7 DNA polymerase: effect of dyes and dNTPs on incorporation of dye-terminators and probability analysis of termination fragments. *Nucl. Acids Res.* **20**: 2471-2483

LEMIEUX, N.; DUTRILLAUX, B.; VIEGAS-PÉQUIGNOT, E. (1992): A simple method for simultaneous R- or G-banding and fluorescence in situ hybridization of small copy genes. *Cytogenet. Cell Genet*. **59**: 311-312

LEWIS, E. B. (1992): Cluster of master control genes regulate the development of higher organisms. *J. Am. Med. Assoc*. **267**: 1524-1531

LI, S.-R.; GYSELMAN, V. G.; DORUDI, S.; BUSTIN, S. A. (2000): Elevated levels of RanBP7 mRNA in colorectal carcinoma are associated with increased proliferation and are similar to the transcription pattern of the protooncogene c-myc. *Biochem. Biophys Res. Comm*. **271**: 537-543

LICHTER, P.; CREMER, T.; BORDEN, J.; MANUELIDIS, L.; WARD, D. C. (1988): Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum. Genet.* **80**: 224-234

LICHY, J. H.; MODI, W. S.; SEUANEZ, H. N.; HOWLEY, P. M. (1992): Identification of a human chromosome 11 gene which is differentially regulated in tumorigenic and nontumorigenic somatic cell hybrids of HeLa cells. *Cell. Growth Diff*. **3**: 541-548

LIDSKY, A. S.; LAW, M. L.; MORSE, H. G.; KAO, F. T.; RABIN, M.; RUDDLE, F. H.; WOO, S. L. (1985): Regional mapping of the phenylalanine hydroxylase gene and the phenylketonuria locus in the human genome. *Proc. Natl. Acad. Sci. USA* **82**: 6221-6225

LIEW, C. C.; HWANG, D. M.; FUNG, Y. W.; LAURENSSEN, C.; CUKERMAN, E.; TSUI, S.; LEE, Y. (1994): A catalogue of genes in the cardiovascular system is identified by expressed sequence tags. *Proc. Natl. Acad. Sci. USA* **91**: 10645-10649

LOOTS, G. G.; LOCKSLEY, R. M.; BLANKESPOOR, C. M.; WANG, Z. E.; MILLER, W.; RUBIN, E. M.; FRAZER, K. A. (2000): Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* **288**: 136-140

LYON, M. F. (1998): X-Chromosome inactivation: a repeat hypothesis. *Cytogenet. Cell Genet.* **80**: 133-137

LYON, M. F. (2000): LINE-1 elements and X-chromosome inactivation: a function for „junk" DNA? *Proc. Natl. Acad. Sci. USA* **97**: 6248-6249

MACGREGOR, H. C.; MIZUNO, S. (1976): In situ hybridization of „nick-translated" 3H-ribosomal DNA to chromosomes from salamanders. *Chromosoma* **54**: 15-25

MAGENIS, R. E.; MASLEN, C. L.; SMITH, L.; ALLEN, L.; SAKAI, L. Y. (1991): Localization of the fibrillin (FBN) gene to chromosome 15, band q21.1. *Genomics* **11**: 346-351

MAKALOWSKI, W.; ZHANG, J.; BOGUSKI, M. S. (1996): Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences. *Genome Res.* **6**: 846-857

MAKALOWSKI, W.; BOGUSKI, M. S. (1998): Evolutionary parameters of the transcribed mammalian genome: an analysis of 2820 orthologous rodent and human sequences. *Proc. Natl. Acad. Sci. USA* **95**: 9407-9412

MAO, M.; FU, G.; WU, J.-S.; ZHANG, Q.-H.; ZHOUR, J.; KANN, L. X.; HUANG, Q.-H.; HE, L.-L.; GU, B.-W.; HAN, Z.-G.; SHEN, Y.; GU, J.; YU, Y.-P.; XU, S.-H.; WANG, Y.-X.; CHEN, S.-J.; CHEN, Z. (1998): Identification of genes expressed in human $CD34^+$ hematopoietic stem/progenitor cells by expressed sequence tags and efficient full-length cDNA cloning. *Proc. Natl. Acad. Sci. USA* **95**: 8175-8180

MARRA, M.; HILLIER, L.; KUCABA, T.; ALLEN, M.; BARSTEAD, R.; BECK, C.; BLISTAIN, A.; BONALDO, M.; BOWERS, Y.; BOWLES, L.; CARDENAS, M.; CHAMBERLAIN, A.; CHAPPELL, J.; CLIFTON, S.; FAVELLO, A.; GEISEL, S.; GIBBONS, M.; HARVEY, N.; HILL, F.; JACKSON, Y.; KOHN, S.; LENNON, G.; MARDIS, E.; MARTIN, J.; WATERSTON, R. et al. (1999): An encyclopedia of mouse genes. *Nature Genet*. **21**: 191-194

MARSHALL, E. (1999): Human Genome Project. Sequencers endorse plan for a draft in 1 year. *Science* **284**: 1439-1441

MARTIN, A. P.; PALUMBI, S. R. (1993): Body size, metabolic rate, generation time and the molecular clock. *Proc. Natl. Acad. Sci. USA* **90**: 4087-4091

MARTINDALE, D. W.; WILSON, M. D.; WANG, D.; BURKE, R. D.; CHEN, X.; DURONIO, V.; KOOP, B. F. (2000): Comparative genomic sequence analysis of the Williams syndrome region (LIMK1-RFC2) of human chromosome 7q11.23. *Mamm. Genome* **11**: 890-898

MARTIN-GALLARDO, A.; LAMERDIN, J.; CARRANO, A. (1994): Shotgun sequencing. In: Automated DNA sequencing and analysis. Adams, M. D.; Fields, C.; Venter, J. C. (Hrsg.), *Academic Press London*

MATSUO, K.; CLAY, O.; TAKAHASHI, T.; SILKE, J.; SCHAFFNER, W. (1993): Evidence for erosion of mouse CpG islands during mammalian evolution. *Somat. Cell Mol. Genet*. **19**: 543-564

MAXAM, A. M.; GILBERT, W. (1977): A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **74**: 560-564

MCGOWAN, C. H.; RUSSEL, P. (1995): Cell cycle regulation of human WEE1. *EMBO J*. **14**: 2166-2175

MCKUSIK, V. A. (1991): Mendelian inheritance in man; 9th edtion; John Hopkins University. Baltimore

MESELSON, M. S.; STAHL, F. W. (1958): The replication of DNA in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **44**: 671-682

MEWES, H. W.; ALBERMANN, K.; BÄHR, M.; FRISHMAN, D.; GLEISSNER, A.; HANI, J.; HEUMANN, K.; KLEINE, K.; MAIERL, A.; OLIVER, S. D.; PFEIFFER, F.; ZOLLNER, A. (1997): Overview of the yeast genome. *Nature* **387**: 7-65

MIRONOV, A. A.; FICKET, J. W.; GELFAND, M. S. (1999): Frequent alternative splicing of human genes. *Genome Res*. **9**: 1288-1293

MONACO, A. P.; NEVE, R. L.; COLLETTI-FEENER, C.; BERTELSON, C. J.; KURNIT, D. M.; KUNKEL, L. M. (1986): Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature* **323**: 646-651

MURNANE, J. P.; MORALES, J. F. (1995): Use of a mammalian interspersed repetitive (MIR) element in the coding and processing sequences of mammalian genes. *Nucl. Acids Res.* **23**: 2837-2839

MYSLINSKI E.; KROL A.; CARBON P. (1998): ZNF76 and ZNF143 are two human homologs of the transcriptional activator Staf. *J. Biol. Chem.* **273**: 21998-22006

NATIONAL RESEARCH COUNCIL (1988): Mapping and sequencing the human genome. *Natl. Acad. Press*, *Washington DC*

NETO, E. D.; CORREA, R. G. et al. (2000): Shutgun sequencing of the human transcriptomes with ORF expressed sequence tags. *Proc. Natl. Acad. Sci. USA* **97**: 3491-3496

NEWMEYER, D. D.; LUCOCQ, J. M.; BURGLIN, T. R.; DEROBERTIS, E. M. (1986): Assembly in vitro of nuclei active in nuclear protein transport: ATP is required for nucleoplasmin accumulation. *EMBO J*. **5**: 501-510

NICOLAS, F.; ZHANG, C.; HUGHES, M.; GOLDBERG, M.; WATTON, S.; CLARKE, P. (1997): Xenopus Ran-binding protein 1: molecular interactions and effects on nuclear assembly in Xenopus egg extracts. *J. Cell Sci*. **110**: 3019-3030

NIRENBERG, M. W.; MATTHAEI, J. H. (1961): The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. USA* **47**: 1588-1602

NORMAN, A. M.; READ, A. P.; CLAYTON SMITH, J.; ANDREWS, T.; DONNAI, D. (1992): Recurrent Wiedemann-Beckwith syndrome with inversion of chromosome (11)(p11.2p15.5). *Am. J. Med. Genet*. **42**: 638-641

O´BRIEN, S. J.; MENOTTI-RAYMOND, M.; MURPHY, W. J.; NASH, W. G.; WIENBERG, J.; STANYON, R.; COPELAND, N. G.; JENKINS, N. A.; WOMACK, J. E.; MARSHALL GRAVES, J. A. (1999): The promise of comparative genomics in mammals. *Science* **286**: 458-481

OELTJEN, J. C.; MALLEY, T. M.; MUZNY, D. M.; MILLER, W.; GIBBS, R. A.; BELMONT, J. W. (1997): Large-scale comparative sequence analysis of the human and murine Bruton´s tyrosine kinase loci reveals conserved regulatory domains. *Genome Res.* **7**: 315-329

OHNO, S. (1970): Evolution by gene duplication. *Springer Verlag*, Berlin

OKADA, N.; HAMADA, M.; OGIWARA, I.; OHSHIMA, K. (1997): SINEs and LINEs share common 3´ sequences: a review. *Gene* **205**: 229-243

OLSON, T. M.; KISHIMOTO, N. Y.; WHITBY, F. G.; MICHELS, V. V. (2001): Mutations that alter the surface charge of alpha-tropomyosin are associated with dilated cardiomyopathy. *J. Mol. Cell Cardiol*. **33**: 723-732

ONYANGO, P.; MILLER, W.; JEHOCZKY, J.; LEUNG, C. T.; BIRREN, B.; WHEELAN, S.; DEWAR, K.; FEINBERG, A. (2000): Sequence and comparative analysis of the mouse 1-megabase region orthologous to the human 11p15 imprinted domain. *Genome Res.* **10**: 1697-1710

OPHIR, R.; GRAUR, D. (1997): Patterns and rates of indel evolution in processed pseudogenes from humans and murids. *Gene* **205**: 191-202

OSOEGAWA, K.; TATENO, M.; WOON, P. Y.; FRENGEN, E.; MAMMOSER, A. G.; CATANESE, J. J.; HAYASHIZAKI, Y.; DE JONG, P. J. (2000): Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res.* **10**: 116-128

OTA, T.; NISHIKAWA, T.; SUZUKI, Y.; ISHII, S.; SAITO, K.; KAWAI, Y.; YAMAMOTO, J.; WAKAMATSU, A.; NAKAMURA, Y.; NAGAI, T.; SUGANO, S.; ISOGAI, T. (2000): HRI human cDNA project. *Unpublished*

PALCA, J. (1986): Human genome. Department of Energy on the map. *Nature* **321**: 371

PARK, J. M.; YANG, E. S.; HATFIELD, D. L.; LEE, B. J. (1996): Analysis of the selenocysteine tRNA $^{(SER)SEC}$ gene transcription *in vitro* using Xenopus oocyte extracts. *Biochem. Biophys. Res. Commun*. **226**: 231-236

PARKER, L. L.; PIWNICA-WORMS, H. (1992): Inactivation of the p34cdc2-cyclin b complex by the human WEE1 tyrosine kinase. *Science* **257**: 1955-1957

PASSARGE, E.; HORSTHEMKE, B.; FARBER, R. A. (1999): Incorrect use of the term synteny. *Nature Genet.* **23**: 387

PAWSON, T.; NASH, P. (2000): Protein-protein interactions define specificity in signal transduction. *Genes & Dev*. **14**: 1027-1047

PEARSON, W. R.; LIPMAN, D. J. (1988): Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**: 2444-2448

PROBER, J. M.; TRAINOR, G. L.; DAM, R. J.; HOBBS, F. W.; ROBERTSON, C. W.; ZAGURSKY, R. J.; COCUZZA, A. J.; JENSEN, M. A.; BAUMEISTER, K. (1987): A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* **238**: 336-341

REDEKER, E.; ALDERS, M.; HOOVERS, J. M.; RICHARD, C. W. 3^{RD}; WESTERVELD, A.; MANNENS, M. (1995): Physical mapping of 3 candidate tumor suppressor genes relative to Beckwith-Wiedemann syndrome asociated chromosomal breakpoints at 11p15.3. *Cytogenet. Cell Genet*. **68**: 222-225

RICHARD, C. W.; BOEHNKE, M.; BERG, D. J.; LICHY, J. H.; MEEKER, T. C.; HAUSER, E.; MYERS, R. M.; COX, D. R. (1993): A radiation hybrid map of the distal short arm of human chromosome 11, containing the Beckwith-Wiedemann and associated embryonal tumor loci. *Am. J. Hum. Genet*. **52**: 915-921

RICHTERICH, P. (1998): Estimation of errors in „raw" DNA sequences : a validation study. *Genome Res.* **8**: 251-259

RINCHIK, E. M.; MAGNUSON, T.; HOLDENER-KENNY, B.; KELSEY, G.; BIANCHI, A.; CONTI, C.; CHARTIER, F.; BROWN, K. A.; BROWN, S. D. M.; PETERS, J. (1992): Mouse chromosome 7. *Mamm. Genome* **3**: 104-120

RINCON, J. C.; ENGLER, S. K.; HARGROVE, B. W.; KUNKEL, G. R. (1998): Molecular cloning of a cDNA encoding human SPH-binding factor, a conserved protein that binds to the enhancer-like region of the U6 small nuclear RNA gene promoter. *Nucl. Acids Res.* **26**: 4846-4852

ROGIC, S.; MACKWORTH, A. K.; OUELLETTE, F. B. (2001): Evaluation of gene-finding programs on mammalian sequences. *Genome Res*. **11**: 817-832

ROMMENS, J. M.; IANNUZZI, M. C.; KEREM, B.; DRUMM, M. L.; MELMER, G.; DEAN, M.; ROZMAHEL, R.; COLE, J. L.; COLLINS, F. S. (1989): Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* **245**: 1059-1065

ROSENBLUM, B. B.; LEE, L. G.; SPURGEON, S. L.; KHAN, S. H.; MENCHEN, S. M.; HEINER, C. R.; CHEN, S. M. (1997): New dye-labeled terminators for improved DNA sequencing patterns. *Nucl. Acids Res.* **25**: 4500-4504

ROYER-POKORA, B.; KUNKEL, L. M.; MONACO, A. P.; GOFF, S. C.; NEWBURGER, P. E.; BAEHNER, R. L.; COLE, F. S.; CURNUTTE, J. T.; ORKIN, S. H. (1986): Cloning the gene for an inherited human disorder – chronic granulomatous disease – on the basis of its chromosomal location. *Nature* **322**: 332-338

RUBIN, G. M.; YANDELL, M. D.; WORTMAN, J. R.; GABOR MIKLOS, G. L.; NELSON, C. R.; HARIHARAN, I. K.; FORTINI, M. E.; LI, P. W.; APWEILER, R.; FLEISCHMANN, W. et al. (2000): Comparative genomics of the eukaryotes. *Science* **287**: 2204-2215

SACCONE, S.; DESARIO, A.; DELLA VALLE, G.; BERNARDI, G. (1992): The highest gene concentrations in the human genome are in telomeric bands of metaphase chromosomes. *Proc. Natl. Acad. Sci. USA* **89**: 4913-4917

SACCONE S.; DE SARIO, A.; WIEGANT, J.; RAAP, A. K.; DELLA VALLE, G.; BERNARDI, G. (1993): Correlations between isochores and chromosomal bands in the human genome. *Proc. Natl. Acad. Sci. USA* **90**: 11929-11933

SAHOO, T.; GOENAGA-DIAZ, E.; SEREBRIISKII, I. G.; THOMAS, J. W.; KOTOVA, E.; CUELLAR, J. G.; PELOQUIN, J. M.; GOLEMIS, E.; BEITINJANEH, F.; GREEN, E. D.; JOHNSON, E. W.; MARCHUK, D. A. (2001): Computational and experimental analyses reveal previously undetected coding exons of the KRIT1 (CCM1) gene. *Genomics* **71**: 123-126

SAIKI, R. K.; BUGAWAN, T. L.; HORN, G. T.; MULLIS, K. B.; ERLICH, H. A. (1986): Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature* **324**: 163-166

SAIKI, R. K.; GELFAND, D. H.; STOFFEL, S.; SCHARF, S. J.; HIGUCHI, R.; HORN, G. T.; MULLIS, K. B.; EHRLICH, H. A. (1988): Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-491

SAIT, S. N.; NOWAK, N. J.; SINGH KAHLON, P.; WEKSBERG, R.; SQUIRE, J.; SHOWS, T. B.; HIGGINS, M. J. (1994): Localization of Beckwith-Wiedemann and rhabdoid tumor chromosome rearrangements to a defined interval in chromosome band 11p15.5. *Genes Chromosomes Cancer* **11**: 97-105

SALAMOV, A. A.; SOLOVYEV, V. V. (2000): Ab initio gene finding in Drosophila genomic DNA. *Genome* Res. **10**: 516-522

SAMBROOK, J.; FRITSCH, E. R.; MANIATIS, T. (1989): Molecular Cloning: A Laboratory Manual. 2nd ed., *Cold Spring Harbor Laboratory Press*, Cold Spring Harbor, NY

SANGER, F.; NICKLEN, S.; COULSON, A. R. (1977): DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463-5467

SASAKI, N.; NAGAOKA, S.; ITOH, M.; IZAWA, M.; KONNO, H.; CARNINCI, P.; YOSHIKI, A.; KUSAKABE, M.; MORIUCHI, T.; MURAMATSU, M.; OKAZAKI, Y.; HAYASHIZAKI, Y. (1998): Characterization of gene expression in mouse blastocyst using single-pass sequencing of 3995 clones. *Genomics* **49**: 167-179

SCHULER, G. D.; BOGUSKI, M. S.; STEWART, E. A.; STEIN, L. D.; GYAPAY, G. et al. (1996): A gene map of the human genome. *Science* **274**: 540-546

SCHULER, G. D. (1997): Pieces of the puzzle: expressed sequence tags and the catalog of human genes. *J. Mol. Med.* **75**: 694-698

SCHULZ, R. A.; BUTLER, B. A. (1989): Overlapping genes of Drosophila melanogaster. Organization of the Z600-gonadal-Eip28/29 gene cluster. *Genes & Dev*. **3**: 232-242

SCHUSTER, C.; MYSLINSKI, E.; KROL, A.; CARBON, P. (1995): Staf, a novel zink finger protein that activates the RNA polymerase III promoter of the selenocystein tRNA gene. *EMBO J*. **14**: 3777-3787

SCHWARTZ , D. C.; CANTOR, C. R. (1986): Separation of yeast chromosome-sized DNAs by pulsed-field gradient electrophoresis. *Cell* **37**: 67-75

SCHWARTZ, S.; ZHANG, Z.; FRAZER, K. A.; SMIT, A.; RIEMER, C.; BOUCK, J.; GIBBS, R.; HARDISON, R.; MILLER, W. (2000): PipMaker: a web server for aligning two genomic DNA sequences. *Genome Res.* **10**: 577-586

SEIPEL, B. (1996): Vergleichende zytogenetische Analyse der humanen Chromosomenregion 11p15 und der entsprechenden Syntäniegruppe auf dem Maus-Chromosom 7 mit Hilfe der Fluoreszenz-in-situ-Hybridisierung (FISH). Diplomarbeit im Fachbereich Biologie der Johannes Gutenberg-Universität Mainz

SHOWS, T. B.; ALDERS, M.; BENNET, S.; BURBEE, D.; CARTWRIGHT, P.; CHANDRASEKHARAPPA, S.; COOPER, P.; COUREAUX, A.; DAVIES, C.; DEVIGNES, M.-D. et al. (1996): Report of the fifth international workshop on human chromosome 11 mapping 1996. *Cytogenet. Cell. Genet*. **74**: 1-54

SINSHEIMER, R. L. (1989): The Santa Cruz Workshop – May 1985. *Genomics* **5**: 954-956

SMIT, A. F. A. (1993): Identification of new abundant superfamily of mammalian LTR-transposons. *Nucl. Acids Res.* **21**: 1863-1872

SMIT, A. F. A.; RIGGS, A. D. (1995): MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. *Nucl. Acids Res.* **23**: 98-102

SMIT, A. F. A. (1996): The origin of interspersed repeats in the human genome. *Curr. Opin. Genet & Dev.* **6**: 743-748

SMIT, A. F. A. (1999): Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev*. **9**: 657-663

SMITH, D. R.; ZHU, Y.; CHENG, J. F.; RUBIN, E. M. (1995): Construction of a panel of transgenic mice containing a contiguous 2-Mb set of YAC/P1 clones from human chromosome 21q22.2. *Genomics* **27**: 425-434

SMITH, L. M.; SANDERS, J. Z.; KAISER, R. J.; HUGHES, P.; DODD, C.; CONNELL, C. R.; HEINER, C.; KENT, S. B. H.; HOOD, L. E. (1986): Fluorescence detection in automated DNA sequence analysis. *Nature* **321**: 674-679

STOCHAJ, U.; ROTHER, K. L. (1999): Nucleocytoplasmic trafficking of proteins: with or without Ran? *BioEssays* **21**: 579-589

STUBBS, L.; RINCHIK, E. M.; GOLDBERG, E.; RUDY, B.; HANDEL, M. A.; JOHNSON, D. (1994): Clustering of six human 11p15 gene homologs within a 500 kb interval of proximal mouse chromosome 7. *Genomics* **24**: 324-332

SULSTON, J. E.; SCHIERENBERG, E.; WHITE, J. G.; THOMSON, J. N. (1983): The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev. Biol.* **100**: 64-119

SUZUKI, Y.; YOSHITOMO-NAKAGAWA, K.; MARUYAMA, K.; SUYAMA, A.; SUGANO, S. (1997): Construction and characterization of a full length-enriched and a 5´-end-enriched cDNA library. *Gene* **200**: 149–156

SUZUKI, Y.; ISHIHARA, D.; SASAKI, M.; NAKAGAWA, H.; HATA, H.; TSUNODA, T.; WATANABE, M.; KOMATSU, T.; OTA, T.; ISOGAI, T.; SUYAMA, A.; SUGANO, S. (2000): Statistical analysis of the 5' untranslated region of human mRNA using 'Oligo-Capped' cDNA libraries. *Genomics* **64**: 286-297

SWEET, D. J.; GERACE, L. (1995): Taking from the cytoplasm and giving to the pore: soluble transport factors in nuclear protein import. *Trends Cell. Biol*. **5**: 444-447

TAUDIEN, S.; RUMP, A.; PLATZER, M.; DRESCHER, B.; SCHATTEVOY, R.; GLOECKNER, G.; DETTE, M.; BAUMGART, C.; WEBER, J.; MENZEL, U.; ROSENTHAL, A. (2000): RUMMAGE – a high throughput sequence annotation system. *Trends Genet*. **11**: 519-520

TAVIAUX, S. A.; DEMAILLE, J. G. (1993): Localization of human cell cycle regulatory genes CDC25C to 5q31 and WEE1 to 11p15.3-11p15.1 by fluorescence in situ hybridization. *Genomics* **15**: 194-196

THE ARABIDOPSIS GENOME INITIATIVE (2000): Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408**: 796 – 815

THE INTERNATIONAL HUMAN GENOME MAPPING CONSORTIUM (2001): A physical map of the human genome. *Nature* **409**: 934-941

THE INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM (2001): Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921

THE INTERNATIONAL SNP MAP WORKING GROUP (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**: 928-933

TOMMERUP, N.; AAGAARD, L.; LUND, C. L.; BOEL, E.; BAXENDALE, S.; BATES, G. P.; LEHRACH, H.; VISSING, H. (1993): A zinc-finger gene ZNF141 mapping at 4p16.3/D4S90 is a candidate gene for the Wolf-Hirschhorn (4p-) syndrome. *Hum. Mol. Genet*. **2**: 1571-1575

TOMMERUP, N.; VISSING, H. (1995): Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders. *Genomics* **27**: 259-264

TRASK, B.; PINKEL, D.; VAN DEN ENGH, G. (1989): The proximity of DNA sequences in interphase cell nuclei is correlated to genomic distance and permits ordering of cosmids spanning 250 kilobase pairs. *Genomics* **5**: 710-717

TRASK, B. J.; MASSA, H.; KENWRICK, S.; GITSCHER, J. (1991): Mapping of human chromosome Xq28 by two-color fluorescence in situ hybridization of DNA sequences to interphase cell nuclei. *Am. J. Hum. Genet*. **48**: 1-15

TURLEAU, C.; DE GROUCHY, J.; CHAVIN COLIN, F.; MARTELLI, H.; VOYER, M.; CHARLAS, R. (1984): Trisomy 11p15 and Beckwith-Wiedemann syndrome. A report of two cases. *Hum. Genet*. **67**: 219-221

UBERBACHER, E. C. (1991): Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl. Acad. Sci. USA* **88**: 11261-11265

VAN HEYNINGEN, V.; LITTLE, P. F. R. (1995): Report of the fourth international workshop on human chromosome 11 mapping 1994. *Cytogen. Cell. Genet*. **69**: 128-155

VANIN, C. C. (1985): Processed pseudogenes: characteristics and evolution. *Ann. Rev. Genet.* **19**: 253-272

VENTER, J. C.; ADAMS, M. D.; SUTTON, G. G.; KERLAVAGE, A. R.; SMITH, H. O.; HUNKAPILLAR, M. (1998): Shotgun sequencing of the human genome. *Science* **280**: 1540-1542

WATANABE, N.; BROOME, M.; HUNTER, T. (1995): Regulation of the human WEE1Hu CDK tyrosine 15-kinase during the cell cycle. *EMBO J*. **14**: 1878-1891

WATSON, J. D.; CRICK, F. H. C. (1953) A structure for deoxyribose nucleic acid. *Nature* **171**: 737-738

WAZIRI, M.; PATIL, S. R.; HANSON, J. W.; BARTLEY, J. A. (1983): Abnormality of chromosome 11 in patients with features of Beckwith-Wiedemann syndrome. *J. Pediatr*. **102**: 873–876

WEBER, J. L.; MYERS, E. W. (1997): Human whole-genome shotgun sequencing. *Genome Res.* **7**: 401-409

WEKSBERG, R.; TESHIMA, I.; WILLIAMS, B. R.; GREENBERG, C. R.; PUESCHEL, S. M.; CHERNOS, J. E.; FOWLOW, S. B.; HOYME, E.; ANDERSON, I. J.; WHITEMAN, D. A. et al. (1993): Molecular characterization of cytogenetic alterations associated with the Beckwith-Wiedemann syndrome (BWS) phenotype refines the localization and suggests the gene for BWS is imprinted. *Hum. Mol. Genet*. **2**: 549-556

WHITE, J. G.; SOUTHGATE, E.; THOMSON, J. N.; BRENNER, S. (1983): Factors that determine connectivity in the nervous system of Caenorhabditis elegans. *Cold Spring Harb. Symp. Quant. Biol.* **48**: 633-40

WIEDEMANN, H. R. (1964): Complexe malformatif familial avec hernie ombilicale et macroglossie – un syndrome nouveau? *J. Genet. Hum*. **13**: 223-232

WIEDEMANN, H. R. (1983): Tumors and hemihypertrophy associated with Wiedemann-Beckwith syndrome. *Eur. J. Pediatr*. **141**: 129

WINZELER, E. A.; SHOEMAKER, D. D.; ASTROMOFF, A.; LIANG, H.; ANDERSON, K.; ANDRE, B.; BANGHAM, R.; BENITO, R.; BOEKE, J. D.; BUSSEY, H.; CHU, A. M.; CONNELLY, C.; DAVIS, K.; DIETRICH, F.; DOW, S. W.; EL BAKKOURY, M.; FOURY, F.; FRIEND, S. H.; GENTALEN, E.; GIAEVER, G. et al. (1999): Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **285**: 901-906

WU, Q.; ZHANG, T.; CHENG, J.-F.; KIM, Y.; GRIMWOOD, J.; SCHMUTZ, J.; DICKSON, M.; NOONAN, J. P.; ZHANG, M. Q.; MYERS, R. M.; MANIATIS, T. (2001): Comparative DNA sequence analysis of mouse and human protocadherin gene clusters. *Genome Res.* **11**: 389-404

ZHANG, M. Q. (1997): Identification of protein coding regions in the human genome by quadratic discriminant analysis. *Proc. Natl. Acad. Sci. USA* **94**: 565-568

ZHENG, B.; MILLS, A. A.; BRADLEY, A. (1999): A system for rapid generation of coat color-tagged knockouts and defined chromosomal rearrangements in mice. *Nucl. Acids Res*. **27**: 2354-2360

7 Anhang

7.1 Veröffentlichungen

1. Cichutek, A.; Brückmann, T.; Seipel, B.; Hauser, H.; Schlaubitz, S.; Prawitt, D.; Hankeln, T.; Schmidt, E. R.; Winterpacht, A.; Zabel, B. U. (2001):
   Comparative architectural aspects of regions of conserved synteny on human chromosome 11p15.3 and mouse chromosome 7 (including genes WEE1 and LMO1).
   Cytogenet. Cell Genet. 3: 277-283

Kurzpublikationen (Poster):

1. Zabel, B.; Löbbert, R.; Prawitt, D., Seipel, B.; Germayer, S.; Brückmann, T.; Cichutek, A.; Munroe, D. J.; Pelletier, J.; Housman, D. E.; Winterpacht (1996):
   Chromosome region 11p15: genomic analysis, transcript mapping, gene identification, comparative sequencing.
   5th International Chromosome 11 Workshop, Niagara-On-The-Lake, Canada, 12.05.-6.05.1996
   Cytogenet. Cell Genet. 74: 56

2. Zabel, B.; Bahr, A.; Amid, C.; Brueckmann, T.; Cichutek, A.; Seipel, B.; Winterpacht, A.; Hankeln, T.; Schmidt, E. R. (1997):
   Comparative sequencing of a 1Mb region in man (chromosome 11p15) and mouse (chromosome 7).
   Abstracts of the Human Genome Meeting, 6.03.-8.03.1997, Toronto, Canada

3. Schmidt, E. R.; Bahr, A.; Amid, C.; Brückmann, T.; Cichutek, A.; Seipel, B.; Winterpacht, A.; Hankeln, T.; Zabel, B. (1997):
   Comparative sequencing of a 1 Mb region in man (chromosome 11) and mouse (chromosome 7).
   Abstract of the International Conference on Molecular Biology and Evolution. Garmisch-Partenkirchen

4. Schmidt. E. R.; Amid, C.; Bahr, A.; Brueckmann, T.; Cichutek, A.; Hankeln, T.; Seipel, B.; Winterpacht, A.; Zabel, B. (1997):
   Comparative genomics: sequencing of a 1 Mb syntenic region (HSAC11p15/MmuC7) in mouse and man.
   2nd International Beutenberg Symposium, Jena, 11.12.-13.12.1997

5. Bahr, A.; Amid, C.; Bikar, S. E.; Brückmann, T.; Cichutek, A.; Hankeln, T.; Seipel, B.; Schmidt, E. R.; Winterpacht, A.; Zabel, B. (1998):
   Comparative genomics: Sequencing of a 1 Mb syntenic region in man (HsaC11p15) and mouse (MmuC7).
   Abstracts of the Human Genome Meeting 1998, Turin, Italy

6. Winterpacht, A.; Amid, C.; Bahr, A.; Brueckmann, T.; Cichutek, A.; Hankeln, T.; Schmidt, E. R.; Seipel, B.; Zabel, B. (1998):
Analysis of a 1 Mb region on chromosome 11p15.3 by comparative sequencing between man and mouse.
10. Jahrestagung der Deutschen Gesellschaft für Humangenetik, Jena, 25.03.-28.03.1998
Med. Genetik 10:113

7. Zabel, B.; Bahr, A.; Amid, C.; Brückmann, T.; Cichutek, A.; Seipel, B.; Winterpacht, A.; Hankeln, T.; Schmidt, E. R. (1998): Comparative sequencing of a 1 Mb region in man (chromosome 11p15) and mouse (chromosome 7).
Abstracts of the 6th International Workshop on Human Chromosome 11 Map 1998, Nizza, France

8. Schmidt, E. R.; Bahr, A.; Amid, C.; Bikar, S. E.; Brückmann, T.; Cichutek, A.; Seipel, B.; Sampson, N.; Winterpacht, A.; Hankeln, T.; Zabel, B. (1998):
Comparative sequencing of a 1 Mb region in man (chromosome 11p15) and mouse (chromosome 7).
Abstracts of the 10th Genome Sequencing and Analysis Conference, Miami, USA
Microbial and Comparative Genomics 3, C-37

9. Schmidt, E. R.; Bahr, A.; Amid, C.; Bikar, S. E.; Brückmann, T.; Cichutek, A.; Seipel, B.; Sampson, N.; Winterpacht, A.; Hankeln, T.; Zabel, B. (1998):
Comparative sequencing of a 1 Mb region in man (chromosome 11p15) and mouse (chromosome 7).
Abstracts of the International Symposium on Genomics and Proteomics. Functional and Computational Aspects and Annual Meeting of the GfG, 4.10.-7.10.1998, Heidelberg

10. Winterpacht, A.; Cichutek, A.; Brückmann, T.; Bahr, A.; Amid, C.; Bikar, S. E.; Hankeln, T.; Zabel, B.; Schmidt, E. R. (1999):
Identification of genes and putative gene regulatory sequences by comparative sequencing between man and mouse.
11. Jahrestagung der Deutschen Gesellschaft für Humangenetik, Nürnberg, 24.03.-27.03.1999
Med. Genetik 11: 222

12. Zabel, B.; Amid, C.; Bahr, A.; Bikar, S.; Brückmann, T.; Cichutek, A.; Mujica, A.; Sampson, N.; Schlaubitz, S.; Hankeln, T.; Winterpacht, A.; Schmidt, E. R. (1999):
Identification of genes and putative gene regulatory sequences by comparative sequencing between man and mouse.
DHGP-Tagung: German Human Genome Project - Implications, Progress, and the Future, München 28.11.-30.11.1999

13. Hankeln, T.; Schmidt, E. R.; Winterpacht, A.; Zabel, B.; Amid, C.; Bahr, A.; Bikar, S. E.; Brückmann, T.; Cichutek, A.; Mujica, A.; Sampson, N.; Schlaubitz, S. (2000):
Die vergleichende Genomananlyse in Mensch und Maus als Werkzeug zur Identifizierung von Genen.
DHGP Xpress 8: 3-7

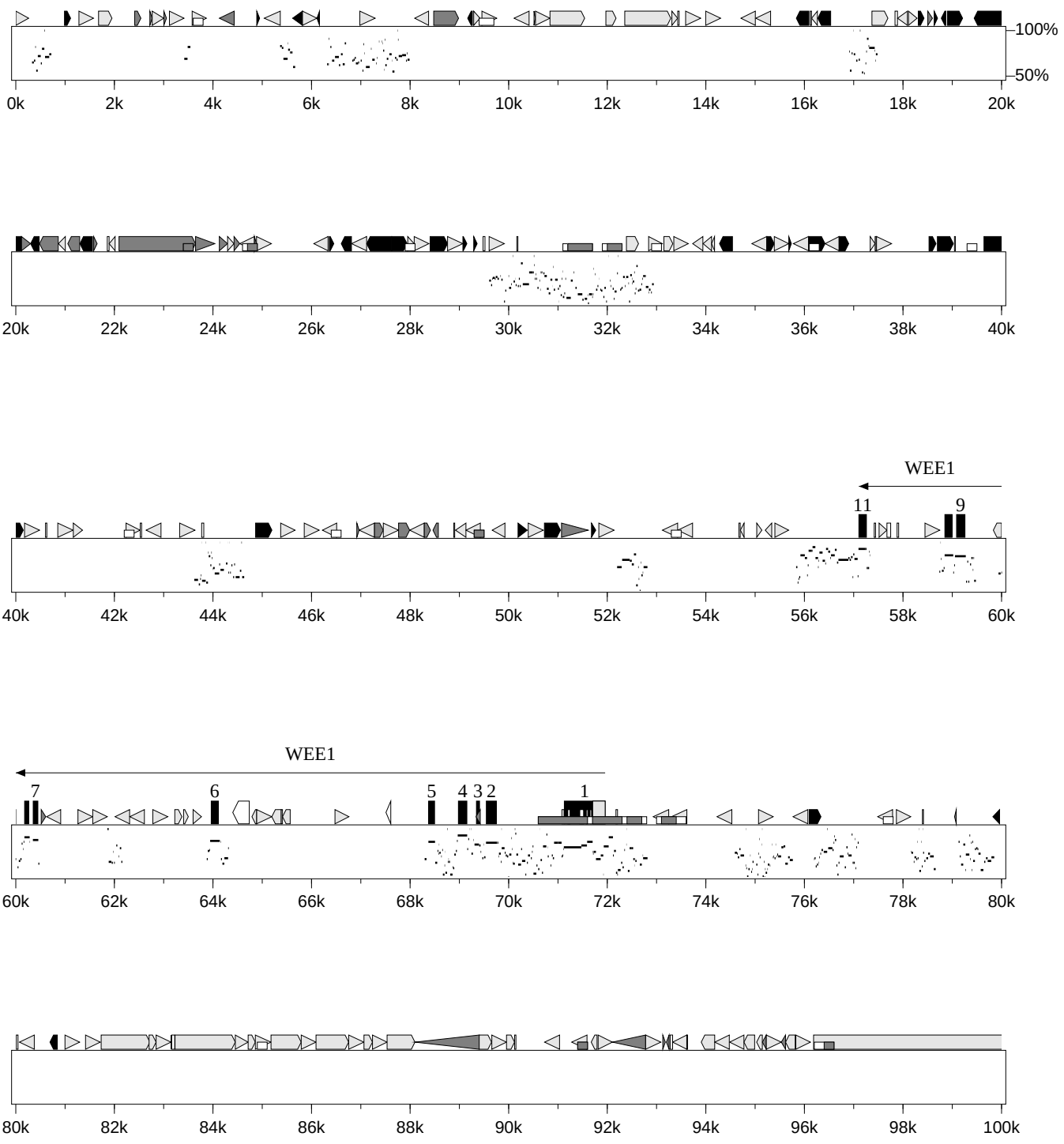
14. Winterpacht, A.; Pfarr, N.; Cichutek, A.; Bahr, A.; Schmidt, E. R.; Hankeln, T.; Zabel, B. (2000):
Novel gene family encoding putative Ca2+-binding proteins with EGF-like modules and a CUB domain.
50th Annual Meeting of the American Society of Human Genetics, Philadelphia 03.-07.10.2000; Am. J. Hum. Genet. 67: 185

7.2 CD

Auf der beigefügten CD sind die alle im Rahmen der Arbeit erzeugten Sequenzdaten, die Informationen der entsprechenden Datenbankeinträgen sowie alle verwendeten Primersequenzen aufgeführt.

7.3 Abbildungen

Abb. 3.17


100%
50%
0k
2k
4k
6k
8k
10k
12k
14k
16k
18k
20k
20k
22k
24k
26k
28k
30k
32k
34k
36k
38k
40k
WEE1
11
9
40k
42k
44k
46k
48k
50k
52k
54k
56k
58k
60k
WEE1
7
6
5
4
3
2
1
60k
62k
64k
66k
68k
70k
72k
74k
76k
78k
80k
80k
82k
84k
86k
88k
90k
92k
94k
96k
98k
100k

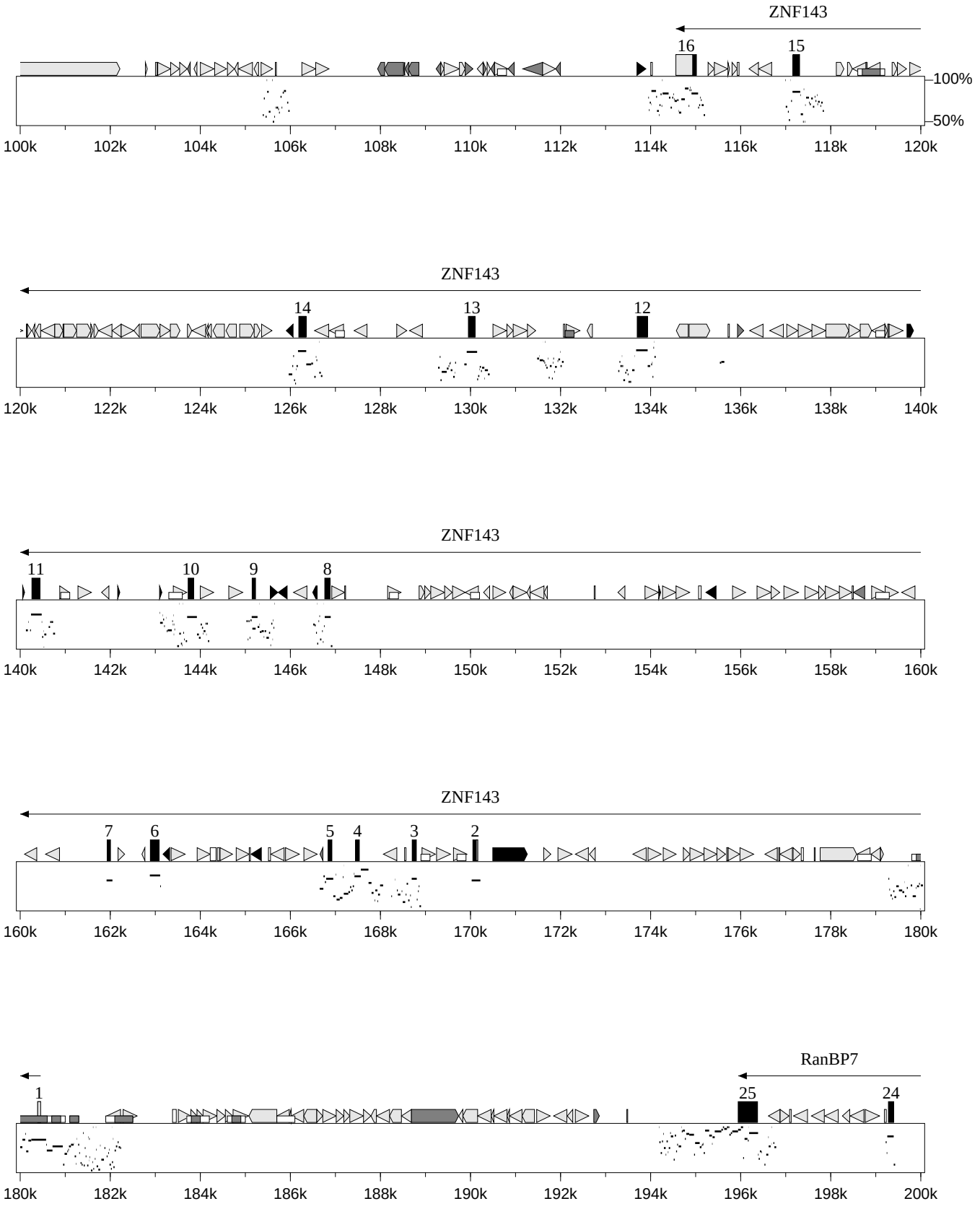
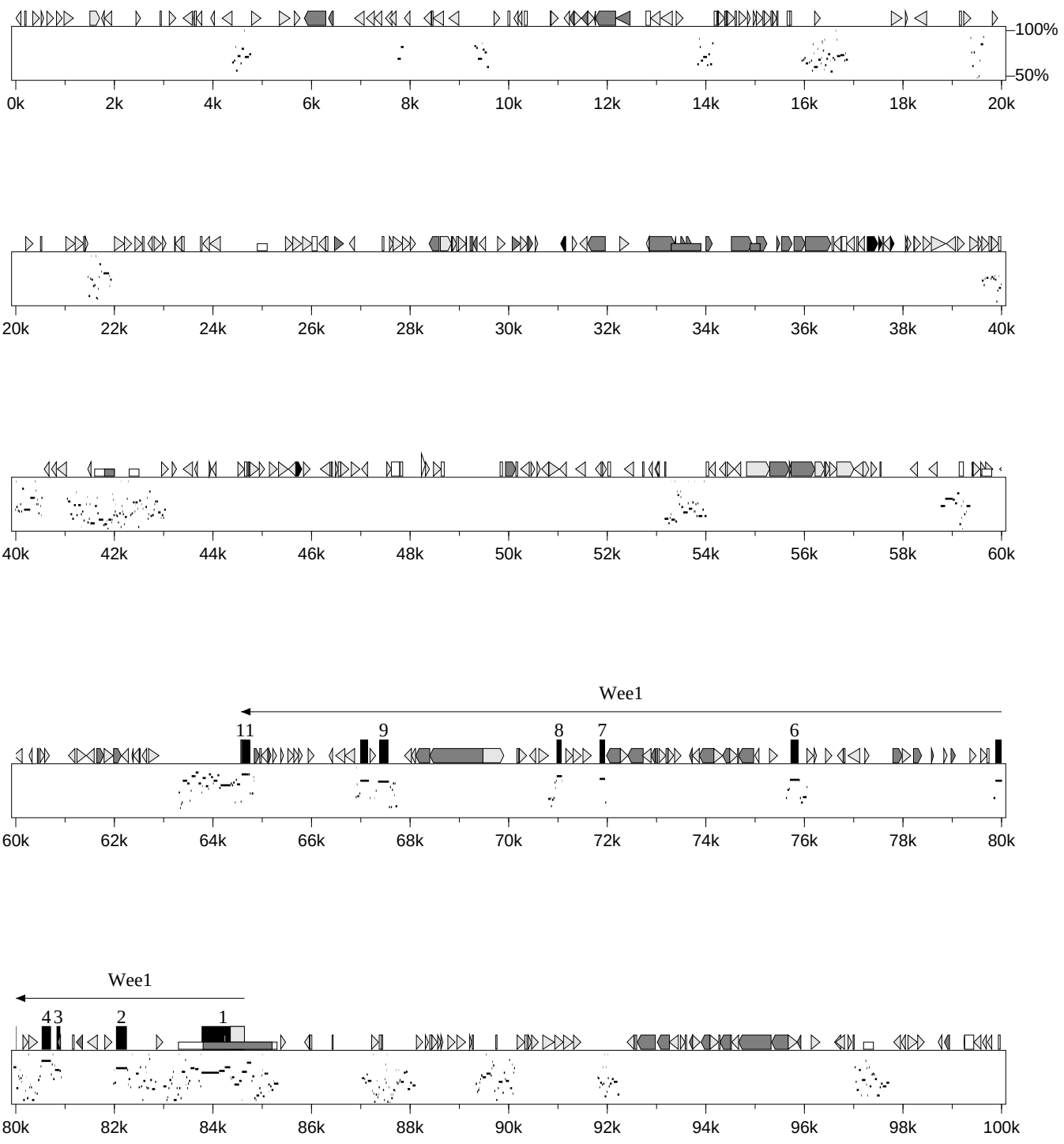
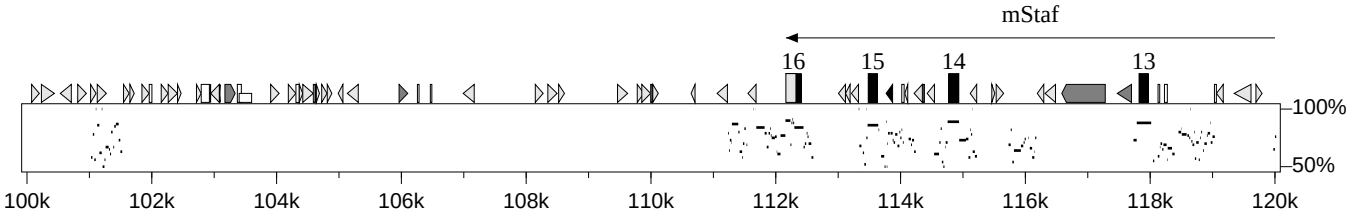
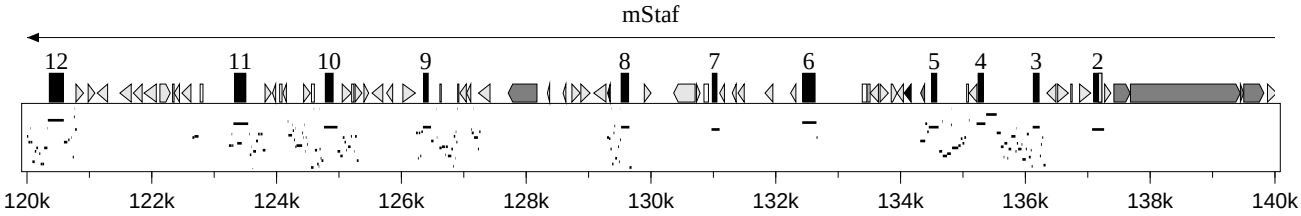
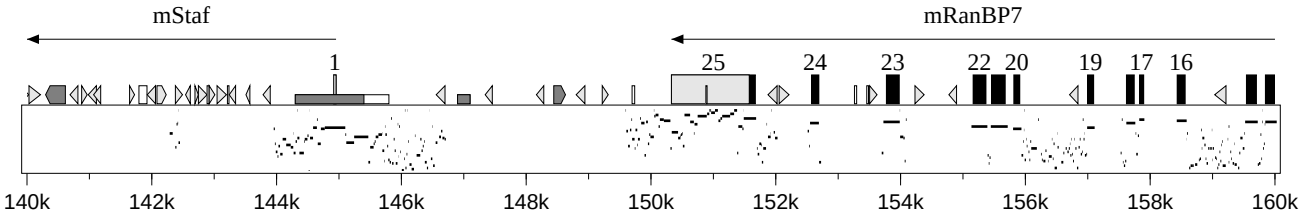

ZNF143
16
15
100%
50%
100k
102k
104k
106k
108k
110k
112k
114k
116k
118k
120k
ZNF143
14
13
12
120k
122k
124k
126k
128k
130k
132k
134k
136k
138k
140k
ZNF143
11
10
9
8
140k
142k
144k
146k
148k
150k
152k
154k
156k
158k
160k
ZNF143
7
6
5
4
3
2
160k
162k
164k
166k
168k
170k
172k
174k
176k
178k
180k
RanBP7
1
25
24
180k
182k
184k
186k
188k
190k
192k
194k
196k
198k
200k

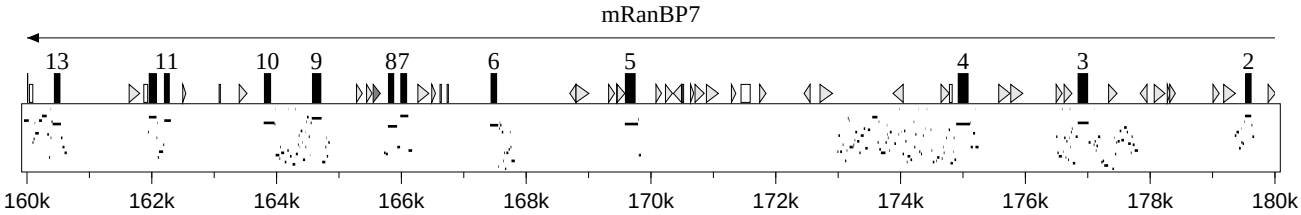
Abb. 3.18


100%
50%
0k
2k
4k
6k
8k
10k
12k
14k
16k
18k
20k
22k
24k
26k
28k
30k
32k
34k
36k
38k
40k
42k
44k
46k
48k
50k
52k
54k
56k
58k
60k
Wee1
11
9
8
7
6
62k
64k
66k
68k
70k
72k
74k
76k
78k
80k
Wee1
4
3
2
1
82k
84k
86k
88k
90k
92k
94k
96k
98k
100k


mStaf
16
15
14
13
100%
50%
100k
102k
104k
106k
108k
110k
112k
114k
116k
118k
120k


mStaf
12
11
10
9
8
7
6
5
4
3
2
120k
122k
124k
126k
128k
130k
132k
134k
136k
138k
140k


mStaf
mRanBP7
1
25
24
23
22
20
19
17
16
140k
142k
144k
146k
148k
150k
152k
154k
156k
158k
160k


mRanBP7
13
11
10
9
87
6
5
4
3
2
160k
162k
164k
166k
168k
170k
172k
174k
176k
178k
180k

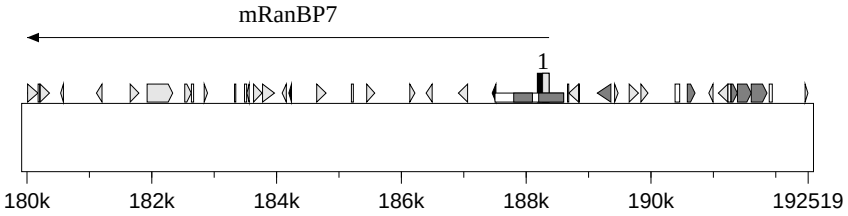
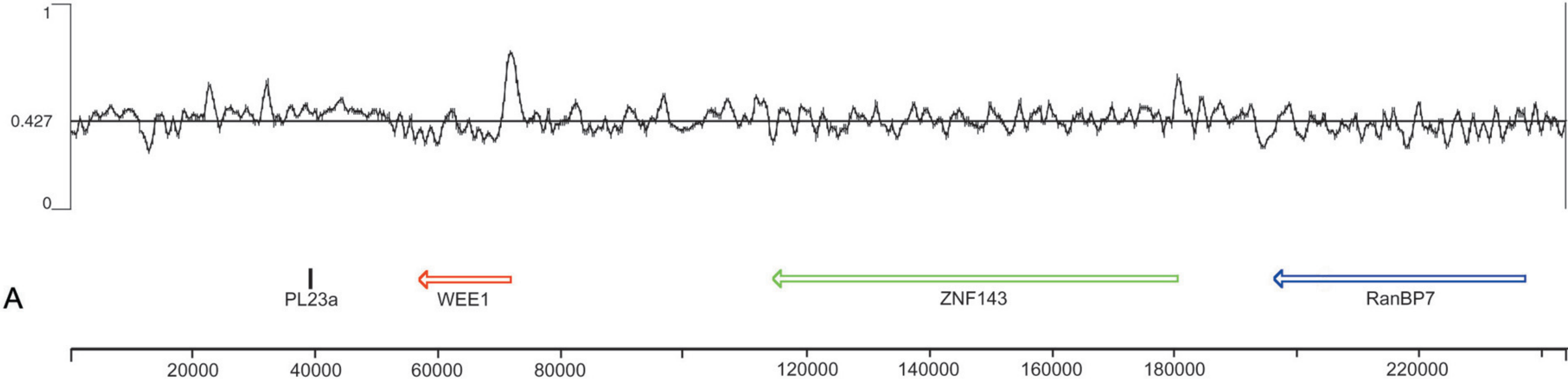
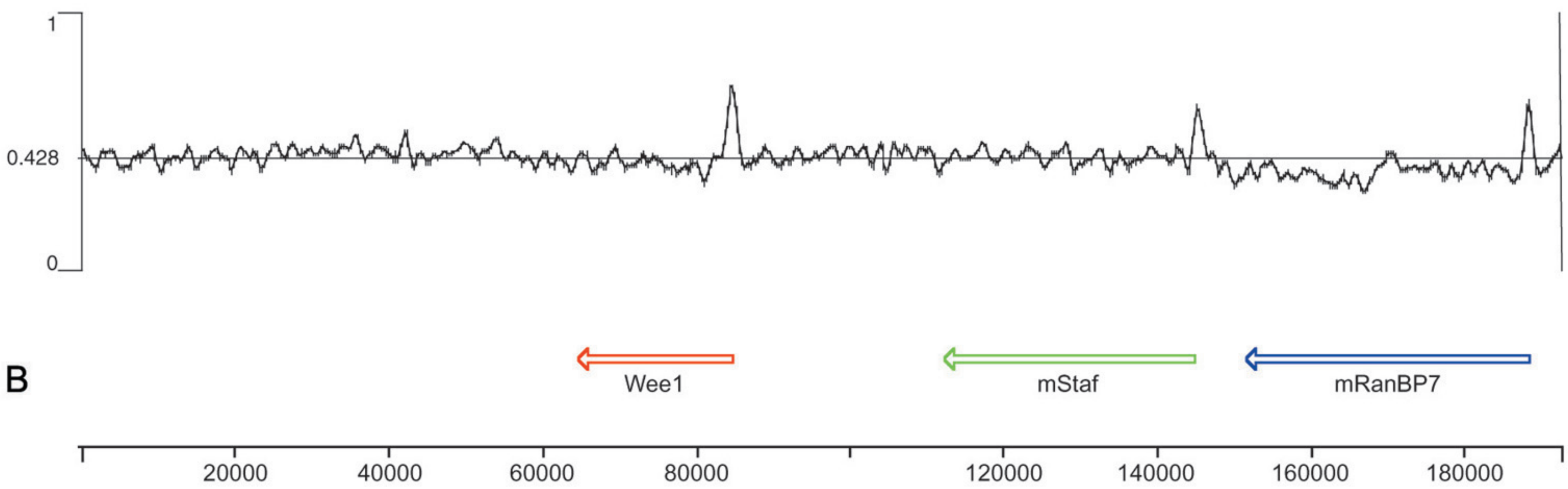

mRanBP7
1
180k
182k
184k
186k
188k
190k
192519

Abb. 3.19


1
0.427
0
A
PL23a
WEE1
ZNF143
RanBP7
20000
40000
60000
80000
120000
140000
160000
180000
220000


1
0.428
0
B
Wee1
mStaf
mRanBP7
20000
40000
60000
80000
120000
140000
160000
180000

Ich versichere hiermit, die vorliegende Arbeit selbständig und nur mit den angegebenen Hilfsmitteln angefertigt zu haben.

Andrea Cichutek