# Systematische Analyse von Transkripten aus humaner fetaler Wachstumsfuge -Klonierung und Charakterisierung von Kandidatengenen der Zelldifferenzierung

Dissertation

Zur Erlangung des Doktorgrades des Fachbereichs Biologie der Universität Hamburg

> vorgelegt von Gerrit Mohrmann aus Hamburg

Hamburg, 2003

Genehmigt vom Fachbereich Biologie der Universität Hamburg auf Antrag von Herm Professor Dr. A. WINTERPACHT Weitere Gutachter der Dissertation: Herr Professor Dr. H.-P. MÜHLBACH

Tag der Disputation: 12. Dezember 2003

Hamburg, den 25. November 2003



Professor Dr. A. Frühwald Dekan

1.	EIN	LEITUNG	1
	1.1	Identifizierung und Charakterisierung differentiell exprimierter Gene	1
	1.2	Differenzierungsvorgänge und Chromatinstruktur	5
	1.3	Nukleosomenstruktur und Transkription	8
	1.4	Subnukleäre Domänen	12
	1.5	Zielsetzung	14
2.	MA	TERIAL UND METHODEN	15
2.	.1	Isolierung von DNA	15
	2.1.1	Isolierung von genomischer DNA aus Blut	15
	2.1.2	2 Isolierung kleiner Mengen Plasmid-DNA	15
	2.1.3	Isolierung von Plasmid-DNA für die Transfektion eukaryotischer Zellen.	16
2.	.2	Isolierung von RNA	16
	2.2.1	Isolierung von RNA aus Zellen	16
	2.2.2	2 Isolierung von RNA aus menschlichen und murinen Geweben	16
	2.2.3	Isolierung von RNA aus menschlichen Knorpelgewebe	17
	2.2.4	DNase I-Behandlung von RNA	17
2.	.3	DNA und RNA Standardmethoden	17
	2.3.1	Spaltung von DNA mit Restriktionsendonukleasen	17
	2.3.2	2 Agarose-Gelelektrophorese von DNA und RNA	18
	2.3.3	8 Wiedergewinnung von DNA aus Agarosegelen	18
	2.3.4	Fällung von DNA	19
	2.3.5	5 Transfer von RNA auf Membranen (Northern Blot)	19
	2.3.6	5 DNA Sequenzierung und Mutationsanalyse von SPOC1	19
2.	.4	Klonierung von DNA-Fragmenten in Plasmid-Vektoren	20
	2.4.1	Klonierung von PCR-Produkten	20
	2.4.2	2 Subklonierung von DNA Fragmenten	20
	2.4.3	B Ligation	20
	2.4.4	Transformation und Selektion positiver Klone	21
2.	.5	DNA Amplifikation durch Polymerasekettenreaktion (PCR)	21
2.	.6	Reverse Transkriptase-Polymerasekettenreaktion (RT-PCR)	22
2.	.7	Radioaktive und nichtradioaktive Markierungstechniken	22
	2.7.1 2.7.2	Markierung von DNA über "random primed oligonucleotide labelling" Endmarkierung von Oligonukleotiden	22 22
2.	.8	DNA-RNA-Hybridisierung (Northern-Hybridisierung)	23
2.	.9	RNA in situ-Hybridisierung von Gewebeschnitten	23
	2.9.1	Herstellung von Schnittpräparaten	23

2.10	2.10 Genomische Lokalisation eines murinen Gens	
2.11	Protein Standardmethoden	25
2.11	Protein-Extraktion aus eukaryotischen Zellen	25
2.11	1.2 SDS-Polyacrylamid-Gelelektrophorese (SDS-PAGE)	25
2.11	1.3 Proteintransfer (Western Blot)	26
2.11	1.4 Immunologischer Nachweis von Proteinen	26
2.11	1.5 Proteinquantifizierung	26
2.12	Zellbiologische Methoden	27
2.12	2.1 Kultur von adhärent wachsenden Zellen	27
2.12	2.2 Einfrieren und Auftauen von Zellen	27
2.12	2.3 Transfektion von adhärenten Zellen mit Lipofectamine <sup>™</sup> 2000 Reagenz	28
2.12	2.4 Herstellung und Verifizierung eines anti-SPOC1-Peptidantikörpers	28
2.12	2.5 Immunfluoreszenz und EGFP-Fluoreszenz	29
2.13	Luziferase Reporter Assays	30
2.14	Computerauswertung von ESTs einer cDNA-Bibliothek	31
2.15	Reagenzien und Materialien	32
2.13	L Duffen und Läsungen	
2.15	5.1 Puller und Losungen	32 25
2.15	5.2 Baktenenstalline	33 25
2.15	5.5 Eukaiyousene Zenen	35 n 35
2.15	5.5 Geräte	37
2.15	5.6 Primärantikörper	38
2.15	5.7 Sekundärantikörper	
2.15	5.8 Synthetische Oligonukleotide	39
3. ER	GEBNISSE	41
3.1	Auswertung und Charakterisierung von EST-Klonen einer humanen	
	Chondrozyten cDNA-Bibliothek	41
2.1	1 Bioinformatische Auswertung von ESTs mit dem Datenhank Analyse Tor	-1
5.1.	FSTsween"	л //3
31	2 Auswertung der Klone aus der Chondrozyten cDNA-Bank	<del>4</del> 5 46
3.1	2 FST-Klone mit Homologien zu Sequenzen bekannter Funktion	40
5.11	(Kategorie I)	49
3.1	2.2 EST-Klone mit Homologien zu Sequenzen unbekannter Funktion oder	17
5.1.	genomischen Sequenzen (Kategorien II und III)	
3.1.	3 Analyse von 4 EST-Klonen aus der Chondrozyten cDNA-Bibliothek	50
3.1.	3.1 Analyse des Klons T01 203	51
3.1.	3.2 Analyse des EST-Klons T01 517	52
3.1.	3.3 Analyse des EST-Klons T01 519	53
3.1.	3.4 Analyse des Klons T01 564	54

3.2	Klonierung und Chromosomale Lokalisation des humanen und murinen	
	SPOC1 (Spoc1) -Gens	56
3.2	Klonierung und Sequenzierung des menschlichen SPOCI-Gens	56
3.2	Klonierung und Sequenzierung des murinen <i>Spoc1</i> -Gens	
3.2	Chromosomale Lokalisation des humanen SPOCI-Gens	60
3.2	2.4 Chromosomale Lokalisation des murinen <i>Spoc1</i> -Gens	61
3.2	Homologievergleiche des humanen und murinen SPOC1-Gens	63
3.2	Promotorsuche im humanen und murinen <i>SPOC1</i> -Gen	65
3.3	Analyse der Expression des humanen und murinen SPOC1-Gens	68
3.3	5.1 SAGEmap-Datenbank-Analyse der Expression des humanen SPOC1-Ger	1s68
3.3	Northern-Analysen des humanen SPOC1-Gens	68
3.3	Northern-Analysen des murinen <i>Spoc1</i> -Gens	71
3.3	RNA-in situ-Hybridisierung des murinen <i>Spoc1</i> -Gens	72
3.4	Subzelluläre Lokalisation des SPOC1-Proteins	74
3.4	.1 Subzelluläre Lokalisation des SPOC1-EGFP-Fusionsproteins	75
3.4	.2 Subzelluläre Lokalisation des FLAG-SPOC1-Fusionsproteins	76
3.4	.3 Spezifität des SPOC1 #10 Peptidantikörpers	77
3.4	.4 Kolokalisation von SPOC1 mit RNA-Polymerase II und E2F-1	81
3.5	Effekt von SPOC1 auf die Transkription	83
3.6	Mutationsanalyse des humanen SPOC1-Gens	85
		~
4. DI	SKUSSION	87
4.1	Analyse und Homologievergleich von EST-Klonen aus einer humanen	
	Chondrozyten-cDNA-Bibliothek	87
4.1	.1 ESTs mit Homologien zu Genen bekannter Funktion	88
4.1	.2 ESTs mit Homologien zu Sequenzen unbekannter Funktion oder	
	genomischen Sequenzen	90
4.2	Vergleich des EST-Projekts mit anderen EST-Projekten	91
4.3	Präliminäre Untersuchung einiger ausgewählter ESTs	92
4.4	Detaillierte Analyse und Charakterisierung des SPOC1-Gens	93
4.4	.1 Expression des humanen und murinen SPOC1-Gens	94
4.4	.2 Subzelluläre Lokalisation des SPOC1-Gens	96
4.4	.3 Funktion der PHD-Domäne bei der Regulation der Transkription	99
4.4	.4 Bedeutung der PHD-Domäne bei der Entstehung von Krankheiten	102
4.4	.5 Expression von SPOC1 in Ovarialkarzinomen	105
4.4	.6 Modell der Funktion von <i>SPOC1</i> in der Zelle	106
5. ZU	JSAMMENFASSUNG	109
7 11	τεριατιστερατιστικίς	111
0. LI	Ι ΕΝΑΙ UK VEKLEIUΠΝΙζ	111

	Inhaltsverzeichnis	IV
7.	ANHANG	

# Verzeichnis der Abbildungen

Abb. 1.1:	Darstellung der chromosomalen Domänen im Zellkern eines Huhn- Fibroblasten	
Abb. 1.2:	Modell der Regulation der Chromatinstruktur	11
Abb. 2.1:	Luziferase-Reporter-Assay	30
Abb. 3.1:	Struktur des EST-Projektes	42
Abb. 3.2:	Startseite von "ESTsweep" auf dem HUSAR-Server	44
Abb. 3.3:	Automatisierte BlastN-Analysen von ESTs mit dem Datenbank Analyse-Tool "ESTsweep"	45
Abb. 3.4:	Ausgabe der Ergebnisse einer "ESTsweep"-Analyse im Tabellenformat	46
Abb. 3.5:	Verteilung der EST-Klone des Gesamtprojektes	47
Abb. 3.6:	Verteilung der 3426 Klone mit Homologien zu bekannten Genen (Kategorie I).	49
Abb. 3.7:	Klonierungsstrategie des menschlichen SPOC1-Gens	57
Abb. 3.8:	Struktur des humanen SPOC1-Gens	58
Abb. 3.9:	Klonierungsstrategie des murinen Spoc1-Gens	59
Abb. 3.10:	Struktur des murinen Spoc1-Gens	60
Abb. 3.11:	Chromosomale Lokalisation und Orientierung des humanen SPOC1- Gens	61
Abb. 3.12:	Chromosomale Lokalisation des murinen Spoc1-Gens	63
Abb. 3.13:	Alignment der abgeleiteten Aminosäuresequenzen des humanen und murinen SPOC1-Proteins	64
Abb. 3.14:	MAP-Alignment der PHD-Domäne von <i>SPOC1</i> mit den entsprechenden Domänen verschiedener humaner Proteine (A); Konsensus Sequenz der PHD-Domäne (B)	65
Abb. 3.15:	Sequenzausschnitt aus dem humanen SPOC1-Gen	66
Abb. 3.16:	Sequenzausschnitt aus dem murinen Spoc1-Gen	67

V

Abb. 3.17:	Northern-Blot-Analyse des humanen <i>SPOC1</i> -Gens (A); "Human Multiple Tissue Northern-Blot" (B)	69
Abb. 3.18:	"Human Multiple Tissue Expression Array"	70
Abb. 3.19:	Northern-Analyse des murinen <i>Spoc1</i> -Gens an Gesamt-RNA aus murinen Testis-Gewebe (A); "Mouse Multiple Tissue Expression Array" (B)	71
Abb. 3.20:	RNA <i>in situ</i> -Hybridisierung zur Analyse der <i>Spoc1</i> -Gen Expression in Kryoschnitten aus Testis- und Uterus-Gewebe aus adulten Balb/C Mäusen	73
Abb. 3.21:	Western-Blot-Analyse der Expression des SPOC1-EGFP-Fusionsproteins in U2OS-Zellen	75
Abb. 3.22:	Subzelluläre Lokalisation des <i>SPOC1</i> -EGFP- und FLAG- <i>SPOC1</i> - Fusionsproteins in U2OS-Zellen	76
Abb. 3.23:	Expression des FLAG-SPOC1-Fusionsproteins in U2OS-Zellen	77
Abb. 3.24:	Detektion des FLAG- <i>SPOC1</i> -Fusionsproteins durch den polyklonalen SPOC1-#10-Peptidantikörper in U2OS-Zellen	78
Abb. 3.25:	Detektion des pM2- <i>SPOC1</i> -Fusionsproteins durch den polyklonalen SPOC1-#10-Peptidantikörper in U2OS-Zellen	79
Abb. 3.26:	Subzelluläre Lokalisation von endogen exprimierten SPOC1-Protein In U2OS-Zellen	79
Abb. 3.27:	Spezifität des polyklonalen SPOC1-#10-Peptidantikörpers in U2OS- Zellen und DeuserPT-Zellen	80
Abb. 3.28:	Partielle Kolokalisation von RNA-Polymerase II und SPOC1 im Zellkern von DeuserPT-Zellen	81
Abb. 3.29:	Signifikante partielle Co-Lokalisation von Transkriptionsfaktor <i>E2F-1</i> und <i>SPOC1</i> in DeuserPT-Zellen	82
Abb. 3.30:	Western-Blot-Analyse der Expression des Gal4-SPOC1-Fusionsproteins in 293T-Zellen	83
Abb. 3.31:	Gal4-Luziferase-Reporter-Assay des Gal4-SPOC1-Fusionsproteins in 293T-Zellen	84
Abb. 3.32:	Nukleotidaustausch (A $\rightarrow$ G) an der Position 305 des <i>SPOC1</i> -Leserahmens	86
Abb. 4.1:	Vergleich der prozentualen Anteile der ESTs der Kategorie I und der zu einer Gruppe zusammengefassten ESTs der Kategorien II und III mit der Gesamtzahl der untersuchten ESTs	91

	Abbildungsverzeichnis	VII
Abb. 4.2:	Konsensus Sequenz der PHD-Domäne (A); Variabilität der 3 Sequenz- abschnitte L1-3 der PHD-Domäne	102
Abb. 4.3:	Modell der putativen Funktion von SPOC1 in der normalen Entwicklung und in der Tumorentstehung	108

# Verzeichnis der Tabellen

Tab. 2.1:	Zusammensetzung der verschiedenen Polyacrylamid-Trenngele	25
Tab. 2.2:	Für die Transfektionsansätze verwendete Mengen an DNA und Lipofektamine 2000	28
Tab. 2.3:	Für die Kotransfektion von 293T-Zellen verwendete Mengen an Reporterplasmid, Fusionskonstrukt, pCDNA3, pLacZ und Lipofectamine 2000	31
Tab. 3.1:	Auszug aus der Ergebnis-Tabelle der bioinformatischen Analyse der EST-Klone	48
Tab. 3.2:	Zusammenfassung der UniGene und BlastN/X-Auswertungen von 4 ausgewählten EST-Klonen aus der Chondrozyten cDNA-Bibliothek	55
Tab. 3.3:	Zusammenfassung der bei der Mutationsanalyse des SPOC1-Gens ermittelten Nukleotidaustausche	86
Tab. 4.1:	Einteilung der EST-Klone, die Homologien zu bekannten oder funktionell charakterisierten cDNA-Sequenzen oder Proteinen zeigen	89

# Abkürzungsverzeichnis

Abb.	Abbildung
A. bidest.	Aqua bidestillata
abs.	absolut
Ac	Acetat
AS	Aminosäure
Вр	Basenpaare
BSA	Rinderserumalbumin
cDNA	komplementäre DNA
Ci	Curie
cM	centi-Morgan
cpm	"counts per minute"
C-terminal	carboxyterminal
DEPC	Diethylpyrocarbonat
DNA	Desoxyribonukleinsäure
dNTP	2'-Desoxynucleosid-5'-triphosphat
dpc	"days postconception"
DTT	Dithiothreitol
E.coli	Escherichia coli
EDTA	Ethylendiamintetraacetat
EST	"expressed sequence tag"
EtOH	Ethanol
FISH	"fluorescence in situ hybridization"
FITC	Fluoresceinisothiocyanat
ges.	gesättigt
h	Stunde
HEPES	2-[4-(2-Hydroxylethyl)-1-piperazinyl]-ethansulfonsäure
IPTG	Isopropyl-β-D-thio-galactopyranosid
kBp	Kilobasenpaare
kDa	Kilodalton
konz.	konzentriert
μ	mikro
М	molar
MBp	Megabasenpaare
min	Minute
MOPS	3-(N-morpholin)propansulfonsäure
N-terminal	aminoterminal
р	piko
PBS	"phosphate buffered saline"-Puffer
PCR	"polymerase chain reaction"
PEG	Polyethylenglycol
RNA	Ribonukleinsäure
RT	Reverse Transkriptase

RT-PCR	"reverse transcriptase polymerase chain reaction"
SDS	Natriumdodecylsulfat
S	Sekunde
SSC	"standard saline citrate"-Puffer
Т	Thymin
Tab.	Tabelle
TBE	Tris-Borat-EDTA-Puffer
TEMED	N,N,N',N'-Trismethyl-ethylendiamin
Tris	Tris(hydroxymethyl)aminomethan
TRITC	Tetramethylrhodaminisothiocyanat
U	"Units" (Enzymeinheiten)
Upm	Umdrehungen pro Minute
3'-UTR	3'-untranslatierte Region
5'-UTR	5'-untranslatierte Region
UV	ultraviolett
V	Volt
vgl.	vergleiche
Vol.	Volumen
X-Gal	5-Bromo-4-chloro-3-indolyl-β-D-galactosid
WT	Wildtyp

# 1. Einleitung

Neben Wachstum und Morphogenese ist die Differenzierung von Zellen eine der drei wesentlichen Phasen während der Entwicklung höherer Organismen. Genetisch-bedingte Fehler in Differenzierungsprozessen können zu zahlreichen Erkrankungen führen, die neben Funktionsstörungen von Organen auch Tumorerkrankungen umfassen. Differenzierung beschreibt die Entstehung von hunderten unterschiedlicher, spezialisierter Zelltypen aus einer einzigen befruchteten Eizelle, was letztlich erst zu Bildung funktionsfähiger Gewebe und Organe führt.

Die fetale Wachstumsfuge während der enchondralen Ossifikation, die für die Bildung und Wachstum des Skelettsystems wichtig ist, stellt ein ausgezeichnetes System dar, an dem Differenzierungsprozesse untersucht werden können. In räumlich geordneter Form liegen hier Zellen in verschiedenen Differenzierungsstadien vor, die von ruhenden (relativ undifferenzierten) Chondrozyten über proliferierende und reifende (hypertrophe) Chondrozyten bis hin zu apoptotischen Chondrozyten reichen. Die einzelnen Differenzierungsstadien sind immer mit der Expression bestimmter Sätze differentiell exprimierter Gene assoziiert.

### 1.1 Identifizierung und Charakterisierung differentiell exprimierter Gene

Die Identifizierung und Charakterisierung differenziell exprimierter Gene ist für das Verständnis entwicklungsbiologischer Fragestellungen und der Entstehung von Krankheiten von zentraler Bedeutung. Die funktionelle Analyse dieser Gene kann eine genauere Untersuchung von Differenzierungsprozessen ermöglichen und so zum Verständnis der molekularen Vorgänge bei Entwicklungsprozessen und der Pathogenese humaner Erkrankungen beitragen. Von besonderem Interesse ist hierbei auch die Beschreibung von neuen "Drug-Targets" oder diagnostischen Markern für verschiedene Erkrankungen. Zur effizienten Identifizierung und Charakterisierung von relevanten Genen werden verschiedene Hochdurchsatzmethoden, wie Z. B. die EST-Sequenzierung, cDNA-Microarray-Hybridisierung oder SAGE-Analyse angewendet. Ein erstes EST-Projekt wurde 1991 von Adams et al. beschrieben. Hierbei wird aus einem definierten Gewebe RNA isoliert und eine cDNA-Bibliothek hergestellt, die zum Teil normalisiert werden (d. h. häufig vorkommende Transkripte werden eliminiert).

Durch Isolierung und Einzelstrang-Sequenzierung willkürlich ausgewählter cDNA-Klone kann eine große Anzahl an Sequenzen mit einer durchschnittliche Länge von 300-400 Bp generiert werden, die jeweils ein bestimmtes Transkript (EST, expressed sequence tag) repräsentieren. Die anschließende bioinformatische Auswertung der ESTs mit Hilfe der ESTund nr-Datenbanken erlaubt eine effiziente Identifikation und Analyse von exprimierten Kandidatengenen (Schuler et al., 1996). Mittlerweile sind in der öffentlichen dbEST-Datenbank (NCBI) mehr als 5.000.000 humane und 3.710.187 murine ESTs aus verschiedenen Geweben und Zelltypen abgelegt (NCBI, release 11.04.03). Die drastisch anwachsende Menge an Sequenzdaten des Humangenomprojekts sowie weiterer großer EST-Projekte (Wiemann et al., 2001; Yudate et al., 2001; Jia et al., 2001) trägt zu einer erhöhten Effizienz bei der Identifizierung von relevanten Genen bei. So enthält die "UniGene Collection", bei der es sich um eine EST-Datenbank handelt, mehr als 48.000 EST-Cluster, bestimmtes von denen jeder einzelne ein Transkript repräsentiert (http://www.ncbi.nlm.nih.gov/Web/Newsltr/aug96.html#advance). Die UniGene-Datenbank enthält 7224 cDNA-Bibliotheken, die für EST-Projekte verwendet wurden. Von diesen cDNA-Bibliotheken wurden 4 Knorpelgewebe jedoch nur aus erstellt (http://www.ncbi.nlm.nih.gov/UniLib/). Weil Gewebe aus wachsenden Knorpel/Knochen eine sehr seltene und problematische Quelle für die Herstellung von cDNA-Bibliotheken darstellen, sind diese bisher nur schlecht charakterisiert worden. Aus diesem Grund bietet eine cDNA-Bibliothek aus humanen fetalen Knorpel eine exklusive Möglichkeit, bisher unbekannte Gene zu identifizieren, die an Differenzierungsprozessen beteiligt sind.

Im Rahmen von EST-Projekten ist es aufgrund der großen Anzahl von ESTs in den verschiedenen öffentlichen Datenbanken häufig möglich, durch überlappende Anordnung von EST-Sequenzen die komplette cDNA-Sequenz eines Transkripts zu identifizieren. Homologievergleiche mit genomischen Sequenzen können darüber hinaus Aussagen über die Intron/Exon Struktur, sowie über alternative und gewebespezifisch-exprimierte Splicevarianten eines Transkripts ermöglichen. Da eine große Anzahl von cDNA-Bibliotheken über Oligo-dT-Priming synthetisiert wird, stellen 65% der in den Datenbanken veröffentlichten EST-Sequenzen das 3'-Ende und somit häufig den 3'-untranslatierten Bereich einer cDNA dar. Nur 26% der EST-Sequenzen repräsentieren das 5'-Ende (Schuler et al., 1997). Um dieses Problem zu umgehen, konzentrieren sich aktuelle Projekte darauf "fulllength"-cDNAs zu generieren und eine sorgfältige Auswahl der Gewebe oder Zelltypen zu treffen, die für die Herstellung von cDNA-Bibliotheken verwendet werden (Dias-Neto et al., 2000). Die Verwendung neuer Methoden, wie die der Laser-Mikrodissektion, ermöglicht es,

kleine Gruppen von Zellen oder einzelne Zellen zu isolieren, die für die Konstruktion einer cDNA-Bibliothek verwendet werden können (Emmert-Buck et al., 1996; Mikulowska-Mennis et al., 2002). Schwierigkeiten bei EST-Projekten betreffen vor allen Dingen Kontaminationen. So kommen cDNAs, welche für Genprodukte kodieren, die in Blutzellen vorkommen, in vielen cDNA-Bibliotheken vor, weil die verwendeten Gewebeproben mit Blutresten kontaminiert sind (Schuler et al., 1997).

Ein weiteres Problem, welches bei der Durchführung von EST-Projekten auftritt, ist die hohe Redundanz bestimmter mRNAs, die im starken Gegensatz zum geringen Vorkommen anderer mRNAs steht. Untersuchungen haben ergeben, dass die mRNAs einer typischen somatischen Zelle in 3 Gruppen (1, 2 und 3) einteilbar sind, die in bestimmten Frequenzen vorkommen (Davidson und Britten, 1979). Durchschnittlich sind in den Gruppen 1 und 2 ungefähr 10 verschiedene mRNA-Spezies mit je ca. 5000 Kopien pro Zelle vertreten. Die Gruppe der komplexen mRNAs (3) ist mit ca. 15.000 verschiedenen Spezies pro Zelle vertreten, von denen jede in ca. 1 bis 15 Kopien pro Zelle vorhanden ist (Soares et al., 1994). Aufgrund des Missverhältnisses in der Anzahl verschiedener mRNA-Spezies in einer cDNA-Bank kann die Identifizierung seltener Transkripte, trotz einer Normalisierung oder Subtraktion der cDNA-Bank mit bereits bekannten cDNA-Klonen, sehr problematisch sein und erfordert die Sequenzierung einer sehr großen Anzahl von EST-Klonen (Soares et al., 1994; Bonaldo et al., 1996). EST-Projekte sind die Basis für eine Reihe weiterer Hochdurchsatzverfahren zur Identifizierung differentiell exprimierter Gene. Die Aufklärung der Funktion von Kandidatengenen kann wirkungsvoll mit der Mikroarray-Hybridisierung verknüpft werden, die Analyse und Vergleich gewebespezifischer und krankheitsspezifischer Expressionsprofile ermöglicht.

Bei der cDNA-Mikroarray Hybridisierung handelt es sich um eine Hochdurchsatzmethode, bei der eine große Anzahl von cDNA-Sequenzen oder Oligonukleotiden auf einem Glas-Objektträger immobilisiert werden. Durch eine anschließende Hybridisierung mit unterschiedlichen fluoreszenzmarkierten RNA-Sonden kann die differenzielle Expression eines Transkripts analysiert und verifiziert werden (Schena et al, 1995; 1996; Lockhart und Winzeler, 2000). Ein besonderer Vorteil dieser Methode ist, dass durch Herstellung mehrerer Objektträger der gleiche Satz von Transkripten unter verschiedenen Bedingungen und mit verschiedenen Sonden untersucht werden kann. Die Mikroarray-Hybridisierung kann im Anschluss an ein EST-Projekt dazu dienen, die Expression identifizierter Kandidatengene in verschiedenen Geweben zu Untersuchen oder zu vergleichen (Lockhart et al., 1996). Die SAGE-Technik ("serial analysis of gene expression"), basiert auf der Herstellung und quantitativen Auswertung von kurzen, mRNA-spezifischen 12 Bp-Sequenzabschnitten (Velculescu et al., 1995). Die Herstellung einer SAGE-Bibliothek geht ebenfalls von einer cDNA-Bibliothek aus, wobei jedes der generierten 12 Bp-Fragmente ("Tags") ein spezifisches Transkript der cDNA-Bibliothek repräsentiert. Ein wesentlicher Vorteil der SAGE-Methode ist, dass eine sehr große Anzahl an Transkripten untersucht werden kann und qualitative sowie quantitative Aussagen möglich sind (Velculescu et al., 1997). Die Erstellung von SAGE-Bibliotheken erlaubt es Expressionsprofile verschiedener Gewebe- und Zelltypen miteinander zu vergleichen und so Aussagen über Unterschiede in der Gen-Expression von Tumorgewebe und gesunden Gewebe zu ermöglichen (Zhang et al., 1997). Aktuelle Untersuchungen haben gezeigt, dass die Erstellung von neuen SAGE-Bibliotheken eine effektive Identifikation von neuen Transkripten und Genen ermöglicht, obwohl eine große Anzahl von SAGE-Tags in den öffentlichen Datenbanken vorliegt (Chen et al., 2002). Ein Problem der Analyse von SAGE-Bibliotheken ist, dass ein SAGE-Tag, welches ein spezifisches Transkript repräsentiert, aufgrund seiner geringen Länge Sequenz-Homologien zu verschiedenen Sequenzen in öffentlichen Datenbanken zeigen kann. Das Problem der Identifikation der richtigen Sequenzen kann jedoch durch die Generierung längerer SAGE-Fragmente umgangen werden, die genauere Homologievergleiche ermöglichen (Chen et al., 2000). Viele SAGE-Bibliotheken können inzwischen über die öffentliche SAGEmap-Datenbank des NCBI genutzt werden (http://www.ncbi.nlm.nih.gov/SAGE/).

Alle Hochdurchsatz-Methoden produzieren eine große Datenmenge, deren Auswertung und Interpretation sehr komplex ist. Die Identifizierung von relevanten Kandidatengenen ist daher sehr schwierig. Vorteilhaft ist an diesen Methoden, dass sie die simultane Analyse oder den Vergleich einer großen Anzahl von Proben zu einem bestimmten Zeitpunkt oder über eine spezifische Periode hinweg erlauben (Lockhart und Winzeler, 2000). Weiterhin ist es möglich komplexe Zusammenhänge, wie z. B. die Interaktionen in Gen-Netzwerken zu untersuchen (Lockhart und Winzeler, 2000; Kanehisa und Bork, 2003). Die bioinformatische Auswertung der Daten ermöglicht eine effektive Analyse, Identifikation und Charakterisierung von medizinisch relevanten Genen und Gen-Netzwerken, die an Entwicklungs- und Differenzierungsprozessen beteiligt sind.

## 1.2 Differenzierungsvorgänge und Chromatinstruktur

Während der Interphase liegen bestimmte Regionen des Chromatins in einem offenen, ausgebreiteten Zustand vor, der verschiedenen Proteinen, die mit der Genexpression assoziiert sind, Zugang ermöglicht. Das Chromatin eines Interphasechromosoms liegt jedoch nicht in allen Bereichen im gleichen Verpackungsgrad vor. So kommen transkriptionell aktive Bereiche in einer dekondensierten Form vor, die als Euchromatin bezeichnet wird, während die am stärksten kondensierte Form von Interphasechromatin als Heterochromatin bezeichnet wird. Heterochromatin ist wie Mitosechromatin transkriptionell inaktiv und besitzt die Fähigkeit transkriptionell aktive genomischer Bereiche, die in räumliche Nähe kommen, in einen transkriptionsinaktiven Zustand zu überführen (Henikoff und Dreesen, 1989; Brown et al., 1997; Francastel et al., 1999). Somit besteht ein direkter Zusammenhang zwischen der räumlichen Organisation des Chromatins und transkriptioneller Aktivität. Die Fähigkeit einer eukaryotischen Zelle, die Packungsdichte des Chromatins zu regulieren, wird während der Interphase zur Kontrolle der Genexpression eingesetzt. Untersuchungen deuten darauf hin, dass sich die Chromatinstruktur bei der Aktivierung der Transkription von eukaryotischen Genen öffnet und so die Interaktion von Promotor- und Enhancer-Elementen mit ubiquitären und gewebespezifischen Proteinen und Proteinkomplexen ermöglicht wird (Felsenfeld, 1996<sup>1,2</sup>; Kramer et al., 1998; Francastel et al., 2000; Labrador und Corces, 2002).

Die Ausprägung verschiedener Zell-Phänotypen mit spezifischen Funktionen ist bei der Entwicklung von Vertebraten von essentieller Bedeutung (Francastel et al., 2000; Mostoslavsky et al., 2003). Bei der Zelldifferenzierung handelt es sich um einen Vorgang, bei dem undifferenzierte Zellen durch die Aktivierung Zelltyp-spezifischer Gene einen neuen gewebespezifischen Phänotyp erlangen. Gene, die wichtig für die Aufrechterhaltung der Identität einer undifferenzierten Zelle sind, werden dabei inaktiviert, während Zelltyp spezifische Gene aktiviert werden (Francastel et al., 2000). Damit eine undifferenzierte Zelle ihre festgelegte, gewebespezifische Identität erlangen kann, unterliegt die Etablierung eines gewebespezifischen Expressionsmusters einer sehr genauen Kontrolle. In Vorläuferzellen liegen Gene, die für die Differenzierung verschiedener Zelltypen benötigt werden, in einer dekondensierten, offenen Chromatinstruktur vor, die Interaktionen mit gewebespezifischen Transkriptionsregulatoren erlaubt (Kramer et al., 1998). Beispielsweise liegt die Chromatinstruktur von hämatopoetischen Stammzellen vor der Festlegung für die Differenzierung in einen spezifischen Zelltyp in einer offenen Konformation vor. Diese erlaubt die Expression von Genen, die für die Möglichkeit der Differenzierung in verschiedene spezifische Zelltypen wichtig sind (Jimenez et al., 1992; Hu et al., 1997). Nach der Festlegung eines Differenzierungsschritts hin zu einem spezifischen Zelltyp (bestehend aus Spezifikation und Determination) kommt es zur Aktivierung von Zelltyp-spezifischen Genen, die im Zusammenhang mit der Inaktivierung von Chromatin-Domänen steht, die für andere Zelltypen spezifisch sind (Jimenez et al., 1992; Hu et al., 1997; Francastel et al., 2000). Es handelt sich hierbei vermutlich um einen generellen Mechanismus der Gen-Regulation bei der Zelldifferenzierung (Francastel et al., 2000). Untersuchungen deuten darauf hin, dass neben Chromatinverpackung die räumliche Organisation Reorganisation der und der Chromatinstruktur im Zellkern für die Modulation der **Gen-Expression** bei Differenzierungsprozessen von großer Wichtigkeit ist (Marshall et al., 1997; Lamond und Earnshaw, 1998; Cockell und Gasser, 1999; Cmarko et al., 2003). Es wird angenommen, dass Veränderungen der räumlichen Lokalisation von Gewebe-spezifischen Genen einen Einfluss auf die Interaktion mit regulatorischen Proteinen und dem Transkriptionsapparat haben. Der räumliche Kontext, in dem sich ein Gen befindet, könnte somit für die Repression oder Expression eines Gens verantwortlich sein (Francastel et al., 2000).

Ein wichtiges Element dieser Regulation ist die räumliche Einteilung des Interphase-Chromatins in unterschiedliche Kompartimente (Francastel et al., 2000), die durch verschiedene Mechanismen vermittelt wird. Jedes Chromosom liegt in einer eigenen subnukleären Domäne ("chromosome territory") vor (Schardin et al., 1985; Cremer et al., 1993; Ferreira et al., 1997; Zink et al., 1998), die nicht im Kontakt mit dem Chromatin von anderen Chromosomen steht (Visser et al., 2000; Verschure et al., 2002) (Abb. 1.1). Weiterhin existieren Hinweise darauf, dass die kondensierten chromosomalen Domänen abhängig von der Größe und Gen-Dichte der enthaltenen Chromosomen im Zellkern organisiert sind (Abb 1.1). So sind Chromosomen mit einer geringen Gen-Dichte mit der Peripherie des Zellkerns assoziiert, während Chromosomen mit einer größeren Gen-Dichte innerhalb des Zellkerns lokalisiert sind (Croft et al., 1999). Dabei sind die Gene innerhalb der kondensierten chromosomalen Domänen in transkriptionell aktiven und inaktiven Kompartimenten verteilt (Francastel et al., 2000). Innerhalb der chromosomalen Domänen ist das Chromatin so gefaltet, dass transkriptionsaktive Bereiche nahe der Oberfläche der kondensierten Chromatin-Domänen lokalisiert sind, die als Interchromatin Domäne bezeichnet wird (Zirbel et al., 1993; Cmarko et al., 1999; Verschure et al., 1999).



Abb. 1.1: Darstellung der chromosomalen Domänen im Zellkern eines Huhn-Fibroblasten mit Hilfe der "multicolour" Fluoreszenz *in situ*-Hybridisierung. Die Chromosomen 1, 2, 3, 4, 5, 6 und Z liegen im Zellkern in voneinander abgegrenzten Territorien vor. Die chromosomalen Domänen homologer Chromosomen sind in verschiedenen Regionen lokalisiert. Von den Chromosomen 4 und 6 ist in dieser Abbildung jeweils nur eine chromosomale Domäne sichtbar. Verändert nach Cremer und Cremer, 2001.

Nicht kodierende Sequenzen zeigen eine zufällige Verteilung oder liegen im inneren Bereich der kondensierten chromosomalen Domänen vor (Kurz et al., 1996; Verschure et al., 1999). Beispielsweise konnte gezeigt werden, dass kondensierte Chromatin-Domänen bei der Aktivierung von großen Gen-Clustern dekondensieren und in Kontakt mit dem transkriptionsaktiven Interchromatin-Kompartiment gelangen (Tumbar et al., 1999; Volpi et al., 2000; Ye et al., 2001). Es wird vermutet, dass nur die Bereiche des Chromatins, die in direkten Kontakt mit dem Interchromatin-Kompartiment stehen, transkriptionell aktiv sind (Verschure et al., 2002). Folglich findet die Transkription von Genen in der Zone zwischen den kondensierten Chromatin-Domänen und dem Interchromatin-Kompartiment statt, die als Perichromatin-Kompartiment bezeichnet wird (Krause et al., 1994). In Übereinstimmung mit dieser Vermutung konnte nachgewiesen werden, dass sich transkriptionell aktive Gen-Cluster auf Chromatin-Schleifen befinden, die aus der kondensierten Chromatin-Domäne hervortreten und mit dem Perichromatin-Kompartiment in Kontakt stehen (Volpi et al., 2000; Mahy et al., 2002). Neu synthetisierte RNA könnte so aus dem Perichromatin-Kompartiment direkt in das Interchromatin-Kompartiment abgegeben werden, in dem Prozessierung, Verpackung und Transport der RNA erfolgen (Verschure et al., 2002).

Wie oben ausgeführt, ist die Modulation der Chromatinstruktur ein wichtiger Bestandteil der Regulation der Transkription in Eukaryoten. Dabei wird die Transkription von Genen durch eine koordinierte Interaktion von bestimmten DNA-Abschnitten, wie Promotoren und Enhancern ("cis-transcriptional elements") mit Proteinen, die für die Transkription benötigt werden ("trans-acting factors"), reguliert (Francastel et al., 2000; Mostoslavsky et al., 2003). Verschiedene Studien zeigen, dass unterschiedliche Multiproteinkomplexe die Transkription, über Modifikationen von Histonen oder Veränderungen der Chromatinstruktur, regulieren (Clark-Adams et al., 1988; Han und Grunstein, 1988; Armstrong und Emerson, 1998; Kornberg und Lorch, 1999<sup>1</sup>; Kornberg und Lorch, 1999<sup>2</sup>). Damit Proteine, die für die Regulation der Transkription benötigt werden, an DNA binden können, muss die Chromatinstruktur durch Chromatin reorganisierende Proteine oder Proteinkomplexe in einen "offenen" Zustand überführt werden (Workman und Kingston, 1998). Umgekehrt wird die Repression der Transkription durch Proteinkomplexe vermittelt, welche die Chromatinstruktur in eine "geschlossene" Konformation überführen und so die Assoziation von Aktivatoren der Transkription mit der DNA verhindern (Workman und Kingston, 1998). Die funktionelle Untereinheit der Chromatinstruktur ist das Nukleosom (Ahmad und Henikoff, 2002). Nukleosomen sind aufgebaut aus 146 Bp DNA, die um ein "Kern"-Oktamer von Histon-Proteinen (je zwei Kopien H2A, H2B, H3 und H4) gewickelt ist (Horn und Peterson, 2002) und durch H1 Histon-Proteine untereinander verbunden sind (Woodcock und Dimitrov, 2001; Hayes und Hansen, 2001). Die lineare, sich wiederholende Anordnung der Nukleosomen, welche die Primärstruktur des Chromatins ausmacht, kann zu Strukturen höherer Ordnung kondensieren und dadurch sowohl aktivierenden wie auch inaktivierenden Einfluss auf die Transkription ausüben (Wolffe und Guschin, 2000; Horn und Peterson, 2002). Aus diesem Grund wird vermutet, dass ein wichtiger Mechanismus in der Modulation der Chromatinstruktur die Modifikation von "Kern"-Nukleosom-Komponenten ist (Mostoslavsky et al., 2003). Die meisten der bisher charakterisierten Chromatin modifizierenden Komplexe können aufgrund ihrer Funktion in zwei Kategorien eingeteilt werden (Vignali et al., 2000; Tsukiyama, 2002). Bei der ersten Kategorie handelt es sich um ATP-abhängige Proteinkomplexe, welche die freiwerdende Energie der Hydrolyse von ATP zur Auftrennung oder Veränderung der lokalen Interaktion von DNA und Histonen nutzen und so Proteinen, die an der Regulation der Transkription beteiligt sind, den Zugang zur DNA ermöglichen (Whitehouse et al., 1999; Längst und Becker, 2001; Tsukiyama, 2002). Alle

bisher bekannten ATP-abhängigen Chromatin modifizierenden Komplexe enthalten eine ATPase-Untereinheit, die zur SW12/SNF2- oder "Imitation" SWI- (ISWI) Proteinfamilie gehört (Eisen et al., 1995; Workman und Kingston, 1998; Peterson, 2000; Vignali et al., 2000). Proteine mit dieser Untereinheit sind an der Chromatin-Reorganisation bei Transkription, Replikation, Reparatur und Rekombination beteiligt (Flanagan und Peterson, 1999; Kwon et al., 2000; Ura et al., 2001; Längst und Becker, 2001). Die Mechanismen, durch welche ATP-abhängige Proteinkomplexe die Nukleosomenstruktur verändern, sind bisher nicht bekannt (Vignali et al., 2000; Längst und Becker, 2001). Es wird jedoch vermutet, dass es aufgrund von Änderungen der Interaktionen zwischen Histonen und DNA zu Modifikationen in der Konformation des Nukleosoms kommt, die zu einer Reorganisation der lokalen Chromatinstruktur führen (Abb. 1.2) (Vignali et al., 2000). Weiter wird angenommen, dass die Reorganisation der Nukleosomenstruktur zu einem Transfer des Histon-Oktamers auf ein anderes DNA-Segment (trans) oder zu einer Verschiebung des Oktamers an eine andere Position auf dem gleichen DNA-Segment (cis) führt (Vignali et al., 2000). Abhängig von der Position des Nukleosoms könnte die lokale Reorganisation sowohl zur Öffnung (Aktivierung der Transkription) als auch zur Blockierung (Repression) eines Promotors führen (Vignali et al., 2000). Die zweite Kategorie beinhaltet Proteinkomplexe, welche die Transkriptionsaktivität bestimmter Gene durch kovalente Modifikationen der aminoterminalen Domänen von Nukleosomen-Histonen durch Acetylierung, Methylierung oder Phosphorylierung regulieren (Strahl und Allis, 2000; Vignali et al., 2000). Die kovalenten Modifikationen von Histonen beeinflussen die Interaktionen zwischen Nukleosom und DNA und haben sowohl in vitro wie auch in vivo eine wichtige Funktion in der Regulation der Transkription (Han und Grunstein, 1988; Lorch et al., 1998; Strahl und Allis, 2000; Eberharter und Becker, 2002; Mostoslavsky et al., 2003). Die Acetylierung der aminoterminalen Domänen von Histonen durch Histon-Acetyltransferasen (HAT) führt zu einer lokalen Dekondensation der Chromatinstruktur. Werden Histone an Promotor-Nukleosomen acetyliert, führt dies zu einer Änderung der Histon-DNA-Interaktion, die eine Öffnung der lokalen Chromatinstruktur zur Folge hat (Abb. 1.2). Proteine, die an der Regulation der Transkription beteiligt sind, können mit der DNA interagieren (Struhl, 1998; Strahl und Allis, 2000; Eberharter und Becker, 2002). Umgekehrt führt die Deacteylierung durch Histon-Deacetylasen (HDAC) zu von Histonen einer Kondensation der Chromatinstruktur (Abb. 1.2). Die Kompaktierung der Chromatinstruktur verhindert die Interaktion von Regulatoren der Transkription mit der DNA und reprimiert so die Transkription (Struhl, 1998; Strahl und Allis, 2000; Eberharter und Becker, 2002; De Ruijter et al., 2003). Eine weitere Funktion der Acetylierung von Nukleosomen könnte in der Rekrutierung von Modulatoren der Chromatinstruktur liegen. Es wird angenommen, dass Veränderungen in der Nukleosomenstruktur von Promotoren nicht durch Acetylierung allein, sondern durch synergistische Effekte, wie z. B. Phosphorylierung, erreicht werden (Eberharter und Becker, 2002). Beispielsweise konnte in Untersuchungen an *S. cerevisiae* festgestellt werden, dass die Transkription des *INO1*-Gens durch eine Histon-Acetyltransferase (*Gcn5*) und eine Kinase (*Snf1*) co-reguliert wird (Lo et al., 2001). In diesem Zusammenhang wird ebenfalls eine Koregulation der Nukleosomenstruktur durch Histon-Acetylierung und Methylierung angenommen (Vandel und Trouche, 2001). Ferner zeigten Untersuchungen von Czermin et al. (2001) in *Drosphila*, dass die Histon-Methyltransferase *SU(VAR)3-9* und die Histon-Deacetylase *HDAC1* in vivo assoziieren und es dadurch zu einer Methylierung von prä-acetylierten Histonen kommt. Es wird ein Modell angenommen, in dem die Deacetylierung und anschließende Methylierung von Histonen eine permanente Repression der Transkription in bestimmten Bereichen des Genoms zur Folge hat (Czermin et al., 2001).

Wahrscheinlich ist, dass die Acetylierung von Histonen allein nicht ausreicht und ein Zusammenwirken der verschiedenen Nukleosom reorganisierenden Proteinkomplexe notwendig ist (Strahl und Allis, 2000; Eberharter und Becker, 2002). So ist bereits mehrfach eine funktionelle Interaktion zwischen ATP-abhängiger Chromatin-Reorganisation und Histon Acetylierung beschrieben worden (Becker, 2002; Becker und Hörz, 2002). In diesem Zusammenhang konnte nachgewiesen werden, dass die Interaktion des ATP-abhängigen Chromatin reorganisierenden *SWI/SNF2* Komplexes (*S. cerevisiae*) durch Histon Acetylierung verstärkt wird (Hassan et al., 2001). Bei der Etablierung einer reprimierenden Chromatins führt und Faktoren rekrutiert, die mit einer Repression der Gen-Expression assoziiert sind (Eberharter und Becker, 2002). Eine Möglichkeit, um die reprimierende Chromatinstruktur dauerhaft aufrecht zu erhalten und erneute Acetylierung zu verhindern besteht in der Methylierung der Histone, die vermutlich eine Interaktion mit Heterochromatin-Protein 1 (*HP1*) unterstützt und das Chromatin in eine nicht zugängliche, die Transkription reprimierende Konfiguration überführt (Eberharter und Becker, 2002).



Abb. 1.2: Modell der Regulation der Chromatinstruktur durch die Aktivität von Histon-Acetyltransferasen (HAT) und Histon-Deacetylasen (HDAC). Die Acetylierung von Histonen führt zu einer Dekondensation der Chromatinstruktur. Dies erlaubt eine Interaktion von ATP-abhängigen Chromatin reorganisierenden Proteinkomplexen (grüner Pfeil) mit Nukleosomen, die zu einer Öffnung von Promotor-Regionen und Aktivierung der Transkription führen kann (schwarzer Pfeil). Die Deacetylierung und Methylierung von Histonen könnte die Chromatinstruktur in Strukturen überführen, welche die Transkription reprimieren (z. B. Heterochromatin). Acetylierte Histon-Reste (Ac) sind als gelbe Kreise dargestellt. Die Methylierung (Me) von Histon-Resten ist als graues Rechteck gezeigt. Histon-Methyltransferase (*HMT*), Heterochromatin Protein 1 (*HP1*). Verändert nach Eberharter und Becker, 2002.

#### 1.4 Subnukleäre Domänen

Die Organisation des Interphase-Zellkerns unterliegt einer strikten Kontrolle, die unter anderem durch die räumliche Einteilung des Genoms in funktionelle Kompartimente, wie z. B. chromosomale Domänen, sowie durch Reorganisation der Chromatinstruktur erreicht wird (s.o.). Eine bedeutende Rolle bei der Zell-spezifischen Genregulation spielen subnukleäre Domänen. Viele im Zellkern lokalisierte Proteine kommen sowohl frei im Nukleoplasma, als auch in bestimmten subnukleären Domänen vor (Lamond und Earnshaw, 1998; Misteli, 2001). In der Fluoreszenz-Mikroskopie erscheinen subnukleäre Domänen als unregelmäßige, in unterschiedlichen Anzahlen auftretende gepunktete Strukturen, welche in Größe und Form variieren (Verschure et al., 2002; Lamond und Spector, 2003; Spector und Gasser, 2003). Subnukleäre Domänen enthalten einen geringen oder keinen DNA-Anteil und eine jeweils spezifische Fraktion nukleärer Proteine. Diese sind an der Regulation der Gen-Expression, der Transkription sowie an der Synthese und Prozessierung von RNA beteiligt (Lamond und Earnshaw, 1998; Verschure et al., 2002; Lamond und Spector, 2003). Subnukleäre Domänen, die Komponenten des prä-mRNA Spliceapparates enthalten, werden als "speckles" oder "interchromatin granule clusters" (ICG) bezeichnet (Spector 1993; Mintz et al., 1999; Mintz und Spector, 2000). Von ICGs wird angenommen, dass sie an der Assemblierung und Modifikation von prä-mRNA Splicefaktoren beteiligt sind, die sowohl in ICGs als auch diffus im Nukleoplasma verteilt vorkommen (Lamond und Earnshaw, 1998; Mintz et al., 1999; Spector, 2001). Bei ICGs handelt es sich um sehr dynamische Strukturen, aus denen an Orten der Transkription verschiedene Proteine rekrutiert werden können (Lamond und Earnshaw, 1998; Mintz et al., 1999; Spector, 2001). Die Untersuchung verschiedener Transkriptionsfaktoren zeigte, dass diese ebenfalls in spezifischen subnukleären Domänen vorliegen können, deren Struktur, Anzahl und subnukleäre Lokalisation der Verteilung der ICGs gleicht (Grande et al., 1997; Verschure et al., 2002). Weiter konnte nachgewiesen werden, dass der größte Bestandteil der Transkriptionsfaktoren nicht mit den Orten der aktiven Transkription kolokalisiert. Aus diesem Grund wird vermutet, dass Transkriptionsfaktoren nicht frei durch das Nukleoplasma diffundieren (Grande et al., 1997; Verschure et al., 2002), sondern in subnukleären Domänen gespeichert werden. Aufgrund der Tatsache, dass verschiedene Transkriptionsfaktoren innerhalb dieser Domänen nicht unbedingt miteinander kolokalisieren, könnten spezifische Domänen jeweils spezifische Transkriptionsfaktoren et al., 1997). enthalten (Grande Bei Bedarf könnten Transkriptionsfaktoren aus diesen Domänen an Orte der Transkription rekrutiert werden (Grande et al., 1997). Von "Cajal bodies" ("coiled bodies") wird angenommen, dass es sich um subnukleäre Domänen handelt, die an der Biogenese von snRNPs ("small nuclear ribonucleoproteins") und snoRNPs ("small nucleolar ribonucleoproteins") beteiligt sind (Gall, 2000; Spector, 2001), die eine Funktion beim Splicen von prä-mRNAs bzw. bei der Prozessierung von rRNAs besitzen (Gall, 2000; Lamond und Spector, 2003). Proteine und Proteinkomplexe, die mit der Prozessierung von prä-mRNAs assoziiert sind, kommen neben einer diffusen Verteilung im Zellkern in subnukleären Kompartimenten vor, die als "cleavage bodies" bezeichnet werden (Schul et al., 1996; Spector et al., 2001). "Cleavage bodies" und "Cajal bodies" kolokalisieren aufgrund der Beteiligung beider Domänen an der Prozessierung von prä-mRNAs. Das auffälligste subnukleäre Kompartiment, der Nukleolus, ist der Ort der rRNA-Synthese, rRNA-Prozessierung und Assemblierung von ribosomalen Untereinheiten (Spector, 1993; Scheer und Weisenberger, 1994; Leung und Lamond, 2003). Bei "nuclear bodies" handelt es sich um eine weitere subnukleäre Domäne, die auch als "PML-bodies", "ND10", "PODs (PML oncogenic domains)" und "Kr-bodies" bezeichnet wird (Goddard et al., 1991; Zhong et al., 2000<sup>1</sup>; Spector et al., 2001). Neben dem *PML*-Protein ("promyelocytic leukemia protein") sind verschiedene Proteine, wie z. B. pRB, p53, CBP, BLM, Daxx, SUMO-1 und Sp100 bekannt, die transient in "nuclear bodies" rekrutiert werden können und mit der Regulation der Transkription assoziiert sind (Alcalay et al., 1998; Zhong et al., 2000<sup>1</sup>; LaMorte et al., 1998; Ishov et al., 1999; Zhong et al., 2000<sup>2</sup>; Müller et al., 1998; Seeler et al., 1998). In vivo- und in vitro-Experimente deuten darauf hin, dass es sich bei PML um einen Tumorsuppressor handelt (Zhong et al., 2000<sup>1</sup>). Wird die RING-finger Domäne des PML-Proteins zerstört, kommt es zu einem Verlust der Tumorsuppressor-Aktivität (Le et al., 1996). Es ist möglich, dass "nuclear bodies" als Depot dienen, welches durch Rekrutierung, Modifikation und Freisetzung die Aktivität von Transkriptionsfaktoren und Kofaktoren für die Transkription reguliert (Zhong et al.,  $2000^{1}$ ).

## 1.5 Zielsetzung

Die Zielsetzung der vorliegenden Arbeit war es, anhand des Beispiels der fetalen humanen Wachstumsfuge, neue Gene zu identifizieren und charakterisieren, die an der Zelldifferenzierung beteiligt sind, oder differentiell exprimiert werden und deren medizinische Relevanz zu untersuchen. Zu diesem Zweck sollten zunächst die Daten aus einem EST-Projekt aus einer Genbibliothek aus fetalen humanen Knorpelgewebe bioinformatisch ausgewertet werden. Um potentielle Kandidatengene zu identifizieren und auszuwählen, die eine grundsätzliche Bedeutung in Entwicklungs- und Differenzierungsprozessen haben, sollten die cDNAs vorläufig charakterisiert und kategorisiert werden. In einem zweiten Schritt sollten ausgewählte Kandidatengene funktionell charakterisiert und deren Einfluss auf die Differenzierung untersucht werden. Ferner sollte die Bedeutung der Kandidatengenen für die Entstehung von möglichen Erkrankungen (genetische bedingte Funktionsstörungen von Organen und Tumorerkrankungen) untersucht werden.

# 2. Material und Methoden

## 2.1 Isolierung von DNA

## 2.1.1 Isolierung von genomischer DNA aus Blut

Die Isolierung von genomischer DNA aus Blut erfolgte mit Hilfe von Qiagen<sup>®</sup> Genomic-tip 20/G Säulen (Qiagen, Hilden) entsprechend den Angaben des Herstellers. 1 ml Vollblut wurden mit 1 ml eiskalten Puffer C1 und 3 ml eiskalten A. bidest. vermischt und 10 min auf Eis inkubiert. Nach Zentrifugation bei 1300 x g für 15 min (4°C) wurde das Pellet in 0,25 ml eiskalten Puffer C1 und 0,75 ml eiskalten A. bidest. resuspendiert und erneut bei 1300 x g für 15 min (4°C) zentrifugiert. Nach Resupension des Pellets in 1 ml Puffer G2 wurden 25  $\mu$ l Proteinase K Lösung (20 mg/ml) zugesetzt und bei 50°C für 1 h inkubiert. Die nun klare Lösung wurde dann auf eine mit 1 ml Puffer QBT äquilibrierte Qiagen Genomic-tip 20/G Säule aufgetragen und mit 3 x 1 ml Puffer QC gewaschen. Die Elution der genomischen DNA von der Säule erfolgte mit 2 x 1 ml Puffer QF. Nach Isopropanolfällung und Waschen der DNA in 70 % Ethanol konnten zwischen 10 und 20  $\mu$ g genomischer DNA isoliert werden.

## 2.1.2 Isolierung kleiner Mengen Plasmid-DNA

Die Präparation von Plasmid-DNA für die Sequenzierung und für Restriktionsanalysen erfolgte in abgewandelter Form nach Feliciello und Chinali (1993). Hierzu wurden 3 ml LB-Medium mit geeigneten Antibiotikumzusatz mit dem gewünschten Plasmidklon angeimpft und bei 37°C im Schüttelinkubator für 12-14 h inkubiert, abzentrifugiert und in 100  $\mu$ l Puffer I (50 mM Glukose, 10 mM Tris-HCl, 1 mM EDTA, pH 8.0) resuspendiert. Die Lyse der Bakterien erfolgte durch Zugabe von 200  $\mu$ l Puffer II (0,2 N NaOH, 1 % SDS). Zur Fällung bakterieller Proteine und chromosomaler DNA wurden 300  $\mu$ l Puffer III (4 M K-Acetat, 2,55 M Essigsäure, pH 4.8) zugegeben und 5 min auf Eis inkubiert. Nach Zentrifugation (5 min, 10.000 x g) wurden 500  $\mu$ l des plasmidhaltigen Überstandes vorsichtig abgezogen und durch eine Phenol/Chloroform/Isoamylalkohol-Extraktion (25:24:1) und eine Chloroform-Extraktion gereinigt, mit Ethanol gefällt und nach einem Waschschritt in 70%

Ethanol getrocknet. Die Plasmid DNA wurde anschließend in einem geeigneten Volumen TE-Puffer oder A.bidest. gelöst.

#### 2.1.3 Isolierung von Plasmid-DNA für die Transfektion eukaryotischer Zellen

Die Isolierung von Endotoxin freier Plasmid-DNA erfolgte mit Hilfe des "EndoFree<sup>TM</sup> Plasmid Maxi Kit" (Qiagen<sup>®</sup>, Hilden) entsprechend den Angaben des Herstellers. Hierzu wurden 100 ml LB-Medium mit geeigneten Antibiotikumzusatz mit dem gewünschten Plasmidklon angeimpft, bei 37°C im Schüttelinkubator für 12-16 h inkubiert, für 15 min bei 6000 x g abzentrifugiert (4°C) und in 10 ml Puffer P1 resuspendiert. Die Bakterien wurden durch Zugabe von 10 ml Puffer P2 lysiert und 5 min bei Raumtemperatur inkubiert. Nach Zugabe von 10 ml eiskalten Puffer P3 wurde das Lysat durch eine QIAfilter Maxi Säule filtriert, 2,5 ml Puffer ER zugegeben und 30 min auf Eis inkubiert. Das Filtrat wurde auf eine mit 10 ml Puffer QBT äquilibrierte Qiagen-tip 500 Säule gegeben und 2 x mit 30 ml Puffer QC gewaschen. Durch Zugabe von 15 ml Puffer QN wurde die Plasmid-DNA von der Säule eluiert, mit 10,5 ml Isopropanol gefällt und 30 min bei 15.000 x g zentrifugiert (4°C). Anschließend wurde das Plasmid-Pellet 1 x mit 70% Ethanol gewaschen, bei Raumtemperatur getrocknet und in 500 µl A.bidest.

# 2.2 Isolierung von RNA

## 2.2.1 Isolierung von RNA aus Zellen

Die Präparation von Gesamt-RNA aus eukaryotischen Zellen erfolgte mit Hilfe des "High Pure RNA Isolation Kit" (Roche, Mannheim). Die Präparationen wurden entsprechend den Angaben des Herstellers durchgeführt.

#### 2.2.2 Isolierung von RNA aus menschlichen und murinen Geweben

Die Isolierung von Gesamt-RNA aus tiefgefrorenen Geweben erfolgte je nach Menge des verwendeten Gewebes bzw. der gewünschten RNA-Ausbeute mit dem "RNeasy<sup>®</sup> Total RNA

Mini-Kit" oder mit dem "RNeasy<sup>®</sup> Total RNA Midi-Kit" (Qiagen<sup>®</sup>, Hilden). Die Präparationen wurden entsprechend den Angaben des Herstellers durchgeführt.

### 2.2.3 Isolierung von RNA aus menschlichen Knorpelgewebe

Bei der Gewinnung von Gesamt-RNA wurden je nach erwünschter Ausbeute 50 bis 100 mg tiefgefrorenes Knorpelgewebe in flüssigen Stickstoff fein zermörsert, in 1 ml Trizol aufgenommen und in einem Glashomogenisator 2 min homogenisiert. Nach 5 min Inkubation bei Raumtemperatur wurden der Probe 0,2 ml Chloroform zugesetzt, das Gemisch stark geschüttelt und erneut 3 min bei Raumtemperatur inkubiert. Nach 15 min Zentrifugation (4°C) bei 10.000 x g wurde der Überstand vorsichtig abgezogen, mit 0,5 ml Isopropanol gefällt und 10 min bei Raumtemperatur stehen gelassen. Nach erneuter Zentrifugation (4°C) bei 10.000 x g wurde das Pellet mit 1 ml 75% Ethanol gewaschen, 5 bis 10 min bei Raumtemperatur getrocknet und in 50  $\mu$ l A.bidest. gelöst.

#### 2.2.4 DNase I-Behandlung von RNA

Um Kontaminationen der RNA mit DNA zu vermeiden, erfolgte eine Inkubation der RNA mit DNase I in einem Reaktionspuffer bestehend aus 1 x TE-Puffer, 10 mM MgCl<sub>2</sub>, 1 mM DTT, 20 Units RNase-Inhibitor (MBI Fermentas, St. Leon Rot) und 10 Units DNase I (Roche, Mannheim) für 2 h bei 25°C. Anschließend wurde die RNA erneut mit Trizol und Chloroform extrahiert und mit Isopropanol gefällt. Die gereinigte RNA wurde nochmals mit 75% Ethanol gewaschen, bei Raumtemperatur getrocknet und in 50 µl A.bidest. gelöst.

## 2.3 DNA und RNA Standardmethoden

#### 2.3.1 Spaltung von DNA mit Restriktionsendonukleasen

Die Restriktonsspaltung von DNA dient entweder der Analyse der aufgereinigten DNA oder der Isolation spezifischer DNA-Fragmente zum Zwecke der Klonierung bzw. Subklonierung (vgl. 2.4.1, 2.4.2). Typischerweise wurden Restriktionsverdaus von Plasmid-DNA oder auch von PCR-Produkten in einem Reaktionsvolumen von 20 bis 50 µl durchgeführt. Dabei

wurden je Mikrogramm DNA mindestens 1 U der jeweiligen Restriktionsendonuklease(n) eingesetzt. Die verwendete Menge an Enzym variierte je nach Aktivität des entsprechenden Enzyms, der Qualität und Menge der zu restringierenden DNA. Die Restriktionsbedingungen wurden entsprechend den Angaben des jeweiligen Enzymlieferanten, unter Verwendung der mitgelieferten Restriktionspuffer gewählt.

## 2.3.2 Agarose-Gelelektrophorese von DNA und RNA

DNA-Fragmente können im analytischen TBE-Agarosegel (0,8-2% (w/v) Agarose / 0,1 µg/ml Ethidiumbromid) mittels Elektrophorese entsprechend ihrer Größe aufgetrennt werden. Dazu wurden die Proben mit 1/6 Vol. DNA-Probenpuffer versetzt und bei einer konstanten Spannung zwischen 75 und 120 V aufgetrennt. Als Molekulargewichtsstandards dienten die 100 bp "Leiter" (Gibco BRL<sup>TM</sup>, Eggenstein) und Hind III restringierte  $\lambda$ -DNA (Roche, Mannheim). Die Dokumentation erfolgte unter UV-Licht mit dem Herolab UVT-28M Videosystem (Wiesloch). Die Auftrennung von RNA-Molekülen erfolgte auf horizontalen 1% igen Agarosegelen (6,5% Formaldehyd). Als Laufpuffer diente 1 x MOPS. Vor der Elektrophorese wurde der RNA 1/10 Vol Ethidiumbromid (5 µg/µl) und 1 Vol. 2 x RNA Loading-Dye Solution (MBI-Fermentas, St. Leon, Rot) zugesetzt, der gesamte Ansatz zur Auflösung von Sekundärstrukturen bei 70°C für 5 min denaturiert und bei einer konstanten Spannung von 75 bis 95 V aufgetrennt. Als Molekulargewichtsstandard diente die "High Range RNA-Ladder" (MBI-Fermentas, St. Leon, Rot). Die Dokumentation erfolgte unter UV-Licht mit dem Herolab UVT-28M Videosystem (Wiesloch).

#### 2.3.3 Wiedergewinnung von DNA aus Agarosegelen

Nach Gelelektrophoretischer Auftrennung von DNA-Fragmenten auf 1-2%igen Agarosegelen wurden die mit Ethidiumbromid gefärbten DNA-Banden unter UV-Licht aus dem Gel ausgeschnitten. Die Isolierung der Fragmente aus dem Gel erfolgte mit Hilfe des QIAEX II Gel Extraction Kit (Qiagen<sup>®</sup>, Hilden) nach Angaben des Herstellers. Zur Wiedergewinnung von PCR-Produkten, die ausschließlich als "Template" für weitere PCR-Reamplifikationen verwendet wurden, wurde die ausgeschnittene Bande mit ca. 1 Vol. A.bidest. überschichtet und 1 h bei Raumtemperatur inkubiert. Ein Aliquot dieser Suspension wurde als "Template" für Reamplifikationen verwendet.

#### 2.3.4 Fällung von DNA

In der Regel wurde DNA durch Zusatz von 1 Volumen 4 M Ammoniumacetat und 6 Volumen Ethanol abs. gefällt. Nach 20 min Zentrifugation bei 25.000 x g wurde das Pellet mit 70% Ethanol gewaschen, erneut für 10 min zentrifugiert, getrocknet und in A.bidest. gelöst. Sequenzierungen wurden durch Zusatz von 1/10 Volumen 3 M Na-Acetat (pH 5,2) und 2,5 Volumen Ethanol abs. gefällt. Nach 30 min Zentrifugation bei 25.000 x g wurde das Pellet mit 70% Ethanol gewaschen, erneut für 10 min zentrifugation bei 25.000 x g wurde das Pellet mit 70% Ethanol gewaschen, erneut für 10 min zentrifugiert, getrocknet und bei  $-20^{\circ}$ C gelagert. Alle Zentrifugationsschritte wurden bei Raumtemperatur durchgeführt.

#### **2.3.5** Transfer von RNA auf Membranen (Northern Blot)

Nach gelelektrophoretischer Auftrennung der RNA (vgl. 2.3.2) wurde das Gel 20 min in A.bidest. gewaschen, um Formaldehyd aus dem Gel zu entfernen, welches den RNA-Transfer behindert. Anschließend wurde das Gel 10 min in 20 x SSC inkubiert. Der Transfer der RNA auf die positiv geladene Nylonmembran erfolgte über Nacht (14-16 h), wobei 20 x SSC als Transferpuffer verwendet wurde. Die Fixierung der RNA an die Membran erfolgte durch "backen" der Membran bei 120°C für 30 min oder durch "UV-Crosslinking".

### 2.3.6 DNA Sequenzierung und Mutationsanalyse von SPOC1

Für die Sequenzierung von DNA Fragmenten und Plasmiden wurde die von Lee et al. (1992) modifizierte DNA-Sequenzierung der von Sanger et al. (1977) beschriebenen Kettenabbruchmethode angewandt. Für die Sequenzierung, die auf einem ABI Prism<sup>®</sup> 377 DNA Sequenzer (Applied Biosystems, USA) erfolgte, wurde das "Big Dye<sup>TM</sup> Terminator Cycle Sequencing Ready Reaction Mix" (PE Biosystems, Warrington) nach Angaben des Herstellers verwendet. Für die Mutationsanalyse des Gens *SPOC1* wurden 31 verschiedene DNA-Proben von Patienten mit endometrialen Hyperplasien und endometrioiden Adenokarzinomen, die sich durch einen Verlust der Heterozygotie (LOH) in der Chromosomenregion 1p36 auszeichneten (Kiechle et al., 2000) untersucht. Dazu wurden die Exons 1, 2, 3 und 4 von *SPOC1* über PCR-Reaktionen aus den Patienten-DNAs amplifiziert (vgl. 2.5) und sequenziert. Die Primer (vgl. 2.15) wurden dabei so gelegt, dass auch die Intron/Exon Übergange mit analysiert werden konnten.

## 2.4 Klonierung von DNA-Fragmenten in Plasmid-Vektoren

#### 2.4.1 Klonierung von PCR-Produkten

Zur Klonierung von PCR-Produkten wurde der "TOPO TA Cloning<sup>®</sup> Kit" (Invitrogen<sup>TM</sup>, Karlsruhe) verwendet. Die Klonierungen in den pCR<sup>®</sup>2.1-TOPO<sup>®</sup> Vektor (Invitrogen<sup>TM</sup>, Karlsruhe) erfolgten nach den Angaben des Herstellers. Zur Klonierung von PCR-Produkten, die mit Restriktionsschnittstellen enthaltenden Primern generiert wurden, wurde der entsprechende Vektor mit dem entsprechenden Restriktionsenzym geschnitten (vgl. 2.3.1) und die überhängenden Enden mit alkalischer Phosphatase aus Kälberdarm (Roche, Mannheim) dephosphoryliert. Die zu klonierenden PCR-Produkte wurden ebenfalls mit dem entsprechenden Restriktionsenzym geschnitten (vgl. 2.3.1) und in den Vektor ligiert (vgl. 2.4.3). Sowohl die Vektor-DNA, als auch die zu klonierenden Fragmente wurden über ein Agarosegel aufgereinigt und die Klonierung über Sequenzierung verifiziert.

#### 2.4.2 Subklonierung von DNA Fragmenten

Zur Subklonierung von DNA-Fragmenten wurden die Plasmid-Vektoren pEGFP-N1 (Clontech, Heidelberg), pFLAG-CMV-4<sup>™</sup> (Sigma<sup>®</sup>, Darmstadt), pM2 (Clontech, Heidelberg) und pGEX-4T-3 (Amersham Biosciences) mit dem entsprechenden Restriktionsenzym verdaut (vgl. 2.3.1) und die überhängenden Enden mit alkalischer Phosphatase aus Kälberdarm (Roche, Mannheim) dephosphoryliert. Die für die Subklonierung vorgesehenen DNA-Fragmente wurden ebenfalls mit einem entsprechenden Restriktionsenzym geschnitten und in den jeweiligen Vektor ligiert (vgl. 2.4.3). Die Aufreinigung der DNA-Fragmente und der Vektor-DNA erfolgte über ein Agarosegel (vgl. 2.3.3). Die Konstrukte wurden über Sequenzierung verifiziert.

## 2.4.3 Ligation

Die T4-DNA-Ligase verknüpft kovalent doppelsträngige DNA-Fragmente mit glatten oder überstehenden, kompatiblen Enden. Zur Durchführung der Ligation wurde geschnittene und dephosphorylierte (vgl. 2.4.2) Vektor-DNA mit der zu inserierenden DNA so gemischt, daß eine dreifach höhere Konzentration der Insertions-DNA vorlag. Dieses Gemisch wurde in Ligationspuffer mit 5 U T4-DNA-Ligase (Gibco BRL<sup>TM</sup>, Karlsruhe) bei RT für 2 h oder ÜN bei 4°C inkubiert. Die Hälfte des Ligationsansatzes wurde zur Transformation (vgl. 2.4.4) kompetenter *E.coli* DH5 $\alpha$  Bakterien eingesetzt.

## 2.4.4 Transformation und Selektion positiver Klone

Für die Transformation wurden 50  $\mu$ l kompetenten *E.coli* DH5 $\alpha$  Bakterien mit 2  $\mu$ l eines TOPO TA Klonierungsansatzes (vgl. 2.4.1) oder der Hälfte eines Ligationsansatzes (vgl. 2.4.4) vermischt, 30 min auf Eis inkubiert und für 30 sec einem 42°C Hitzeschock ausgesetzt. Nach Zugabe von 200  $\mu$ l SOC-Medium wurde der Transformationsansatz für 30 min bei 37°C geschüttelt und auf LB-Agarplatten mit dem entsprechenden Antibiotikum ausgestrichen. Die Selektion positiver Klone erfolgte über PCR-Analyse (vgl. 2.5).

## 2.5 DNA Amplifikation durch Polymerasekettenreaktion (PCR)

Die PCR (Saiki et al., 1988) wurde in 10-100 µl Ansätzen mit 10 mM Tris-HCl (pH 8.3), 1,5 mM MgCl<sub>2</sub>, 40 mM KCl und 10 pmol eines jeden Primers durchgeführt. Die Amplifikationsbedingungen wurden entsprechend der verwendeten Primerpaare sowie der DNA-Matrizen, ausgehend von einem Standardprogramm (erster Denaturierungsschritt 3 min 94°C, 25-35 Zyklen 1 min 94°C, 1 min 52-68°C, 1-3 min 72°C, Extensionsschritt 10 min 72°C) variiert. Die Amplifikationen wurden mit Taq-Polymerase (Gibco BRL<sup>™</sup>, Karlsruhe) entsprechend den Angaben des Herstellers durchgefürt. Bei sehr GC-reichen oder problematischen Matrizen wurde dem PCR Ansatz 10% 5 M Betain zugesetzt. Als weitere Auftreten unspezifischer PCR-Produkte Zusätze wurden bei Glycerol (5-10%) Endkonzentration) oder Ammoniumsulfat (60 mM Endkonzentration) verwendet. Als Matrize für die PCR-Reaktion wurden 10-100 ng DNA eingesetzt. Die Amplifikation erfolgte in einem PTC-200 Thermocycler (MJ Research, Inc., USA) oder Eppendorf Mastercycler Gradient (Eppendorf, Deutschland). Zur Überprüfung der PCR-Reaktion wurde 1/10 des Reaktionsansatzes auf einem entsprechend konzentrierten Agarosegel aufgetrennt.

## 2.6 Reverse Transkriptase-Polymerasekettenreaktion (RT-PCR)

Für die Herstellung größerer Mengen cDNA wurde ein Reaktionsgemisch bestehend aus 1 x Reaktionspuffer (Gibco BRL<sup>TM</sup>, Karlsruhe), 50 mM MgCl<sub>2</sub>, 10 mM dNTPs, 5 μg Hexanucleotide (Gibco BRL<sup>TM</sup>, Karlruhe), 1 U/μl RNase-Inhibitor (MBI Fermentas, St.Leon, Rot) und 4-5 μg Gesamt-RNA verwendet. Das Reaktionsgemisch wurde zunächst 5 min auf 65°C erhitzt und 1 min auf Eis inkubiert. Nach Zugabe von 1 μl (200 U) Superscript II Reverse Transcriptase (Gibco BRL<sup>TM</sup>, Karlsruhe) wurde das Reaktionsgemisch für 1 h bei 37°C inkubiert und die Reverse Transkriptase anschließend durch 10minütiges Erhitzen auf 94°C inaktiviert. Für die anschließende PCR wurden 2-3 μl der cDNA als Matrize für spezifische Primer verwendet.

# 2.7 Radioaktive und nichtradioaktive Markierungstechniken

### 2.7.1 Markierung von DNA über "random primed oligonucleotide labelling"

Die radioaktive Markierung von DNA Fragmenten wurde mit Hilfe des "Ready-To-Go<sup>™</sup> DNA labelling (-dCTP) Kit" (Amersham Pharmacia Biotech Inc, USA) nach Angaben des Herstellers durchgeführt. Für die Markierung wurden 50 ng über ein Agarosegel aufgereinigtes PCR-Produkt verwendet. Um die nicht eingebauten Nukleotide von den markierten DNA-Fragmenten abzutrennen, wurden die markierten DNA-Fragmente über TE-äquilibrierte Sephadex G-50 Säulen (Pharmacia, Freiburg) aufgereinigt.

#### 2.7.2 Endmarkierung von Oligonukleotiden

Die 3'-Endmarkierung von Oligonukleotiden für die RNA *in situ*-Hybridisierung an Gewebeschnitten (vgl. 2.9) erfolgte mit Hilfe des "Digoxigenin (DIG) Oligonucleotide Tailing Kit" (Roche, Mannheim) nach Angaben des Herstellers. Um die optimale Konzentration zu ermitteln wurde eine Verdünnungsreihe (1:10, 1:100, 1:1000) des markierten Oligonukleotids hergestellt, auf eine Nylonmembran aufgetropft und diese 30 min bei 120°C gebacken. Die Membran wurde 5 min in Puffer I (100 mM Tris-HCl, 150 mM NaCl, pH 7.5) gewaschen und 15 min in Blocking-Puffer (1% w/v Blocking Reagenz, Roche Mannheim, in Puffer I) inkubiert. Anschließend erfolgte eine 45minütige Inkubation mit Anti-

Digoxygenin-AP-Fab-Fragmenten (Roche, Mannheim) in Blocking Puffer (Verdünnung 1:5000). Die Membran wurde 2 x 5 min in Puffer I gewaschen und 5 min in Puffer NBT äquilibriert. Die Detektion der markierten Oligonukleotide erfolgte mit Hilfe von NBT/BCIP Substrat (Roche, Mannheim). Die Substratreaktion wurde mit PBS abgestoppt.

## 2.8 DNA-RNA-Hybridisierung (Northern-Hybridisierung)

Für die Northern-Hybridisierung wurden 15 µg Gesamt-RNA aus Gewebe oder Zellkultur in einem 1 %igen Agarosegel elektrophoretisch aufgetrennt (vgl. 2.3.2) und auf Nylonmembranen transferiert (vgl. 2.3.5). Neben selbst Hergestellten Northern-Blots wurden auch kommerziell erhältliche "Multiple Tissue Expression Arrays" und "Multiple Tissue Northern (MTN<sup>™</sup>) Blots" (Clontech, USA) verwendet. Alle Membranen wurden mit 5-10 ml "ExpressHyb<sup>™</sup>-Hybridisierungslösung (Clontech, USA) nach Angaben des Herstellers hybridisiert. Bei bereits mehrfach verwendeten Membranen wurde die Hybridisierungszeit auf 12 h ausgedehnt und die letzten Waschschritte (0,1 x SSC, 0,12% SDS für 2 x 20 min) bei 68°C durchgeführt. Die noch feuchten Filter wurden anschließend in Frischaltefolie eingeschweißt und mit einem Röntgenfilm (Hyperfilm<sup>™</sup> RNP1674, Amersham, England) bedeckt. Die Exposition der Autoradiographie erfolgte in einer mit Verstärkerfolie ausgekleideten Kassette (Hypercassette<sup>™</sup>, Amersham, England) bei -80°C für 1 Stunde bis zu 14 Tagen.

# 2.9 RNA in situ-Hybridisierung von Gewebeschnitten

Die Untersuchung der Genexpression erfolgte an Hoden- und Uterusgewebe adulter Mäuse (C57BL/6J).

## 2.9.1 Herstellung von Schnittpräparaten

Für die RNA-*in situ*-Hybridisierung von Gewebeschnitten mittels Digoxigenin (DIG) markierter Oligonukleotidsonden (vgl. 2.7.2) wurden die präparierten Gewebe direkt auf Trockeneis eingefroren, als Blöckchen in Tissue-Tek (Sakura, Japan) eingebettet und anschließend bei –80°C gelagert. Von den Gewebe-Blöckchen wurden mit einem Kryostat

(Microm HM 560, Microm, Deutschland) Schnitte von 8-14 um Dicke angefertigt, auf Objektträger aufgebracht (HistoBond, Marienfeld, Deutschland) und für 15 min bei 40°C auf einem Heiztisch getrocknet. Die Aufbewahrung der Schnitte bis zur weiteren Verwendung erfolgte bei -80°C. Für die in situ-Hybridisierung wurden die Schnitte 30 min bei Raumtemperatur aufgetaut, 15 min bei Raumtemperatur in 4 % PFA/1 x PBS fixiert und 2 x 5 min in Puffer PTW gewaschen. Anschließend wurden 50 µl der 1:50-1:100 in 2 x SSC/deion. Formamid verdünnten Oligonukleotidsonden luftblasenfrei auf den entsprechenden Gewebeschnitt pipettiert und für 3 h bei 33°C in einer feuchten Kammer inkubiert. Die Objektträger wurden dann für 3 x 5 min in 2 x SSC (33°C), 2 x 10 min in 0,1 x SSC (33°C) gewaschen und für 5 min bei Raumtemperatur in Puffer PBT inkubiert. Nach 15minütiger Inkubation in Blockierungslösung (1% w/v) "Blocking-Reagenz" (Roche, Mannheim) in Puffer I wurden pro Objektträger 50 µl 1:100 in Blockierungslösung verdünnte APkonjugierte anti-Digoxigenin Fab-Fragmente (Roche, Mannheim) zugegeben und für 1 h bei 37°C inkubiert. Vor der Durchführung der Substratreaktion wurden die Schnitte 2 x 5 min in Puffer PBT und 1 x 5 min in Puffer NBT gewaschen. Die Substratreaktion wurde nach Angaben des Herstellers mit einer NBT/BCIP-haltigen Lösung (Roche, Mannheim) für ca. 30 bis 60 min inkubiert und schließlich durch Zugabe von 1 x PBS abgestoppt. Das Eindeckeln der Objektträger erfolgte mit Crystal/Mount<sup>™</sup> und Clarion<sup>™</sup> (Biomeda, USA) nach Angaben des Herstellers.

## 2.10 Genomische Lokalisation eines murinen Gens

Die Feststellung der Chromosomalen Lokalisation eines Gens in der Maus erfolgte mit Hilfe eines "RH04.02 Mouse/Hamster Radiation Hybrid Panel" (Research Genetics, Inc., USA). Zu diesem Zweck wurden die 99 Hybridzell-DNAs mit unterschiedlichen Anteilen von murinen Chromosomenbruchstücken sowie Donor- (Hamster) und Akzeptor-DNA (Maus) über PCR mit einem Paar genspezifischer Primer getestet. Vorrausetzung dabei war, dass das gewählte Primerpaar (MM12fw und MM4brev, vgl. 2.15.8) nur PCR-Produkte mit den Hybridzell-DNAs und der Donor-DNA, nicht aber mit der Akzeptor-DNA ergab. Die genomische Lokalisation des untersuchten Gens konnte schließlich durch Auswertung der Kombination der positiven und negativen PCR-Ergebnisse für jede Hybrid-DNA über den Server des Jackson Lab (http://www.genome.wi.mit.edu/cgi-bin/mouse\_rh/rhmap-auto/rhmapper.cgi) ermittelt werden. Von den einzelnen Hybrid-DNAs wurden für jede PCR-Reaktion 50 ng für einen 50 µl Ansatz eingesetzt.

## 2.11 Protein Standardmethoden

#### 2.11.1 Protein-Extraktion aus eukaryotischen Zellen

Zur Isolierung von Gesamt-Proteinextrakt wurden die Zellen in der Kulturschale 2 x mit PBS-Puffer gewaschen. Anschließend wurde pro Kulturschale 1 ml IP Lysis-Puffer oder RIPA Lysis-Puffer zugegeben und kurz geschwenkt. Der Zellrasen wurde mit einem Zellschaber von der Kulturschale gelöst und das Lysat in einem 1,5 ml Reaktionsgefäß unter gelegentlichen Schütteln 5 min auf Eis inkubiert. Die Lagerung des Lysats bis zur weiteren Verwendung erfolgte bei –80°C.

#### 2.11.2 SDS-Polyacrylamid-Gelelektrophorese (SDS-PAGE)

Proteine können aufgrund ihrer relativen Molekülmasse mit Hilfe der SDS-Polyacrylamid-Gelelektrophorese aufgetrennt werden (Laemmli, U. K., 1970). Für das Trenngel, welches für die eigentliche Auftrennung der Proteine verantwortlich ist, wurde abhängig von der Größe des zu untersuchenden Proteins eine Polyacrylamidkonzentration von 7,5 bis 12% (Tab. 2.1) verwendet. Für das Sammelgel wurde eine PAA-Konzentration von 3% verwendet. Die Protein-Proben (vgl. 2.11.1) wurden zunächst mit 5 x SDS-Probenpuffer versetzt und das Gemisch 3 x durch eine 20 G Kanüle gepresst. Das Proteinlysat wurde anschließend für 10 min bei 95°C denaturiert und 20 bis 40 µl auf das PAA-Gel geladen. Die Elektrophorese der Proben erfolgte im Sammelgel bei 90 Volt und im Trenngel bei 150 Volt. Die Dauer der Elektrophorese richtete sich nach der Größe der untersuchten Proteine. Als Protein-Standard wurde der "Recombinant protein molecular weight marker RPN 800" der Firma Amersham Life Science (Freiburg) verwendet.

Polyacrylamidgel (PAA-Gel)	12%	10%	7,5%
A.bidest.	3,35 ml	2,79 ml	4,85 ml
1,5 M Tris-HCl, pH 8,8	2,5 ml	2,5 ml	2,5 ml
10% SDS	100 µl	100 µl	100 µl
Acrylamid/Bis (30%)	4 ml	3,33 ml	2,5 ml
10% Ammonium Persulfat (APS)	150 µl	150 µl	150 µl
TEMED	5 μl	5 μl	5 µl

Tab.	2.1:	Zusammensetzung	der verschiedenen 🛛	Polyacr	vlamid-Trenng	ele
						,
#### 2.11.3 Proteintransfer (Western Blot)

Der Western Blot dient dem Nachweis von Proteinen nach deren gelelektrophoretischer Auftrennung und Transfer auf eine polymere Membran mittels spezifischer Antikörper. Die getrennten Proteine wurden dazu mit Hilfe einer "Semidry"-Transferapparatur (Fastblot B 44 Elektro-Blotting) der Firma Biometra (Göttingen) im elektrischen Feld auf eine Polyvinyldifluoriden (PVDF)-Membran (Immobilon P<sup>®</sup>, Ø 0,45 µm, Millipore, Frankreich) übertragen. Der Aufbau des Blots wurde nach Angaben des Herstellers durchgeführt. Der Proteintransfer erfolgte mit Transferpuffer für 15-25 min bei 350-500 mA.

## 2.11.4 Immunologischer Nachweis von Proteinen

Für den Nachweis des untersuchten Proteins auf der Membran wurde diese zunächst 1 h bei Raumtemperatur oder ÜN bei 4°C mit Blocking-Puffer inkubiert. Anschließend wurden verschiedene Verdünnungen des ersten Antikörpers in Antikörper-Puffer verdünnt und für 2 h oder ÜN bei 4°C auf die Membran gegeben. Die Membran wurde dann 3 x 5 min mit TTBS gewaschen und 1 h mit dem entsprechenden Peroxidase-gekoppelten zweiten Antikörper in Antikörper-Puffer inkubiert (Schaf anti-Maus: Verdünnung 1:5000, Esel anti-Hase: Verdünnung 1:3000). Für die Detektion wurde die Membran 1 x 15 min und 3 x 5 min in TTBS gewaschen. Die Detektion erfolgte mit dem "ECL<sup>™</sup> Western Blotting Detection Kit" (Amersham Pharmacia, Freiburg) nach Angaben des Herstellers. Dazu wurden gleiche Teile der Reagenzien 1 und 2 des Kits gemischt, die Membran darin 1 min inkubiert, in eine Plastikhülle gelegt und bei Raumtemperatur für 5-35 min in einer Kassette (Hypercassette<sup>™</sup>, Amersham, England) auf Röntgenfilmen exponiert.

## 2.11.5 Proteinquantifizierung

Die Quantifizierung von Protein erfolgte mit dem "Micro BCA Protein Assay Kit" (Pierce, USA) nach Angaben des Herstellers. Dazu wurde zunächst mit dem im Kit enthaltenen Proteinstandard eine Eichreihe erstellt. Anschließend erfolgte die photometrische Auswertung der Eichreihe und der zu untersuchenden Protein-Proben in einem "Novaspec II"-Photometer (Amersham Pharmacia, Freiburg).

## 2.12 Zellbiologische Methoden

Alle Tätigkeiten im Zusammenhang mit der Kultivierung von Säugerzellen erfordern sterile Arbeitstechniken zur Vermeidung bakterieller oder sonstiger Kontaminationen. Aus diesem Grunde wurden alle diesbezüglichen Arbeiten in eigens dafür vorgesehenen Sterilwerkbänken durchgeführt. Darüber hinaus wurden ausschließlich gammabestrahlte Zellkulturmaterialien wie Kulturflaschen, Kulturschalen oder Zentrifugenröhrchen der Firmen Falcon<sup>™</sup> (Heidelberg) und Nunc (Wiesbaden) verwendet.

## 2.12.1 Kultur von adhärent wachsenden Zellen

Adhärente Zellinien wurden in einer humiden Atmosphäre (95% relative Luftfeuchtigkeit) bei 37°C und einer CO<sub>2</sub>-Konzentration von 5% in 2 ml / Well (6-Well Platte) oder 10 ml (10 cm Platte) DMEM-Medium (Gibco BRL<sup>™</sup>, Eggenstein) kultiviert. Zum Teilen bzw. ausplattieren wurden die Zellen einmalig mit PBS (Gibco BRL<sup>™</sup>, Eggenstein) gewaschen und anschließend für ca. 2 Minuten mit Trypsin-EDTA-Lösung (Gibco BRL<sup>™</sup>, Eggenstein) von den Zellkulturflaschen oder Kulturschalen abgelöst. Überschüssiges Trypsin wurde im Anschluss daran durch Zugabe von frischen vorgewärmten Vollmedium inaktiviert und die resuspendierten Zellen dann in einer Dichte von 10-20% ausplattiert. Adhärente Zelllinien: 293, 293T, COS7, U2OS, SW 1353 (HCS); Hela, NCI-H1299, Huh7.

## 2.12.2 Einfrieren und Auftauen von Zellen

Als Einfriermedium wurde eine Mischung aus DMEM-Medium (Gibco BRL<sup>™</sup>, Eggenstein) und 10% DMSO verwendet. Jeweils etwa 10<sup>7</sup> Zellen wurden in 2 ml des Einfriermediums aufgenommen und nach einer maximal einwöchigen Zwischenlagerung bei –80°C dauerhaft in flüssigem Stickstoff aufbewahrt. Das Auftauen der Zellen erfolgte möglichst rasch durch Inkubation im Wasserbad bei 37°C und sofortigem Überführen der Zellsuspension in vorgewärmtes Vollmedium.

## 2.12.3 Transfektion von adhärenten Zellen mit Lipofectamine<sup>™</sup> 2000 Reagenz

Für die transiente Transfektion von adhärenten Zellen wurde Lipofectamine<sup>™</sup> 2000 Reagenz (Invitrogen<sup>™</sup>, Karlsruhe) nach Angaben des Herstellers verwendet. Zur Durchführung wurden die zur Transfektion vorgesehene Plasmid-DNA und Lipofectamine<sup>™</sup> 2000 jeweils mit serum- und antibiotikafreiem OptiMEM1-Medium (Gibco BRL<sup>™</sup>, Eggenstein) gemischt (Mengenangaben siehe Tabelle 2.2), 5 min bei RT inkubiert und anschließend miteinander vermischt und für 20 min bei RT stehen gelassen. In der Zwischenzeit wurden die am Vortag ausplattierten Zellen einmalig mit 2-5 ml OptimMEM1 gewaschen. Nach vorsichtigen Aufbringen der Lipofectamine<sup>™</sup> 2000-DNA Mischung auf die Zellen wurden diese für 4-6 h im Brutschrank in Optimem1-Medium kultiviert, mit vorgewärmten DMEM-Medium 1 x gewaschen und für 14-24 h bis zur Weiterverarbeitung in DMEM-Medium kultiviert. In Abhängigkeit vom Zellkulturformat wurden für die Transfektionsansätze folgende Mengen eingesetzt:

Zellkulturformat	Gesamt-DNA	Lipofectamine <sup>™</sup> 2000	Serumfreies Medium
6-Well Platte	2 μg / Well	6 µl / Well	in 50 µl / Well
10 cm Platte	12 µg	36 µl	in 200 µl

#### Tab. 2.2: Für die Transfektionsansätze verwendete Mengen an DNA und Lipofektamine™ 2000

Die Zellen wurden mittels Western-Blot (vgl. 2.11.3) auf die Expression der transfizierten DNA überprüft.

## 2.12.4 Herstellung und Verifizierung eines anti-SPOC1-Peptidantikörpers

Der affinitätsaufgereinigte polyklonale *SPOC1*-Peptidantikörper aus Hase (SPOC1-#10) wurde durch die Firma Eurogentec S. A. (Belgien) gegen das Peptid H<sub>2</sub>N-CRSNRSRTGSRKLFLD-COOH aus der *SPOC1*-Aminosäuresequenz hergestellt. Die Spezifität des Peptidantikörpers wurde mit Hilfe von Western-Blot-Analysen (vgl. 2.11.3) mit Zell-Lysaten von transient mit FLAG-*SPOC1*- und pM2-*SPOC1*-Konstrukten transfizierten U2OS- und 293T-Zellen und Prä-Immunserum aus Hase verifiziert. Weiterhin wurden Western-Blot-Analysen von untransfizierten Zell-Lysaten von U2OS-, 293T-, DeuserPT-, BG1p2-, Hela, NCI-H1299, Huh7-Zellen durchgeführt (vgl. 2.11.3). Dazu wurde jeweils das Lysat von 1 x 10<sup>6</sup> Zellen in einer Tasche des PAA-Western-Gels aufgetragen.

## 2.12.5 Immunfluoreszenz und EGFP-Fluoreszenz

Immunfluoreszenz-Experimente wurden zur Untersuchung der intrazellulären Lokalisation von Proteinen eingesetzt. Dazu wurden COS7- und U2OS-Zellen auf 20 x 20 mm Deckgläsern (Superior, Marienfeld, Deutschland) ausplattiert und am darauffolgenden Tag mittels Lipofectamine<sup>™</sup> 2000 transfiziert (vgl. 2.12.3). Nach 24 - 48 Stunden wurden die Zellen dann 1 x mit PBS gewaschen und anschließend mit einer -20°C kalten 1:1 Methanol/Aceton-Mischung für 3 Minuten fixiert. Nach Absaugen des Fixierungsgemisches wurden die Zellen kurz bei Raumtemperatur getrocknet und 10 min in PBS rehydriert. Die Inkubation mit den entsprechenden Primärantikörpern (Konzentrationen: vgl. 2.15) erfolgte für 1-2 h bei RT. Nach drei 5minütigen Waschschritten in PBS erfolgte die eigentliche Fluoreszenzfärbung durch Zugabe des entsprechenden in PBS verdünnten fluoreszenzmarkierten Sekundärantikörpers (Konzentrationen: vgl. 2.15). Die chromosomale DNA wurde mittels DAPI (Roche, Mannheim) (je 4 µg/ml in PBS) für 10 min bei Raumtemperatur angefärbt. Anschließend folgten drei weitere Waschschritte in PBS und das Eindeckeln der Deckgläser auf Objektträgern (HistoBond, Marienfeld, Deutschland) in Glycerol-Gelatine (Sigma<sup>®</sup>, Deutschland). Die mit EGFP-Fusionsproteinen transfizierten Zellen wurden entweder in 4% PFA für 10 min fixiert, einmal in PBS gewaschen und anschließend direkt die chromosomale DNA gefärbt oder 2 x mit PBS gewaschen und ohne Fixierung in PBS eingedeckelt und mit Fixogum (Marabu, Deutschland) abgedichtet. Bei EGFP ("Enhanced Green Fluorescent Protein") handelt es sich um eine durch Punktmutationen optimierte Version des GFP-Proteins aus der Qualle Aequorea victoria, welches verbesserte fluoreszierende Eigenschaften aufweist. Die Analyse der Färbungen wurde mit einem Fluoreszenz-Mikroskop (Leica-DMRA) durchgeführt, welches zur unmittelbaren EDV-gestützten Weiterverarbeitung (Cytovision Ultra, Applied Biosystems, England) der aufgenommenen Bilder mit einer Digitalkamera (Pieper, Deutschland) gekoppelt war.

## 2.13 Luziferase Reporter Assays

Biolumineszenz-Assays können zur Untersuchung von in vivo Interaktionen zwischen Proteinen verwendet werden. Das Enyzym Luziferase aus der Weichkoralle Renilla reniformis ist in der Lage, durch die Katalyse der ATP-abhängigen Oxidation seines Substrates Luziferin eine bioluminiszente Lichtemission zu erzeugen. In Gegenwart überschüssigen Luziferins ist diese Lichtemission dabei proportional zur Menge der im Zellextrakt vorhandenen Luziferase, was die Luziferase zu einem guten Reporterenzym macht. Zur Durchführung des Assays fusioniert man die zu untersuchende cDNA an die im pM2-Plasmid (Clontech, Heidelberg) enthaltene GAL4-Bindedomäne. Bei der anschließenden **Co-Transfektion** pM2-Fusionskonstruktes (pM2-SPOC1) des mit dem pG5Luc-Reporterplasmid (Promega<sup>™</sup>, Mannheim), welches fünf GAL4-Bindestellen stromaufwärts vom Reportergen enthält, kann das pM2-Fusionsprotein an die GAL4-Bindestellen des Luziferase-Reporters binden und der Einfluß des zu untersuchenden Gens auf die Transkriptionsrate des Luziferase-Gens luminometrisch bestimmt werden (Abb. 2.1). Die Expression des Fusionsproteins in 293T-Zellen wurde über Western-Blot (vgl. 2.11.3) mit dem GAL4(DBD) Antikörper (Santa Cruz Biotechnology, USA) verifiziert (vgl. 2.15).



Abb. 2.1: Luziferase-Reporter-Assay. Das zu untersuchende Gen (Gen X) wird als Fusionsprotein mit der GAL4-DNA bindenden Domäne (GAL4-DBD) exprimiert. Über diese Domäne kann das Fusionsprotein an die GAL4-Bindestellen (BS) des Reportergens binden und der Einfluss des Gens X auf das Reportergen luminometrisch untersucht werden.

Nach der Co-Transfektion (vgl. 2.12.3) von 293T Zellen in 6-Well Platten (Tab.2.3) wurden diese 24 Stunden später geerntet. Allen Ansätzen wurden zur Kontrolle der Transfektionseffizienz 50 ng pLacZ-Plasmid zugesetzt und die Gesamt-DNA Menge mit dem Plasmid pCDNA3 (Clontech, Heidelberg) auf 2  $\mu$ g aufgefüllt. Zum Ernten der Zellen wurden diese nach einmaligem Waschen in PBS in 120  $\mu$ l Reporter-Lysepuffer (Promega<sup>TM</sup>, Mannheim) lysiert. Unlösliche Bestandteile wurden durch eine 15minütige Zentrifugation (13.000 rpm, 4°C) entfernt und die Aktivität der im Überstand enthaltenen Luziferase mit Hilfe eines Luminometers (Berthold Microlumat LB 96 P) bestimmt. Die Injektion von 50 µl der das Luziferin enthaltenden Substratlösung (40 mM Tricin, 2,14 mM (MgCO<sub>3</sub>)<sub>4</sub>Mg(OH)<sub>2</sub> x 5 H2O, 5,34 mM MgSO<sub>4</sub>, 0,2 mM EDTA, 66,6 mM DTT, 540 µM Acetyl-Coenzym A, 940 µM Luziferin, 1,06 mM ATP) wurde automatisch von dem Luminometer vorgenommen. Die Aufzeichnung der Lichtemission im Emissionsmaximum  $(\lambda = 562 \text{ nm})$  erfolgte für ein Zeitintervall von 20 Sekunden. Für jede Konzentration des pM2-SPOC1-Fusionskonstrukts wurden jeweils 3 voneinander unabhängige Assays durchgeführt. Die Ergebnisse von 3 voneinander unabhängigen Messungen wurden gemittelt und die Standardabweichung für die jeweils transfizierte Konzentration des pM2-SPOCI-Fusionskonstrukts berechnet. Die Luziferase-Aktivität der Messungen wurde mit der Expression eines kotransfizierten LacZ-Konstruktes normalisiert.

Reporterplasmid (pG5Luc)	Fusionskonstrukt (pM2/SPOC1)	pCDNA3	pLacZ	Lipofectamine <sup>™</sup> 2000
750 ng	5 ng	1195 ng	50 ng	5 µl
750 ng	10 ng	1190 ng	50 ng	5 µl
750 ng	50 ng	1150 ng	50 ng	5 µl
750 ng	200 ng	1000 ng	50 ng	5 µl
750 ng	500 ng	500 ng	50 ng	5 µl

Tab. 2.3: Für die Kotransfektion von 293T-Zellen verwendete Mengen an Reporterplasmid, Fusionskonstrukt, pCDNA3, pLacZ und Lipofectamine™ 2000.

## 2.14 Computerauswertung von ESTs einer cDNA-Bibliothek

Die verwendete humane Chondrozyten-cDNA-Bibliothek wurde von B. Lee et al. (Houston) hergestellt (Ahn et al., 1995). Als Ausgangsmaterial für die Erstellung der Bibliothek diente humanes Knorpelgewebe (20. Schwangerschaftswoche bis 2. Lebensjahr). Die Sequenzierung von ca. 5000 ESTs aus der cDNA-Bibliothek erfolgte durch die Firma GENterprise (Mainz). Die Auswertung der Sequenzdaten erfolgte auf dem Zentralrechner (Genius) des Deutschen Krebsforschungszentrums (DKFZ) unter Verwendung des Programmpakets HUSAR. Um eine automatisierte Datenbank-Analyse der ESTs in großer Menge zu ermöglichen, wurde in Gemeinschaftsarbeit mit der Firma GENterprise (Mainz) und der "Bioinformatic Service Group" im DKFZ (http://genome.dkfz-heidelberg.de) (Heidelberg) das Datenbank-Analyse-Programm "ESTsweep" entwickelt, mit dem die EST-Analysen, ausgehend vom Genius Zentralrechner, in den BLASTN- und BLASTX-Datenbanken (Pearson und Lipman, 1988; Altschul et al., 1990, Altschul et al., 1994; Altschul et al., 1997) durchgeführt wurden. Als Exon- und Genvorhersage-Programme wurden "GRAIL II"-Analysen ("gene recognition and analysis internet link", Uberbacher und Mural, 1991, Xu et al.,1994) und "GENSCAN"-Analysen (Burge und Karlin, 1997; Burge und Karlin 1998) verwendet. Das Programm "PromotorInspector" (Genomatix, München) wurde für die Vorhersage von putativen Promotoren verwendet. Weitere BLAST-Analysen wurden auf dem BLAST-Server des NCBI durchgeführt (http://www.ncbi.nlm.nih.gov/BLAST/).

## 2.15 Reagenzien und Materialien

## 2.15.1 Puffer und Lösungen

Agar-Platten	15 g Agar-Agar ad 1000 ml LB
Antikörperlösung	0,2% Milchpulver in TTBS
Blocking-Puffer (für Western Blot)	1 x TBS 0,1% (v/v) Tween 20 1% (w/v) Magermilchpulver
Blocking-Puffer (für RNA in situ-Hybridisierung)	1% (w/v) Blocking Reagenz in Puffer 1
DNA-Ladepuffer I	0,25% Bromphenolblau 0,25% Xylencyanol 30% Glycerol
Fixierungslösung	50% (v/v) Methanol 50% (v/v) Aceton
Kinase-Puffer	20 mM Tris-HCl, pH 7,5 10 mM MnCl <sub>2</sub> 5 mM MgCl <sub>2</sub> (5 μCi [γ- <sup>32</sup> P] ATP)
LB-Medium	10 g Trypton 5 g Hefeextrakt 10 g NaCl ad 1000 ml Aqua bidest., pH 7,5

Luziferin-Substratlösung	40 mM Tricin 2,14 mM (MgCO <sub>3</sub> ) <sub>4</sub> Mg(OH) <sub>2</sub> x 5 H2O 5,34 mM MgSO <sub>4</sub> 0,2 mM EDTA 66,6 mM DTT 540 μM Acetyl-Coenzym A 940 μM Luziferin 1,06 mM ATP
MOPS (10x)	0,2 M 3-( <i>N</i> -orpholin)propansulfonsäure 0,05 M NaOAc 0,01 M EDTA, pH 7,0
NBT-Puffer	100 mM Tris-HCl, pH 9,5 100 mM NaCl 50 mM MgCl <sub>2</sub>
PBS	137 mM NaCl 8,1 mM Na <sub>2</sub> HPO4 2,7 mM KCl 1,5 mM KH <sub>2</sub> PO4
PBT-Puffer	0,2 % (w/v) BSA 0,1 % (v/v) Tween 20 in PBS
Puffer 1 (für RNA in situ-Hybridisierung)	100 mM Tris-HCl, pH 7,5
Puffer I	50 mM Glukose 50 mM Tris-HCl
Puffer II	0,2 M NaOH 1% SDS
Puffer III	4 M Kaliumacetat, 2,55 M Essigsäure, pH 4,8
PTW-Puffer	0,1% Tween 20 in PBS
RIPA-Lysis-Puffer	25 mM Tris/HCl, pH 8.2 50 mM NaCl 1% Natriumdeoxycholat 1 mM EDTA 10 mg/ml Leupeptin 10 mg/ml Aprotinin 1% SDS 1% NP40 (IGEPAL)
RNA-Ladepuffer	500 μl Formamid (deionisiert) 100 μl 10x MOPS-Puffer 150 μl Formaldehyd (filtriert, 37 %)

SDS-Laufpuffer (5 x)	125 mM Tris-Base 960 mM Glycin 0,5% (w/v) SDS
SDS-Probenpuffer (5 x, reduz.)	312,5 Mm Tris-HCl, pH 6,8 25% (v/v) β-Mercaptoethanol 50% (v/v) Glycerol 10% (w/v) SDS 0,01% (w/v) Bromphenolblau
SDS-Sammelgel (3 %)	125 mM Tris-HCl, pH 6,8 3% (v/v) Acrylamid/Bis 0,1% (w/v) SDS 0,3% (v/v) TEMED
SDS-Trenngel	350 mM Tris-HCl, pH 8,8 7,5-12% Acrylamid 0,1% (w/v) SDS 0,04% (v/v) APS 0,075% (v/v) TEMED
Standard-Saline-Citrat-Puffer (20 x SSC)	3 M NaCl 0,3 M Natriumcitrat, pH 7,0
TBS (10 x für Western Blot)	250 mM Tris 1,87 M NaCl 50 mM KCl 7 mM CaCl <sub>2</sub> 1 mM MgCl <sub>2</sub> , pH 7,4
TBE (5 x)	445 mM Tris 445 mM Borsäure 5 mM EDTA
TE-Puffer	10 mM Tris 1 mM EDTA
Transferpuffer Westernblot	25 mM Tris 190 mM Glycin 20% (v/v) Methanol 0,038% (w/v) SDS
Tris-Borat-Elektrophoresepuffer (1x TBE)	90 mM Tris 90 mM Borsäure
TBST	1 x TBS 0,1% (v/v) Tween 20

## 2.15.2 Bakterienstämme

E. coli Stamm	Verwendung	Referenz/Bezugsquelle
DH5a-T1 <sup>R</sup>	Vermehrung von Plasmiden	Invitrogen <sup>™</sup> , Karlsruhe
DH10b	Vermehrung von Plasmiden	Gibco BRL <sup>™</sup> , Eggenstein
XL1 blue	Mutagenese	Stratagene, Heidelberg

## 2.15.3 Eukaryotische Zellen

Zelllinie	Herkunft
293	humane embryonale Nierenzelllinie
293T	humane embryonale Nierenzelllinie, transformiert mit dem
	langen Arm des SV-40 Virus
COS-7	Affennierenzellen
U2OS	humane Osteosarcomzelllinie
SW 1353 (HCS)	humane Chondrosarcomzelllinie
Hela	humane Adenokarzinomzelllinie
NCI-H1299	humane Lungenkarzinomzelllinie
Huh7	humane hepatozelluläre Karzinomzelllinie
BG1p2	Östrogen-Rezeptor positive humane Ovarialkarzinomzelllinie
DeuserPT	Zelllinie aus einem humanen Plattenepithelkarzinom

## 2.15.4 Radioisotope, Enzyme, Nukleinsäuren und Zellkulturmedien, Chemikalien

 $[\alpha^{32}P] dCTP, [\gamma^{32}P] dATP$ Albumin bovine Fraction V Alkalische Phosphatase Blocking Reagent DNaseI Lysozym Maus Cot-1 DNA Non-Fat Dry Milk *Pfu* DNA Polymerase Proteinase K Restriktionsendonukleasen

RNase A RNase ERASE Salmon Sperm DNA Superscript II<sup>™</sup> Reverse Transcriptase Taq-DNA Polymerase Trizol Reagent T4-DNA Ligase

Amersham (Braunschweig) Gibco BRL<sup>TM</sup> (Eggenstein) Roche (Mannheim) Roche (Mannheim) Roche (Mannheim) Roche (Mannheim) Roche (Mannheim) **BioRad** (Deutschland) Stratagene (Heidelberg) Roche (Mannheim) Roche (Mannheim), New England Biolabs Inc (USA) Roche (Mannheim) ICN (USA) Pharmacia (Schweden) Gibco BRL<sup>TM</sup> (Eggenstein) Gibco BRL<sup>TM</sup> (Eggenstein) Gibco BRL<sup>TM</sup> (Eggenstein) Roche (Mannheim)

Alle Zellkulturmedien sowie benötigte Zusätze (Serum, Antibiotika) wurden von der Firma GibcoBRL<sup>™</sup> (Eggenstein) erworben.

## Molekulargewichtsstandards

Gibco BRL<sup>™</sup> (Eggenstein) Roche (Mannheim) Gibco BRL<sup>™</sup> (Eggenstein) MBI-Fermentas (St. Leon, Rot) Amersham (Braunschweig)

## Chemikalien

Acrylamid/Bis Agar-Agar Agarose Ampicillin BCIP Chloramphenicol Clarion<sup>TM</sup> Mounting Media Crystal/Mount<sup>TM</sup> Desoxycholinsäure Diethyl Pyrocarbonat EDTA Ethanol Ethidiumbromid Fixogum Formaldehyd Formamid Hefeextrakt **IPTG** Isobutanol Kanamycin-Sulfate Lipofectamine<sup>TM</sup> 2000 Methanol NBT Paraformaldehyd Phenol SDS Trypton Tris Tween 20 X-Gal

Carl-Roth (Karlsruhe) Difco (USA) Gibco BRL<sup>TM</sup> (Eggenstein) Ratiopharm (Ulm) Roche (Mannheim) Roche (Mannheim) Biomeda corp. (USA) Biomeda corp. (USA) Sigma<sup>®</sup> (Darmstadt) Sigma<sup>®</sup> (Darmstadt) Roche (Mannheim) Riedel-de-Haen (Seelze) Oncor (USA) Marabu (Deutschland) Merck (Darmstadt) Merck (Darmstadt) Difco (USA) Sigma<sup>®</sup> (Darmstadt) Merck (Darmstadt) Roche (Mannheim) Invitrogen<sup>™</sup> (Karlsuhe) Merck (Darmstadt) Roche (Mannheim) Sigma<sup>®</sup> (Darmstadt) Roth (Karlsruhe) ICN (USA) Difco (USA) Boehringer (Mannheim) BioRad (Deutschland) Sigma<sup>®</sup> (Darmstadt)

## 2.15.5 Geräte

DNA-Sequenziergerät	ABI PRISM ™ 377 Applied Biosystems GmbH (Weiterstadt)
Geldokumentation	Herolab UVT-28M (Wiesloch)
Hybridisierungsofen	Biometra OV2 (Göttingen)
Luminometer	Microlumat LB 96 P, Berthold (Bad Wildbad)
PCR-Geräte	PTC 200 <sup>TM</sup> (MJ Research, Inc. (USA) Mastercycler Gradient Eppendorf (Deutschland) Gene-Amp 2400 (Perkin Elmer, USA)
Photometer	Novaspec II, Amersham (Freiburg)
Szintillationszähler	LS 2800, Beckmann (USA)
Zellkultur-Inkubator	NUaire <sup>™</sup> CO <sub>2</sub> Water Jacketed Incubator Zapf (Sarstedt)
Zentrifugen	Centrifuge 5417R, Eppendorf (Deutschland) Megafuge 1.0, Heraeus (Hanau) Varifuge 3.0 R, Heraeus (Hanau) Sorvall RC-5B, Du Pont Instruments (USA)
Kryostat	Microm HM 560 (Walldorf)

Bezeichnung	Spezies	Antigen	Herkunft/	Eingesetzte
			Referenz	Konzentration
anti-FLAG	Maus	DYKDDDDK	Sigma®, Darmstadt	8 µg/ml (1:100)
anti-FLAG	Hase	DYKDDDDK	Sigma®, Darmstadt	8 µg/ml (1:100)
PML (PG-M3)	Maus	PML, N-Terminus	Santa Cruz, USA	2 µg/ml (1:100)
Pol II (A-10)	Maus	RNA-Polymerase II,	Santa Cruz, USA	2 µg/ml (1:100)
		Aminosäuren 1-224 (human)		
E2F-1 (KH95)	Maus	Humanes E2F-1 p60 Protein	Santa Cruz, USA	2 µg/ml (1:100)
Living colors <sup>®</sup> A.v.	Hase	GFP	Clontech,	1 μg/ml (1:100)
			Heidelberg	
anti-GAL4 (DBD)	Hase	GAL4, N-Terminus	Santa Cruz, USA	0,2 μg/ml (1:1000)
anti-Digoxigenin-AP	Schaf	Digoxigenin	Roche, Mannheim	1:5000
Fab fragments				
SPOC1-#10	Hase	SPOC1-Peptid	Eurogentec S. A.,	4,2 μg/ml (1:100)
		CRSNRSRTGSRRKLFLD	Belgien	

## 2.15.6 Primärantikörper

## 2.15.7 Sekundärantikörper

Bezeichnung	Spezies	Konjugat	Herkunft/Referenz	Eingesetzte
				Konzentration
Alexa Fluor® 488 anti-Hase	Ziege	Fluorescein 488	Molecular Probes, USA	1,3 μg/ml (1:1500)
Alexa Fluor® 546 anti-Hase	Ziege	Fluorescein 546	Molecular Probes, USA	1,3 μg/ml (1:1500)
Alexa Fluor® 488 anti-	Ziege	Fluorescein 488	Molecular Probes, USA	1,3 μg/ml (1:1500)
Maus				
Alexa Fluor® 546 anti-	Ziege	Fluorescein 546	Molecular Probes, USA	1,3 μg/ml (1:1500)
Maus				
Cy <sup>™</sup> 3 AffiniPure F(ab') <sub>2</sub>	Ziege	Cy3 F(ab') <sub>2</sub>	Jackson ImmunoResearch	1 μg/ml (1:1500)
anti-Hase			Laboratories, INC., USA	
Cy <sup>TM</sup> 3 AffiniPure $F(ab')_2$	Ziege	Cy3 F(ab') <sub>2</sub>	Jackson ImmunoResearch	1 μg/ml (1:1500)
anti-Maus			Laboratories, INC., USA	
anti-mouse Ig	Schaf	Peroxidase	Amersham Pharmacia,	1:5000
			England	
anti-rabbit Ig	Esel	Peroxidase	Amersham Pharmacia,	1:3000
			England	

## 2.15.8 Synthetische Oligonukleotide

## **Allgemeine Primer**

$T_3A$ :	5'-ATTAACCCTCACTAAAGGG-3'
$T_7A$ :	5'-AATACGACTCACTATAGGG-3'
M13 for:	5'-GTAAAACGACGGCCAGT-3'
M13 rev:	5'-GGAAACAGCTATGACCATG-3'
GAPDH 1 forw:	5'-GTGGAGTCCACTGGCGTCTTC-3'
GAPDH 1 rev:	5'-CTCCGACGCCTGCTTCACCAC-3'
pEGFP_fw:	5'-GTGAACCGTCAGATCCGCTAG-3'
pEGFP_rev:	5'-CGGACACGCTGAACTTGTGGC-3'

## Primer für die Klonierung des humanen SPOC1-Gens

Schnittstellen für Restriktionsendonukleasen sind fett hervorgehoben

K 203 F1:	5'-AGTCCCACCTTGCAGGATATCC-3'
K 203 R1:	5'-GCAGTCAGTCCAGGAACAGCTT-3'
K 203 II F:	5'-CTCCGATACTCCCTCGAGTGG-3'
K 203 II R:	5'-CCAGCAGTCAGTCCAGGAACAGC-3'
K 203 Ex 0 fw:	5'-CACGAGCCTTCGCCTCCGCTC-3'
K 203 Ex 1 fw:	5'-CAGTTGCAAGAGGCGCAGGAC-3'
K 203 Ex 2 rev:	5'-GCGCATCACTGTCCCTGCGCTTC-3'
Stop1 rev:	5'-GCAGTCAGTCCAGGAACAGC-3'
c0 forw:	5'-GCTCCAGCATCCCAGCTCCTG-3'
c0 rev:	5'-GCTGCAAGAGAGTAGGCTGCAGG-3'
c1 forw:	5'-CTCAGACATCGCGTCCTCAGTG-3'
c1 rev:	5'-GGTGTGGCACTCATTACACTCG-3'
c2 forw:	5'-GACTCAGATGACGATTCCTG-3'
c2 rev:	5'-CACTGCAGGGAACGCAGCTCTG-3'
c3 forw:	5'-CTGTACATAGCCGGTGACCG-3'
c3 rev:	5'-CCAACTCCCTTTCCACAGAGAAG-3'
c4 forw:	5'-GGTCAGCTGGTGGTTCTTAGG-3'
c4 rev:	5'-GAGATTTAAGGAACTCGGCACAG-3'
c5 forw:	5'-GGTACAGTTAGGACTTGAGTC-3'
c5 rev:	5'-CAAGTGTGAAAGATACTAGGAC-3'
c6 forw:	5'-GCAGACTCTTCCCACCTTCTC-3'
c6 rev:	5'-GTGTCCGACTTCTTAGACACCAC-3'
c6 rev II:	5'-CAGCACAGAACGCAATCGCC-3'
K203 E fw BamHI:	5'-CGCGGATCCCATGGACTCTGACTCTTGCG-3'
K203 E rev BamHI:	5'-GGCGGATCCGGGTCCAGGAACAGCTTCCGGGAGC-3'
Gal4/SPOC1_BamHI fw:	5'-CGGGGGATCCATGGACTCTGACTCTTGCG-3'
Gal4/SPOC1_BamHI rev:	5'-GGCGGATCCGGTCAGTCCAGGAACAGCTT-3'
K203fw_bp1:	5'-CATGGACTCTGACTCTTGCGCC-3'
K203rev_bp900:	5'-GTCCAGGAACAGCTTCCGGGAG-3'
k203S f1:	5'-CATGATCGAGTGTAATGAGTGC-3'
k203S r1:	5'-GTCAATGGTACCAGCAGTGCTG-3'

1F:	5'-CGAGCCGAACTGGGCGTCAGG-3'
1FN:	5'-GGACTACCACTCCCACAAGGC-3'
1R:	5'-CTGGCTGCCTGAGGACGGCAG-3'
2F:	5'-CTGTGATTCCGCCTGGAAGCC-3'
2R:	5'-GTCAGAAGCCCACGAACTTGCG-3'
3F:	5'-GCTCAAAGTTGTGCAGATGTC-3'
3R_int:	5'-CACTCGAGGGAGTATCGGAGTC-3'
3R:	5'-GAAGGCACCAGACCCTGTGCC-3'
4F:	5'-CATGGAAGCCCAGCTGATGGCG-3'
4R:	5'-GATCTGCCCGCACCAGAGCTGAG-3'

## Primer für die Amplifikation einzelner Exons zur Mutationsanalyse in humanen SPOC1

## Primer zur Klonierung des murinen Spoc1-Gens mittels RT-PCR

MM 2 rev:	5'-CGATGTTGTGTCTGGCCTGGC-3'
Nuc 2 forw:	5'-CGGACAAGCTGAAGAAGAAG-3'
PHD 1 rev:	5'-GTGGCACTCATTACACTCGATC-3'
MM 3 forw:	5'-CATGACAGTTGTTGTCCAGTG-3'
MM 4 rev:	5'-CTCTGTAGATTGTGCAGACCTC-3'
MM 4b rev:	5'-CCAGAGGCATTGATCTCTGAG-3'
MM 5 forw:	5'-CACTGGCAATGATGAAGACATC-3'
MM 6 forw:	5'-GATCGAGTGTAACGAGTGCC-3'
MM 7 rev:	5'-GATCTGCTCACTAGCTATGTACAG-3'
MM 8 forw:	5'-CAGGTCTGCTTCAGCTAACAG-3'
MM 9 forw:	5'-CACTTGTAAGAGGCGCCGGAC-3'
MM 10 rev:	5'-GGATGTCGCGTCGCAGGCTCG-3'
MM 11 rev:	5'-CCAAGGACAGTGTCTTCTGTG-3'
MM 12 forw:	5'-GTGTCCCAGCTCTGCTGATG-3'
MM 13 forw:	5'-CGCTGTGGACACTCCAGTTC-3'
MM 14 forw:	5'-CTGCTCAGGCTGGGACTCTGAC-3'
MM 14 rev:	5'-GTCAGAGTCCCAGCCTGAGCAG-3'
MM 15 rev:	5'-GCTGCAGGTGGCTGAAGCATC-3'
MM 16 forw:	5'-CATGGACTCCGACTCCTGCGCC-3'

## Primer für die RNA in situ-Hybridisierung von murinen Spoc1

AS1:	5'-GCTTGAGGGAGTCTCCACATATGGGTCTGCAGCTTCCAGC-3'
SS1:	5'-GCTGGAAGCTGCAGACCCATATGTGGAGACTCCCTCAAGC-3'
AS2:	5'-CTTCCGGGAGCCCATTCTGGACCGGTTGGAGCGACGGATATC-3'
SS2:	5'-GATATCCGTCGCTCCAACCGGTCCAGAATGGGCTCCCGGAAG-3'
AS3:	5'-CCTCCACTCCCGAGCACAGCGGAGACAGGACAGCTAGCAC-3'
SS3:	5'-GTGCTAGCTGTCCTGTCTCCGCTGTGCTCGGGAGTGGAGG-3'

## 3. Ergebnisse

## 3.1 Auswertung und Charakterisierung von EST-Klonen einer humanen Chondrozyten cDNA-Bibliothek

Die Wachstumsfuge ist ein ideales Gewebe, um Prozesse und Komponenten der Zelldifferenzierung zu studieren. Die Untersuchung von neuen Genen und Gen-Expressionsprofilen, die an der Differenzierung der Chondrozyten beteiligt sind und zur Entwicklung des humanen Skelettsystems führen, wird durch die geringe Verfügbarkeit von menschlichen Knorpelgewebe erschwert. Knorpelgewebe setzt sich aus nur wenigen Zellen (Chondrozyten, Osteoblasten, Osteoklasten) zusammen, die in einer extrazellulären Matrix eingebettet sind. Da die RNA-Extraktion aus Knorpelgewebe sehr schwierig ist und in der Regel nur geringe Ausbeuten erzielt werden, kann nur durch Vereinigung mehrerer Knorpelfragmente eine ausreichende Menge an RNA zur Herstellung einer cDNA-Bibliothek gewonnen werden. Die Erzeugung einer großen Anzahl von ESTs (expressed sequence tags) durch teilweise Sequenzierung von zufällig ausgewählten cDNA-Klonen aus einer geeigneten cDNA Bank stellt eine mögliche Strategie zur systematischen Identifizierung und Analyse von Genen, die an der Entwicklung und Differenzierung beteiligt sind, dar (Adams et al., 1991, Adams et al., 1993).

Ausgangspunkt für das vorliegende EST-Projekt war die zur Zeit einzige verfügbare cDNA-Bibliothek aus humanem, fetalen Knorpel (20. Schwangerschaftswoche bis 2. Lebensjahr), die von Dr. B. Lee (Houston, Texas) mit Hilfe des ZAP"-cDNA Synthese Kits (Stratagene, Heidelberg) hergestellt wurde (Ahn et al., 1995). Die cDNA wurde mit Hilfe eines modifizierten oligo-dT Primers hergestellt und konnte gerichtet in einen "Uni-ZAP" Vektor kloniert werden. Voruntersuchungen hatten bereits gezeigt, dass die cDNA-Bank knorpelspezifische Gene wie z. B. Kollagen Typ II oder Kollagen Typ X enthält. Ferner konnten Ahn et al. (1995) das für die erbliche multiple Exostose verantwortliche Gen (EXT1) aus dieser cDNA-Bibliothek isolieren. Weitere Untersuchungen bestätigten die Qualität der cDNA-Bibliothek. Stelzer (2000), konnte zeigen, dass bei 384 analysierten Klonen die durchschnittliche Integratgröße 550 bp betrug und zahlreiche Knorpel-spezifische Gene enthalten sind. Bei dem EST-Projekt zur Skelettentwicklung (Abb. 3.1) handelt es sich um eine Kooperation der Arbeitsgruppe von Prof. Winterpacht (Hamburg/Erlangen), der Firma GENterprise (Mainz) und der Arbeitsgruppe von Prof. Zabel (Mainz). Die Generierung von ca. 5000 ESTs aus der Chondrozyten cDNA-Bibliothek wurde von der Firma GENterprise (Mainz) übernommen. Vor der Sequenzierung der cDNA-Klone wurden diese durch Amplifikation mit vektorspezifischen T<sub>3</sub>A- und T<sub>7</sub>A-Primern auf die Integratgröße hin untersucht und nur solche Klone verwendet, deren durchschnittliche Integratgröße >350 bp betrug. Die unbearbeiteten Sequenzdaten wurden anschließend auf die drei Kooperationspartner aufgeteilt und über einen FTP-Server verfügbar gemacht (Abb. 3.1).



Abb. 3.1: Struktur des EST-Projektes. Nach der Sequenzierung von ca. 5000 cDNA-Klonen wurden die unbearbeiteten ESTs auf einem FTP-Server bereitgestellt und von den drei Kooperationspartnern weiter bearbeitet.

## 3.1.1 Bioinformatische Auswertung von ESTs mit dem Datenbank-Analyse-Tool "ESTsweep"

Das Datenbank-Analyse-Tool "EST-Sweep" (Abb. 3.2), welches in Gemeinschaftsarbeit der Arbeitsgruppen Prof. Winterpacht, Prof. Zabel und der Firma GENterprise mit der "Bioinformatic Service Group" von HUSAR (Heidelberg Unix Sequence Analyses Resources) im DKFZ entwickelt wurde (Hotz-Wagenblatt et al., 2003), ermöglicht über den HUSAR-Server eine automatisierte BlastN-Datenbank Analyse (NCBI, release 2.2.4, August 2002, http://www.ncbi.nlm.nih.gov/BLAST/) (Altschul et al., 1997) von ESTs in großer Anzahl. Über ein FTP-Transferprotokoll ist es möglich, eine größere Anzahl von Rohsequenzen auf den HUSAR-Server zu transferieren. Die Vorbereitung der Rohsequenzen für die Analyse mit "ESTsweep" erfolgte mit der HUSAR-Anwendung "Phredplus", mit der Sequenzabschnitte mit geringer Komplexität sowie Vektorbestandteile in den Rohsequenzen automatisiert maskiert wurden. Die bearbeiteten Sequenzen können dann als gesamter Block direkt in das Programm "ESTsweep" (Abb. 3.2) importiert und verschiedene BlastN-Datenbank Analysen des gesamten Blocks durchgeführt werden. Dabei besteht die Möglichkeit, aus 117 verschiedenen Datenbanken bis zu sechs verschiedene Datenbanken auszuwählen, mit denen die einzelnen Sequenzen des gesamten Sequenz-Blocks einer BlastN-Analyse unterzogen werden können (http://genius.embnet.dkfz-heidelberg.de/menu/cgibin/w2h/w2h.start).

Go to Go to Go to Advanced Show Extra items Exit						
inside			Application	estsweep		Help on estsweep
		Run	Defaults Restore Rom. © jast now C daily	Save Batcl y C weekly C monthly	h-Queue: long 💌	
C Nud. C Froi	🗌 Giraılar	Begin norange	End norange		List item	
Command Line ( <i>parameters</i> estsweep -NUCdb2="Re	v/options can only be set i efseq_dna (local,	n the section(s) below): nuc) " -NUCdb3="Nr	nuc (l			
			Choose one of the two fi	unctions of ESTSWEE	IP	
use the selecte Use trace (AB) File Specification	d sequences from the 1 1) file n for trace (ABI) file	dain Window Using:				
r			Which genomic nucleotide i	DB to search with Bla	rstN2	
Available databases fo GeAll (local, nuc) Bacterial (local, nuc) EST (local, nuc) Fungi (local, nuc) Human (local, nuc) -Invertebrate (local, nuc)	r Blast search	uman_assembled				
e			Which mrna nucleotide Di	B to search with Blas	IN2	
Available databases for Blast search human_mma (local, nuc) human_inonest (local, nuc) mouse_ensemble (local, nuc) mouse_ensemble_cdna (local, nuc) Using: Refseq_dma (local, n) Which EST DB to search with BlastN2						
Available databases fo	r Blact coarsh			ore meers 20000212		
Viral (local, nuc) EMEL (local, nuc) GenBank (local, nuc) HonEST (local, nuc) EMNew (local, nuc) GBNew (local, nuc)	Using:	(rnuc (local, nuc)				

Abb. 3.2: Startseite von "ESTsweep" auf dem HUSAR-Server. Über "pull down" Menüs können die verschiedenen Datenbanken ausgewählt werden, mit denen eine BlastN-Analyse (z. B. human\_assembled und RefSeq\_dna) eines importierten EST-Blocks durchgeführt werden soll. Es besteht auch die Möglichkeit mit unmaskierten Rohsequenzen eine ESTsweep-Analyse durchzuführen.

Im vorliegenden EST-Projekt zur Skelettentwicklung wurden die EST-Sequenzen einer BlastN-Analyse gegen die Datenbanken "human\_assembled" ("genomische" Datenbank), "RefSeq\_dna" (mRNA Datenbank) und "Nrnuc" ("nicht-redundante" Datenbank) unterzogen (Abb. 3.3).



Abb. 3.3: Automatisierte BlastN-Analysen von ESTs mit dem Datenbank Analyse-Tool "ESTsweep". Über ein FTP-Transferprotokoll werden die ESTs auf dem HUSAR-Server gespeichert, mit der HUSAR-Anwendung "Phredplus" maskiert und in "ESTsweep" importiert. Die Analyse der ESTs erfolgt mit den BlastN-Datenbanken "human\_assembled", "RefSeq\_DNA" und "Nrnuc". Die Ausgabe der Ergebnisse erfolgt im Tabellenformat (Abb. 3.4).

Die Datenbanken "human assembled" und "RefSeq\_dna" enthalten Sequenzen des humanen Genoms und deren entsprechende mRNA- und Proteinsequenzen. Die Datenbank "Nrnuc" enthält alle nicht redundanten GENBANK- und EMBL-Sequenzen, aber keine ESTs oder Sequenzen aus der "htgs" (high throughput genomic sequences) Datenbank. Das Ergebnis der BlastN-Analysen wird in Form einer Tabelle im Blast-Format zusammengefasst, in der alle relevanten Daten, Sequenzen, Alignments und Grafiken der verschiedenen Analysen über Mausklick abrufbar sind (Abb. 3.4). Für jedes Nutzerkonto auf dem HUSAR-Server stehen ca. 60 MB Speicherplatz zur Verfügung. Die Anzahl der Sequenzen, die in einem Block

List of Results							
Refresh Mark All UnMark All							
Remove from the table Delete files Rename file							
				mark jobs "en bloc"			
		C	C	0 · · · T"	<b>G</b> !	m	
	Application	Started	Status	Output File	Size	Туре	
	ESTSWEEP	10:39:25	done	ch67_A01.abd.bin_01_15386.estsweep	59644	Т	
	<u>ESTSWEEP</u>	10:39:25	done	ch67_A02.abd.bin_02_15386.estsweep	80950	Т	
	ESTSWEEP	10:39:25	done	ch67_A04.abd.bin_03_15386.estsweep	493659	Т	
	ESTSWEEP	10:39:25	done	ch67_A07.abd.bin_04_15386.estsweep	203754	Т	
	ESTSWEEP	10:39:25	done	ch67_A08.abd.bin_05_15386.estsweep	371105	Т	
	ESTSWEEP	10:39:25	done	ch67_A09.abd.bin_06_15386.estsweep	65360	Т	
	ESTSWEEP	10:39:25	done	ch67_A10.abd.bin_07_15386.estsweep	1691009	Т	
	ESTSWEEP	10:39:25	done	ch67_A11.abd.bin_08_15386.estsweep	264730	Т	
	ESTSWEEP	10:39:25	done	ch67_B01.abd.bin_09_15386.estsweep	129822	Т	
	ESTSWEEP	10:39:25	done	ch67_B02.abd.bin_10_15386.estsweep	681920	Т	
	ESTSWEEP	10:39:25	done	ch67 B03.abd.bin 11 15386.estsweep	112195	Т	

bearbeitet werden können liegt zwischen 60 und 100, abhängig von der Länge der einzelnen EST-Sequenzen.

Abb. 3.4: Ausgabe der Ergebnisse einer "ESTsweep"-Analyse im Tabellenformat. In der Spalte "Output File" sind die Ergebnisse für jeden einzelnen analysierten EST in Form von Alignments und Sequenzen abrufbar.

## 3.1.2 Auswertung der Klone aus der Chondrozyten cDNA-Bank

Im Rahmen der vorliegenden Doktorarbeit erfolgte die bioinformatische Auswertung von 450 Klonen aus der Chondrozyten cDNA-Bank bis zur Verwendung von "ESTsweep" mit Hilfe des Programms BlastN (http://www.ncbi.nlm.nih.gov/BLAST/) gegen die Datenbanken "human\_assembled", RefSeq\_dna" und "Nrnuc". Die Analyse von weiteren 600 Klonen der cDNA-Bank wurde dann automatisiert mit "ESTsweep" gegen die gleichen Datenbanken durchgeführt. Die detaillierte bioinformatische Auswertung der Klone umfasste nicht nur eine ausgedehnte Homologiesuche in verschiedenen Datenbanken, sondern auch die umfassende Analyse der korrespondierenden chromosomalen Regionen und die Identifizierung von EST-Clustern mit Hilfe der UniGene-Datenbank, in der überlappende ESTs automatisch zu einer Gruppe (Cluster) zusammengefasst werden (http://www.ncbi.nlm.nih.gov/UniGene/). Auf diese Weise konnte für nahezu alle cDNA-Klone die genomische Lokalisation aufgeklärt

werden. Darüber hinaus ermöglichten die in silico-Analysen bei einem Teil der Klone die Ermittlung weiter Bereiche der putativen Protein-kodierenden Region. Bei den Datenbank-Suchen wurden Sequenz-Homologien mit einem P-Wert (Statistisches Signifikanzniveau) kleiner als 10<sup>-10</sup> als Signifikant angesehen (Claudio et al., 1998). Die Ergebnisse der bioinformatischen Auswertungen wurden für weitere Untersuchungen in einer Datenbank (Bsp. s. Tab. 3.1) zusammengefasst, die für jeden Klon alle wichtigen Daten der in silico-Analysen enthält. Nach den Ergebnissen der Homologievergleiche lassen sich die ESTs in unterschiedliche Kategorien einteilen. Die erste Kategorie (I) beinhaltet Klone, die eine signifikante Homologie zu bekannten cDNA-Sequenzen oder Proteinen bekannter Funktion besitzen. (Bsp. s. Tab. 3.1). In der zweiten Kategorie (II) sind Klone zusammengefasst, die Homologien zu cDNA-Sequenzen und Proteinen mit unbekannter Funktion zeigen (Bsp. s. Tab. 3.1). Die dritte Kategorie (III) enthält Klone, die keine signifikanten Homologien aufweisen (P-Wert größer 10<sup>-10</sup>), oder nur Homologien zu ESTs unbekannter Funktion zeigen. Weiterhin enthält diese Kategorie Klone, für die ausschließlich Homologien zu genomischen Sequenzen gefunden wurden (Bsp. s. Tab. 3.1). Die Analyse der eigenen Daten im Vergleich zu den gesammelten Daten für das Gesamtprojekt zeigte, dass die Verteilung der 1050 von mir analysierten Klone in die drei Kategorien (I-III) der Verteilung der Klone des Gesamtprojekts entspricht. Daher werden im Folgenden die Ergebnisse der EST-Analysen des Gesamtprojektes dargestellt (Abb. 3.5).



Abb. 3.5: Verteilung der EST-Klone des Gesamtprojektes. Von den insgesamt analysierten 4748 EST-Klonen können 3426 Klone (72,2%) cDNA-Sequenzen oder Proteinen bekannter Funktion zugeordnet werden<sup>1</sup>. 667 EST-Klone (14%) zeigen ausschließlich Homologien zu genomischen Sequenzen oder ESTs unbekannter Funktion<sup>2</sup>. 490 EST-Klone (10,3%) besitzen Homologien zu cDNAs/Proteinen unbekannter Funktion<sup>3</sup>. Bei 165 Sequenzen (3,5%) handelt es sich um Kontaminationen mit *E. coli*- oder Vektorsequenzen<sup>4</sup>.

Sequence name	Length (bp)	1, v A	mRNA	Description of the mRNA	har accession to 4	Chromosomal localization	RNA or EST accession no.	Genomic clone	TuiCana chictor 7
	D	ŧ	accession no. (RefSeq dna) <sup>2</sup>	product <sup>3</sup>			(Nrmuc) <sup>5</sup>	(Nrnuc) <sup>6</sup>	
Kategorie I									
ch3c9	338		XIM_003989	Homo sapiens secreted I protein, acidic, cysteine-rich (osteonectin) (SPARC),	NT_006951	5	BF968209	Ac011374	Hs.111779
ch16d10	415		NM_00228	mrwA Human c-jun proto oncogene (JUN), complete	NT_004460	1	J04111	A1136985	Hs.78465
				cds					
ch19a3	524		NM_000967	Homo sapiens ribosomal   protein L3 (RPL3) mRNA	NT_007 <i>577</i>	6	BF969718	no match	Hs.119598
ch30a6	506		NM_001844	Homo sapiens collagen, 1 type II, alpha 1	NT_009785	12	XD6268	no match	Hs.81343
Kategorie II	-								
1T0203	446		no match	Novel human gene mapping   to chomosome 1.	NT_019267	1	AL121733	AL031447	Hs.7299
ch15d2	570		NM_015629	Homo sapiens DKFZP566J153 protein (DKFZP566J153) , mRNA	NT_011148	19	Be740206	no match	Hs.183438
ch19h8	365		NM_016209	Homo sapiens unknown   (LOC51693), mRNA (auch: AF161524)	NT_010404	16	no match	Ac009149	Hs.27445
ch20f2	439	polyA	no match	NIH_MGC_83 Homo sapiens cDNA clone	NT_011520	22	Bf680434	A c023979	no match
ch31d9	274	polyA	no match	NCI_CGAP_Lym12Homo 1 sapiens cDNA clone	NT_019508	10	A <i>w77</i> 0011	A1513174	Hs.48516
ch31fl	356	polyA	NM_025203	Homo sapiens hypothetical   protein FLJ21945	NT_005204	2	Af218013	Ac034295	Hs.24624
ch55E08	290	polyA	NM_014805	Homo sapiens KIAA0766 1 gene product	NT_028117	3	Aul 44230	Ac011816	Hs.28020
ch55F04	416	polyA	NM_030571	Homo sapiens hypothetical   protein	NT_006489	5	Be881059	Ac004752	Hs.9788
Kategorie III									
ch3c5	362		no match	no match	NT_009308.2	11	no match	AF125183	no match
ch15d1	687		no match	no match	no match	15	no match	AC087433	no match
ch55C09	571	polyA	no match	no match	NT_004623	1	AL117644	Aq316084	no match

Kategorien. Kategorie I beinhaltet Klone, die signifikante Homologien zu bekannten cDNA-Sequenzen oder Proteinen bekannter Funktion besitzen. Kategorie II enthält Klone, die Homologien zu cDNA-Sequenzen und Proteinen unbekannter Funktion besitzen. Kategorie III enthält alle Klone, die keine Tab. 3.1: Auszug aus der Ergebnis-Tabelle der bioinformatischen Analyse der EST-Klone. Die Einteilung der Klone erfolgte in drei verschiedene signifikanten Homologien, oder nur Homologien zu ESTs unbekannter Funktion zeigen. Außerdem enthält diese Kategorie Klone, die ausschließlich Homologien zu genomischen Sequenzen besitzen. Poly-Adenylierungssignal<sup>1</sup>, mRNA Accession-Nr. aus der "RefSeq"-Datenbank<sup>2</sup>, Beschreibung der mRNA<sup>3</sup>, Sequenz Accession-Nr. aus der "htgs"-Datenbank<sup>4</sup>, RNA- oder EST Accession-Nr. aus der "Nrnuc"-Datenbank<sup>5</sup>, Accession-Nr. von genomischen Klonen<sup>6</sup>, "Unigene" Accession-Nr<sup>7</sup>.

#### 3.1.2.1 EST-Klone mit Homologien zu Sequenzen bekannter Funktion (Kategorie I)

Von den 4748 analysierten EST-Klonen konnten 3426 ESTs (72,2%) der Kategorie I zugeordnet werden, da sie signifikante Homologien zu cDNA-Sequenzen oder Proteinen bekannter Funktion besitzen (Abb. 3.5). Diese Klone lassen sich in verschiedene Gruppen einteilen (Abb. 3.6). Den größten Anteil dieser Gruppe zeigen 1303 (38,1%) EST-Klone, die Homologien zu verschiedenen bekannten Genen oder Proteinen besitzen. 887 (25,9%) der EST-Klone zeigen Homologien zu der Gruppe der ribosomalen Proteine. Weitere 886 (25,8%) der EST-Klone können der Gruppe der Haushaltsgene zugeordnet werden. Für 165 Klone (3,5%) ergab der Homologievergleich, dass es sich um Kontaminationen der cDNA-Bank mit Vektor- und Bakteriensequenzen handelt (Abb. 3.5). 350 EST-Klone (10,2%) zeigen signifikante Homologien zu Genen oder Proteinen, die spezifisch für Knorpel- oder Knochengewebe sind.



25,8% Haushaltsgene<sup>3</sup>

Abb. 3.6: Verteilung der 3426 Klone mit Homologien zu bekannten Genen (Kategorie I). 1303 EST-Klone (38,1%) zeigen Homologien zu verschiedenen bekannten Genen und Proteinen<sup>1</sup>. 887 EST-Klone (25,9%) können der Gruppe der ribosomalen Proteine zugeordnet werden<sup>2</sup>. Bei 886 EST-Klonen (25,8%) handelt es sich um Sequenzen mit Homologien zu Haushaltsgenen<sup>3</sup>. 350 EST-Klone (10,2%) aus dieser Kategorie zeigen Homologien zu Knochen- und knorpelspezifischen Genen<sup>4</sup>.

## 3.1.2.2 EST-Klone mit Homologien zu Sequenzen unbekannter Funktion oder genomischen Sequenzen (Kategorien II und III)

Der Kategorie II konnten 490 EST-Klone (10,3%) zugeordnet werden, die signifikante Homologien zu cDNA-Sequenzen und Proteinen unbekannter Funktion zeigten (Abb. 3.5). Weitere 667 EST-Klone (14%) wurden der Kategorie III zugeordnet, da sie ausschließlich Homologien zu genomischen Sequenzen oder ESTs unbekannter Funktion oder keine signifikanten Homologien zeigten (Abb. 3.5).

#### 3.1.3 Analyse von 4 EST-Klonen aus der Chondrozyten cDNA-Bibliothek

Die Auswahl von Klonen für weitergehende Untersuchungen erfolgte aus den Kategorien II und III. Beide enthalten cDNAs und Proteine unbekannter Funktion, die putative Kandidatengene von medizinischer Relevanz darstellen können. Abhängig von den Expressionsdaten können diese putativen Kandidatengene auch an Differenzierungsprozessen während der Skelettentwicklung sowie an der Homöostase des Skelettsystems beteiligt sein. Kategorie III enthält EST-Klone, die gar keine Homologien, oder nur Homologien zu ESTs unbekannter Funktion zeigen. Weiterhin enthält diese Kategorie EST-Klone, die ausschließlich Homologien zu genomischen Sequenzen zeigen. Wichtig für die Auswahl von Kandidatengenen war, dass diese keine signifikanten Homologien zu bereits bekannten Genen zeigten. Weitere wichtige Auswahlkriterien waren im Gen enthaltene Domänen, die an Entwicklungs- und Differenzierungsprozessen beteiligt sein können (wie z. B. Domänen, die auf einen Transkriptionsfaktor hinweisen), die chromosomale Lokalisation (z. B. in einer Region die mit der Entstehung verschiedener Tumoren oder anderer Erkrankungen assoziiert ist) und, falls bekannt, das Expressionsmuster. Mit Hilfe der durch die UniGene-Datenbank ermittelten EST-Gruppen (Cluster) wurden die für genauere Untersuchungen ausgewählten Klone weiter analysiert. Weitere Informationen bezüglich der chromosomalen Lokalisation und der Expression in verschiedenen Geweben sind angegeben. Mit den erweiterten Sequenzdaten wurden erneut Blast-Analysen auf Nukleotidebene (BlastN) gegen die Datenbanken "Nrnuc", "est" (enthält alle GENEBANK und EMBL EST-Sequenzen) und "htgs" durchgeführt. Für den Vergleich der abgeleiteten Aminosäuresequenzen der EST-Klone mit Proteindatenbanken wurde das Programm "BlastX" verwendet.

Die Auswertung der Sequenzdaten des EST-Klons T01 203 (Kategorie II, 446 Bp) mit der UniGene-Datenbank führte zur Identifizierung des UniGene-Clusters Hs.7299, der signifikante Homologien zu einem putativen neuen menschlichen Gen unbekannter Funktion auf Chromosom 1 zeigte (Tab. 3.2). Anhand der Expressionsdaten der ESTs aus dem genannten Cluster war ersichtlich, dass das putative Gen in sehr vielen verschiedenen menschlichen Geweben, unter anderem aber auch in Tumoren des Knorpels (Chondrosarkom) und osteoarthritischen Knorpel exprimiert wird. BlastN-Analysen des EST-Klons mit der Datenbank "Nrnuc" zeigten eine 98%ige Homologie zu einer unvollständigen mRNA-Sequenz (Accession Nr.: AL121733, 3158 Bp), die für ein neues humanes Gen unbekannter Funktion auf Chromosom 1 kodiert (Tab. 3.2) und einen längeren offenen Leserahmen (762 Bp) enthält. Die mRNA-Sequenz war im 5' Bereich unvollständig, da sie kein Startcodon enthielt. Die Analyse der Nukleotidsequenz des EST-Klons mit der "htgs" Datenbank ergab ebenfalls eine signifikante Homologie des EST-Klons zu einem genomischen Klon (Accession Nr.: NT 019267.3), der auf Chromosom 1p36.21-36.33 lokalisiert ist. Ein Homologievergleich der Aminosäuresequenz des offenen Leserahmens mit der Proteindatenbank "PROSITE" (http://www.expasy.org/cgi-bin/scanprosite) zeigte, dass das putative Protein eine bipartite Kern-Lokalisationssequenz und eine PHD-Domäne (Plant-Homeodomain) enthält. Der Klon T01 203 wurde für weitere funktionelle Untersuchungen ausgewählt, weil es sich bei diesem um ein unbekanntes humanes Gen handelt, welches in einer chromosomalen Region Lokalisiert ist, die mit Tumorentstehung (z. B. Chondrosarkom) assoziiert ist. Weiterhin ist die enthaltene PHD-Domäne charakteristisch für Gene, die an der Regulation der Transkription beteiligt sind. Das korrespondierenden Genprodukt könnte aus diesem Grund an Entwicklungs- und Differenzierungsprozessen beteiligt sein. Die Isolierung und Charakterisierung des neuen humanen Gens ist in Abschnitt 3.2 beschrieben.

## Sequenz des EST-Klons T01 203 Länge: 411 Bp

#### 3.1.3.2 Analyse des EST-Klons T01 517

Die Auswertung der Sequenzdaten des EST-Klons T01 517 (Kategorie III, 310 Bp) mit Hilfe der UniGene-Datenbank (Tab. 3.2) ergab eine signifikante Homologie zu dem UniGene-Cluster Hs.49272, der 3 ESTs (W87823, N66846 und R47302) enthält und auf Chromosom 6 lokalisiert ist. Die BlastN-Analyse des EST-Klons (Tab. 3.2) ergab eine einzige signifikante Sequenzübereinstimmung zu dem genomischen Klon HS91J24 (Accession Nr.: AL024474), der ebenfalls auf Chromosom 6q24 lokalisiert ist und einen Teil des Utropin Gens, einen Teil des Cytochrom C Oxidase Gens und verschiedene ESTs enthält. Da über die UniGene- und Blast-Analysen keine weiteren Sequenzdaten gewonnen werden konnten, wurde beim Deutschen Ressourcenzentrum für Genomforschung (RZPD) (http://www.rzpd.de/) über Datenbank-Sequenzvergleiche der entsprechende IMAGE-Klon (IMAGp998B23660Q2) herausgesucht, bestellt und sequenziert. Mit Hilfe dieses Klons konnte die Sequenz des Klons T01 517 im 3' Bereich auf gesamt 911 Bp verlängert werden. Erneute UniGene- und Blast-Analysen ergaben keine weiteren Homologien zu neuen Sequenzen. Aufgrund der Tatsache, dass sich die 3 genannten ESTs relativ genau überlappen und in einem Sequenzbereich des genomischen Contigs liegen, der repetitive Elemente und keine Exons enthält, wurde der EST-Klon T01 517 nicht weiter bearbeitet, da es sich bei diesem Klon mit hoher Wahrscheinlichkeit um eine Kontamination der cDNA-Bibliothek mit genomischer DNA handelt. Diese Annahme wird durch das Fehlen eines offenen Leserahmens in allen Leserastern unterstützt.

#### Sequenz des EST-Klons T01 517 Länge: 310 Bp

TAAGCAGCAGCCAGTGGGATGTGAAAAGAAATCAGGAATGCAACCTGAGGGCCA TGCACATTTTAAAAAATTGCTTATTTTTACTTCTCTTTTCTCTTTGGACTAG AACTTGACATAGTGTTGAGGGGGCAAGCCTCATCAAGAAAATGGAACAACACTTT GGAAAGGCAGAACAAGAGAGAAGGAACCAGGATCCCTGAGAGACCATATGAAG CGGAGCCACCTTCCCACCCTCCATCACCCACCAACTCTGAACTGTAAGGAGAGAA AGAAATAAACTCTGATTTTATTTGAGCCACTGTATTT

#### 3.1.3.3 Analyse des EST-Klons T01 519

Der Vergleich der Sequenz des Klons T01 519 (Kategorie III, 408 Bp) mit der UniGene-Datenbank (Tab. 3.2) zeigte eine Homologie zu dem UniGene-Cluster Hs.13872, der zunächst 6 ESTs enthielt, die in verschiedenen Geweben, aber auch in Knorpel, osteoarthritischen Knorpel und Knorpel-Tumoren (Chondrosarkom) exprimiert werden. Die ESTs dieses Clusters, der auf Chromosom 4 lokalisiert ist, zeigten zunächst keine Homologie zu einer bekannten cDNA oder mRNA. Die Sequenzierung von zwei RZPD-Klonen (IMAGp998P151625Q2 und IMAGp998D221965Q2) führte zu einer Verlängerung des 5' Bereichs der EST-Sequenz auf 801 Bp. Erneute Datenbank Analysen der verlängerten Sequenz mit dem Programm BlastN (Tab.3.2) zeigten, dass der Klon T01 519 eine signifikante Homologie zu der mRNA des "Homo sapiens cytokine-like protein C17" besitzt (Accession Nr.: NM 018659). Durch einen Homologievergleich der erweiterten EST-Sequenz mit der mRNA des "Homo sapiens cytokine-like protein C17" mit dem Programm "ALIGN" (GENESTREAM Server, http://www2.igh.cnrs.fr/bin/align-guess.cgi) konnte gezeigt werden, dass es sich bei dem Klon T01 519 um einen Sequenzabschnitt aus dem 3' untranslatierten Bereich des "Homo sapiens cytokine-like protein C17" handelt und dieser daher nicht weiter bearbeitet wurde. Ein mögliches Polyadenylierungssignal ist unterstrichen.

#### Sequenz des Klons T01 519 Länge: 310 Bp

## 3.1.3.4 Analyse des Klons T01 564

Die Analyse des EST-Klons T01 564 (Kategorie III, 338 Bp) mit Hilfe der UniGene Datenbank ergab zunächst keine Homologie zu einem UniGene-Cluster, einer unbekannten cDNA oder mRNA (Tab. 3.2). Bei der BlastN-Analyse des Klons mit der Datenbank "est" konnten Homologien zu 9 verschiedenen ESTs (AA687387, H25549, AA083211, AI765633, R88138, AW574971, AW572575, AA083310 und H25593) festgestellt werden, die aber zu keiner signifikanten Verlängerung der bereits vorliegenden cDNA-Sequenz beitragen 5 konnten. Die Sequenzierung von RZPD Klonen (IMAGp998N073104O2, IMAGp998M15260Q2, IMAGp998C091320Q2, IMAGp9980095961Q2 und IMAGp998P07349) führte zu einer Verlängerung der Sequenz des EST-Klons im 5' Bereich auf 1578 Bp. Durch einen erneuten Homologievergleich der verlängerten Sequenz des EST-Klons mit der UniGene-Datenbank konnte der UniGene-Cluster Hs.391653 identifiziert werden, bei dem es sich um das Gen EPS15R handelt ("epidermal growth factor receptor substrate EPS15R", Accession Nr.: BC037558), welches auf Chromosom 19 lokalisiert ist. Die signifikante Sequenzhomologie zwischen dem erweiterten EST-Klon und der mRNA des EPS15R-Gens wurde mit Hilfe des "ALIGN"-Programms verifiziert. Da es sich bei dem Klon T01 564 ebenfalls um einen Sequenzabschnitt eines bekannten Gens handelte, wurde dieser nicht weiter bearbeitet.

#### Sequenz des Klons T01 564 Länge: 338 Bp

Klon	UniGene	Chromosomale	Homologievergleich	Homologievergleich
	Cluster	Lokalisation	mit BlastN	mit BlastX
T01 203	Hs.7299	1p36.21-36.33	AL121733 Novel human	CAB57324
			gene mapping to	hypothetical protein
			chromosome 1	[Homo sapiens]
T01 517	Hs.49272	6q24	AL024474 Human DNA	keine Homologie
			sequence from clone	
			91J24 on chromosome	
			6q24	
T01 519	Hs.13872	4	NM_018659 Homo	NP_061129 cytokine-
			sapiens cytokine-like	like protein C17 [Homo
			protein C17 (C17),	sapiens]
			mRNA	
T01 564	Hs.391653	19	BC037558 Homo	keine Homologie
			sapiens, Similar to	
			epidermal growth factor	
			receptor substrate	
			EPS15R	

Tab. 3.2: Zusammenfassung der UniGene und BlastN/X-Auswertungen von 4 ausgewählten EST-Klonen aus der Chondrozyten cDNA-Bibliothek.

# 3.2 Klonierung und Chromosomale Lokalisation des humanen und murinen SPOC1 (Spoc1) -Gens

## 3.2.1 Klonierung und Sequenzierung des menschlichen SPOC1-Gens

Wie bereits in Abschnitt 3.1.3.1 beschrieben, zeigte der EST-Klon T01 203 in ersten UniGene- und BlastN-Analysen auf Nukleotidebene eine Homologie von 100% zu einem humanen cDNA-Klon (Accession Nr.: AL121733) und zu einem genomischen Contig (Accession Nr.: NT\_019267.3). Da der cDNA-Klon nicht die vollständige kodierende Region des unbekannten Gens enthielt und der 5' Bereich nicht durch EST-Sequenzen aus den Datenbanken verlängert werden konnte, wurde der genomische Contig NT 019267.3 mit dem Genvorhersage-Programm "GENESCAN" (http://genes.mit.edu/GENSCAN.html) analysiert. Mit Hilfe der durch "GENESCAN" vorhergesagten putativen Exonsequenzen konnten Primer-Sequenzen abgeleitet werden und die kodierende Region des unbekannten Gens durch RT-PCR Experimente an RNA aus humanen Fibroblasten vollständig kloniert und sequenziert werden. Die Klonierungsstrategie des SPOCI-Gens ist in Abb. 3.7 dargestellt. Das Gen besteht aus 4 Exons (1, 2, 3 und 4), von denen Exon 1 ein Startcodon, Exon 3 eine bipartite Kernlokalisationssequenz und Exon 4 eine PHD-Domäne (plant homeodomain) sowie das Stopcodon enthält (Abb. 3.8). Drei weitere, im 5' Bereich des neuen Gens gelegene, putative Exons (a, b und c), die im gleichen Leseraster wie die Exons 1-4 liegen, konnten dabei nicht durch RT-PCR Experimente verifiziert werden. Exon a enthält ein Startcodon, welches mit dem Startcodon der verifizierten kodierenden Sequenz in einem Leseraster liegt. Exon c enthält eine weitere Kernlokalisationssequenz (Abb. 3.8). Das Gen wurde als SPOC1 (survival time - associated PHD protein in ovarian cancer) bezeichnet und unter der Accession Nr.: AY069975 in die GenBank-Datenbank eingegeben. Die vollständige Nukleotidsequenz wurde mittels RT-PCR an RNAs aus Fibroblasten, Patella-, und Plazentagewebe analysiert und durch Sequenzierung beider Stränge verifiziert. Während der Entstehung dieser Arbeit wurde die komplette cDNA des SPOCI-Gens (Accession Nr.: XM\_086204) durch das "NCBI Annotation Project" (2002) annotiert, so dass mit Hilfe eines neuen genomischen Contigs (Accession Nr.: NT 028054.9) die vollständige genomische Struktur des SPOC1-Gens ermittelt werden konnte (Abb. 3.8). Die kodierenden Sequenzen der einzelnen Exons des SPOC1-Gens und der annotierten cDNA XM 086204 sind vollständig identisch (Daten nicht gezeigt).



Abb. 3.7: Klonierungsstrategie des menschlichen *SPOC1*-Gens. Der EST-Klon T01 203, der als Ausgangspunkt für die Identifizierung des cDNA-Klons AL121733 (gepunktete Linie) diente, ist als rote Linie abgebildet. Die Nukleotidsequenzen der RT-PCR-Produkte (Ex0f/c2r, c2f/c4r, K203F1/K203R1, c5f/c6rII und c3f/c5r) sind als schwarze Linien dargestellt. Zur Orientierung sind neben dem möglichen Polyadenylierungssignal das verifizierte Startcodon (2), sowie das den offenen Leserahmen begrenzende Stopcodon eingezeichnet. Die Kernlokalisationssequenz ist als graues Kästchen, die PHD-Domäne als blaues Kästchen dargestellt.

Die abgeleitete cDNA-Konsensus-Sequenz des *SPOC1*-Gens besitzt eine Länge von 3681 Bp und enthält einen offenen Leserahmen von 900 Bp, der für ein putatives, aus 300 Aminosäuren bestehendes Protein kodiert (Abb. 3.13). Die Nukleotidsequenz um das putative Startcodon in Exon 1 von *SPOC1* (CGG<u>A</u>ACATG<u>G</u>) stimmt mit dem Konsensus-Motif für eukaryotische Translationsstartpunkte GCC(A/G)CCATGG (Position -6 bis +4) (Übersichtsartikel: Kozak, 1996) überein. Die stromaufwärts gelegene Position –3 wird durch den hochkonservierten Purinrest A eingenommen, an der stromabwärts gelegenen Position +4 befindet sich ein hochkonservierter Purinrest (G).



Abb. 3.8: Struktur des humanen *SPOC1*-Gens. Die durch RT-PCR-Experimente verifizierten Exons (1-4) sind als blaue Kästchen dargestellt. Das Startcodon von *SPOC1* liegt in Exon 1, das Stopcodon in Exon 4. Die PHD-Domäne ist als rotes Kästchen, die Kernlokalisationssequenz als graues Kästchen dargestellt. Die nicht verifizierten putativen Exons (a-c) sind als graue Kästchen abgebildet. Die cDNA-Konsensus-Sequenz hat eine Länge von 3681 Bp und enthält einen durchgehenden offenen Leserahmen von 900 Bp.

## 3.2.2 Klonierung und Sequenzierung des murinen Spoc1-Gens

Um weiterführende Untersuchungen des *SPOC1*-Gens in der Maus durchführen zu können, wurde das orthologe, murine *Spoc1*-Gen wie in Abb. 3.9 dargestellt kloniert. Über Homologievergleiche der *SPOC1*-cDNA-Sequenz gegen die "est\_mouse" Datenbank (Programm "BlastN") konnten mehrere murine EST-Klone (AI505407, BF101276, BF583303, AA624523, BI133875, W83552 und C87828) identifiziert werden, die Sequenzübereinstimmungen von mehr als 80% zum humanen *SPOC1*-Gen zeigten und mit großer Wahrscheinlichkeit dem murinen SPOC1 entsprechen. Aus der Sequenz der genannten EST-Klone konnten Primer abgeleitet werden, die für RT-PCR-Experimente an RNA aus Maus-Testis und die nachfolgende Sequenzierung der PCR-Produkte verwendet wurden (Abb. 3.9). Zur Verifizierung des 5' Bereichs der cDNA des murinen Gens wurde über Homologievergleiche mit der "mouse\_est"-Datenbank der korrespondierende "IMAGE"-Klon (IMAGp998M202777) ausgewählt und sequenziert. Die Klone enthielten vornehmlich das 5'-Ende der cDNA.



Abb. 3.9: Klonierungsstrategie des murinen *Spoc1*-Gens. Die durch Homologievergleiche identifizierten murinen EST-Klone AI505407, BF101276, BF583303, AA624523, BI133875, W83552 und C87828 (gepunktete Linien) wurden für die Ableitung von Primersequenzen verwendet. Die Nukleotidsequenzen der RT-PCR-Produkte (MM13fw/MM7rev, MM6fw/MM11rev, MM12fw/MM2rev und MM3fw/MM2rev) sind als schwarze Linien dargestellt. Neben dem möglichen Polyadenylierungssignal sind das verifizierte Startcodon, sowie das Stopcodon eingezeichnet. Die Kernlokalisationssequenz ist als graues Kästchen, die PHD-Domäne als blaues Kästchen abgebildet. Der zur Verifizierung des 5' Bereichs des *Spoc1*-Gens verwendete "IMAGE"-Klon (IMAGp998M202777) ist rot dargestellt.

Die vollständige, durch überlappende PCR-Produkte erhaltene Nukleotidsequenz wurde mittels RT-PCR an RNAs aus Maus Testis, Gehirn, Leber, Niere und ganzer Maus-Embryo (Tag 17) amplifiziert und durch Sequenzierung beider Stränge verifiziert (Daten nicht gezeigt). Die Genstruktur des murinen Gens konnte durch Homologievergleiche der verifizierten cDNA-Sequenz mit der Mausgenomsequenz (Accession Nr.: AL611931) aufgeklärt werden. Im Gegensatz zur humanen Sequenz konnten in der murinen Sequenz mit Genvorhersage-Programmen keine weiteren, 5' gelegenen, putativen Exons identifiziert werden. Ein Vergleich mit der Struktur des *SPOC1*-Gens zeigt, dass der Aufbau des murinen Gens dem Aufbau des humanen Gens entspricht. Das murine Gen besteht aus 4 Exons (1, 2, 3 und 4), von denen Exon 1 ein Startcodon, Exon 3 eine bipartite Kernlokalisationssequenz und Exon 4 eine PHD-Domäne und das Stopcodon enthält (Abb. 3.10). Das neue murine Gen wurde als *Spoc1* bezeichnet und unter der Accession Nr.: AY069976 in die GenBank-Datenbank eingegeben. Die abgeleitete cDNA-Konsensus-Sequenz des *Spoc1*-Gens besitzt

eine Länge von 3127 Bp und enthält einen offenen Leserahmen von 888 Bp, der für ein putatives, aus 296 Aminosäuren bestehendes Protein kodiert (Abb. 3.13). Die Nukleotidsequenz um das putative Startcodon von *Spoc1* (CGG<u>A</u>ACATG<u>G</u>) stimmt wie bei dem humanen *SPOC1*-Gen mit dem Konsensus-Motif für eukaryotische Translationsstartpunkte GCC(A/G)CCATGG (Position -6 bis +4) (Kozak, 1996) überein. Die stromaufwärts gelegene Position –3, sowie die stromabwärts gelegene Position +4 um das Startcodon sind durch die hochkonservierten Purinreste A bzw. G besetzt und erfüllen dadurch die Bedingungen eukaryotischer Translationsstartpunkte.



Abb. 3.10: Struktur des murinen *Spoc1*-Gens. Die durch RT-PCR Experimente verifizierten Exons (1-4) sind als blaue Kästchen dargestellt. Das Startcodon von *Spoc1* liegt in Exon 1, das Stopcodon in Exon 4. Die cDNA-Konsensus-Sequenz hat eine Länge von 3127 Bp und besitzt einen durchgehenden offenen Leserahmen von 888 Bp, der eine bipartite Kernlokalisations-Sequenz (graues Kästchen) und eine PHD-Domäne (rotes Kästchen) enthält.

## 3.2.3 Chromosomale Lokalisation des humanen SPOC1-Gens

Durch "BlastN" Homologievergleiche der cDNA-Sequenz des humanen *SPOC1*-Gens mit der "htgs" Datenbank konnte gezeigt werden, dass *SPOC1* auf dem genomischen Contig NT\_019267.3 in der Chromosomenregion 1p36.21-36.33 kartiert. Die Region 1p36 ist mit der Entstehung verschiedener humaner Tumoren und Leukämien assoziiert, wobei zum Teil bisher unbekannte Gene hier eine Rolle spielen müssen. Der Contig NT\_019267.3 wurde im Januar 2003 durch den genomischen Contig NT\_028054.12 ersetzt (International Human Genome Sequencing Consortium), der eine genaue Lokalisation des humanen *SPOC1*-Gens in der Region 1p36.23 ermöglichte (Abb. 3.11). Zusätzlich enthält das humane *SPOC1*-Gen einen intragenischen Marker (SHGC-150), der im 3' untranslatierten Bereich von *SPOC1* liegt (Abb. 3.11).



Abb. 3.11: Chromosomale Lokalisation und Orientierung des humanen *SPOC1*-Gens in der Region 1p36.23. Die Ausschnittvergrößerung zeigt die genaue Lokalisation von *SPOC1* im genomischen Contig NT\_028054.12. Der im 3' UTR von *SPOC1* liegende intragenische Marker SHGC-150 ist durch einen blauen Pfeil markiert. *SPOC1* liegt wie das benachbarte Gli-Kruppel-family-member-Gen *HKR3* und das hypothetische Protein BC005114 auf dem (+) Strang, während das KIAA0469-Genprodukt und das hypothetische Protein FLJ10737 auf dem (-) Strang liegen.

## 3.2.4 Chromosomale Lokalisation des murinen Spoc1-Gens

Die Feststellung der chromosomalen Lokalisation des murinen *Spoc1*-Gens erfolgte mit Hilfe eines "Maus/Hamster Radiation Hybrid Panels" (Research Genetics, Inc., USA), da zum Zeitpunkt des Experiments die murine genomische Sequenz noch nicht durchgehend in den
Datenbanken vorlag. Bei dem Hybrid-Panel handelt es sich um 100 DNAs, die aus verschiedenen Maus/Hamster-Hybridzellen isoliert wurden und mit deren Hilfe eine genaue Bestimmung der chromosomalen Lokalisation des murinen Spoc1-Gens möglich ist. Um die genaue Lokalisation festzustellen, wird mit Hilfe von PCR-Experimenten mit einem spezifischen Primerpaar untersucht, in welcher der 100 DNAs ein PCR-Produkt nachgewiesen werden kann (positiv) und in welchen nicht (negativ). Mit den durch die PCR-Experimente ermittelten Ergebnissen kann in einer Datenbank die chromosomale Lokalisation ermittelt werden. Durch Verwendung der Primerkombination MM12fw/MM4brev konnten 30 positive und 70 negative Hybridklone identifiziert werden. Die Analyse der PCR-Ergebnisse erfolgte Jackson Laboratory (Bar Harbour, USA) (http://www-genome.wi.mit.edu/cgiam bin/mouse rh/rhmap-auto/rhmapper.cgi). Die Auswertung der chromosomalen Lokalisation in der Maus ergab einen LOD-Score (Summe der dekadischen Logarithmen der Wahrscheinlichkeit auf Kopplung) von 19,8 zu den Markern D4Mit63 und D4Mit180 auf Maus Chromosom 4, zwischen 79 und 80,1 cM (Abb. 3.12). Diese Region ist entsprechend der "human mouse homology map" (NCBI, http://www.ncbi.nlm.nih.gov/Homology/) syntän zu dem humanen Chromosom 1p36.2 und bestätigt die Klonierung des orthologen Maus-Gens von SPOC1.



Abb. 3.12: Chromosomale Lokalisation des murinen *Spoc1*-Gens auf Maus Chromosom 4 (rotes Kästchen). Die Ausschnittvergrößerung zeigt die Lokalisation von *Spoc1* mit einem LOD-Score von 19,8 zwischen den Markern D4Mit63 und D4Mit180 (rot markiert) in einem Bereich von 79 bis 80,1 cM.

#### 3.2.5 Homologievergleiche des humanen und murinen SPOC1-Gens

Die Homologievergleiche der Nukleotidsequenzen des humanen und murinen *SPOC1*-Gens zeigten große Übereinstimmung zwischen beiden Genen. So ergab der Vergleich der cDNAs beider Gene im Bereich der offenen Leserahmen eine Sequenzidentität von 87,3%. Die untranslatierten Regionen beider cDNAs zeigten im Homologievergleich erwartungsgemäß eine wesentlich geringere Homologie von 57,8%. Auch der Vergleich der Organisation des humanen und murinen *SPOC1*-Gens ergab große Übereinstimmungen (*SPOC1*: Abb. 3.8, *Spoc1*: Abb. 3.10). So sind beide Gene aus 4 Exons aufgebaut, enthalten eine bipartite

Kernlokalisationssequenz (Exon 3) sowie eine PHD-Domäne (Exon 4). Beide Gene kodieren für ein putatives Protein mit einer Größe von 300 AS (*SPOC1*) und 296 AS (*Spoc1*) und einem kalkulierten Molekulargewicht (http://www.expasy.org/tools/peptide-mass.html) von 33,5 kDa (Mensch) bzw. 33,4 kDa (Maus). Die abgeleiteten Aminosäuresequenzen beider Proteine sind zu 92,3% identisch (Abb. 3.13).

SPOC1	1	MDSDSCAAAFHPEEYSPSCKRRRTVEDFNKFCTFVLAYAGYIPYPKEELPLRSSPSPANS
Spoc1	1	MDSDSCAAAFHPEEYSPTCKRRRTVEDFNKFCTFVLAYAGYIPYPKEELPLRSSPSPANS
Konsensus	1	MDSDSCAAAFHPEEYSPSCKRRRTVEDFNKFCTFVLAYAGYIPYPKEELPLRSSPSPANS
SPOC1	61	TAGTIDSDGWD <mark>AGFSDI</mark> AS <mark>SVPLPVSDRCFSHLQPT</mark> LLQRAKPSNFLLDRK <mark>K</mark> TDKLKKKK
Spoc1	61	TAGTIDSDGWDTGFSDITPSVPDRCFSHLQPSLLQRAKPSNYLLDRKTTDKLKKKK
Konsensus	61	TAGTIDSDGWD·GFSDI··SVPlpvsDRCFSHLQPtLLQRAKPSNfLLDRK·TDKLKKKK
SPOC1	121	KRKRRDSDAPG <mark>KEGYR</mark> GG <mark>LLKLEAADPYVETPTSPTLQDIPQA</mark> PS <mark>DPCSGWDSDTPSSGS</mark>
Spoc1	117	RRKRRDSDIPVKEGFRESLLKLEAADPYVETPSSPTMQDIPQASADPCSGWDSDTPSSGS
Konsensus	121	<b>kRKRRDS</b> D·P·KEGyR··LLKLEAADPYVETPtSPTlQDIPQA··DPCSGWDSDTPSSGS
SPOC1	181	CATVSPDQVKEIKTEGKRTIVRQGKQVVFRDEDSTGNDEDIMVDSDDDSWDLVTCFCMKP
Spoc1	177	CATVSPDQVTEIKTEGKRTIVRQGKQVVFRDEDSTGNDEDIMVDSDDDSWDLVTCFCMKP
Konsensus	181	CATVSPDQV·EIKTEGKRTIVRQGKQVVFRDEDSTGNDEDIMVDSDDDSWDLVTCFCMKP
<i>SPOC1</i>	241	FAGRPMIECNECHTWIHLSCAKIRKSNVPEVFVCQKCRDSKFDIRRSNRSRTGSRKLFLD
Spoc1	237	FAGRPMIECNECHTWIHLSCAKIRKSNVPEVFVCQKCRDSKFDIRRSNRSRMGSRKLFLD
Konsensus	241	FAGRPMIECNECHTWIHLSCAKIRKSNVPEVFVCQKCRDSKFDIRRSNRSR·GSRKLFLD

Abb. 3.13: Alignment der abgeleiteten Aminosäuresequenzen des humanen und murinen *SPOC1*-Proteins (Pattern Induced local Multiple Alignment, *PIMA 1.4*, BCM). Die abgeleiteten Aminosäuresequenzen beider Proteine weisen eine Sequenzidentität von 92,3% auf. Identische Aminosäuren sind schwarz schattiert, unterschiedliche Aminosäuren gleicher abgeleiteter Funktionalität grau. Nicht schattierte Aminosäuren besitzen keine Homologie zueinander. Die bipartite Kernlokalisationssequenz (dunkelgrau) ist mit einer gepunkteten Linie markiert, die PHD-Domäne (rot) ist mit einer gestrichelten Linie gekennzeichnet.

Die Aminosäuresequenzen der PHD-Domänen beider Gene zeigen im Sequenzvergleich eine Identität von 100% (Abb. 3.13). Bei der PHD-Domäne (Plant-Homeodomain) handelt es sich um ein kurzes Sequenzmotif von ca. 50-80 Aminosäuren, welches durch eine hochkonservierte Anordnung von bestimmten Aminosäuren (Cys<sub>4</sub>-His-Cys<sub>3</sub>) charakterisiert ist, die in mehr als 400 verschiedenen eukaryotischen Proteinen gefunden wurden (Aasland et al., 1995; Capili et al., 2001). Der Vergleich der Konsensussequenz der PHD-Domäne von *SPOC1* mit den PHD-Domänen anderer humaner Proteine (ClustalW 1.8, BCM Search Launcher) zeigte, dass die in *SPOC1* enthaltene PHD-Domäne denen anderer Proteine entspricht (Abb. 3.14 A). Ein Vergleich mit der Konsensus-Sequenzen der PHD-Domäne (Capili et al., 2001) bestätigte dieses Ergebnis (Abb. 3.14 B).



Abb. 3.14: (A) MAP-Alignment (BCM Search Launcher) der PHD-Domäne von SPOC1 mit den entsprechenden Domänen verschiedener humaner Proteine. Die konservierten Cysteinreste (C) der PHD-Domäne sind durch schwarze Pfeile markiert. Der hochkonservierte Histidinrest (H) ist durch einen roten Pfeil markiert. Identische Aminosäuren sind schwarz schattiert, verschiedene Aminosäuren mit abgeleiteter gleicher Funktion sind grau schattiert. DDBJ/EMBL/GenBank "Accession Numbers" und "PDB codes": AF-17 (P55198), KAP-1 (U78773), AIRE (O43918), CHD3 (Q12873), ATRX (P46100), ING1 (AAG02578), RBB2 (P29375), CAC09389.1 (novel human protein containing a putative PHD finger domain), JUNE-1 (NP\_084340), SHL1 (Q9FEN9), YM42 (Q03214), SPOC1 (AY069975). (B) Konsensus Sequenz der PHD-Domäne (verändert nach Capili et al., 2001). Die konservierten Cysteinreste und der hochkonservierte Histidinrest (H) sind blau umrahmt. Das nicht charakterisierte humane Gen CAC09389.1 und das murine June-1 Gen zeigen große Homologien zu SPOC1.

#### 3.2.6 Promotorsuche im humanen und murinen SPOC1-Gen

Mit Hilfe der Programme "PromotorInspector" und "MatInspector" der Firma Genomatix (http://www.genomatix.de/) ist es möglich, genomische Sequenzen von bis zu 100.000 Bp auf das Vorhandensein verwendeter Promotorsequenzen und Transkriptionsfaktor-Bindestellen zu untersuchen (Quandt et al., 1995; Werner, 2001 und Werner, 2003). Da im Laufe der Arbeit für das humane und murine *SPOC1*-Gen jeweils die genomischen Sequenzen in die

Datenbanken eingegeben wurden (vgl. 3.2.3 und 3.2.4), konnten die Sequenzabschnitte vor dem verifizierten Exon 1 beider Gene einer Promotor-Analyse unterzogen werden. Für das humane *SPOC1*-Gen wurde ein putativer Promotor mit einer Länge von 548 Bp vorhergesagt, der 94 Bp stromaufwärts vor Exon 1 liegt (Abb. 3.15). Zusätzlich wurden innerhalb der vorhergesagten Promotorsequenz 71 Transkriptionsfaktor-Bindestellen angegeben, von denen exemplarisch zwei in Abb 3.15 eingezeichnet sind.

ACTGACAACG CTGGCCCGGT GTTTTTAATG CACGAACCCG GCCACCTCGA GTGAATCCCG CAGATGCCTA GAGAAGAGCC CGGTCCAGTC CCGGCCCCGC GGGCCCCGGAG CAGGGAGCCG CACGAAGCCC GCAGCCGCGA AAACAACTTC GCTGAGCCCG GCCTGGAGCG GGGGCCGGGC GECGACTCCE CCCTCACCEG GEAGCEGEGEA GCGGEGAGCE ATECCCEGEGE CEECCECEAC STCAGCGTCC CCGCCCTGC CGCCCCTACT CCCGGCCAGG CCTCGGGGGAC CCCCGCCCCA CAARGGGGCG CGAACCCCAA GTAGGAGACG AGGAAGGGTA GGTGCCAGCC GCCGCCGCCC TTCGCAACCC GGGCCGTCAG CCCCGCCCCC ACCGCCGCAG CGACTCCCGC CCCCGCGCCC GCCGAGACCA AGGCGCTACC GCTCACACTG CCCCTCCCCC CGCGGCCGAG CCGGCGCGGG CCCCCGCACC TCCCCCGCTC GCGCGCCTC GAGTCCAGAA GACCCGCCTC CACACCCGG GGCCGCCGCC GCGGAGCCTC ATGGGGGTTG GAGTCCCCAA GGTTTCCTTT GTGCGCAGTA TTGGCGGGGC CTTGGACTAC CACTCCCACA AGGCACCGCG CCCGCCTCCC GCGCCACGCC CCCTCGCCGC TGGCTTCCAG TCGCCACCCA GACTACATTT CCCGACAGGC CTCCCGGCTC TCCCGCCCTC CCTCCCGAGA CACGAGCCGA ACTGGGCGTC AGGTCGGGGA GCCGGTCGGG TTCCCGCTCA CCGCCGCCGC CGCCGCCCCC TGCAGCCACT CTCCCGCCTC TACCGCCGCG GGAGCTGCAT CGTCCACTCC GGTCGGCGGT GGAACCGCCA GTCCGGGGTC ACAGAGCTTG AGAAGCGACG CGCTGAGCCC CCCATCACCT CCAGCCCGGG CGACCCCTCC CGGGTCCGCC CTCGCCCTGC GCAGCCGCCC GAGCCCCCAG CCCCGGGCGG CCCCGCTCCA GCATCCCAGC TCCTGCACTC TCGCAGCCGC CGCCGCCCC CGCCCGGAAC ATGGACTCTG ACTCTTGCGC CGCCGCCTTC CACCCGGAGG TGAGTGAAGC TGTGCGTCCG AATCGCCCCC GACCACCCCC

Abb. 3.15: Sequenzausschnitt aus dem humanen *SPOC1*-Gen, der den durch "PromotorInspector" vorhergesagten putativen *SPOC1*-Promotor (rot) mit einer Länge von 548 Bp enthält. Das Exon 1 des *SPOC1*-Gens (grau) liegt 94 Bp stromabwärts hinter der Promotorsequenz und ist durch schwarze Pfeile markiert. Das verifizierte Startcodon von *SPOC1* ist unterstrichen. Genomische Sequenzabschnitte sind blau dargestellt. Von den durch "MatInspector" vorhergesagten 71 Transkriptionsfaktor-Bindestellen sind exemplarisch zwei (E2F, grauer Kasten und SP1, gepunkteter Kasten) dargestellt.

Für das murine *Spoc1*-Gen konnte ebenfalls ein putativer Promotor vorhergesagt werden, der eine Länge von 220 Bp besitzt und 465 Bp stromaufwärts vor Exon 1 liegt (Abb. 3.16). Innerhalb der vorhergesagten Promotorsequenz wurden 25 Transkriptionsfaktor-Bindestellen angegeben, von denen exemplarisch zwei in Abb 3.15 dargestellt sind.

GAAGCCGGCG	CGTGTGTGGC	Gecectecec	TGTCCCGGAG	TTCGGTCCCA	TCTTCCGCTG
TGCGCACCGG	GCGCCCACAA	GTGACCCCGC	CAGGCCCAAC	<b>ecceceetce</b>	CCGGGAACCT
CCAGCTAAAC	eccceceecc	GCGAAAACAA	<b>etecectcce</b>	CCCTCGCGGA	CGTCAGTGGC
CCGCCCCAT	CCTGCTCCAG	CAGCCCGGAC	ACCCCCGCCC	CACAAAGGCC	GAAGCTGAGG
CCCCACGGGA	GAAAGCGGAG	cccgggccgc	CGCCCTCCGA	GACCCAGGCC	GCTGCTCGAC
cccgcccgc	GGAGCCCGCG	GGCTGCCGCC	CCTCCCCAC	CAGCTGGGCG	GTGGGAGCCC
CCGCACCTCC	CCCGCTCTGA	AGCTCCGGAA	TCCTGTCAGC	ACCGCCTCCG	CATTCGTGCG
CTGCCGCAGA	GCCTCTTGGG	AGTTGGAGTT	CTGAGATAGT	CTCTAGACGC	CGGGGGGGCAA
CGCTAGGGAC	TACCACTCCC	AGAAGGCAGC	GCGCCGAGCG	GGCCACACCC	CCACATGACG
AAGTCCAGTC	AACCCGGACT	ACATTTCCCG	ACGAACCTCG	CGACGCTCCC	GCCCTCCCTC
CCGAGACACG	AGCCGAACTG	GGCGTCAGGT	CGGGGAGCCG	GTCGGGTTCC	CGCTCGCTGT
CTCCGCCGCC	GCCCCCACA	GTGCCTCTCC	CGCCTCTCCA	CAGCGCGGGT	CGTGGGAGCT
GCATCCTCCA	CTCAGTTCGC	TGTGGACACT	CCAGTTCGGG	ATCAGCGAGC	CTGCGACGCG
ACATCCCGAC	ACCCCCGTCA	CCTCCAGCCC	GGGCGACCCC	TCCCCGGCCG	GGTCCGCCCT
Cecctetec	GGTGCCGGAG	CCCCCTGGCC	CTGAGCGTCC	CGTGGGCCCG	CTCCCGCACT
ссеесстсее	CACTCCCACA	ecceccecce	CCCCCCACCC	GGAAC <mark>ATG</mark> GA	CTCCGACTCC
Tececcecce	CCTTCCACCC	CGAGGTGAGT	CCGAGCTTCG	CTTCGCCGCC	CCCTCCCCGG

Abb. 3.16: Sequenzausschnitt aus dem murinen *Spoc1*-Gen, der den durch "PromotorInspector" vorhergesagten putativen *Spoc1*-Promotor enthält. Die Promotorsequenz ist in rot dargestellt, genomische Sequenzabschnitte blau. Exon 1 (grau) ist durch Pfeile markiert, das verifizierte Startcodon ist unterstrichen. Von den durch "MatInspector" vorhergesagten 25 Transkriptionsfaktor-Bindestellen sind exemplarisch zwei (E2F, grauer Kasten und SP1, gepunkteter Kasten) eingezeichnet.

# 3.3 Analyse der Expression des humanen und murinen SPOC1-Gens

#### 3.3.1 SAGEmap-Datenbank-Analyse der Expression des humanen SPOC1-Gens

Aufgrund der Expressionsdaten der im UniGene-Cluster Hs.7299 enthaltenen ESTs war ersichtlich, dass das humane *SPOC1*-Gen in vielen verschiedenen humanen Geweben und Tumoren exprimiert wird. Erste, weitergehende Expressionsanalysen wurden ebenfalls *in silico* durchgeführt. Hierzu wurde die "SAGEmap"-Datenbank am NCBI verwendet (http://www.ncbi.nlm.nih.gov/SAGE/) (Lash et al., 2000). Mit Hilfe dieser Datenbank ist es möglich, die Expression eines bestimmten kurzen Sequenzabschnitts (Tag), der ein Gen-Transkript eines UniGene-Clusters repräsentiert, in einem bestimmten Gewebe oder Tumor zu quantifizieren. Enthält der untersuchte UniGene-Cluster ein oder mehrere Tags, kann die Häufigkeit, mit der diese Sequenzabschnitte in einer bestimmten cDNA-Bank vorkommen, bestimmt werden. Die "SAGEmap"-Analyse des korrespondierenden UniGene-Clusters Hs.7299 ergab Hinweise auf eine verstärkte Expression des *SPOC1*-Gens in Ovarialkarzinom-Gewebe (cDNA-Bibliothek: SAGE OC14), einer Ovarialkarzinom-Zelllinie (cDNA-Bibliothek: SAGE A2780-9) sowie in Gewebe der Brustdrüse (cDNA-Bibliothek: SAGE Duke 48N) (Daten nicht gezeigt).

#### 3.3.2 Northern-Analysen des humanen SPOC1-Gens

Mit Hilfe von RT-PCR-Analysen an RNA aus humanen Knorpel- und Patella-Gewebe konnte in Vorversuchen gezeigt werden, dass *SPOC1* in humanen Knorpel-Gewebe exprimiert wird (Daten nicht gezeigt). Zur weitergehenden Untersuchung der Expression und der Ermittlung der Transkriptgrösse des humanen *SPOC1*-Gens wurden Northern-Blot-Analysen mit selbst hergestellten Northern-Blots sowie kommerziell erhältlichen "Multiple Tissue Northern (MTN/E<sup>TM</sup>) Blots/Arrays" durchgeführt. Als Sonden wurden zwei verschiedene RT-PCR-Fragmente (Sonde 1: Primerkombination K 203 F1/K203 R1, Sonde 2: Primerkombination c3 fw/c3 rev) verwendet, die in der kodierenden Region (Sonde 1, 448 Bp) bzw. im 3' untranslatierten Bereich (Sonde 2, 512 Bp) der *SPOC1*-cDNA liegen. Mit beiden Sonden konnte ein gleiches Expressionsmuster nachgewiesen werden. Northern-Analysen von *SPOC1* an Gesamt-RNA zeigten eine starke Expression eines 3,7 kB Transkripts in Gesamt-RNA aus Plazenta und mittlere bzw. schwache Expression des 3,7 kB Transkripts in Gesamt-RNA aus Fibroblasten und einer Chondrosarkom-Zelllinie (Abb. 3.17). Die Größe des *SPOC1*-Transkripts entspricht der Größe der klonierten *SPOC1*-cDNA. Weitere Northern-Analysen mit einem kommerziellen "Multiple Tissue Northern Blot" bestätigten zum einen die Transkriptgröße von 3,7 kB und zeigten zum anderen, dass *SPOC1* in unterschiedlicher Stärke in vielen verschiedenen humanen Geweben und auch während der Entwicklung exprimiert wird (Abb. 3.17, 3.18). Die stärkste Expression von *SPOC1* wurde an RNA aus humanen Testis detektiert (Abb. 3.17, 3.18). Eine erhöhte Expression von humanen *SPOC1* konnte in RNA aus Plazenta, Harnblase, Uterus, Brustdrüse, Speicheldrüse, fötaler Niere und fötalen Herz festgestellt werden (Abb. 3.18).



Abb. 3.17: (A) Die Northern-Blot-Analyse des humanen *SPOC1*-Gens zeigt eine starke Expression des 3,7 kB *SPOC1*-Transkripts an Gesamt-RNA aus humaner Plazenta (3) und eine schwächere Expression an Gesamt RNA aus einer humanen Chondrosarkom-Zelllinie (1) und humanen Fibroblasten (2). Für die Hybridisierung wurde die radioaktiv markierte Sonde c3 fw/rev verwendet. (B) Das Autoradiogramm eines "Human Multiple Tissue Northern-Blots" (Clontech), der mit der radioaktiv markierten *SPOC1*-Sonde c3 fw/rev hybridisiert wurde, zeigt eine starke Expression eines 3,7 kB *SPOC1*-Transkripts in Testis (4) und geringere Expression in Milz (1), Thymus (2), Prostata (3), Ovar (5), Dünndarm (6), Dickdarm (7) und peripheren Blut-Leukozyten (8). Für eine relative Abschätzung der Hybridisierungssignale wurden die Northern-Blots anschließend mit einer *GADPH*-Sonde (A) oder einer  $\beta$ -Actin-Sonde (B) hybridisiert.

	1	2	3	4	5	6	7	8	9	10	11	12
A	whole brain	cerebellum, left	substantia nigra	heart	esophagus	colon, transverse	kidney	lung	liver	leukemia, HL-60	fetal brain	yeast total RNA
В	cerebral cortex	cerebellum, right	accumbens nucleus	aorta	stomach	colon, desending	skeletal muscle	placenta	pancreas	HeLa S3	fetal heart	yeast tRNA
C	frontal lobe	corpus callosum	thalamus	atrium, left	duodenum	rectum	spleen	bladder	adrenal gland	leukemia, K-562	fetal kidney	<i>E. coli</i> rRNA
D	parietal lobe	amygdala	pituitary gland	atrium, right	jejunum		thymus	uterus	thyroid gland	leukemia, MOLT-4	fetal liver	E. coli DNA
E	occipital lobe	caudate nucleus	spinal cord	ventricle, left	ileum		peripheral blood leukocyte	prostate	salivary gland	Burkitt's lymphoma, Raji	fetal spleen	Poly r(A)
F	temporal lobe	hippo- campus		ventricle, right	ilocecum		lymph node	testis	mammary gland	Burkitt's Iymphoma, Daudi	fetal thymus	human C <sub>0</sub> t <sub>-</sub> 1 DNA
G	p. g.* of cerebral cortex	medulla oblongata		inter- ventricular septum	appendix		bone morrow	ovary		colorectal adeno- carcinoma, SW480	fetal lung	human DNA 100 ng
H	pons	putamen		apex of the heart	colon, ascending		trachea			lung carcinoma, A549		human DNA 500 ng
	* paracen	itral gyrus										
	1	2	3	4	5 (	6	78		9	10 1	1	12
A		*		•			• •	•	*		•	
B				•	•	•	• •	•	•		•	
С		*		*	•		• •	•	•	•	•	
D				•	•			•		*	•	
E				*	•	3	• •		•		•	
F					•		•	•	•		•	
G	•				•			•		• •	•	
H												

Abb. 3.18: Das Autoradiogramm eines humanen "Multiple Tissue Expression Array" (Clontech), der mit der radioaktiv markierten *SPOC1*-Sonde c3 fw/rev hybridisiert wurde, zeigt eine ubiquitäre Expression von *SPOC1* in humanen Geweben. Die stärkste Expression von *SPOC1* konnte in RNA aus Testis nachgewiesen werden. Das *SPOC1*-Transkript wird erhöht exprimiert in Plazenta, Harnblase, Uterus, Brustdrüse, Speicheldrüse, fetalen Herz, Niere und Lunge.

#### 3.3.3 Northern-Analysen des murinen Spoc1-Gens

Für die Expressions-Analysen des murinen *Spoc1*-Gens wurde ein selbst hergestellter Northern-Blot mit Gesamt-RNA aus Testis sowie ein kommerziell erhältlicher "Mouse Multiple Tissue Expression Array" (Clontech) verwendet. Als Sonde wurde ein radioaktiv markiertes RT-PCR-Produkt (MM12 fw/MM4b rev) aus dem 3' untranslatierten Bereich der *Spoc1*-cDNA verwendet. Die Hybridisierung des Northern-Blots mit der *Spoc1*-Sonde ergab eine Transkriptgröße von 3,1 kB, die mit Größe der klonierten cDNA von *Spoc1* übereinstimmt (Abb. 3.19 A). Die Hybridisierung des "Mouse Multiple Tissue Expression Array" zeigte die stärkste Expression von *Spoc1* in RNA aus Testis, sowie eine erhöhte Expression in RNA aus Uterus und Prostata, sowie aus den embryonalen Stadien Tag 7 dpc und Tag 11 dpc (Abb. 3.19 B). In allen anderen Geweben konnte eine schwache Expression beobachtet werden (Abb. 3.19 B).



B



Abb. 3.19: (A) Northern-Analyse des murinen *Spoc1*-Gens an Gesamt-RNA aus murinen Testis-Gewebe. Die Hybridisierung mit der radioaktiv markierten *Spoc1*-Sonde (MM12 fw/MM4b rev) zeigt eine starke Expression eines 3,1 kB *Spoc1*-Transkripts. Zur relativen Abschätzung des Hybridisierungssignals wurde der Northern-Blots anschließend mit einer murinen *Gapdh*-Sonde hybridisiert. (B) Autoradiogramm eines "Mouse Multiple Tissue Expression Arrays" (Clontech), der mit der radioaktiv markierten *Spoc1*-Sonde hybridisiert wurde. Starke Expression von *Spoc1* zeigt sich in Testis, Ovar und Prostata sowie in den embryonalen Stadien Tag 7 dpc und Tag 11 dpc. Ein Vergleich der Expressionsdaten beider Gene zeigte, dass humanes und murines *SPOC1* jeweils am stärksten in Testis-Gewebe exprimiert wird. Während die Expression von humanen *SPOC1* in Ovar und Prostata eine normale Stärke besitzt ist sie in Ovar- und Prostata-Gewebe der Maus stark erhöht. Hinweise auf eine erhöhte Expression des *SPOC1*-Transkripts in humanen Ovar-Gewebe ergaben sich lediglich aus in der "SAGEmap"-Datenbank in der in Gewebe aus einem Ovarialkarzinom und aus einer Ovarialkarzinom-Zelllinie eine erhöhte Expression angegeben wurde. Die starke Expression des *SPOC1*-Transkripts in humaner Plazenta konnte sowohl in einem selbst hergestellten Northern-Blot (Abb. 3.17 A) wie auch in einem humanen "Multiple Tissue Expression Array" (Abb. 3.18) nachgewiesen werden. Der Hinweis auf eine verstärkte Expression des *SPOC1*-Transkripts in humaner Brustdrüse ("SAGEmap"-Datenbank) konnte durch den "Multiple Tissue Expression Array" (Abb. 3.18) bestätigt werden.

#### 3.3.4 RNA-in situ-Hybridisierung des murinen Spoc1-Gens

Für weitergehende Untersuchungen der Expression des *Spoc1*-Gens an Geweben mit der stärksten Expression wurden RNA *in situ*-Hybridisierungen an transversalen Kryoschnitten aus adulten Testis- und Uterus-Gewebe aus Balb/C Mäusen durchgeführt. Als Antisense- und Sense-Sonden wurden mit Digoxigenin endmarkierte Oligonukleotide (vgl. 2.7.2 und 2.9) verwendet. Spezifische Hybridisierungssignale konnten in Testis-Schnitten in den Spermatogonien (Abb. 3.20 A, E) und in Uterus-Schnitten in Zellen des Stratum funktionale des Endometriums (Abb. 3.20 B) nachgewiesen werden. Von beiden Zell-Typen ist bekannt, dass sie an der Entwicklung von Tumoren (Keimzell-Tumoren und Tumoren des Endometriums) beteiligt sind. Die Hybridisierung mit den Sense-Sonden ergab keine spezifischen Signale (Abb. 3.20 C, D).



Abb. 3.20: RNA *in situ*-Hybridisierung zur Analyse der *Spoc1*-Gen Expression in Kryoschnitten aus Testis- und Uterus-Gewebe aus adulten Balb/C Mäusen. Als Antisense- und Sense-Sonden wurden mit Digoxigenin endmarkierte Oligonukleotide verwendet. Spezifische Signale von *Spoc1*-Transkripten konnten in den Spermatogonien in Testis (A, E) und in den Zellen des Stratum funktionale in Uterus (B) mit den Antisense-Sonden AS1-3 nachgewiesen werden. Keine Signale konnten mit den Sense-Sonden SS1-3 in Testis (C) und Uterus (D) beobachtet werden. Die Vergrößerung ist 63 X für A, B, C und D; 10 X für E. *Spoc1* exprimierende Spermatogonien (A) und Zellen des Stratum funktionale (B) sind durch Pfeile markiert.

## 3.4 Subzelluläre Lokalisation des SPOC1-Proteins

Die Analyse der intrazellulären Lokalisation des humanen SPOC1-Proteins erfolgte mit Hilfe SPOC1-EGFP-(enhanced fluorescent protein) und FLAG-SPOC1von green Fusionskonstrukten. Hierfür wurde der SPOC1-Leserahmen mit den Primern K203 E fw BamHI und K203 E rev BamHI, die jeweils eine BamHI-Restriktionsschnittstelle enthalten, in PCR-Experimenten amplifiziert. Die PCR-Produkte wurden mit dem Restriktionsenzym BamHI verdaut und das aufgereinigte Produkt in die BamHI-Restriktionsschnittstelle der Vektoren pEGFP-N1 und pFLAG-CMV-4<sup>™</sup> kloniert (vgl. 2.4.2). Die Untersuchung der subzellulären Lokalisation der Fusionsproteine erfolgte durch transiente Transfektion (vgl. 2.12.3) in COS-7-, U2OS-, DeuserPT- und BG1p2-Zellen und Fluoreszenz-Mikroskopie 12-24 Stunden nach der Transfektion (vgl. 2.12.5). Durch Prof. Hengstler (Leipzig) durchgeführte Untersuchungen der Expression von SPOC1 in verschiedenen Zelllinien (Taqman-Analysen) hatten ergaben, dass SPOC1 besonders stark in der Östrogen-Rezeptor positiven humanen Ovarialkarzinom-Zelllinie (BG1p2) und einer Zelllinie aus einem humanen Plattenepithelkarzinom (DeuserPT) exprimiert wird. Als Negativ-Kontrollen wurden transiente Transfektionen mit den Vektoren pEGFP-N1 und pFLAG-CMV-4™ unter gleichen Bedingungen durchgeführt.

#### 3.4.1 Subzelluläre Lokalisation des SPOC1-EGFP-Fusionsproteins

Die Expression des *SPOC1*-EGFP-Fusionsproteins mit einem kalkulierten Molekulargewicht von 61,2 kDa konnte in allen Zelllinien durch Western Blot-Analysen (vgl. 2.11.3 und 2.11.4) mit einem Living colors<sup>®</sup> A.v-Antikörper (Clontech, Heidelberg) bestätigt werden (Ergebnisse für COS-7, DeuserPT und BG1p2 nicht dargestellt). Exemplarisch sind die Ergebnisse der Western Blot-Analysen für U2OS-Zellen dargestellt (Abb. 3.21). Die Untersuchung der Expression des *SPOC1*-EGFP-Fusionskonstrukts in COS-7-, U2OS-, DeuserPT- und BG1p2 Zellen zeigte, dass das Fusionsprotein fast ausschließlich im Zellkern in Form von kleinen über den ganzen Zellkern verteilten Punkten oder "speckles" unterschiedlicher Größe und Anzahl lokalisiert ist (Abb. 3.22). Die Lokalisation des Fusionsproteins deutet auf ein Vorkommen von *SPOC1* in Chromatin assoziierten subnukleären Domänen hin. Das Expressionsmuster von *SPOC1*-EGFP ist in allen untersuchten Zelllinien identisch, daher sind nur die Ergebnisse der transienten Transfektionen in U2OS-Zellen abgebildet. Die Kontroll-Transfektionen mit dem pEGFP-N1 Vektor zeigen eine gleichmäßige über die ganze Zelle verteilte Lokalisation des EGFP-Proteins (Abb. 3.22 G).



Abb. 3.21: Western-Blot-Analyse der Expression des *SPOC1*-EGFP-Fusionsproteins in U2OS-Zellen. Das kalkulierte Molekulargewicht des Fusionsproteins liegt bei 61, 2 kDa (1). Negativkontrolle mit Zell-Lysat von untransfizierten U2OS-Zellen (2).



Abb. 3.22: Subzelluläre Lokalisation des *SPOC1*-EGFP- (A-C) und FLAG-*SPOC1*-Fusionsproteins (D-F) in U2OS-Zellen. Beide Fusionsproteine sind ausschließlich im Zellkern in Form von kleinen Punkten oder "speckles" lokalisiert. DAPI-Färbung des Zellkerns (A, D), *SPOC1*-EGFP (B), *SPOC1*-EGFP und DAPI (C), FLAG-*SPOC1* (E), FLAG-*SPOC1* und DAPI (F). Transiente Transfektionen mit den Vektoren EGFP-N1 und pFLAG-CMV-4 zeigen eine gleichmäßig über die Zelle verteilte Expression von EGFP (G) und keine Expression des FLAG-Konstrukts (H).

#### 3.4.2 Subzelluläre Lokalisation des FLAG-SPOC1 Fusionsproteins

Die Expression des FLAG-*SPOC*1-Fusionsproteins mit einem kalkulierten Molekulargewicht von 36,6 kDa konnte in allen Zelllinien durch Western Blot-Analysen (vgl. 2.11.3 und 2.11.4) mit einem anti-FLAG-Antikörper (Sigma) verifiziert werden und ist exemplarisch für U2OS-Zellen in Abb. 3.23 dargestellt. Die Analyse der Lokalisation des FLAG-*SPOC1*-Fusionsproteins in COS-7-, U2OS-, DeuserPT- und BG1p2-Zellen zeigte, ein über den

ganzen Zellkern verteiltes granuläres, gepunktetes Erscheinungsbild des Fusionsproteins (Abb. 3.22). Um die Lokalisation des FLAG-*SPOC1*-Fusionsproteins in den Zellkernen nachzuweisen, wurden diese mit DAPI angefärbt. Die Lokalisation des FLAG-*SPOC1*-Fusionsproteins im Zellkern bestätigt somit das Erscheinungsbild und die Lokalisation des *SPOC1*-EGFP-Fusionsproteins in den getesteten Zelllinien. Ferner deutet die Lokalisation des Fusionsproteins auf eine Assoziation mit den durch DAPI angefärbten Bereichen des Euchromatins hin. Kontroll-Transfektionen mit dem pFLAG-CMV-4<sup>™</sup> Vektor zeigten keine spezifische Lokalisation (Abb. 3.22 H).



Abb. 3.23: Expression des FLAG-*SPOC1*-Fusionsproteins in U2OS-Zellen. Die Western Blot-Analyse des FLAG-*SPOC1*-Fusionsproteins zeigt ein leicht aberrantes Migrationsverhalten des Fusionsproteins, welches aufgrund von Modifikationen des Proteins vorkommen kann (2). Das kalkulierte Molekulargewicht des FLAG-*SPOC1*-Fusionsproteins liegt bei 36,6 kDa. Negativkontrolle mit Zell-Lysat von untransfizierten U2OS-Zellen (1).

#### 3.4.3 Spezifität des SPOC1 #10 Peptidantikörpers

Um endogen exprimiertes *SPOC1*-Protein nachweisen zu können, wurde ein polyklonaler Peptidantikörper gegen eine Peptidsequenz des *SPOC1*-Proteins hergestellt (vgl. 2.12.4). Die Spezifität des Antikörpers wurde in Western-Blot-Analysen mit Prä-Immunserum aus Kaninchen (Daten nicht gezeigt) und Zell-Lysaten von transient mit FLAG-*SPOC1*- und pM2-*SPOC1*-Konstrukten transfizierten U2OS- und 293T-Zellen verifiziert. Ebenso wurden untransfizierte Zell-Lysate von U2OS-, 293T-, DeuserPT-, und BG1p2-Zellen und in Western-Blot-Analysen untersucht. In Kooperation mit Dr. Jan Hengstler (Mainz) am Institut für Toxikologie durchgeführte semiquantitative RT-PCR Analysen der Expression von *SPOC1*-mRNA hatten gezeigt, dass *SPOC1* in den DeuserPT- und BG1p2-Zelllinien im Vergleich zu weiteren getesteten Tumorzelllinien (MCF7, HeLa, EFO2, EFO21, Lu I, HIH3T3, NIH3T3-HER2) überexprimiert wird. Die Western-Blot-Analysen zeigten, dass der SPOC1-Peptidantikörper überexprimiertes SPOC1-Protein in transient transfizierten U2OSund 293T-Zellen detektieren kann (Daten für 293T-Zellen nicht gezeigt) (Abb. 3.24. und 3.25). Die Untersuchung der nicht transfizierten Zellen ergab, dass endogen exprimiertes SPOC1 in der Western-Analyse nicht durch den SPOC1-Peptidantikörper detektiert wird (Abb. 3.24. und 3.25, jeweils Spur 2). Im Gegensatz dazu zeigte die Auswertung von Immunfluoreszenz-Experimenten, dass der SPOC1-Peptidantikörper hier sowohl endogen exprimiertes, wie auch überexprimiertes SPOC1 in U2OS-, DeuserPT- und BG1p2-Zellen nachweisen kann. Exemplarisch sind die Ergebnisse für U2OS- und DeuserPT-Zellen dargestellt. Das durch den SPOC1-#10 erkannte Expressionsmuster im Zellkern entspricht den mit dem SPOC1-EGFP- und dem FLAG-SPOC1-Konstrukten erhaltenen Ergebnissen (Abb. 3.26). SPOC1 kommt im Zellkern in Form von kleinen, über den ganzen Kern verteilten Punkten vor. Zur Verifizierung der Spezifität des SPOCI-Antikörpers wurden Kolokalisationsexperimente mit transient mit dem FLAG-SPOC1-Konstrukt transfizierten U2OS-, DeuserPT- und BG1p2-Zellen durchgeführt (Ergebnisse für BG1p2 nicht gezeigt) (Abb. 3.27). Beide Signale zeigen eine Kolokalisation (Abb. 3.27 D, H). Die Auswertung der Immunfluoreszenzen zeigte, dass der polyklonale SPOC1-#10-Peptidantikörper das gleiche Protein erkennt, wie der anti-FLAG Antikörper.



Abb. 3.24: Nachweis des FLAG-*SPOC1*-Fusionsproteins durch den polyklonalen SPOC1-#10-Peptidantikörper in U2OS-Zellen (1). Negativkontrolle mit Zell-Lysat von untransfizierten U2OS-Zellen (2). Das kalkulierte Molekulargewicht des FLAG-*SPOC1*-Fusionsproteins liegt bei 36,6 kDa. Die Western Blot-Analyse des FLAG-*SPOC1*-Fusionsproteins zeigt wie in Abb. 3.23 ein leicht aberrantes Migrationsverhalten des Fusionsproteins, welches aufgrund von Modifikationen des Proteins vorkommen kann.



Abb. 3.25: Detektion des pM2-*SPOC1*-Fusionsproteins durch den polyklonalen SPOC1-#10-Peptidantikörper in U2OS-Zellen (1). Das kalkulierte Molekulargewicht des Fusionsproteins liegt bei 51,1 kDa (1). Negativkontrolle mit Zell-Lysat von untransfizierten U2OS-Zellen (2). Unspezifische Banden in den Spuren 1 und 2 sind durch einen Pfeil angedeutet.



Abb. 3.26: Subzelluläre Lokalisation von endogen exprimierten *SPOC1*-Protein. *SPOC1* kommt im Zellkern von untransfizierten U2OS-Zellen in Form von kleinen über den ganzen Zellkern verteilten Punkten ("speckles") vor. DAPI (A), SPOC1-#10 (B), SPOC1-#10 und DAPI (C).



Abb. 3.27: Spezifität des polyklonalen SPOC1-#10-Peptidantikörpers. Das FLAG-SPOC1-Fusionsprotein wird in transient transfizierten Zellen sowohl durch den anti-FLAG Antikörper, wie auch durch den polyklonalen SPOC1-#10-Peptidantikörper in U2OS Zellen (A-D) und DeuserPT-Zellen (E-H) erkannt. DAPI (A, E, I), FLAG-SPOC1 (B, F), SPOC1-#10-Peptidantikörper (C, G). Die Überlagerung der Bilder mit dem anti-FLAG-Antikörper und dem polyklonalen SPOC1-#10-Peptidantikörper zeigt die Kolokalisation der Signale (gelb) in U2OS-Zellen (D) und DeuserPT-Zellen (H). Transiente Transfektionen mit dem pFLAG-CMV4-Konstrukt zeigen kein Signal (I, J). Negativ Kontrollen mit dem SPOC1-Antikörper zeigen ebenfalls kein Signal (Daten nicht gezeigt).

#### 3.4.4 Kolokalisation von SPOC1 mit RNA-Polymerase II und E2F-1

Die Verteilung von *SPOC1* im Zellkern zeigt große Ähnlichkeiten zu den subnukleären Domänen verschiedener Proteine, die mit der Transkription assoziiert sind. Da für PHD-Proteine eine Funktion in der Regulation der Chromatinstruktur angenommen wird, wurden Immunfluoreszenz-Experimente mit dem polyklonalen SPOC1-#10-Peptidantikörper und monoklonalen Antikörpern gegen *RNA-Polymerase II* und den Transkriptionsfaktor *E2F-1* in U2OS- und DeuserPT Zellen durchgeführt (Daten für U2OS-Zellen nicht gezeigt). Die Auswertung der Immunfluoreszenzen zeigte, dass *SPOC1* mit *RNA-Polymerase II* partiell, aber deutlich kolokalisiert (Abb. 3.28). Ein Vergleich der räumlichen Verteilung deutet darauf hin, dass *SPOC1* und *RNA-Polymerase II* teilweise im gleichen subnukleären Kompartiment vorliegen (Abb. 3.28). Immunfluoreszenz-Experimente mit dem Transkriptionsfaktor *E2F-1* ergeben ebenfalls eine partielle, aber deutliche Kolokalisation mit *SPOC1* und bestätigen zumindest teilweise die Assoziation mit transkriptionsaktiven Bereichen (Abb. 3.29). Die Ergebnisse der Kolokalisationsexperimente zeigen somit eine deutliche Assoziation von *SPOC1* zu subnukleären Domänen, die mit der Regulation der Transkription assoziiert sind.



Abb. 3.28: Deutliche partielle Kolokalisation von *RNA-Polymerase II* und *SPOC1* im Zellkern von DeuserPT-Zellen. Der Vergleich der Expressionsmuster zeigt Übereinstimmungen in der räumlichen Verteilung von *SPOC1* und *RNA-Polymerase II*. DAPI (A), *RNA-Polymerase II* (B, E) und *SPOC1* (C, F). Die Überlagerung der Bilder mit dem anti-*RNA-Polymerase II* Antikörper und dem polyklonalen *SPOC1*-Peptidantikörper zeigt die partielle Kolokalisation der Signale in gelb (D). Der durch ein Rechteck markierte Bereich (D) ist 4-fach vergrößert für *RNA-Polymerase II* (E), *SPOC1* (F) und die Überlagerung beider Aufnahmen (G) dargestellt.



Abb. 3.29: Deutliche partielle Kolokalisation von Transkriptionsfaktor *E2F-1* und *SPOC1* im Zellkern von DeuserPT-Zellen. DAPI (A), E2F-1 (B), und SPOC1 (C). Die Überlagerung der Bilder mit dem anti-*E2F-1*-Antikörper und dem polyklonalen *SPOC1*-Peptidantikörper zeigt die signifikante partielle Kolokalisation der Signale in gelb (D). Das Rechteck (D) markiert einen 4-fach vergrößerten Ausschnitt für *E2F-1* (grün), *SPOC1* (rot) und die Überlagerung der Aufnahmen (gelb). Beispiele weiterer Kolokalisationen sind durch Pfeile markiert (D).

#### 3.5 Effekt von SPOC1 auf die Transkription

Um den Einfluss des *SPOC1*-Proteins auf die Transkription zu untersuchen, wurde der Leserahmen des *SPOC1*-Gens C-terminal an die GAL4-DNA bindende Domäne des pM2-Vektors (Clontech, Heidelberg) fusioniert (vgl. 2.13). Hierfür wurde der *SPOC1*-Leserahmen mit den Primern Gal4/PHF5\_BamHI fw und Gal4/PHF5\_BamHI rev, die jeweils eine *BamHI*-Restriktionsschnittstelle enthalten, in PCR-Experimenten amplifiziert. Die PCR-Produkte wurden mit dem Restriktionsenzym *BamHI* geschnitten und das aufgereinigte Produkt in die *BamHI*-Restriktionsschnittstelle des Vektors pM2 kloniert. Die Expression des Gal4-*SPOC1*-Fusionskonstruktes in 293T-Zellen wurde durch Western-Blot-Analysen mit dem Gal4 (DBD)-Antikörper (Santa Cruz) verifiziert (Abb 3.30).



Abb. 3.30: Western-Blot-Analyse der Expression des Gal4-*SPOC1*-Fusionsproteins in 293T-Zellen. Das Fusionsprotein besitzt ein kalkuliertes Molekulargewicht von 51,1 kDa (1). Negativkontrolle mit Zell-Lysat von untransfizierten 293T-Zellen (2).

Gal4-SPOC1-Fusionskonstruktes Durch Kotransfektionen des mit dem pG5Luc-Reporterplasmid, welches Gal4-Bindestellen und ein Luziferase-Reportergen enthält, konnte der Einfluss des SPOCI-Proteins auf die Transkriptionsrate des Luziferase-Gens untersucht werden (vgl. 2.13). Der in 293T-Zellen durchgeführte Gal4-Luziferase-Reporter-Assay ergab, dass steigende Mengen des Gal4-SPOC1-Fusionsproteins eine zunehmende starke Repression der Transkription des Gal4-Luziferase-Reportergens zur Folge haben (Abb. 3.31). So führte die Transfektion des Gal4-SPOC1-Fusionskonstruktes zu einer Repression von 48% (5 ng Repression Gal4-SPOC1-Fusionskonstrukt) Gal4-SPOC1 bis zu 74% (500 ng Fusionskonstrukt) im Verhältnis zur Expression eines reinen Luziferase-Reporterkonstrukts (Kontrolle) (Abb. 3.31). Die Luziferase-Aktivität der Messungen wurde mit der Expression eines kotransfizierten LacZ-Konstrukts normalisiert. Die Ergebnisse der Messungen des Einflusses von *SPOC1* auf die Expression des Luziferase-Gens weisen darauf hin, dass *SPOC1* indirekt oder direkt als ein starker Repressor der Transkription wirkt.



Abb. 3.31: Gal4-Luziferase Reporter Assay des Gal4-SPOC1-Fusionsproteins in 293T-Zellen. Die Expression von 1 μg des Reporterkonstruktes mit der Gal4-DNA bindenden Domäne (Gal4-DBD) als Negativkontrolle zeigt keine Repression und wurde als Referenz gleich 1 gesetzt. Transfektionen mit steigenden Mengen des Gal4-*SPOC1*-Fusionskonstruktes (5, 10, 50, 200, 500 ng) ergeben eine steigende Repression (48, 51, 57, 66, 74%) des Gal4-Luziferase-Reportergens. Die Mittelwerte von 3 unabhängigen Experimenten sind mit Standardabweichung als % Repression im Verhältnis zum Reporterkonstrukt dargestellt. Die Luziferase Aktivität der Messungen wurde mit der Expression eines kotransfizierten LacZ-Konstruktes normalisiert.

#### 3.6 Mutationsanalyse des humanen SPOC1-Gens

Aufgrund der Lokalisation von SPOC1 in der Chromosomenregion 1p36, die mit der Entwicklung verschiedener Tumorarten assoziiert ist und der Expression von SPOC1 in Endometrium-Gewebe, wurde eine Mutationsanalyse des SPOCI-Gens an insgesamt 31 endometrialen Hyperplasien bzw. endometrioiden Adenokarzinomen durchgeführt. Für alle Proben war ein Verlust der Heterozygotie (LOH) in der Chromosomenregion 1p36 nachgewiesen worden (Kiechle et al., 2000). Für das Mutationsscreening wurden die Exons 1, 2, 3 und 4 von SPOC1 aus den Patienten-DNAs amplifiziert, (vgl. 2.5) sequenziert (vgl. 2.3.6) und mit der genomischen Sequenz verglichen. Die Primer (vgl. 2.15) wurden dabei so ausgewählt, dass auch die Sequenzen der Intron/Exon Übergange analysiert werden konnten. Die Mutationsanalyse der 31 Patienten-DNAs ergab 17 Sequenzvariationen, die allerdings ausserhalb der kodierenden Region des SPOC1-Gens liegen (Tab. 3.3) und somit Polymorphismen, vermutlich ohne pathologische Bedeutung darstellen. In der Probe des Patienten 28 konnte ein Nukleotidaustausch (G  $\rightarrow$  A) an Position 354 der kodierenden Region des SPOC1-Gens (Exon 3) gefunden werden, der allerdings nicht zu einem Austausch einer Aminosäure führt (Tab. 3.3). Weiterhin wurde in Patienten-DNA Nr. 30 ein Nukleotidaustausch (A  $\rightarrow$  G) an Position 305 der kodierenden Region des SPOC1-Gens ermittelt (Tab. 3.3). Dieser Nukleotidaustausch führt zum Austausch der Aminosäure Lysin gegen Arginin (K102R) in Exon 3 (Abb. 3.32). Da die Aminosäuren Lysin (K) und Arginin (R) der Gruppe der basischen Aminosäuren angehören und auch strukturell große Ähnlichkeiten aufweisen, handelt es sich bei dem Nukleotidaustausch mit großer Wahrscheinlichkeit ebenfalls um einen Polymorphismus ohne funktionelle Bedeutung, der nicht weiter untersucht wurde. Der Nukleotidaustausch liegt zudem außerhalb der Kernlokalisationssequenz und der PHD-Domäne.

Patienten-DNA	Nukleotidaustausch	Position	Intron/ORF/3'-UTR
2	$C \rightarrow T$	43	3' UTR
5	$C \rightarrow T, C \rightarrow T$	2930, 43	Intron 2, 3'-UTR
7	$C \rightarrow T, C \rightarrow T$	2930, 43	Intron 2, 3'-UTR
8	$C \rightarrow T$	43	3' UTR
9	$C \rightarrow T$	43	3' UTR
10	$C \rightarrow T, C \rightarrow T$	2930, 43	Intron 2, 3'-UTR
13	$C \rightarrow T$	43	3' UTR
14	$C \rightarrow T, C \rightarrow T$	2930, 43	Intron 2, 3'-UTR
19	$C \rightarrow T$	43	3' UTR
20	$C \rightarrow T$	43	3' UTR
21	$C \rightarrow T$	43	3' UTR
22	$C \rightarrow T$	43	3' UTR
28	$G \rightarrow A$	354	ORF (Exon 3)
29	$C \rightarrow T$	43	<b>3' UTR</b>
30	$A \rightarrow G$	305	ORF (Exon 3), K102R

Tab. 3.3: Zusammenfassung der bei der Mutationsanalyse des *SPOC1*-Gens ermittelten Nukleotidaustausche. Mit Ausnahme der Patienten-Proben 28 und 30 liegen alle identifizierten Nukleotidaustausche in Intron 2 oder in der 3'-untranslatierten Region (3'-UTR) des *SPOC1*-Gens. Der in der Patienten-Probe 28 an der Position 354 der kodierenden Region des *SPOC1*-Gens (Exon 3) gelegene Nukleotidaustausch führt zu keinem Austausch einer Aminosäure. Der in der Probe 30 an der Position 305 der kodierenden Region des *SPOC1*-Gens gelegene Nukleotidaustausch (Exon 3) führt zu einem Austausch der Aminosäure Lysin (K) gegen Arginin (R) an der Position 102 der *SPOC1*-Aminosäuresequenz (s. Abb. 3.32).

#### **K102R**



Abb. 3.32: Nukleotidaustausch (A  $\rightarrow$  G) an der Position 305 des *SPOC1*-Leserahmens, der zu einem Austausch der Aminosäure Lysin (K) gegen Arginin (R) an der Position 102 der *SPOC1*-Aminosäuresequenz (K102R) führt.

# 4. Diskussion

# 4.1 Analyse und Homologievergleich von EST-Klonen aus einer humanen Chondrozyten-cDNA-Bibliothek

Ziel des Projekts war es, am Beispiel der Wachstumsfuge Gene und Gen-Netzwerke zu identifizieren und zu charakterisieren, die an Differenzierungsprozessen beteiligt und medizinisch relevant sind. Es sollten hierbei neben allgemein Differenzierungs-relevanten Genen auch Gene identifiziert werden, die an den Prozessen der Skelettentwicklung beteiligt sind. Eine wesentliche Voraussetzung für die Analysen war die umfassende Auswertung der generierten Daten. In der vorliegenden Arbeit wurden die Sequenzen von 1050 EST-Klonen einer humanen fetalen Chondrozyten-cDNA-Bibliothek mit Hilfe von Datenbank-Analysen auf Homologien zu bereits bekannten Nukleotidsequenzen untersucht. Insgesamt wurden von allen Projektpartnern im gesamten EST-Projekt 4748 EST-Klone bioinformatisch ausgewertet. Im folgenden werden die Gesamt-Daten des Projekts ausgewertet und besprochen, da so eine bessere Interpretation der Ergebnisse möglich ist. Entsprechend der Funktion der zugrundeliegenden Genprodukte wurden die EST-Sequenzen in verschiedenen Kategorien (I-III) eingeordnet und tabellarisch aufgelistet (Tab. 4.1).

Mit Hilfe des Datenbank-Analyse-Programms "EST-Sweep" (http://genome.dkfzheidelberg.de/), welches in Gemeinschaftsarbeit der Arbeitsgruppen Prof. Winterpacht, Prof. Zabel und der Firma GENterprise mit der "Bioinformatic Service Group" von HUSAR (Heidelberg Unix Sequence Analyses Resources) im DKFZ entwickelt wurde, kann eine große Anzahl an EST-Sequenzen simultan in verschiedenen BlastN-Nukleotid-Datenbanken analysiert werden. Die Verwendung von ESTsweep zur Analyse von EST-Sequenzen bietet den Vorteil, dass nicht jede Sequenz einzeln bearbeitet werden muss, sondern 60 bis 100 Sequenzen als "Block" automatisiert nacheinander analysiert werden. Dabei ist es möglich, alle Sequenzen eines Blocks in bis zu 6 verschiedenen BlastN-Datenbanken, wie z. B. "RefSeq\_dna" (mRNA Datenbank) oder "Nrnuc" ("nicht-redundante" Datenbank) gleichzeitig zu untersuchen. Die Verwendung von ESTsweep bietet somit die Möglichkeit alle relevanten Datenbanken, in denen ein EST-Klon einem Homologievergleich unterzogen werden soll, mit einer einzelnen Gesamtanalyse zu bearbeiten, ohne das weitere Analysen in Nukleotid-Datenbanken notwendig sind. Gleichzeitig können neben der Bearbeitung einer abgeschlossenen Datenbank-Analyse weitere ESTsweep-Analysen durchgeführt werden und so eine zeitlich dicht gestaffelte Charakterisierung der EST-Klone erlauben. Bei herkömmlichen BlastN-Datenbank-Analysen (http://www.ncbi.nlm.nih.gov/blast) ist es nicht möglich, eine Sequenz in mehreren Datenbanken gleichzeitig zu untersuchen, so dass für jede neu ausgewählte Datenbank eine neue Analyse gestartet werden muss. Die Ergebnisse der ESTsweep-Datenbank-Analysen wurden in einer Ergebnis-Tabelle zusammengefasst und gespeichert, in der alle relevanten Ergebnisse und Alignments für jeden EST-Klon über Mausklick als "Output File" jederzeit abrufbar sind (Abb. 3.4). Aufgrund der automatisierten Durchführung und Speicherung der Daten können zeitaufwendige Analysen eines Sequenz-Blocks ohne Überwachung durch den Benutzer über Nacht durchgeführt werden. Durch die Verwendung des Datenbank-Analyse-Programms ESTsweep bei der Durchführung eines EST-Projekts wird die bioinformatische Einordnung, Charakterisierung und Verwaltung der sequenzierten EST-Klone bedeutend und effizient vereinfacht. Günstig ist, dass die Anzahl der für jeden EST-Klon notwendigen verschiedenen Datenbank-Analysen durch eine Gesamt-Analyse in den relevanten Nukleotid-Datenbanken ersetzt wird. Die bioinformatische Untersuchung der Klone ermöglicht die effiziente Einordnung der ESTs in verschiedene Kategorien, aus denen Kandidatengene für genauere Untersuchungen ausgewählt werden können.

#### 4.1.1 ESTs mit Homologien zu Genen bekannter Funktion

Von den 4748 analysierten EST-Klonen konnten 3426 (72,2%) in der Kategorie I zusammengefasst werden, da sie so große Identität zu bereits bekannten, charakterisierten cDNA- oder Protein-Sequenzen zeigten, dass von einer Identität zu diesen Sequenzen ausgegangen werden musste. Die grobe Einteilung dieser Klone in 4 verschiedene Untereinheiten erfolgte entsprechend ihrer Funktion (Tab. 4.1). Bei 3,5% der untersuchten EST-Klone handelte es sich um Kontaminationen der cDNA-Bibliothek mit Vektor- oder Bakteriensequenzen.

Einteilung der EST-Klone der Kategorie I	Anzahl der EST-Klone (%)
Verschiedene charakterisierte Gene und Proteine	1303 (38,1%)
Ribosomale Proteine	887 (25,9%)
Haushaltsgene	886 (25,8%)
Transkripte/Proteine aus Knorpel/Knochen	350 (10,2%)

Tab. 4.1: Einteilung der EST-Klone, die Homologien zu bekannten oder funktionell charakterisiertencDNA- oder Protein-Sequenzen zeigen.

Die Auswertung der EST-Sequenzen aus der Kategorie I zeigt, dass neben einer großen Anzahl an verschiedenen bekannten oder funktionell charakterisierten cDNA-Sequenzen (38,1%) die Hälfte der sequenzierten EST-Klone (51,7%) auf verschiedene ribosomale Gene und Haushaltsgene entfällt. Dieses Ergebnis stimmt mit den Erwartungen überein, da sowohl ribosomale Proteine wie auch Proteine von Haushaltsgenen in der Zelle in einer großen Anzahl vertreten und dadurch in einer cDNA-Bibliothek im Verhältnis zu anderen Genen häufiger nachzuweisen sind. Weiterhin liegt der Anteil bekannter Gen-Sequenzen in der Größenordnung vergleichbarer EST-Projekte (Okihana et al., 1999; Kumar et al., 2001; Zhang et al., 2003). Der Nachweis von 10,2% typischer Knorpel- und Knochenspezifischer Transkripte, wie z. B. Kollagen Typ II, Osteonektin (SPARC), Fibronektin, oder "Cartilage oligomeric matrix protein" (COMP) bestätigt das spezifische Expressionsmuster von Knorpelgewebe, welches sich in der Chondrozyten-cDNA-Bibliothek wiederspiegelt. Beispielsweise macht Kollagen Typ II (COL2A1) mehr als 50% der extrazellulären Knorpelmatrix aus und wird speziell von proliferierenden Chondrozyten in der Wachstumsfuge exprimiert (Mundlos, 1994). Mittlerweile sind verschiedene humane Erkrankungen beschrieben, die durch Mutationen im Kollagen Typ II-Gen ausgelöst werden (Spranger et al., 1994). Die Expression von Osteonektin, welches in Knorpel an der Regulation der Mineralisierung beteiligt ist, konnte in verschiedenen Zonen der Wachstumsfuge von Knochen, aber auch in weiteren Geweben (fetale Haut, Gefäße, Hoden Ovarien, Nebenniere) nachgewiesen werden (Mundlos., 1994). Beim COMP-Protein handelt es sich um einen weiteren Bestandteil der Knorpelmatrix. Mutationen innerhalb des COMP-Gens konnten als Ursache verschiedener dysproportionierter Kleinwuchssyndrome, wie der autosomal dominant vererbten Pseudoachondroplasie und multiplen epiphysären Dysplasie (Hecht et al., 1995; Briggs et al., 1995) nachgewiesen werden. Ferner konnten auch zahlreiche Transkriptionsfaktoren, wie z. B. des LMX1B-Gen, welches zu der Gruppe der LIM- Homöobox-Transkriptionsfaktoren gehört, identifiziert werden. Mutationen im *LMX1B*-Gen konnten als Ursache für das Nagel-Patella-Syndrom identifiziert werden (Dreyer et al., 1998).

# 4.1.2 ESTs mit Homologien zu Sequenzen unbekannter Funktion oder genomischen Sequenzen

In den Kategorien II und III sind 1157 EST-Klone zusammengefasst, die keine Homologien bzw. Homologien zu genomischen Sequenzen (III), oder Homologien zu nicht charakterisierten cDNA-, EST- oder Protein-Sequenzen der nr- oder der EST-Datenbanken (II) zeigen (Bsp. Tab. 3.1). Über die funktionelle Bedeutung der in dieser Gruppe enthaltenen EST-Klone kann somit keine Aussage gemacht werden. Da die Kategorien II und III uncharakterisierte, unbekannte Sequenzen und somit potentiell neue Gene enthalten, wurden aus dieser Gruppe EST-Klone zu weiteren Untersuchungen ausgewählt. Bezogen auf die Gesamtzahl der analysierten EST-Klone besitzt die aus den Kategorien II und III zusammengesetzte Gruppe einen relativ großen Anteil von 24,3% (Abb. 4.1). Bei den 490 (10,3%) EST-Klonen der Kategorie II, die Homologien zu cDNAs und Proteinen unbekannter Funktion besitzen, kann es sich um putative, medizinisch relevante Kandidatengene handeln, die abhängig von der Expression in Knorpelgewebe auch Kandidatengene für die Skelettentwicklung sein können. Bei den 667 (14%) EST-Klonen der Kategorie III, die keine Homologien, oder nur Homologien zu uncharakterisierten ESTs und genomischen Sequenzen zeigen, kann es sich zum einen um sehr seltene Transkripte handeln, so dass diese nicht durch ESTs oder cDNAs in den öffentlichen Datenbanken vertreten sind. In diesem Fall sind auch diese Klone als putative Kandidatengene interessant. Zum anderen ist es aber möglich, dass es sich bei einigen dieser Klone um Kontaminationen der cDNA-Bibliothek mit genomischer DNA oder um Sequenz-Artefakte (z. B. repetitive Sequenzen) handelt.

Eine genauere Untersuchung einer kleinen Anzahl von Klonen der Kategorie III zeigte, dass es sich bei EST-Klonen mit Homologien zu uncharakterisierten EST-Sequenzen auch um Abschnitte der 3'-untranslatierten Region von bekannten Genen handeln kann (vgl. 3.1.2.2). Dieses Beispiel zeigt, dass eine genaue Analyse der einzelnen EST-Klone nach der automatisierten bioinformatischen Vorauswahl von großer Wichtigkeit ist, um interessante Kandidatengene zu identifizieren und unwichtige EST-Klone herauszufiltern.



Abb. 4.1: Vergleich der prozentualen Anteile der ESTs der Kategorie I (blau) und der zu einer Gruppe zusammengefassten ESTs der Kategorien II und III (rot) mit der Gesamtzahl der untersuchten ESTs. Der prozentuale Anteil der Kontaminationen der EST-Sequenzen ist in gelb dargestellt. Kategorie I enthält bereits charakterisierte ESTs. Die Kategorien II und III beinhalten EST-Klone, die keine Homologien bzw. Homologien zu genomischen Sequenzen oder Homologien zu nicht charakterisierten cDNA- oder EST-Sequenzen zeigen.

## 4.2 Vergleich des EST-Projekts mit anderen EST-Projekten

Aufgrund der problematischen Verfügbarkeit und der komplexen biochemischen Eigenschaften von humanen Knorpelgewebe gibt es bisher erst wenige Projekte, die sich mit der Analyse dieses Gewebes im großen Maßstab befassen. Bis zum Zeitpunkt dieser Arbeit sind erst zwei Vergleichbare EST-Projekte beschrieben, die sich mit der Identifizierung und Charakterisierung von ESTs aus cDNA-Bibliotheken aus humanen fetalen Knorpel (Zhang et al., 2003) und humanen normalen und osteoarthritischen Knorpel (Kumar et al., 2001) befassen. Ein weiteres cDNA-Projekt beschäftigt sich mit der Analyse einer murinen Knorpel cDNA-Bibliothek (Okihana et al., 1999). Die im vorliegenden Projekt verwendete fetale Knorpel-cDNA-Bibliothek unterscheidet sich von den beschriebenen Knorpel-cDNA-Bibliotheken, weil sie aus Knorpelgewebe der Wachstumsfuge hergestellt wurde, um Differenzierungsprozesse zu untersuchen. Der Vergleich der wenigen bisher publizierten Knorpel-cDNA Projekte zeigt, dass die Anzahl der nicht charakterisierten Sequenzen, die ermittelt werden können, stark variiert. Der Anteil der uncharakterisierten Sequenzen liegt dabei in einem Bereich zwischen 0,4% (Zhang et al., 2003) und 40% der untersuchten Klone (Kumar et al., 2001). Die bioinformatische Analyse der humanen fötalen Chondrozyten

92

24,3% an uncharakterisierten EST-Sequenzen. Durch eine weitergehende Analyse der in den Kategorien II und III zusammengefassten EST-Klone wird es, wie bereits in Abschnitt 4.1.3 besprochen, zu einer Verminderung der Anzahl der in dieser Gruppe enthaltenen EST-Klone kommen. Untersuchungen der EST-Klone der Kategorie III haben gezeigt, dass es sich bei der Mehrzahl der ESTs mit Homologien zu genomischen Sequenzen um Kontaminationen der cDNA-Bibliothek mit genomischen und repetitiven Sequenzen handelt, die auch in anderen cDNA-Projekten beobachtet wurde. Der Anteil von DNA-Kontaminationen und repetitiven Sequenzen liegt mit einem Wert von 3,5% (Abb. 4.1) in der Größenordnung vergleichbarer EST-Projekte (Adams et al., 1995; Zhang et al., 2003). Der Anteil der uncharakterisierten ESTs und cDNAs in der untersuchten cDNA-Bibliothek liegt mit 24,3% in der Größenordnung von vergleichbaren EST-Projekten (Mao et al., 1998; Sasaki et al., 1998; Kumar et al., 2001; Zhang et al., 2003). Der Vergleich der ESTs mit Homologien zu bereits charakterisierten Sequenzen ermöglicht auch, das Expressionsprofil bekannter Gene in der Wachstumsfuge zu bestimmen. So konnte durch Kumar et al. (2001) eine große Anzahl von bekannten Genen identifiziert werden, von denen bis zum Zeitpunkt der Analyse keine Expression in Knorpel beschrieben worden war. Weiterhin konnten neben uncharakterisierten Sequenzen auch unbekannte Homologe von bekannten Sequenzen identifiziert werden (Kumar et al., 2001; Zhang et al., 2003). Die Ergebnisse des vorliegenden EST-Projekts und der verschiedenen beschriebenen EST-Projekte unterstreichen, dass es trotz einer sehr großen Anzahl von ESTs in den Datenbanken sinnvoll ist, spezifische cDNA-Bibliotheken zu untersuchen, um differentiell exprimierte und neue Gene zu identifizieren.

# 4.3 Präliminäre Untersuchung einiger ausgewählter ESTs

Nach der bioinformatischen Auswertung der 1050 EST-Klone wurden aufgrund der Expressionsdaten in den öffentlichen Datenbanken und der chromosomalen Lokalisation vier EST-Klone (T01 203; T01 517, T01 519 und T01 564) ausgewählt, die keine Homologien zu bereits bekannten Genen zeigten. Weitere Untersuchungen der Klone T01 517, T01 519 und T01 564 zeigten, dass es sich bei diesen Klonen um repetitive DNA-Sequenzen bzw. um bekannte Gene handelt (vgl. 3.1.3.2 bis 3.1.3.4). Diese Klone wurden nicht weiter bearbeitet. Die Auswertung der Sequenzdaten des Klons T01 203 (vgl. 3.1.3.1) führte zur Identifizierung

eines neuen humanen Gens, welches aufgrund der chromosomalen Lokalisation in 1p36, der Expression in Ovarialkarzinomen (öffentliche Datenbanken) sowie der enthaltenen PHD-Domäne, die mit der Transkriptionsregulation in Verbindung gebracht wird, detailliert untersucht wurde. Die Analyse und Charakterisierung des neuen Gens ist in Abschnitt 4.4 beschrieben.

#### 4.4 Detaillierte Analyse und Charakterisierung des SPOC1-Gens

Die weitergehende bioinformatische Analyse der Sequenz des EST-Klons T01 203 in der vorliegenden Arbeit führte zur Identifizierung und Klonierung eines neuen humanen (SPOCI) und murinen (Spoc1) Gens. Aufgrund der im SPOC1 Protein enthaltenen PHD-Domäne (plant homeo domain) und einer Assoziation mit der Überlebenszeit von Patienten mit Ovarialkarzinom (vgl. 4.4.5) wurde das Gen SPOC1 (survival time -associated PHD protein in ovarian cancer) genannt. Das humane SPOC1-Gen besitzt eine Größe von ca. 10,9 kB, das korrespondierende murine Gen eine Größe von ca. 7,1 kB. Der Vergleich des humanen und murinen SPOC1-Gens zeigt große Übereinstimmungen, sowohl auf Nukleotidebene, wie auch in der Genstruktur. Beide Gene sind aus 4 Exons aufgebaut und enthalten eine PHD-Domäne und eine bipartite Kernlokalisationssequenz. Das humane und das murine Gen besitzen im Bereich des offenen Leserahmens eine Sequenzidentität von 87,3%. Ein Vergleich der abgeleiteten Aminosäuresequenzen zeigt eine Aminosäure-Identität von 92,3% (Abb. 3.13) liegt die hochkonserviert sind. Im Bereich der PHD-Domäne Identität der Aminosäuresequenzen beider Gene sogar bei 100% (Abb. 3.13). Das murine SPOC1-Gen ist auf Maus-Chromosom 4 bei 79 bis 80,1 cM lokalisiert ist. Bei dem murinen Gen handelt es sich somit um das Ortholog zu SPOCI (Abb. 3.11 und 3.12). Interessanterweise konnten zwischen SPOC1 und einem uncharakterisierten PHD-Protein (334 AS) aus Maus (June1) partielle signifikante Homologien im N-terminalen Bereich (66% Identität in einem Bereich von 64 AS) und in der PHD-Domäne (85%) identifiziert werden. Zusätzlich konnte ein hypothetisches humanes PHD-Protein (MGC 2941, 265 AS) mit sehr großen Homologien (93,8% Identität auf AS-Ebene) zum murinen June1-Gen identifiziert werden. Unter Umständen handelt es sich bei diesen Genen um funktionell mit humanen und murinen SPOC1 verwandte Proteine.

Die Untersuchung der Expression des *SPOC1*-Gens mit Hilfe von RT-PCR Analysen an RNAs aus verschiedenen humanen Geweben zeigten, dass es sich bei *SPOC1* um ein ubiquitär exprimiertes Gen handelt, welches auch in Knorpel und Patella vorkommt. Die genauere Untersuchung des UniGene EST-Clusters des *SPOC1*-Gens bestätigt die RT-PCR Experimente, weil der Cluster (Hs.7299) neben einer großen Anzahl von ESTs aus verschiedenen cDNA-Bibliotheken auch ESTs aus Knorpelgewebe enthält. Das Vorkommen von ESTs aus osteoarthritischen Knorpel sowie aus Knorpeltumoren (Chondrosarkom) deutet darauf hin, dass *SPOC1* auch an den komplexen Prozessen der Skelettentwicklung und Homöostase beteiligt sein könnte. Besonders auffällig ist, dass nahezu die Hälfte aller ESTs des *SPOC1*-UniGene-Clusters aus cDNA-Bibliotheken stammt, die aus Tumorgewebe oder Tumorzelllinien hergestellt wurden. Dies weißt darauf hin, dass es sich bei *SPOC1* um ein neues Gen handelt, dessen Expression mit der Entstehung und Progression von Tumoren in verschiedenen Geweben assoziiert sein könnte. Es sollte jedoch berücksichtigt werden, dass cDNA-Bibliotheken aus Tumorgeweben in den öffentlichen Datenbanken überrepräsentiert sind.

Neben der Expression in einer cDNA-Bibliothek aus Ovarialkarzinom-Gewebe (UniGene) zeigt sich auch in der SAGEmap-Datenbank (vgl. 3.3.1) eine signifikante Verstärkung der Expression von *SPOC1* in zwei SAGE cDNA-Bibliotheken aus Ovarialkarzinom-Gewebe und aus einer Ovarialkarzinom-Zelllinie (vgl. 3.3.1). Aufgrund der Ergebnisse der SAGEmap Analysen wurde die Expression von *SPOC1* von Prof. Jan Hengstler (Leipzig) in verschiedenen humanen Ovarialkarzinomen und Zelllinien quantitativ mittels der "real time"-PCR untersucht. Die Ergebnisse dieser Untersuchungen werden in Abschnitt 4.4.5 besprochen.

Northern-Analysen mit kommerziellen "Multiple Tissue Northern Blots/Arrays" und selbst hergestellten Northern-Blots bestätigten die ubiquitäre Expression von *SPOC1* in Maus und Mensch, sowie in humanen Knorpel (vgl. 3.3.2). Vergleicht man die Northern-Analysen miteinander, so ist offensichtlich, dass *SPOC1* am stärksten in murinen und humanen Testis-Gewebe exprimiert wird. Dagegen liegt das Niveau der Expression von humanen *SPOC1* in Ovar und Prostata im Durchschnittsbereich, während das Expressionsniveau von *SPOC1* in Ovar- und Prostata-Gewebe der Maus erhöht ist. In Gewebe aus Maus-Uterus ist dagegen eine schwache Expression zu beobachten. Da zum Zeitpunkt dieser Arbeit kein Gewebe aus humanen Ovarialkarzinomen verfügbar war, konnte die Expression von *SPOC1* in diesen

Geweben nicht per RNA-*in situ*-Hybridisierung mit den Expressionsniveaus von Ovar-Geweben in Maus und Mensch verglichen werden. Eine starke Expression von *SPOC1* in humaner Plazenta-RNA konnte im selbst hergestellten Northern-Blot, wie auch im humanen "Multiple Tissue Expression Array" verifiziert werden. Das erhöhte Expressionsniveau von *SPOC1* in der Brustdrüse, welches in der SAGEmap cDNA-Bibliothek nachgewiesen wurde, konnte ebenfalls im humanen "Multiple Tissue Expression Array" bestätigt werden, weil auch dort die Expression über dem Durchschnitt liegt. Da das *SPOC1*-Gen sowohl in fetalen Herz, der Niere, und der Lunge von Mensch wie auch in Gewebe von Maus Embryonen der Tage 7 dpc und 11 dpc deutlich verstärkt exprimiert wird, scheint es auch während der Entwicklung eine wichtige Funktion zu besitzen (s. Abb. 3.17 und 3.18).

Die in den kommerziellen und selbst hergestellten Northern-Blots nachgewiesene Größe des humanen (3,7 kB) und murinen (3,1 kB) Transkripts des SPOC1 Gens stimmt mit den abgeleiteten cDNA Konsensus-Sequenzen überein und deutet darauf hin, dass die komplette cDNA kloniert wurde (vgl. 3.2.1). Das Ergebnis zeigt gleichzeitig, dass die 3 putativen 5'-Exons (vgl. 3.2.1) und das in ihnen enthaltene 1. Startcodon vermutlich nicht genutzt werden, sondern das 2. Startcodon, welches mit dem Konsensus-Sequenz für eukaryotische Translationsstartpunkte (Kozak, 1996) übereinstimmt. Die Verwendung des putativen Startcodons 1 würde den Leserahmen um 411 Bp (putative Exons a, b und c) verlängern und eine weitere bipartite Kernlokalisationssequenz zu der Proteinsequenz hinzufügen (vgl. 3.2.1). Die Addition der zusätzlichen Exons a-c würde dabei nicht zu einer Änderung des SPOCI-Leserahmens führen. Promotor Analysen mit dem Programm PromotorInspector deuten aber ebenfalls auf eine Nutzung des 2. Startcodons hin, da für dieses Startcodon ein Promotor in einer zu erwartenden Entfernung von 94 Bp vor dem ersten Exon von SPOC1 vorhergesagt wurde. Der für das 1. Startcodon vorhergesagte Promotor liegt dagegen in einer relativ großen Entfernung von 8748 Bp vom putativen Exon a (vgl. 3.2.1). Grundsätzlich ist jedoch nicht auszuschließen, dass die Transkriptvariante, in der das erste Startcodon genutzt wird, hochgradig Gewebe- oder Zelltyp-spezifisch und daher schwer nachzuweisen ist. Zur Bestätigung der Ergebnisse des humanen SPOC1-Gens wurden Promotoranalysen mit dem Programm PromotorInspector auch an der Maus durchgeführt und führten zur Vorhersage eines Promoters in einer Entfernung von 465 Bp vor Exon 1, welches auch das einzige mögliche Startcodon enthält. Dieses entspricht im Vergleich dem 2. Startcodon des humanen SPOC1-Gens. Es ist somit sehr wahrscheinlich, dass im humanen und murinen SPOC1-Gen das 2. Startcodon genutzt wird, welches in Exon 1 liegt, da in beiden Genen ein Promotor in

geringer Distanz vor diesem Exon vorhergesagt wurde und die putativen Exons a-c in der Maus nicht konserviert sind.

Eine weitergehende Analyse der Expression von *SPOC1* mit Hilfe von RNA-*in situ*-Hybridisierungen wurde in den Geweben durchgeführt, die in Northern-Analysen eine starke Expression zeigen. Die Expression von *SPOC1* konnte hier in den Spermatogonien in adulten murinen Testis-Gewebe wie auch in den Zellen des Endometriums in adulten Maus Uterus-Gewebe nachgewiesen werden (vgl. 3.3.4). Von beiden Zelltypen ist bekannt, dass sie an der Entwicklung von Keimzell-Tumoren und Tumoren des Endometriums beteiligt sein können (Oliver, 1996; Kiechle et al., 2000). In diesem Zusammenhang ist bemerkenswert, dass erst kürzlich ein Protein, *NYD-SP6* (Xiao et al., 2002), publiziert wurde, das wie *SPOC1* eine PHD-Domäne und eine Kernlokalisationssequenz aufweist und ebenfalls stark in Testis exprimiert wird. Allerdings wird *NYD-SP6* nicht wie *SPOC1* in den Spermatogonien und RT-PCR Experimenten nehmen die Autoren an, dass *NYD-SP6* als putativer Aktivator der Transkription an der Testis-Entwicklung beteiligt sein könnte.

Die Ergebnisse der Northern-Analysen und der RNA-*in situ*-Hybridisierungen bestätigen das in den Datenbanken UniGene und SAGEmap ermittelte Expressionsmuster des *SPOC1*-Gens und weisen unabhängig voneinander auf eine Assoziation von SPOC1 mit Tumorbildung und Entwicklung hin.

#### 4.4.2 Subzelluläre Lokalisation des SPOC1-Gens

Die Untersuchung der subzellulären Lokalisation des *SPOC1*-Proteins mit den transient transfizierten *SPOC1*-EGFP und FLAG-*SPOC1* Konstrukten zeigt übereinstimmend eine exklusive Lokalisation beider Fusionsproteine im Zellkern. Das *SPOC1* Protein kommt dabei in allen getesteten Zelllinien (COS-7, U2OS, DeuserPT und BG1p2) im gesamten Zellkern in Form von kleinen, homogen über den ganzen Zellkern verteilten Punkten ("dots", "speckles") unterschiedlicher Anzahl und Größe vor (vgl. 3.4.1 und 3.4.2). Vergleiche mit dem Immunfluoreszenz-Muster eines polyklonalen *SPOC1* Peptidantikörpers ergaben ein übereinstimmendes Bild (vgl. 3.4.3). Zur Verifizierung der Spezifität des *SPOC1* Peptidantikörpers wurden Kolokalisationsexperimente durchgeführt. Immunfluoreszenzen mit U2OS- und DeuserPT-Zellen, die transient mit dem FLAG-*SPOC1* Konstrukt transfiziert wurden, bestätigen, dass das durch den anti-FLAG-Antikörper nachgewiesene FLAG-*SPOC1* Fusionsprotein auch durch den polyklonalen *SPOC1*-Peptidantikörper erkannt wird. Die

verschiedenen getesteten Zelllinien zu erkennen (vgl. 3.4.3).

Signale beider Antikörper zeigen in den transient transfizierten Zelllinien eine exakte Kolokalisation (vgl. 3.4.3). Neben dem überexprimierten Protein ist der polyklonale *SPOC1*-Peptidantikörper auch in der Lage, endogen exprimiertes *SPOC1*-Protein in den

Die nukleäre Lokalisation des SPOC1-Proteins wird zum einen durch das Vorkommen einer zweigeteilten Kernlokalisationssequenz (NLS) und zum anderen durch die PHD-Domäne (PHD-Finger) in der SPOC1-Aminosäuresequenz unterstützt. Bisherige Untersuchungen verschiedener PHD-Finger-Proteine haben gezeigt, dass es sich hierbei fast ausschließlich um Chromatin-assoziierte Proteine handelt, die an der Chromatin-vermittelten Regulation der Transkription beteiligt sind (Aasland et al., 1995; Bochar et al., 2000; Loewith et al., 2000; Capili et al., 2001; Xia et al., 2003). Dementsprechend weist das Verteilungsmuster von SPOC1 im Zellkern große Ähnlichkeiten zu den gepunkteten Expressionsmustern verschiedener transkriptionsassoziierter Proteine und Transkriptionsfaktoren auf. Es wird angenommen, das es sich bei den gepunkteten Strukturen um Proteinkomplexe handelt, die in den transkriptionsaktiven Domänen des Chromatins lokalisiert sind. Bei diesen subnukleären Domänen handelt es sich um Kompartimente des Zellkerns, die der Aufbewahrung verschiedener Transkriptionsassoziierter Proteine dienen und aus denen bei Bedarf Proteine an Orte der Transkription rekrutiert werden können. Das Verteilungsmuster dieser Domänen spiegelt dabei die transkriptionelle Aktivität der Zelle wieder (Spector et al., 1993; Rinderle et al., 1999). Zur genaueren Bestimmung der subnukleären Domänen, in denen SPOC1 vorkommt, wurden Kolokalisationen mit Antikörpern gegen verschiedene nukleäre Proteine und dem SPOC1-Peptidantikörper durchgeführt. Die Immunfluoreszenz-Experimente mit dem SPOC1-Peptidantikörper und einem monoklonalen Antikörper gegen RNA-Polymerase II zeigen, dass beide Proteine partiell Kolokalisieren. Dies deutet, wie bereits aufgrund der SPOC1-Aminosäuresequenz vermutet, auf ein Vorkommen von SPOC1 in nukleären Domänen hin, die mit Transkription oder der Regulation der Transkription assoziiert sind. Dieses Ergebnis unterstreicht die Möglichkeit der Beteiligung von SPOC1 an der Regulation der Transkription, da für RNA-Polymerase II und die basalen Transkriptionsfaktoren TFIIH und TFIIF gezeigt werden konnte, dass sie in den gleichen subnukleären Domänen vorliegen, die als Interchromatin Kompartiment bezeichnet werden (Grande et al., 1997; Verschure et al., 2002). Unter dem Interchromatin-Kompartiment versteht man transkriptionell aktive Bereiche des Chromatins, die Nahe der Oberfläche der kondensierten Chromatin-Domänen lokalisiert sind. Transkriptionell aktive Gene lokalisieren an der Oberfläche der kondensierten chromosomalen Domänen und stehen mit dem Interchromatin-Kompartiment in direkten
Kontakt. Proteine, die an der Transkription beteiligt sind und im Interchromatin-Kompartiment vorliegen, könnten somit direkten Zugang zu den regulatorischen Sequenzen des transkriptionsaktiven Chromatins erhalten (Krause et al., 1994; Grande et al., 1997; Verschure et al., 2002). RNA, die an der Oberfläche der Chromatin-Kompartimente synthetisiert wird, könnte so direkt in die Interchromatin-Domäne abgegeben werden, in der auch sämtliche Faktoren für Prozessierung, Verpackung und Transport der RNA vorliegen (Politz et al., 1999; Verschure et al., 2002). Immunfluoreszenz-Experimente mit dem SPOC1 Peptidantikörper und einem Antikörper gegen den Transkriptionsfaktor E2F-1, der an der Regulation der Zellproliferation beteiligt ist (Helin et al., 1998), ergaben ebenfalls eine partielle, aber deutliche Kolokalisation (vgl. 3.4.4), wobei Untersuchungen durch Grande et al. (1997) zeigten, dass E2F-1 ebenfalls im Interchromatin-Kompartiment des Zellkerns lokalisiert ist. Dies bestätigt die Assoziation von SPOC1 zu den transkriptionell aktiven Domänen des Zellkerns. Immunfluoreszenz-Experimente mit PML (promyelocytic leukemia protein) und HIPK2 (homeodomain-interacting protein kinase-2) zeigten, dass SPOC1 nicht mit diesen Proteinen kolokalisiert. Übereinstimmend mit den Ergebnissen für SPOC1 haben Untersuchungen von Komponenten des transkriptionell aktiven Chromatins gezeigt, dass an der Transkription beteiligte Proteine nicht zwingend mit den Zonen aktiver Transkription oder RNA-Polymerase II kolokalisieren müssen, aber in der gleichen Domäne vorliegen (Verschure et al., 1999). Es konnte in diesem Zusammenhang nachgewiesen werden, dass die spezifischen Transkriptionsfaktoren BRG1, Oct1, E2F-1 und GR weder Teil eines aktiven noch Teil eines inaktiven Transkriptionskomplexes sind, der RNA-Polymerase II enthält (Grande et al., 1997). Die Funktion der verschiedenen subnukleären Domänen, die unterschiedliche Transkriptionsfaktoren, aber keine RNA-Polymerase II, enthalten, ist unklar (Verschure et al., 1999). Es wird vermutet, dass es sich um unvollständige Transkriptions-Initiationskomplexe oder inhibitorische Komplexe handelt (Grande et al., 1997). Eine weitere Möglichkeit ist, dass es sich bei den mit Transkriptionsfaktoren angereicherten subnukleären Domänen um Aufbewahrungsorte für spezifische Transkriptionsfaktoren handelt, von denen bei Bedarf Proteine rekrutiert werden könnten (Grande et al., 1997). Diese Vermutung würde erklären, warum die verschiedenen Transkriptionsfaktoren trotz gleichen Expressionsmusters nur eine partielle oder keine Kolokalisationen untereinander zeigen, obwohl nachgewiesen ist, dass sie mit den Domänen des transkriptionell aktiven Chromatins assoziiert sind. Unterstützt wird diese Annahme durch Untersuchungen von Van Steensel et al. (1996), die zeigten, dass in Hippokampus-Neuronen aus Ratte subnukleäre Domänen, die den Glukokortikoid-Rezeptor-Transkriptionsfaktor (GR) oder Mineralokortikoid-Rezeptor-Transkriptionsfaktor

(MR) enthalten, partiell kolokalisieren. Es wird vermutet, dass beide Transkriptionsfaktoren bei der Regulation der Genexpression zusammenwirken. Ein ähnlicher Mechanismus ist auch für die Interaktionen anderer Transkriptionsfaktoren denkbar. Auch für Proteine, die an der Deaktivierung von aktiven Genen beteiligt sind, konnte gezeigt werden, dass sie untereinander und mit konstitutivem Heterochromatin kolokalisieren (Brown et al., 1997; Wreggett et al., 1994; Hendrich et al., 1998). Von einigen bereits beschriebenen subnukleären Domänen, wie z. B. "interchromatin granule clusters" (Mintz et al., 1999; Mintz und Spector, 2000; Lamond und Spector, 2003), "Cajal bodies" (Gall et al., 2000), "nuclear bodies" (Zhong<sup>1</sup> et al., 2000) und "cleavage bodies" (De Jong et al., 1996), ist bekannt, dass sie einen geringen oder keinen DNA-Anteil besitzen und eine spezifische Fraktion nukleärer Proteine, die an der Transkriptionsregulation, der Synthese oder der Prozessierung von RNA beteiligt sind, enthalten (Verschure et al., 2002). Die subzelluläre Lokalisation von SPOC1 zeigt große Übereinstimmungen mit der Lokalisation der genannten Transkriptionsfaktoren. Weiterhin offenbart SPOC1 ähnlich wie die Transkriptionsfaktoren BRG1, Oct1, E2F-1 und GR eine partielle Assoziation zu der Lokalisation von RNA-Polymerase II im Zellkern. Die Vermutung, dass es sich bei SPOC1 um ein Chromatin assoziiertes Protein handelt, welches an der Regulation der Transkription beteiligt ist, wird durch Analogien im Expressionsmuster von SPOC1 und anderen Transkriptionsfaktoren, sowie durch die deutliche Kolokalisation mit dem Transkriptionsfaktor E2F-1 unterstützt (Grande et al., 1997).

### 4.4.3 Funktion der PHD-Domäne bei der Regulation der Transkription

Die Analyse der SPOC1-Aminosäuresequenz ergab u. a. eine PHD-Domäne. Bei der PHD-Domäne (plant homeodomain oder PHD-Finger), auch Leukämie assoziierter Protein-Domäne (LAP) genannt, handelt es sich um ein kurzes Sequenzmotiv von ca. 50-80 Aminosäuren, welches zuerst im Homeobox Protein *HAT3.1* aus *Arabidopsis thaliana* beschrieben wurde (Schindler et al., 1993). Die PHD-Domäne kommt in mehr als 600 verschiedenen eukaryotischen Proteinen vor (Aasland et al., 1995; Capili et al., 2001; Kwan et al., 2003) und ist durch eine hochkonservierte Cys<sub>4</sub>-His-Cys<sub>3</sub> Konsensus-Sequenz (vgl. 3.2.5 und Abb. 4.2) charakterisiert (Aasland et al., 1995; Capili et al., 2001). Die 8 hochkonservierten Zink-Liganden der PHD-Domäne binden 2 Zink Atome, indem jeweils 2 sequentiell aufeinander folgende Paare von Zink-Liganden 1 Zink Atom binden (Pascual et al., 2000; Capili et al., 2001) (Abb. 4.2). Jedes Paar von Zink-Liganden ist dabei durch einen Sequenzabschnitt (L1, L2 und L3) voneinander getrennt, wobei L1 und L3 nur einen geringen Grad der

Konservierung zeigen (Abb. 4.2). Bei der Mehrheit der bisher beschriebenen PHD-Finger Proteine handelt es sich um mögliche Transkriptionsfaktoren oder an der Reorganisation der Chromatinstruktur beteiligte Proteine, die mit chromosomalen Proteinen und Proteinkomplexen assoziiert sind (Aasland et al., 1995; Jacobson and Pillus, 1999; Bochar et al., 2000; Loewith et al., 2000; Linder et al., 2000; Aaravind et al., 2003). Die genaue Funktion der PHD-Domäne ist nicht vollständig aufgeklärt. Aufgrund von verschiedenen Untersuchungen der Struktur und Funktion wird jedoch davon ausgegangen, dass es sich bei der PHD-Domäne um eine Protein-Protein Interaktionsdomäne handelt (Aasland et al 1995; Linder et al., 2000; Loewith et al., 2000; Capili et al., 2001; Kwan et al., 2003). Neuere Untersuchungen bestätigen diese Annahmen. Analysen des humanen MLL-Proteins ("myeloid/lymphoid or mixed-lineage leukemia protein") zeigen beispielsweise, dass beide im Protein enthaltenen PHD-Domänen die Bildung von MLL-Homodimeren, wie auch die Bindung an ein nukleäres Cyclophilin (Cyp33) vermitteln können (Fair et al., 2001). Auch für die PHD-Domäne des humanen AF-10-Proteins konnte eine Homo-Oligomerisation nachgewiesen werden (Linder et al., 2000). Weiterhin konnte durch Schultz et al. (2001) bewiesen werden, dass die PHD-Domäne des KAP-1 Repressors eine Interaktion mit dem Mi- $2\alpha$  Protein vermittelt. Da es sich bei der ATPase *Mi*- $2\alpha$  um eine Komponente des NuRD Histon-Deacetylase-Komplex handelt (Wade et al., 1998; Brehm et al., 1999; Vignali et al., 2000), ist es möglich, dass KAP-1 eine wichtige Funktion bei der Erhaltung der Struktur des Repressor-Komplexes besitzt. Es ist nicht unwahrscheinlich, dass der durch die PHD-Domäne vermittelte Kontakt zwischen verschiedenen Proteinen den Aufbau und die Aufrechterhaltung von Proteinkomplexen ermöglicht, die an der Reorganisation der Chromatin-Struktur beteiligt sind (Kwan et al., 2003). Homologievergleiche der Proteine Yng1, Yng2 und Pho23 aus S. cerevisiae zeigen im Bereich der PHD-Domäne signifikante Homologien untereinander und zu der PHD-Domäne des humanen ING1-Proteins ("inhibitor of growth family, member 1"). Funktionelle Untersuchungen in S. cerevisiae haben ergeben, dass Yng1, Yng2 und Pho23 mit der Aktivität von Histon-Acetyl-Transferasen (HAT) assoziiert sind. Beispielsweise konnte für Yng2 gezeigt werden, dass es mit Tra1, einer Komponente der SAGA-, SLIK- und NuA4-HAT-Komplexe interagiert (Grant et al., 1997 und 1998; Loewith et al., 2000). Interessanterweise wird die spezifische Interaktion von Yng2 mit Tral und dem HAT-Komplex nicht durch die PHD-Domäne von Yng2 vermittelt, sondern durch die N-terminalen Domänen des Proteins. Es wird daher angenommen, dass die Funktion der PHD-Domäne in der Assoziation oder Erhaltung der Aktivität des gesamten HAT-Komplexes liegt (Loewith et al., 2000). Für das humane ING1-Protein wird eine entsprechende Funktion angenommen, weil humanes *ING1*-Protein die Funktion von *Yng2* in *Yng2Δ*-Zellen komplementieren kann. Zusätzlich bestehen große strukturelle Übereinstimmungen zwischen dem humanen *ING1*-Protein und den homologen Proteinen aus *S. cerevisiae* und *S. pombe* (Loewith et al., 2000). Aufgrund der Ergebnisse verschiedener Untersuchungen wird ein Modell postuliert, in dem PHD-Finger-Proteine wichtige Funktionen bei der Strukturerhaltung und beim Aufbau verschiedener Multiprotein-Komplexe (wie z. B. HAT- oder HDAC-Komplexe) besitzen, welche die Chromatin-Struktur modulieren und so die transkriptionelle Aktivität regulieren (Aasland et al., 1995; Loewith et al., 2000). In diesem Zusammenhang ergaben Untersuchungen des MOZ-Genproduktes ("monocytic leukemia zinc finger protein"), dass die Acetyl-Transferase-Aktivität durch die MOZ-MYST-Domäne und nicht durch die PHD-Domäne vermittelt wird (Champagne et al., 2001; Deguchi et al., 2003).

Die für das *SPOC1*-Protein ermittelten Ergebnisse lassen sich gut mit dem Modell der putativen Funktion von PHD-Proteinen vereinbaren. So besitzt die *SPOC1*-PHD-Domäne große Ähnlichkeit zu PHD-Domänen von verschiedenen, an der Regulation der Transkription beteiligten Proteinen, wie z. B. AIRE oder ATRX (vgl. 3.2.5). Weiterhin korrespondiert die subzelluläre Lokalisation des *SPOC1*-Proteins in transkriptionsaktiven Bereichen des Chromatins (vgl. 4.4.2) mit der für diese Proteine vermuteten Funktion in der Regulation der Chromatinstruktur. Übereinstimmend damit zeigen Untersuchungen des Effektes des *SPOC1*-Proteins auf die Transkription, dass die Expression eines Gal4DBD-*SPOC1* Fusionsproteins eine stark reprimierende Wirkung auf die Expression eines Luziferase-Reportergens besitzt (vgl. 3.5). Aufgrund der Ergebnisse kann vermutet werden, dass *SPOC1* an der Assemblierung von Multiproteinkomplexen beteiligt sein könnte, die durch Reorganisation der Chromatinstruktur eine Repression der Transkription vermitteln können.

Die große Anzahl der verschiedenen beschriebenen PHD-Domänen deutet auf unterschiedliche Funktionen in der Zelle hin. Die Variabilität innerhalb der Domäne lässt die Vermutung zu, dass die Protein-Protein Interaktionen, die durch die PHD-Domäne vermittelt werden, sehr spezifisch sind (Kwan et al., 2003). Alignments der PHD-Domänen verschiedener Proteine (vgl. 3.2.5) unterstreichen die geringe Konservierung in den Bereichen der Sequenzabschnitte L1 und L3 (Abb. 4.2). Die große Variationsbreite in diesen Bereichen ist unter Umständen für die spezifischen Bindungseigenschaften der jeweiligen PHD-Domäne verantwortlich (Pascual et al., 2000; Capili et al., 2001). Aufgrund der großen Identität wurde vermutet, dass es sich bei PHD-Domäne, wie bei der RING-Domäne, um eine E3-Ubiquitin-Ligase-Domäne handelt (Lu et al., 2002; Coscoy und Ganem, 2003). Aktuelle Untersuchungen durch Aravind et al. (2003) zeigten jedoch, dass es sich bei diesen PHD-

Domänen nur um spezielle Varianten der RING-Domäne handelte. Im Vergleich zeigen beide Domänen sowohl strukturelle wie auch funktionelle Unterschiede (Capili et al., 2001; Aravind et al., 2003). Ein wichtiges Unterscheidungsmerkmal ist beispielsweise das Fehlen eines in der PHD-Domäne hochkonservierten hydrophoben Aminosäurerests direkt N-terminal vor den konservierten Cysteinresten 7 und 8 in der RING-Domäne (Aravind et al., 2003). Die Analyse der *SPOC1*-PHD-Domäne zeigte, dass diese mit der Konsensus-Sequenz übereinstimmt. Aus diesem Grund handelt es sich bei *SPOC1* mit großer Wahrscheinlichkeit nicht um eine Ubiquitin-Ligase, sondern um ein Protein, welches eine Funktion in der Chromatinregulation besitzt.





Abb. 4.2: (A) Die Konsensus Sequenz der PHD-Domäne zeigt die charakteristische und hochkonservierte Cys<sub>4</sub>-His-Cys<sub>3</sub> Anordnung der Zink-Liganden. Die Bindung von 2 Zink Atomen wird durch die 8 hochkonservierten Zink-Liganden der PHD-Domäne vermittelt. Jeweils 2 sequentiell aufeinander folgende Paare von Zink-Liganden (1-2, 3-4 (blau) und 5-6, 7-8 (orange) binden dabei 1 Zink-Atom. (B) Die Paare der Zink-Liganden sind durch die 3 Sequenzabschnitte L1-3 (rote Kästen) voneinander getrennt. Die Variabilität der Sequenzabschnitte L1 und L3 könnte für die spezifischen Bindungseigenschaften der PHD-Domänen verschiedener Proteine verantwortlich sein.

### 4.4.4 Bedeutung der PHD-Domäne bei der Entstehung von Krankheiten

Die bedeutende Funktion von PHD-Domänen in der Zelle wird durch die Beteiligung von Proteinen mit solchen Domänen an der Pathogenese verschiedener humaner Erkrankungen bekräftigt. So resultieren klinisch relevante "missense"-Mutationen der PHD-Domäne des *ATRX*- (X-linked mental retardation with  $\alpha$ -thalassemia) Proteins in der Entstehung von  $\alpha$ -

Thalassemie und mentaler Retardierung (Gibbons et al., 2000; Rinderle et al., 1999). Aufgrund der im ATRX-Gen enthaltenen Domänen und der Lokalisation im Zellkern wird vermutet, dass ATRX durch Modulation der Chromatinstruktur an der Regulation der Transkription beteiligt ist (Berube et al., 2000). Mutationen der PHD-Domäne des Transkriptionsfaktors AIRE sind ursächlich für das APECED-Syndrom ("autoimmune polyendocrinopathy candidiasis ectodermal dystrophy syndrome type 1") aus (The Finnish-German APECED Consortium, 1997). Dabei führt die Mehrzahl der beschriebenen Mutationen zu einem Funktionsverlust des AIRE-Proteins durch Deletion einer der beiden im Protein enthaltenen PHD-Domänen (Rinderle et al., 1999). Weiterhin konnten Gene, die PHD-Finger-Proteine kodieren, in den kritischen Regionen von verschiedenen Mikrodeletions-Syndromen ("contiguous gene deletion syndromes"), wie z. B. Wolf-Hirschhorn Syndrom (WHS), Williams Syndrom (WSTF) und Immunschwäche Syndrom (ICF) identifiziert werden (Stec et al., 1998; Lu et al., 1998; Aapola et al., 2000). Aufgrund der Funktion von PHD-Proteinen in der Reorganisation der Chromatinstruktur ist es sehr wahrscheinlich, dass diese Proteine auch bei der Entstehung von Tumoren eine wichtige Funktion besitzen. Unterstützt wird diese Annahme dadurch, dass die PHD-Domäne in Genen enthalten ist, die bereits mit Tumorentstehung in Verbindung gebracht werden konnten. So können chromosomale Umlagerungen, die zu einer Deletion der PHD-Domäne in den Proteinen MLL, CBP ("CREB binding protein"), MOZ und AF-10 führen, zur Entwicklung von akuten myeloiden Leukämien führen (Jacobson und Pillus, 1999; Deguchi et al., 2003). Ferner konnten Lu et al. (1999) zeigen, dass das PLU-1-Gen, welches 3 PHD-Domänen enthält, in vielen Brustkrebsfällen überdurchschnittlich stark exprimiert wird. Weiterhin wurde für die Gene GASC1 ("gene amplified in squamous cell carcinoma 1") und NSD3 ("new SET domain-containing gene") eine Amplifikation in Tumoren des Kopf- und Hals-Bereichs (Yang et al., 2000) sowie in einer Brustkrebs-Zelllinie (Angrand et al., 2001) nachgewiesen. Ebenso konnte eine Beteiligung des humanen Pygo-(Pygopus)-Proteins bei der Entstehung von kolorektalen Karzinomen belegt werden (Thompson et al., 2002). Außerdem wurden "missense"-Mutationen im ING1-Gen, wie auch ein Verlust der Heterozygotie (LOH) der chromosomalen Region, in der das ING1-Gen liegt, in Tumoren des Kopf- und Hals-Bereichs beschrieben (Gunduz et al., 2000). Neben der Assoziation der PHD-Domäne mit der Entstehung von Tumoren weist auch die chromosomale Lokalisation des SPOC1-Gens auf

eine mögliche Beteiligung an der Entstehung und Progression von Tumoren hinweisen. Der Chromosomenbereich 1p36, in dem das *SPOC1*-Gen lokalisiert ist, wird mit zahlreichen Tumorerkrankungen wie z. B. Ovarialkarzinom, Prostatakarzinom, Neuroblastom, Karzinomen des Endometriums, Tumoren des Knorpel- und Knochensystems, Keimzelltumoren und myeloiden Leukämien in Verbindung gebracht. Verschiedene Studien haben gezeigt, dass es abhängig von der Art des Tumors sowohl zu einem Verlust als auch einem Zugewinn an genetischen Material in der Chromosomenregion 1p36 kommen kann (Ragnarsson et al., 1999; Gibbs et al., 1999; Shimizu et al., 2000; Arnold<sup>1</sup> et al., 2001; Mora et al, 2002; Mandahl et al., 2002; Zielenska et al., 2001; Tri et al., 2002). Es wird daher angenommen, dass der Bereich 1p36 mehrere Gene enthält, die als mögliche Tumorsuppressorgene oder Onkogene wirken können.

Aufgrund der Proteinstruktur, der Lokalisation in der chromosomalen Region 1p36 und der Expression in einer großen Anzahl von Tumorgeweben und Tumorzelllinien ist eine Funktion von SPOC1 als Onkogen oder Tumorsuppressorgen nicht unwahrscheinlich. Eine aus diesem Grund durchgeführte Mutationsanalyse der 4 Exons des SPOC1 Gens in 31 Proben aus endometrialen Hyperplasien und endometrioiden Adenokarzinomen mit Verlust der Heterozygotie (LOH) in der Chromosomenregion 1p36 (vgl. 3.6) führte allerdings nicht zur Identifikation einer krankheitsauslösenden Mutation, welche auf eine Funktion von SPOC1 als Tumorsuppressor hingewiesen hätte. Dieses Ergebnis schließt aber eine Beteiligung des SPOC1-Gens an der Entstehung verschiedener Tumoren nicht aus. Es ist möglich, dass Mutationen im SPOCI-Gen nicht identifiziert wurden, oder das Mutationen im Promotorbereich von SPOC1 vorliegen. Möglicherweise handelt es sich bei SPOC1 auch um ein Onkogen. Aufgrund der putativen Funktion von SPOC1 in der Reorganisation der Chromatinstruktur ist es möglich, dass Störungen in der Regulation oder Funktion von SPOC1 schwerwiegende Folgen für die Chromatinstruktur und die Transkription haben und so die Entstehung von malignen Entartungen ermöglichen oder unterstützen. Ein Beispiel für eine solche Fehlregulation ist das MOZ-TIF2-Fusionsprotein. Während das MOZ-wt-Protein an Nukleosomen bindet und die Acetylierung von Histonen durch die HAT-Aktivität der MYST-Domäne vermittelt, rekrutiert das MOZ-TIF2 Fusionsprotein CBP an das Nukleosom. Dies führt zu einer aberranten Acetylierung von Histonen. Unter Umständen wird auch die Regulation von assoziierten Transkriptionsfaktoren durch die aberrante Acetylierung gestört. Die gestörte Regulation der Chromatinstruktur führt zu einer aberranten Transkription und führt zur Entstehung einer akuten myeloiden Leukämie (Deguchi et al., 2003).

#### 4.4.5 Expression von SPOC1 in Ovarialkarzinomen

Aufgrund der in öffentlichen Datenbanken ermittelten Expression von *SPOC1* in einer cDNA-Bibliothek aus humanen Ovarialkarzinom und einer Ovarialkarzinom-Zelllinie (UniGene, SAGEmap) wurde die Expression von *SPOC1* mit Taqman-Analysen in verschiedenen humanen Zelllinien und Ovarialkarzinomen von Prof. Hengstler (Leipzig) untersucht.

Die statistischen Auswertungen der Daten ergaben, dass die Expression der *SPOC1*-mRNA mit der Überlebenszeit von Patienten mit primären und rezidivierenden epithelialen Ovarialkarzinomen assoziiert ist. Es konnte dabei gezeigt werden, dass Patienten mit primären Ovarialkarzinomen, die eine schwache Expression von *SPOC1*-mRNA zeigen (n=63), eine längere Überlebenszeit besitzen, als Patienten die eine erhöhte Expression von *SPOC1* zeigen (N=21). Die mediane Überlebenszeit für Patienten mit einer schwachen *SPOC1* Expression liegt bei 1596 Tagen, die mediane Überlebenszeit von Patienten mit rezidivierenden Ovarialkarzinom zeigte den gleichen Einfluss von *SPOC1* auf die Überlebenszeit. So lag die mediane Überlebenszeit der Patienten mit schwacher *SPOC1*-mRNA Expression (n=14) bei 56 Tagen. In Patienten, die eine erhöhte *SPOC1*-mRNA Expression zeigten (n=5), lag die mediane Überlebenszeit bei 10 Tagen.

Zusammenfassend konnte gezeigt werden, dass eine erhöhte *SPOC1*-mRNA Expression mit einer schlechten Prognose für Patienten mit primären und rezidivierenden Ovarialkarzinomen assoziiert ist. Die Überlebenszeit von Patienten mit einer erhöhten *SPOC1* Expression ist erheblich geringer, als die Überlebenszeit von Patienten mit einer schwachen *SPOC1*-Expression.

Aufgrund der Ergebnisse der Untersuchungen vermuten wir, dass SPOC1 als ein prognostischer Marker für das Ovarialkarzinom verwendbar sein könnte. Weiterhin ist es in diesem Zusammenhang möglich, dass SPOC1 wegen seiner Überexpression im Tumorgewebe als Tumormarker Einsatz finden könnte. Wegen der Assoziation von SPOC1 mit der Prognose von Ovarialkarzinomen, der chromosomalen Lokalisation in der Region 1p36, der Kolokalisation mit dem Proto-Onkogen *E2F-1* und der Assoziation der PHD-Domäne mit der Tumorentstehung vermuten wir, dass SPOC1 auch an der Entwicklung, Progression und Therapieresistenz von Tumoren in Knorpel, Knochen, Gehirn, Endometrium, Keimzellen, Zervix, Schilddrüse, Lunge, Prostata, Darm und verschiedenen Leukämieformen beteiligt ist, die in die Chromosomenregion 1p36, in der auch SPOC1 lokalisiert ist, kartiert werden (http://www.ncbi.nlm.nih.gov/mapview/maps.cgi). Unter Umständen kann SPOC1

auch für diese Tumorarten als prognostischer Marker eingesetzt werden und eine Zielstruktur für die Therapie darstellen.

#### 4.4.6 Modell der Funktion von SPOC1 in der Zelle

Die Organisation von Nukleosomen, die durch post-translationale Modifikationen von Histonen, wie z. B. Methylierung, Phosphorylierung und Acetylierung reguliert wird, ist von zentraler Bedeutung für die Genexpression. Beispielsweise ist die Hyperacetylierung von Histonen eines Nukleosoms durch Histon-Acetyl-Transferasen (HAT) mit einer gesteigerten Transkriptionsaktivität assoziiert, während eine Hypoacetylierung von Histonen durch Histon-Deacetylasen (HDAC) zu einer Repression der Transkription führt (Wade et al., 2001; Forsberg et al., 2001).

Weiterhin konnte gezeigt werden, dass die Promotor-Regionen von transkriptionsinaktiven Genen stark methyliert sind und so beispielsweise den Zugang für Transkriptionsfaktoren versperren. Im Gegensatz dazu sind die Promotor-Regionen von transkriptionell aktiven Genen nicht methyliert und ermöglichen so die Bindung von Transkriptionsfaktoren (Ferreira et al., 2001; Jones und Baylin, 2002).

Es wird angenommen, das PHD-Finger-Proteine eine Funktion bei der Assemblierung und Aktivität von Multiproteinkomplexen haben, welche die Transkription durch Modulation der Chromatinstruktur regulieren (vgl. 4.4.3). Funktionelle Untersuchungen des *SPOC1*-Proteins deuten darauf hin, dass *SPOC1* eine reprimierende Wirkung auf die Genexpression hat (vgl. 3.5). Die Repression von transkriptionsaktiven Bereichen durch *SPOC1* könnte dabei durch die Rekrutierung von HDAC- oder Methyltransferase-Aktivität vermittelt werden (Abb. 4.3). Weiterhin ist es aufgrund der PHD-Domäne denkbar, dass *SPOC1* eine wichtige Funktion bei der Assemblierung von Multiprotein-Komplexen besitzt, welche die Transkription reprimieren (Abb. 4.3).

Zieht man in Betracht, dass die Expression von *SPOC1* mit der Überlebenszeit von Patienten mit Ovarialkarzinom assoziiert ist, könnte man vermuten, dass *SPOC1* die Repression von regulatorischen Proteinen wie z. B. des Retinoblastom-Tumor-Suppressor-Gens (*RB*) vermittelt, die für die Aufrechterhaltung des post-mitotischen Status (G<sub>0</sub>) ausdifferenzierter Zellen verantwortlich sind. Das *RB*-Protein besitzt beispielsweise eine wichtige Funktion in der Repression der *E2F*-Transkriptionsfaktoren, welche die Expression einer Reihe von "checkpoint" Genen für den Übergang der Zelle von der G<sub>0</sub>- in die G<sub>1</sub>/S-Phase regulieren (Lipinski et al., 1999; Ferreira et al., 2001). Die aberrante Repression eines solchen Regulators durch *SPOC1* könnte zur Folge haben, dass ausdifferenzierte Zellen aus der G<sub>0</sub>-Phase wieder in die G<sub>1</sub>/S-Phase eintreten und es zur Bildung von Tumoren kommt (Abb. 4.3). In Übereinstimmung mit dieser Vermutung konnte gezeigt werden, dass ausdifferenzierte  $(RB^{-/-})$  Muskelzellen nicht in der Lage sind, den post-mitotischen Status aufrecht zu erhalten, wenn sie mit hohen Serumkonzentrationen stimuliert werden (Lipinski et al., 1999). Ferner konnte gezeigt werden, dass der Verlust von *RB* zu einer aberranten Proliferation der Zelle führen kann (Nahle et al., 2002). Jedoch kann die durch den Verlust von *RB* verursachte verstärkte Expression von *E2F* durch *p53* abhängige und unabhängige Mechanismen auch zur Apoptose führen (Nahle et al., 2002). Da eine aberrante Expression von *SPOC1* andere Mechanismen betreffen kann, steht die durch *E2F* ausgelöste Apoptose nicht im Wiederspruch zu der angenommenen Funktion von *SPOC1*. Weiterhin kann davon ausgegangen werden, dass Störungen in der Funktion von SPOC1 allein nicht ausreichend sind, um eine aberrante Proliferation der Zelle auszulösen, sondern auch andere Regulationswege der Zellproliferation gestört sein müssen.

In situ-Hybridisierungen an murinen Testis-Schnitten hatten gezeigt, dass *SPOC1* in den Spermatogonien exprimiert wird (vgl. 3.3.4). Somit könnte *SPOC1* eine wichtige Funktion in der Embryonalentwicklung oder in stark proliferierenden Zellen, wie z. B. den Spermatogonien besitzen, in denen die Aufrechterhaltung eines embryonalen Zustandes der Zellen von großer Wichtigkeit ist.



Abb. 4.3: Modell der möglichen Funktion von SPOC1 in der normalen Entwicklung und in der Tumorentstehung. SPOC1 wirkt als Repressor der Genexpression, indem es Komponenten (HDAC/Methyltransferasen) rekrutiert, welche die Chromatinstruktur in einen transkriptionsinaktiven Zustand überführen. Eine weitere Möglichkeit ist, dass SPOC1 an der Assemblierung von Multiproteinkomplexen beteiligt ist, die Repression vermitteln. Eine wichtige Funktion von SPOC1 in der Embryonalentwicklung könnte in der Aufrechterhaltung des embryonalen Zustands proliferierender Zellen liegen. Dabei reprimiert SPOC1 Regulatoren (wie z. B. RB), welche die Aktivität von Transkriptionsfaktoren negativ regulieren, die an der Kontrolle und Aufrechterhaltung des Zellzyklus beteiligt sind (z. B. E2F). Dagegen könnte die aberrante Expression von SPOC1 in einem ausdifferenzierten Gewebe zur Entstehung von Tumoren führen. Die Repression negativer Regulatoren (z. B. RB) von Zellzyklus stimulierenden Transkriptionsfaktoren (z. B. E2F) könnte zu einem Rückschritt ausdifferenzierter Zellen aus der G<sub>0</sub>-Phase in die G<sub>1</sub>/S-Phase führen. Es kommt zu einer Dedifferenzierung der Zellen und zur Tumorentstehung durch Proliferation.

## 5. Zusammenfassung

Das Ziel meiner Doktorarbeit am Institut für Humangenetik war die Identifizierung und funktionelle Charakterisierung medizinisch relevanter neuer Gene aus einer cDNA-Bibliothek aus menschlichen fetalen Knorpelgewebe. Für die Auswertung der Gensequenzen der cDNA-Bibliothek wurden im Rahmen der vorliegenden Doktorarbeit 1050 EST-Klone untersucht. Insgesamt wurden im Rahmen einer Kooperation (Prof. Winterpacht, Hamburg/Erlangen; Prof. Zabel, Mainz und Firma GENterprise) 4748 EST-Klone der cDNA-Bibliothek durch die drei Kooperationspartner analysiert. Um bereits bekannte Gensequenzen in der Genbibliothek von unbekannten Gensequenzen zu trennen, wurden alle Sequenzen (EST-Klone) in bioinformatischen Analysen in Nukleinsäure- und Proteindatenbanken auf Homologien zu bekannten Gen- und Proteinsequenzen untersucht. Der Vergleich der Sequenzanalysen von 4748 EST-Klonen des gesamten Projekts zeigte, dass 72,2% der analysierten Klone signifikante Homologien zu bekannten cDNA Sequenzen und Genen besitzen (Kategorie I). Weitere 10,3% der Klone zeigten Homologie zu uncharakterisierten cDNAs und Proteinen (Kategorie II) während 14% der Klone keine Homologien oder Homologien zu genomischen Sequenzen oder uncharakterisierten ESTs zeigten (Kategorie III). Aus den Gensequenzen, die keine Homologien zu bekannten Genen und Proteinen aufweisen (Kategorie II und III), wurden im Rahmen dieser Arbeit einzelne Klone ausgewählt und weitergehend untersucht. Die detaillierte Analyse eines Klons mit auffälligen strukturellen Merkmalen führte dabei zur Identifizierung und Klonierung eines bisher unbekannten menschlichen und murinen Gens, welches auf Chromosomen 1p36 bzw. Maus Chromosom 4 lokalisiert ist und als SPOC1 (survival time - associated PHD protein in ovarian cancer) bezeichnet wurde. Das neu identifizierte Gen besteht aus 4 Exons und enthält eine PHD-Domäne (Plant Homeodomain). Proteine mit dieser Domäne haben häufig eine Funktion bei der Transkriptionsregulation. Expressionsanalysen in öffentlichen Datenbanken zeigten, dass SPOC1 in vielen Geweben und auch während der Entwicklung abgeschrieben wird. Eine verstärkte Expression konnte auch in Geweben aus Ovarialkarzinomen und verschiedenen anderen Tumor-Arten nachgewiesen werden. Northern-Analysen bestätigten die Expression von SPOC1 in zahlreichen menschlichen und murinen Geweben und embryonalen Stadien. Weiterhin konnte mit Hilfe von RNA in situ-Hybridisierungen nachgewiesen werden, dass murines SPOC1 sowohl in den Spermatogonien als auch in Uterus-Gewebe in den Zellen des Stratum functionale des Endometriums exprimiert wird. Von beiden Zelltypen ist bekannt, dass sie im Menschen an der Entwicklung von Tumoren (Keimzell-Tumore bzw. Tumore des Endometriums) beteiligt sind. Untersuchungen der subzellulären Lokalisation mit Hilfe von FLAG-SPOC1-SPOC1-EGFP-Fusionsproteinen und und einem polyklonalen Peptidantikörper gegen SPOC1 zeigten in menschlichen Zelllinien, dass es sich bei SPOC1, wie aufgrund der Proteindomänen vermutet, um ein im Zellkern lokalisiertes Protein handelt. In Kolokalisationsexperimenten konnte zudem nachgewiesen werden, dass SPOC1 mit RNA-Polymerase II und dem Transkriptionsfaktor E2F-1 eine partielle, aber deutliche Kolokalisation zeigt. Die Assoziation mit den transkriptionsaktiven Bereichen des Zellkerns unterstreicht die mögliche Beteiligung von SPOC1 an der Regulation der Transkription. Es wird vermutet, dass es sich bei der im SPOC1-Protein enthaltenen PHD-Domäne um eine Protein-Protein Interaktionsdomäne handelt, welche Interaktionen verschiedener Proteine und Proteinkomplexe vermittelt, die an der Reorganisation der Chromatinstruktur beteiligt sind. Untersuchungen des Effektes des SPOC1-Proteins auf die Transkription, mit Hilfe des Gal4-Luziferase-Reporter-Assays, unterstützen diese Vermutung. Die Expression eines Gal4DBD-SPOC1-Fusionsproteins zeigte eine stark reprimierende Wirkung auf die Expression des Luziferase-Reportergens. Es ist daher möglich, dass SPOC1 an der Assoziation von Proteinkomplexen beteiligt ist, die durch Reorganisation der Chromatinstruktur eine Repression der Transkription vermitteln können.

Untersuchungen der Expression von *SPOC1* in Tumorgewebe aus humanen Ovarialkarzinomen und rezidivierenden Ovarialkarzinomen durch Prof. Hengstler (Leipzig) zeigten, dass eine erhöhte *SPOC1*-mRNA Expression mit einer schlechten Prognose für Patienten mit primären und rezidivierenden Ovarialkarzinom assoziiert ist. Die Überlebenszeit von Patienten mit einer erhöhten *SPOC1*-Expression ist erheblich geringer, als die Überlebenszeit von Patienten mit einer schwachen *SPOC1*-Expression.

Aufgrund der Ergebnisse unserer Untersuchungen vermuten wir, dass *SPOC1* als ein prognostischer Marker für das Ovarialkarzinom verwendbar sein könnte.

Aufgrund der möglichen Funktion von *SPOC1* als prognostischer Marker und der Bedeutung als mögliches Therapieziel, haben wir *SPOC1* als Patent angemeldet: (K203 (*SPOC1*): ein neuer Transkriptionsregulator als "drug target" und molekularer Marker für Prognose, Therapieresistenz und Residualtumor bei Ovarialkarzinomen und anderen Tumorerkrankungen).

# 6. Literaturverzeichnis

- Aapola, U., Liiv, I. and Peterson, P. (2002) Imprinting regulator DNMT3L is a transcriptional repressor associated with histone deacetylase activity. *Nucleic Acids Res*, **30**, 3602-3608.
- Aasland, R., Gibson, T.J. and Stewart, A.F. (1995) The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem Sci*, **20**, 56-59.
- Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F. and et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, 252, 1651-1656.
- Adams, M.D., Kerlavage, A.R., Fields, C. and Venter, J.C. (1993) 3,400 new expressed sequence tags identify diversity of transcripts in human brain. *Nat Genet*, **4**, 256-267.
- Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O. and et al. (1995) Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature*, **377**, 3-174.
- Ahmad, K. and Henikoff, S. (2002) Epigenetic consequences of nucleosome dynamics. *Cell*, **111**, 281-284.
- Ahn, J., Ludecke, H.J., Lindow, S., Horton, W.A., Lee, B., Wagner, M.J., Horsthemke, B. and Wells, D.E. (1995) Cloning of the putative tumour suppressor gene for hereditary multiple exostoses (EXT1). *Nat Genet*, **11**, 137-143.
- Alcalay, M., Tomassoni, L., Colombo, E., Stoldt, S., Grignani, F., Fagioli, M., Szekely, L., Helin, K. and Pelicci, P.G. (1998) The promyelocytic leukemia gene product (PML) forms stable complexes with the retinoblastoma protein. *Mol Cell Biol*, **18**, 1084-1093.
- Altschul, S.F., Boguski, M.S., Gish, W. and Wootton, J.C. (1994) Issues in searching molecular sequence databases. *Nat Genet*, **6**, 119-129.
- Altschul, S.F. and Lipman, D.J. (1990) Protein database searches for multiple alignments. *Proc Natl Acad Sci U S A*, **87**, 5509-5513.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25, 3389-3402.
- Angrand, P.O., Apiou, F., Stewart, A.F., Dutrillaux, B., Losson, R. and Chambon, P. (2001) NSD3, a new SET domain-containing gene, maps to 8p12 and is amplified in human breast cancer cell lines. *Genomics*, **74**, 79-88.
- Aravind, L., Iyer, L.M. and Koonin, E.V. (2003) Scores of RINGS but no PHDs in ubiquitin signaling. *Cell Cycle*, **2**, 123-126.
- Armstrong, J.A. and Emerson, B.M. (1998) Transcription of chromatin: these are complex times. *Curr Opin Genet Dev*, **8**, 165-172.
- Arnold, J.T., Kaufman, D.G., Seppala, M. and Lessey, B.A. (2001) Endometrial stromal cells regulate epithelial cell growth in vitro: a new co-culture model. *Hum Reprod*, 16, 836-845.
- Arnold<sup>1</sup>, J.M., Mok, S.C., Purdie, D. and Chenevix-Trench, G. (2001) Decreased expression of the Id3 gene at 1p36.1 in ovarian adenocarcinomas. *Br J Cancer*, **84**, 352-359.
- Becker, P.B. (2002) Nucleosome sliding: facts and fiction. Embo J, 21, 4749-4753.
- Becker, P.B. and Horz, W. (2002) ATP-dependent nucleosome remodeling. *Annu Rev Biochem*, **71**, 247-273.

- Berube, N.G., Smeenk, C.A. and Picketts, D.J. (2000) Cell cycle-dependent phosphorylation of the ATRX protein correlates with changes in nuclear matrix and chromatin association. *Hum Mol Genet*, **9**, 539-547.
- Bochar, D.A., Savard, J., Wang, W., Lafleur, D.W., Moore, P., Cote, J. and Shiekhattar, R. (2000) A family of chromatin remodeling factors related to Williams syndrome transcription factor. *Proc Natl Acad Sci U S A*, **97**, 1038-1043.
- Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996) Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res*, **6**, 791-806.
- Brehm, A., Nielsen, S.J., Miska, E.A., McCance, D.J., Reid, J.L., Bannister, A.J. and Kouzarides, T. (1999) The E7 oncoprotein associates with Mi2 and histone deacetylase activity to promote cell growth. *Embo J*, **18**, 2449-2458.
- Briggs, M.D., Hoffman, S.M., King, L.M., Olsen, A.S., Mohrenweiser, H., Leroy, J.G., Mortier, G.R., Rimoin, D.L., Lachman, R.S., Gaines, E.S. and et al. (1995) Pseudoachondroplasia and multiple epiphyseal dysplasia due to mutations in the cartilage oligomeric matrix protein gene. *Nat Genet*, **10**, 330-336.
- Brown, K.E., Guest, S.S., Smale, S.T., Hahm, K., Merkenschlager, M. and Fisher, A.G. (1997) Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell*, **91**, 845-854.
- Burge, C. and Karlin, S. (1997) Prediction of complete gene structures in human genomic DNA. *J Mol Biol*, **268**, 78-94.
- Burge, C.B. and Karlin, S. (1998) Finding the genes in genomic DNA. *Curr Opin Struct Biol*, **8**, 346-354.
- Capili, A.D., Schultz, D.C., Rauscher, I.F. and Borden, K.L. (2001) Solution structure of the PHD domain from the KAP-1 corepressor: structural determinants for PHD, RING and LIM zinc-binding domains. *Embo J*, **20**, 165-177.
- Champagne, N., Pelletier, N. and Yang, X.J. (2001) The monocytic leukemia zinc finger protein MOZ is a histone acetyltransferase. *Oncogene*, **20**, 404-409.
- Chen, J., Sun, M., Lee, S., Zhou, G., Rowley, J.D. and Wang, S.M. (2002) Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. *Proc Natl Acad Sci U S A*, **99**, 12257-12262.
- Chen, J.J., Rowley, J.D. and Wang, S.M. (2000) Generation of longer cDNA fragments from serial analysis of gene expression tags for gene identification. *Proc Natl Acad Sci U S A*, **97**, 349-353.
- Clark-Adams, C.D., Norris, D., Osley, M.A., Fassler, J.S. and Winston, F. (1988) Changes in histone gene dosage alter transcription in yeast. *Genes Dev*, **2**, 150-159.
- Claudio, J.O., Liew, C.C., Dempsey, A.A., Cukerman, E., Stewart, A.K., Na, E., Atkins, H.L., Iscove, N.N. and Hawley, R.G. (1998) Identification of sequence-tagged transcripts differentially expressed within the human hematopoietic hierarchy. *Genomics*, **50**, 44-52.
- Cmarko, D., Verschure, P.J., Martin, T.E., Dahmus, M.E., Krause, S., Fu, X.D., van Driel, R. and Fakan, S. (1999) Ultrastructural analysis of transcription and splicing in the cell nucleus after bromo-UTP microinjection. *Mol Biol Cell*, **10**, 211-223.
- Cmarko, D., Verschure, P.J., Otte, A.P., Van Driel, R. and Fakan, S. (2003) Polycomb group gene silencing proteins are concentrated in the perichromatin compartment of the mammalian nucleus. *J Cell Sci*, **116**, 335-343.
- Cockell, M. and Gasser, S.M. (1999) Nuclear compartments and gene regulation. *Curr Opin Genet Dev*, **9**, 199-205.
- Coscoy, L. and Ganem, D. (2003) PHD domains and E3 ubiquitin ligases: viruses make the connection. *Trends Cell Biol*, **13**, 7-12.

- Cremer, T. and Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*, **2**, 292-301.
- Cremer, T., Kurz, A., Zirbel, R., Dietzel, S., Rinke, B., Schrock, E., Speicher, M.R., Mathieu, U., Jauch, A., Emmerich, P. and et al. (1993) Role of chromosome territories in the functional compartmentalization of the cell nucleus. *Cold Spring Harb Symp Quant Biol*, 58, 777-792.
- Croft, J.A., Bridger, J.M., Boyle, S., Perry, P., Teague, P. and Bickmore, W.A. (1999) Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol*, **145**, 1119-1131.
- Czermin, B., Schotta, G., Hulsmann, B.B., Brehm, A., Becker, P.B., Reuter, G. and Imhof, A. (2001) Physical and functional association of SU(VAR)3-9 and HDAC1 in Drosophila. *EMBO Rep*, **2**, 915-919.
- Davidson, E.H. and Britten, R.J. (1979) Regulation of gene expression: possible role of repetitive sequences. *Science*, **204**, 1052-1059.
- de Jong, L., Grande, M.A., Mattern, K.A., Schul, W. and van Driel, R. (1996) Nuclear domains involved in RNA synthesis, RNA processing, and replication. *Crit Rev Eukaryot Gene Expr*, **6**, 215-246.
- de Ruijter, A.J., van Gennip, A.H., Caron, H.N., Kemp, S. and van Kuilenburg, A.B. (2003) Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem J*, **370**, 737-749.
- Deguchi, K., Ayton, P.M., Carapeti, M., Kutok, J.L., Snyder, C.S., Williams, I.R., Cross, N.C., Glass, C.K., Cleary, M.L. and Gilliland, D.G. (2003) MOZ-TIF2-induced acute myeloid leukemia requires the MOZ nucleosome binding motif and TIF2-mediated recruitment of CBP. *Cancer Cell*, **3**, 259-271.
- Dias Neto, E., Correa, R.G., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W., Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J. (2000) Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proc Natl Acad Sci U S A*, 97, 3491-3496.
- Eberharter, A. and Becker, P.B. (2002) Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. *EMBO Rep*, **3**, 224-229.
- Eisen, J.A., Sweder, K.S. and Hanawalt, P.C. (1995) Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. *Nucleic Acids Res*, 23, 2715-2723.
- Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Zhuang, Z., Goldstein, S.R., Weiss, R.A. and Liotta, L.A. (1996) Laser capture microdissection. *Science*, 274, 998-1001.
- Fair, K., Anderson, M., Bulanova, E., Mi, H., Tropschug, M. and Diaz, M.O. (2001) Protein interactions of the MLL PHD fingers modulate MLL target gene regulation in human cells. *Mol Cell Biol*, **21**, 3589-3597.
- Felsenfeld<sup>1</sup>, G. (1996) Chromatin unfolds. *Cell*, **86**, 13-19.
- Felsenfeld<sup>2</sup>, G., Boyes, J., Chung, J., Clark, D. and Studitsky, V. (1996) Chromatin structure and gene expression. *Proc Natl Acad Sci U S A*, **93**, 9384-9388.
- Ferreira, J., Paolella, G., Ramos, C. and Lamond, A.I. (1997) Spatial organization of largescale chromatin domains in the nucleus: a magnified view of single chromosome territories. *J Cell Biol*, **139**, 1597-1610.
- Ferreira, R., Naguibneva, I., Pritchard, L.L., Ait-Si-Ali, S. and Harel-Bellan, A. (2001) The Rb/chromatin connection and epigenetic control: opinion. *Oncogene*, **20**, 3128-3133.

- Flanagan, J.F. and Peterson, C.L. (1999) A role for the yeast SWI/SNF complex in DNA replication. *Nucleic Acids Res*, **27**, 2022-2028.
- Forsberg, E.C. and Bresnick, E.H. (2001) Histone acetylation beyond promoters: long-range acetylation patterns in the chromatin world. *Bioessays*, **23**, 820-830.
- Francastel, C., Schubeler, D., Martin, D.I. and Groudine, M. (2000) Nuclear compartmentalization and gene activity. *Nat Rev Mol Cell Biol*, **1**, 137-143.
- Francastel, C., Walters, M.C., Groudine, M. and Martin, D.I. (1999) A functional enhancer suppresses silencing of a transgene and prevents its localization close to centrometric heterochromatin. *Cell*, **99**, 259-269.
- Gall, J.G. (2000) Cajal bodies: the first 100 years. Annu Rev Cell Dev Biol, 16, 273-300.
- Gibbons, R.J., McDowell, T.L., Raman, S., O'Rourke, D.M., Garrick, D., Ayyub, H. and Higgs, D.R. (2000) Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nat Genet*, **24**, 368-371.
- Gibbs, M., Stanford, J.L., McIndoe, R.A., Jarvik, G.P., Kolb, S., Goode, E.L., Chakrabarti, L., Schuster, E.F., Buckley, V.A., Miller, E.L., Brandzel, S., Li, S., Hood, L. and Ostrander, E.A. (1999) Evidence for a rare prostate cancer-susceptibility locus at chromosome 1p36. *Am J Hum Genet*, 64, 776-787.
- Goddard, A.D., Borrow, J., Freemont, P.S. and Solomon, E. (1991) Characterization of a zinc finger gene disrupted by the t(15;17) in acute promyelocytic leukemia. *Science*, **254**, 1371-1374.
- Grande, M.A., van der Kraan, I., de Jong, L. and van Driel, R. (1997) Nuclear distribution of transcription factors in relation to sites of transcription and RNA polymerase II. *J Cell Sci*, **110** (**Pt 15**), 1781-1791.
- Grant, P.A., Duggan, L., Cote, J., Roberts, S. M., Brownell, J. E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C. D., Winston, F., Berger, S. L., Workman, J. L. (1997) Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes Dev*, **11**, 1640-1650.
- Grant, P.A., Schieltz, D., Pray-Grant, M. G., Yates, J. R. 3rd, Workman, J. L. (1998) The ATM-related cofactor Tra1 is a component of the purified SAGA complex. *Mol Cell*, 6, 863-867.
- Gubler, U. and Hoffman, B.J. (1983) A simple and very efficient method for generating cDNA libraries. *Gene*, **25**, 263-269.
- Gunduz, M., Ouchida, M., Fukushima, K., Hanafusa, H., Etani, T., Nishioka, S., Nishizaki, K. and Shimizu, K. (2000) Genomic structure of the human ING1 gene and tumorspecific mutations detected in head and neck squamous cell carcinomas. *Cancer Res*, 60, 3143-3146.
- Han, M., Kim, U.J., Kayne, P. and Grunstein, M. (1988) Depletion of histone H4 and nucleosomes activates the PHO5 gene in Saccharomyces cerevisiae. *Embo J*, **7**, 2221-2228.
- Hassan, A.H., Neely, K.E., Vignali, M., Reese, J.C. and Workman, J.L. (2001) Promoter targeting of chromatin-modifying complexes. *Front Biosci*, **6**, D1054-1064.
- Hayes, J.J. and Hansen, J.C. (2001) Nucleosomes and the chromatin fiber. *Curr Opin Genet Dev*, **11**, 124-129.
- Hecht, J.T., Nelson, L.D., Crowder, E., Wang, Y., Elder, F.F., Harrison, W.R., Francomano, C.A., Prange, C.K., Lennon, G.G., Deere, M. and et al. (1995) Mutations in exon 17B of cartilage oligomeric matrix protein (COMP) cause pseudoachondroplasia. *Nat Genet*, 10, 325-329.
- Helin, K. (1998) Regulation of cell proliferation by the E2F transcription factors. *Curr Opin Genet Dev*, **8**, 28-35.

- Hendrich, B. and Bird, A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol*, **18**, 6538-6547.
- Henikoff, S. and Dreesen, T.D. (1989) Trans-inactivation of the Drosophila brown gene: evidence for transcriptional repression and somatic pairing dependence. *Proc Natl Acad Sci U S A*, **86**, 6704-6708.
- Horn, P.J. and Peterson, C.L. (2002) Chromatin Higher Order Folding: Wrapping up Transcription. *Science*, **297**, 1824-1827.
- Hu, M., Krause, D., Greaves, M., Sharkis, S., Dexter, M., Heyworth, C. and Enver, T. (1997) Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev*, **11**, 774-785.
- Ishov, A.M., Sotnikov, A.G., Negorev, D., Vladimirova, O.V., Neff, N., Kamitani, T., Yeh, E.T., Strauss, J.F., 3rd and Maul, G.G. (1999) PML is critical for ND10 formation and recruits the PML-interacting protein daxx to this nuclear structure when modified by SUMO-1. J Cell Biol, 147, 221-234.
- Jacobson, S. and Pillus, L. (1999) Modifying chromatin and concepts of cancer. *Curr Opin Genet Dev*, **9**, 175-184.
- Jia, L., Ho, N.C., Park, S.S., Powell, J. and Francomano, C.A. (2001) Comprehensive resource: Skeletal gene database. *Am J Med Genet*, **106**, 275-281.
- Jimenez, G., Griffiths, S.D., Ford, A.M., Greaves, M.F. and Enver, T. (1992) Activation of the beta-globin locus control region precedes commitment to the erythroid lineage. *Proc Natl Acad Sci U S A*, **89**, 10618-10622.
- Jones, P.A. and Baylin, S.B. (2002) The fundamental role of epigenetic events in cancer. *Nat Rev Genet*, **3**, 415-428.
- Kanehisa, M. and Bork, P. (2003) Bioinformatics in the post-sequence era. *Nat Genet*, **33** Suppl, 305-310.
- Kiechle, M., Hinrichs, M., Jacobsen, A., Luttges, J., Pfisterer, J., Kommoss, F. and Arnold, N. (2000) Genetic imbalances in precursor lesions of endometrial cancer detected by comparative genomic hybridization. *Am J Pathol*, **156**, 1827-1833.
- Kornberg<sup>1</sup>, R.D. and Lorch, Y. (1999a) Chromatin-modifying and -remodeling complexes. *Curr Opin Genet Dev*, **9**, 148-151.
- Kornberg<sup>2</sup>, R.D. and Lorch, Y. (1999b) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell*, **98**, 285-294.
- Kozak, M. (1996) Interpreting cDNA sequences: some insights from studies on translation. *Mamm Genome*, **7**, 563-574.
- Kramer, J.A., McCarrey, J.R., Djakiew, D. and Krawetz, S.A. (1998) Differentiation: the selective potentiation of chromatin domains. *Development*, **125**, 4749-4755.
- Krause, S., Fakan, S., Weis, K. and Wahle, E. (1994) Immunodetection of poly(A) binding protein II in the cell nucleus. *Exp Cell Res*, **214**, 75-82.
- Kumar, S., Connor, J.R., Dodds, R.A., Halsey, W., Van Horn, M., Mao, J., Sathe, G., Mui, P., Agarwal, P., Badger, A.M., Lee, J.C., Gowen, M. and Lark, M.W. (2001) Identification and initial characterization of 5000 expressed sequenced tags (ESTs) each from adult human normal and osteoarthritic cartilage cDNA libraries. *Osteoarthritis Cartilage*, 9, 641-653.
- Kurz, A., Lampel, S., Nickolenko, J.E., Bradl, J., Benner, A., Zirbel, R.M., Cremer, T. and Lichter, P. (1996) Active and inactive genes localize preferentially in the periphery of chromosome territories. *J Cell Biol*, **135**, 1195-1205.
- Kwan, A.H., Gell, D.A., Verger, A., Crossley, M., Matthews, J.M. and Mackay, J.P. (2003) Engineering a protein scaffold from a PHD finger. *Structure (Camb)*, **11**, 803-813.
- Kwon, J., Morshead, K.B., Guyon, J.R., Kingston, R.E. and Oettinger, M.A. (2000) Histone acetylation and hSWI/SNF remodeling act in concert to stimulate V(D)J cleavage of nucleosomal DNA. *Mol Cell*, **6**, 1037-1048.

- Labrador, M. and Corces, V.G. (2002) Setting the boundaries of chromatin domains and nuclear organization. *Cell*, **111**, 151-154.
- Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-685.
- Lamond, A.I. and Earnshaw, W.C. (1998) Structure and function in the nucleus. *Science*, **280**, 547-553.
- Lamond, A.I. and Spector, D.L. (2003) Nuclear Speckles: A Model for Nuclear Organelles. *Nature Reviews Molecular Cell Biology*, **4**, 605-612.
- LaMorte, V.J., Dyck, J.A., Ochs, R.L. and Evans, R.M. (1998) Localization of nascent RNA and CREB binding protein with the PML-containing nuclear body. *Proc Natl Acad Sci U S A*, **95**, 4991-4996.
- Längst, G. and Becker, P.B. (2001) Nucleosome mobilization and positioning by ISWIcontaining chromatin-remodeling factors. *J Cell Sci*, **114**, 2561-2568.
- Lash, A.E., Tolstoshev, C.M., Wagner, L., Schuler, G.D., Strausberg, R.L., Riggins, G.J. and Altschul, S.F. (2000) SAGEmap: a public gene expression resource. *Genome Res*, **10**, 1051-1060.
- Leung, A.K. and Lamond, A.I. (2003) The dynamics of the nucleolus. *Crit Rev Eukaryot Gene Expr*, **13**, 39-54.
- Linder, B., Newman, R., Jones, L.K., Debernardi, S., Young, B.D., Freemont, P., Verrijzer, C.P. and Saha, V. (2000) Biochemical analyses of the AF10 protein: the extended LAP/PHD-finger mediates oligomerisation. *J Mol Biol*, **299**, 369-378.
- Lipinski, M.M. and Jacks, T. (1999) The retinoblastoma gene family in differentiation and development. *Oncogene*, **18**, 7873-7882.
- Lo, W.S., Duggan, L., Tolga, N.C., Emre, Belotserkovskya, R., Lane, W.S., Shiekhattar, R. and Berger, S.L. (2001) Snf1--a histone kinase that works in concert with the histone acetyltransferase Gcn5 to regulate transcription. *Science*, **293**, 1142-1146.
- Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. and Brown, E.L. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol*, 14, 1675-1680.
- Lockhart, D.J. and Winzeler, E.A. (2000) Genomics, gene expression and DNA arrays. *Nature*, **405**, 827-836.
- Loewith, R., Meijer, M., Lees-Miller, S.P., Riabowol, K. and Young, D. (2000) Three yeast proteins related to the human candidate tumor suppressor p33(ING1) are associated with histone acetyltransferase activities. *Mol Cell Biol*, **20**, 3807-3816.
- Lorch, Y., Cairns, B.R., Zhang, M. and Kornberg, R.D. (1998) Activated RSC-nucleosome complex and persistently altered form of the nucleosome. *Cell*, **94**, 29-34.
- Lu, P.J., Sundquist, K., Baeckstrom, D., Poulsom, R., Hanby, A., Meier-Ewert, S., Jones, T., Mitchell, M., Pitha-Rowe, P., Freemont, P. and Taylor-Papadimitriou, J. (1999) A novel gene (PLU-1) containing highly conserved putative DNA/chromatin binding motifs is specifically up-regulated in breast cancer. *J Biol Chem*, **274**, 15633-15645.
- Lu, X., Meng, X., Morris, C.A. and Keating, M.T. (1998) A novel human gene, WSTF, is deleted in Williams syndrome. *Genomics*, **54**, 241-249.
- Lu, Z., Xu, S., Joazeiro, C., Cobb, M.H. and Hunter, T. (2002) The PHD domain of MEKK1 acts as an E3 ubiquitin ligase and mediates ubiquitination and degradation of ERK1/2. *Mol Cell*, **9**, 945-956.
- Mahy, N.L., Perry, P.E. and Bickmore, W.A. (2002) Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J Cell Biol*, **159**, 753-763.

- Mandahl, N., Gustafson, P., Mertens, F., Akerman, M., Baldetorp, B., Gisselsson, D., Knuutila, S., Bauer, H.C. and Larsson, O. (2002) Cytogenetic aberrations and their prognostic impact in chondrosarcoma. *Genes Chromosomes Cancer*, **33**, 188-200.
- Mao, M., Fu, G., Wu, J.S., Zhang, Q.H., Zhou, J., Kan, L.X., Huang, Q.H., He, K.L., Gu, B.W., Han, Z.G., Shen, Y., Gu, J., Yu, Y.P., Xu, S.H., Wang, Y.X., Chen, S.J. and 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 U S A*, 95, 8175-8180.
- Marshall, W.F., Fung, J.C. and Sedat, J.W. (1997) Deconstructing the nucleus: global architecture from local interactions. *Curr Opin Genet Dev*, **7**, 259-263.
- Mikulowska-Mennis, A., Taylor, T.B., Vishnu, P., Michie, S.A., Raja, R., Horner, N. and Kunitake, S.T. (2002) High-quality RNA from cells isolated by laser capture microdissection. *Biotechniques*, **33**, 176-179.
- Mintz, P.J., Patterson, S.D., Neuwald, A.F., Spahr, C.S. and Spector, D.L. (1999) Purification and biochemical characterization of interchromatin granule clusters. *Embo J*, **18**, 4308-4320.
- Mintz, P.J. and Spector, D.L. (2000) Compartmentalization of RNA processing factors within nuclear speckles. *J Struct Biol*, **129**, 241-251.
- Misteli, T. (2001) Protein dynamics: implications for nuclear architecture and gene expression. *Science*, **291**, 843-847.
- Mora, J., Cheung, N.K., Oplanich, S., Chen, L. and Gerald, W.L. (2002) Novel regions of allelic imbalance identified by genome-wide analysis of neuroblastoma. *Cancer Res*, 62, 1761-1767.
- Mostoslavsky, R., Alt, F.W. and Bassing, C.H. (2003) Chromatin dynamics and locus accessibility in the immune system. *Nat Immunol*, **4**, 603-606.
- Müller, S., Matunis, M.J. and Dejean, A. (1998) Conjugation with the ubiquitin-related modifier SUMO-1 regulates the partitioning of PML within the nucleus. *Embo J*, **17**, 61-70.
- Mundlos, S. (1994) Expression patterns of matrix genes during human skeletal development. *Prog Histochem Cytochem*, **28**, 1-47.
- Okihana, H. and Yamada, K. (1999) Preparation of a cDNA library and preliminary assessment of 1400 genes from mouse growth cartilage. *J Bone Miner Res*, **14**, 304-310.
- Oliver, R.T. (1996) Testicular cancer. Curr Opin Oncol, 8, 252-258.
- Pascual, J., Martinez-Yamout, M., Dyson, H.J. and Wright, P.E. (2000) Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor. *J Mol Biol*, **304**, 723-729.
- Pearson, W.R. and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A*, **85**, 2444-2448.
- Peterson, C.L. and Workman, J.L. (2000) Promoter targeting and chromatin remodeling by the SWI/SNF complex. *Curr Opin Genet Dev*, **10**, 187-192.
- Politz, J.C., Tuft, R.A., Pederson, T. and Singer, R.H. (1999) Movement of nuclear poly(A) RNA throughout the interchromatin space in living cells. *Curr Biol*, **9**, 285-291.
- Ragnarsson, G., Eiriksdottir, G., Johannsdottir, J.T., Jonasson, J.G., Egilsson, V. and Ingvarsson, S. (1999) Loss of heterozygosity at chromosome 1p in different solid human tumours: association with survival. *Br J Cancer*, **79**, 1468-1474.
- Rinderle, C., Christensen, H.M., Schweiger, S., Lehrach, H. and Yaspo, M.L. (1999) AIRE encodes a nuclear protein co-localizing with cytoskeletal filaments: altered sub-cellular distribution of mutants lacking the PHD zinc fingers. *Hum Mol Genet*, **8**, 277-290.

- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487-491.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, **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. and Hayashizaki, Y. (1998) Characterization of gene expression in mouse blastocyst using single-pass sequencing of 3995 clones. *Genomics*, **49**, 167-179.
- Schardin, M., Cremer, T., Hager, H.D. and Lang, M. (1985) Specific staining of human chromosomes in Chinese hamster x man hybrid cell lines demonstrates interphase chromosome territories. *Hum Genet*, **71**, 281-287.
- Scheer, U. and Weisenberger, D. (1994) The nucleolus. Curr Opin Cell Biol, 6, 354-359.
- Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**, 467-470.
- Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O. and Davis, R.W. (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci U S A*, **93**, 10614-10619.
- Schindler, U., Beckmann, H. and Cashmore, A.R. (1993) HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region. *Plant J*, **4**, 137-150.
- Schul, W., Groenhout, B., Koberna, K., Takagaki, Y., Jenny, A., Manders, E.M., Raska, I., van Driel, R. and de Jong, L. (1996) The RNA 3' cleavage factors CstF 64 kDa and CPSF 100 kDa are concentrated in nuclear domains closely associated with coiled bodies and newly synthesized RNA. *Embo J*, **15**, 2883-2892.
- Schuler, G.D. (1997) Pieces of the puzzle: expressed sequence tags and the catalog of human genes. *J Mol Med*, **75**, 694-698.
- Schuler, G.D., Boguski, M.S., Stewart, E.A., Stein, L.D., Gyapay, G., Rice, K., White, R.E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E., Bentolila, S., Birren, B.B., Butler, A., Castle, A.B., Chiannilkulchai, N., Chu, A., Clee, C., Cowles, S., Day, P.J., Dibling, T., Drouot, N., Dunham, I., Duprat, S., East, C., Hudson, T.J. and et al. (1996) A gene map of the human genome. *Science*, **274**, 540-546.
- Schultz, D.C., Friedman, J.R. and Rauscher, F.J., 3rd. (2001) Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. *Genes Dev*, 15, 428-443.
- Seeler, J.S., Marchio, A., Sitterlin, D., Transy, C. and Dejean, A. (1998) Interaction of SP100 with HP1 proteins: a link between the promyelocytic leukemia-associated nuclear bodies and the chromatin compartment. *Proc Natl Acad Sci U S A*, **95**, 7316-7321.
- Shimizu, S., Suzukawa, K., Kodera, T., Nagasawa, T., Abe, T., Taniwaki, M., Yagasaki, F., Tanaka, H., Fujisawa, S., Johansson, B., Ahlgren, T., Yokota, J. and Morishita, K. (2000) Identification of breakpoint cluster regions at 1p36.3 and 3q21 in hematologic malignancies with t(1;3)(p36;q21). *Genes Chromosomes Cancer*, 27, 229-238.
- Soares, M.B., Bonaldo, M.F., Jelene, P., Su, L., Lawton, L. and Efstratiadis, A. (1994) Construction and characterization of a normalized cDNA library. *Proc Natl Acad Sci* USA, **91**, 9228-9232.
- Spector, D.L. (1993) Nuclear organization of pre-mRNA processing. *Curr Opin Cell Biol*, **5**, 442-447.
- Spector, D.L. (2001) Nuclear domains. J Cell Sci, 114, 2891-2893.
- Spector, D.L. and Gasser, S.M. (2003) A molecular dissection of nuclear function. Conference on the dynamic nucleus: questions and implications. *EMBO Rep*, **4**, 18-23.

- Spranger, J., Menger, H., Mundlos, S., Winterpacht, A. and Zabel, B. (1994) Kniest dysplasia is caused by dominant collagen II (COL2A1) mutations: parental somatic mosaicism manifesting as Stickler phenotype and mild spondyloepiphyseal dysplasia. *Pediatr Radiol*, 24, 431-435.
- Stec, I., Wright, T.J., van Ommen, G.J., de Boer, P.A., van Haeringen, A., Moorman, A.F., Altherr, M.R. and den Dunnen, J.T. (1998) WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a Drosophila dysmorphy gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. *Hum Mol Genet*, 7, 1071-1082.
- Stelzer, C. (2000) Molekulargenetische Untersuchungen von Komponenten des Knorpel-/Knochengewebes. *Dissertation*.
- Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications. *Nature*, **403**, 41-45.
- Struhl, K. (1998) Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev*, **12**, 599-606.
- The Finnish-German APECED Consortium. (1997). An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy. *Nat Genet*, **17**, 399-403.
- Thompson, B., Townsley, F., Rosin-Arbesfeld, R., Musisi, H. and Bienz, M. (2002) A new nuclear component of the Wnt signalling pathway. *Nat Cell Biol*, **4**, 367-373.
- Tri, N.K., Xinh, P.T., Nagao, H., Izumi, T., Ozawa, K., Toyoda, A., Hattori, M., Sakaki, Y., Tokunaga, K. and Sato, Y. (2002) Identification of the breakpoints at 1p36.2 and 3p21.3 in an AML(M3) patient who had t(1;3)(p36.2;p21.3) at the third relapse. *Genes Chromosomes Cancer*, **35**, 365-367.
- Tsukiyama, T. (2002) The in vivo functions of ATP-dependent chromatin-remodelling factors. *Nat Rev Mol Cell Biol*, **3**, 422-429.
- Tumbar, T., Sudlow, G. and Belmont, A.S. (1999) Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. *J Cell Biol*, **145**, 1341-1354.
- Uberbacher, E.C. and Mural, R.J. (1991) Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc Natl Acad Sci U S A*, **88**, 11261-11265.
- Ura, K., Araki, M., Saeki, H., Masutani, C., Ito, T., Iwai, S., Mizukoshi, T., Kaneda, Y. and Hanaoka, F. (2001) ATP-dependent chromatin remodeling facilitates nucleotide excision repair of UV-induced DNA lesions in synthetic dinucleosomes. *Embo J*, 20, 2004-2014.
- van Steensel, B., van Binnendijk, E.P., Hornsby, C.D., van der Voort, H.T., Krozowski, Z.S., de Kloet, E.R. and van Driel, R. (1996) Partial colocalization of glucocorticoid and mineralocorticoid receptors in discrete compartments in nuclei of rat hippocampus neurons. J Cell Sci, 109 (Pt 4), 787-792.
- Vandel, L. and Trouche, D. (2001) Physical association between the histone acetyl transferase CBP and a histone methyl transferase. *EMBO Rep*, **2**, 21-26.
- Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. (1995) Serial analysis of gene expression. *Science*, **270**, 484-487.
- Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E., Jr., Hieter, P., Vogelstein, B. and Kinzler, K.W. (1997) Characterization of the yeast transcriptome. *Cell*, 88, 243-251.
- Verschure, P.J., Van Der Kraan, I., Enserink, J.M., Mone, M.J., Manders, E.M. and Van Driel, R. (2002) Large-scale chromatin organization and the localization of proteins involved in gene expression in human cells. *J Histochem Cytochem*, **50**, 1303-1312.

- Verschure, P.J., van Der Kraan, I., Manders, E.M. and van Driel, R. (1999) Spatial relationship between transcription sites and chromosome territories. *J Cell Biol*, **147**, 13-24.
- Vignali, M., Hassan, A.H., Neely, K.E. and Workman, J.L. (2000) ATP-dependent chromatinremodeling complexes. *Mol Cell Biol*, **20**, 1899-1910.
- Visser, A.E., Jaunin, F., Fakan, S. and Aten, J.A. (2000) High resolution analysis of interphase chromosome domains. *J Cell Sci*, **113** ( **Pt 14**), 2585-2593.
- Volpi, E.V., Chevret, E., Jones, T., Vatcheva, R., Williamson, J., Beck, S., Campbell, R.D., Goldsworthy, M., Powis, S.H., Ragoussis, J., Trowsdale, J. and Sheer, D. (2000) Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. J Cell Sci, 113 (Pt 9), 1565-1576.
- Wade, P.A., Jones, P.L., Vermaak, D., Veenstra, G.J., Imhof, A., Sera, T., Tse, C., Ge, H., Shi, Y.B., Hansen, J.C. and Wolffe, A.P. (1998) Histone deacetylase directs the dominant silencing of transcription in chromatin: association with MeCP2 and the Mi-2 chromodomain SWI/SNF ATPase. *Cold Spring Harb Symp Quant Biol*, **63**, 435-445.
- Whitehouse, I., Flaus, A., Cairns, B.R., White, M.F., Workman, J.L. and Owen-Hughes, T. (1999) Nucleosome mobilization catalysed by the yeast SWI/SNF complex. *Nature*, 400, 784-787.
- Wiemann, S., Weil, B., Wellenreuther, R., Gassenhuber, J., Glassl, S., Ansorge, W., Bocher, M., Blocker, H., Bauersachs, S., Blum, H., Lauber, J., Dusterhoft, A., Beyer, A., Kohrer, K., Strack, N., Mewes, H.W., Ottenwalder, B., Obermaier, B., Tampe, J., Heubner, D., Wambutt, R., Korn, B., Klein, M. and Poustka, A. (2001) Toward a catalog of human genes and proteins: sequencing and analysis of 500 novel complete protein coding human cDNAs. *Genome Res*, 11, 422-435.
- Wolffe, A.P. and Guschin, D. (2000) Review: chromatin structural features and targets that regulate transcription. *J Struct Biol*, **129**, 102-122.
- Woodcock, C.L. and Dimitrov, S. (2001) Higher-order structure of chromatin and chromosomes. *Curr Opin Genet Dev*, **11**, 130-135.
- Workman, J.L. and Kingston, R.E. (1998) Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu Rev Biochem*, **67**, 545-579.
- Wreggett, K.A., Hill, F., James, P.S., Hutchings, A., Butcher, G.W. and Singh, P.B. (1994) A mammalian homologue of Drosophila heterochromatin protein 1 (HP1) is a component of constitutive heterochromatin. *Cytogenet Cell Genet*, **66**, 99-103.
- Xia, Z.B., Anderson, M., Diaz, M.O. and Zeleznik-Le, N.J. (2003) MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein. *Proc Natl Acad Sci U S A*, **100**, 8342-8347.
- Xiao, J., Xu, M., Li, J., Chang Chan, H., Lin, M., Zhu, H., Zhang, W., Zhou, Z., Zhao, B. and Sha, J. (2002) NYD-SP6, a novel gene potentially involved in regulating testicular development/spermatogenesis. *Biochem Biophys Res Commun*, **291**, 101-110.
- Xu, Y., Einstein, J.R., Mural, R.J., Shah, M. and Uberbacher, E.C. (1994) An improved system for exon recognition and gene modeling in human DNA sequences. *Proc Int Conf Intell Syst Mol Biol*, **2**, 376-384.
- Yang, Z.Q., Imoto, I., Fukuda, Y., Pimkhaokham, A., Shimada, Y., Imamura, M., Sugano, S., and Nakamura, Y., Inazawa, J. (2000) Identification of a novel gene, GASC1, within an amplicon at 9p23-24 frequently detected in esophageal cancer cell lines. *Cancer Res*, 60, 4735-4739.

- Ye, Q., Hu, Y.F., Zhong, H., Nye, A.C., Belmont, A.S. and Li, R. (2001) BRCA1-induced large-scale chromatin unfolding and allele-specific effects of cancer-predisposing mutations. *J Cell Biol*, **155**, 911-921.
- Yudate, H.T., Suwa, M., Irie, R., Matsui, H., Nishikawa, T., Nakamura, Y., Yamaguchi, D., Peng, Z.Z., Yamamoto, T., Nagai, K., Hayashi, K., Otsuki, T., Sugiyama, T., Ota, T., Suzuki, Y., Sugano, S., Isogai, T. and Masuho, Y. (2001) HUNT: launch of a fulllength cDNA database from the Helix Research Institute. *Nucleic Acids Res*, 29, 185-188.
- Zhang, H., Marshall, K.W., Tang, H., Hwang, D.M., Lee, M. and Liew, C.C. (2003) Profiling genes expressed in human fetal cartilage using 13,155 expressed sequence tags. *Osteoarthritis Cartilage*, **11**, 309-319.
- Zhang, L., Zhou, W., Velculescu, V.E., Kern, S.E., Hruban, R.H., Hamilton, S.R., Vogelstein, B. and Kinzler, K.W. (1997) Gene expression profiles in normal and cancer cells. *Science*, 276, 1268-1272.
- Zhong<sup>1</sup>, S., Salomoni, P. and Pandolfi, P.P. (2000a) The transcriptional role of PML and the nuclear body. *Nat Cell Biol*, **2**, E85-90.
- Zhong<sup>2</sup>, S., Salomoni, P., Ronchetti, S., Guo, A., Ruggero, D. and Pandolfi, P.P. (2000b) Promyelocytic leukemia protein (PML) and Daxx participate in a novel nuclear pathway for apoptosis. *J Exp Med*, **191**, 631-640.
- Zielenska, M., Bayani, J., Pandita, A., Toledo, S., Marrano, P., Andrade, J., Petrilli, A., Thorner, P., Sorensen, P. and Squire, J.A. (2001) Comparative genomic hybridization analysis identifies gains of 1p35 approximately p36 and chromosome 19 in osteosarcoma. *Cancer Genet Cytogenet*, **130**, 14-21.
- Zink, D. and Cremer, T. (1998) Cell nucleus: chromosome dynamics in nuclei of living cells. *Curr Biol*, **8**, R321-324.
- Zirbel, R.M., Mathieu, U.R., Kurz, A., Cremer, T. and Lichter, P. (1993) Evidence for a nuclear compartment of transcription and splicing located at chromosome domain boundaries. *Chromosome Res*, **1**, 93-106.

# 7. Anhang

## 7.1 Präsentationen und Publikationen

Winterpacht A, <u>Mohrmann G</u>, Schlaubitz S, Stelzer C, et al. (2001). Molecular identification of genes and pathways involved in skeletogenesis – potential candidates for human disorders. 10th International Congress of Human Genetics, Wien 15.-19.05. Eur J Hum Genet 9 (Suppl.), 252.

Hankeln T, Luft F, Schmidt ER, <u>Mohrmann G</u>, Tagariello A, Winterpacht A, Hermann P, Lee B, Schlaubitz S, Stelzer C, Zabel B (2001). Molecular identification of genes and pathways involved in skeletogenesis by EST sequence analysis and expression studies. German Human Genome Meeting, Braunschweig 07.-09.11.

Zabel B, Schlaubitz S, Stelzer C, Luft F, Schmidt ER, Hankeln T, Hermanns P, Lee B, Jakob F, Noeth U, <u>Mohrmann G</u>, Tagariello A, Winterpacht A (2002). Molecular identification of genes and pathways involved in skeletogenesis by EST sequence analyses and microarray expression profiling of human mesenchymal stem cell differentiation. GfH Jahrestagung 2002, Leipzig 29.09.-02.10.

Tagariello A, <u>Mohrmann G</u>, Schlaubitz S, Luft F, Schmidt ER, Hankeln T, Lee B, Winterpacht A (2003). Molecular identification of genes and pathways involved in skeletogenesis by EST sequencing analysis and expression profiling. Joint Annual Meeting of the German and Swiss Connective Tissue Society, Ulm 27.-29.03.

Patent in Beantragung:

Winterpacht A, <u>Mohrmann G</u>, Hengstler JG, Will H, Hofmann TG. K203 (SPOC1) – Ein neuer Transkriptionsregulator als "drug target" und molekularer Marker für Prognose, Therapieresistenz und Residualtumor bei Ovarialkarzinomen und anderen Tumorerkrankungen.

Publikation in Vorbereitung, aus patentrechtlichen Gründen bisher nicht publiziert: <u>Mohrmann G</u>, Hengstler JG, Hofmann TG, Stelzer C, Lee B, Zabel BU, Will H, Winterpacht A (2003).

Enhanced expression of SPOC1, a novel nuclear PHD zinc finger protein, is a prognostic marker for ovarian carcinoma patient survival.

In Vorbereitung zur Publikation im J Clin Invest

### 7.3 Danksagung

Ich danke Herrn Prof. Winterpacht für die Bereitstellung des Themas der vorliegenden Doktorarbeit sowie für die fortwährende, engagierte Betreuung und Unterstützung und Diskussionsbereitschaft, die wesentlich zum Gelingen dieser Arbeit beigetragen haben.

Herrn Prof. Will danke ich für die engagierte Betreuung, Unterstützung und Diskussionsbereitschaft, die ebenfalls einen wesentlichen Beitrag zum Gelingen dieser Arbeit beigetragen haben.

Herrn Prof. Hengstler gilt besonderer Dank für die Durchführung der Studien zur Assoziation von SPOC1 mit Ovarialkarzinomen. Weiterhin bedanke ich mich für die engagierte Unterstützung bei der Durchsicht verschiedener Manuskripte.

Herrn Prof. Gal danke ich ebenfalls für die nachhaltige Unterstützung dieser Arbeit.

Herrn Prof. Mühlbach danke ich für die Übernahme des Zweitgutachtens und für die engagierte Unterstützung dieser Arbeit. Prof. Böttger Danke ich für die Übernahme des Vorsitzes des Prüfungsausschusses.

Herrn Dr. Thomas Hofmann und Frau Dr. Sabine Endele danke ich besonders für die konstruktiven Anregungen und die kontinuierliche Hilfe bei der Planung und Durchführung verschiedener Untersuchungen und Experimente, die für die Anfertigung dieser Arbeit von großer Wichtigkeit waren.

Ferner danke ich allen Freunden und Kollegen und meiner Familie, die mich die ganze Zeit mit viel Geduld unterstützt haben:

Hanno Bolz, Odo, Steffi Schlickum, Anja Prause, Andreas Tagariello, Cordula Steglich, Sandra Fleischer und Hannah Staege sowie Christian und Carsten, Jan, Ina, Astrid, Sverre und Roman & Steffi.

Pro Markétu