

Aus dem Institut für Humangenetik  
Theoretische Medizin und Biowissenschaften  
der Medizinischen Fakultät  
der Universität des Saarlandes, Homburg/Saar

**Identifizierung und Charakterisierung transkribierter Loci der  
humanen endogenen Retrovirus-Gruppen HERV-K(HML-2)  
und HERV-W im Kontext von Tumorerkrankungen und  
Multipler Sklerose**

*Dissertation zur Erlangung des Grades eines Doktors der Naturwissenschaften*

**der Medizinischen Fakultät**

der UNIVERSITÄT DES SAARLANDES

2014

vorgelegt von

**Katja Schmitt**

geboren am 22.03.1986 in St. Wendel



*Who are you who are so wise in the ways of science?*

(Monty Python and the Holy Grail)

## Inhaltsverzeichnis

Abkürzungsverzeichnis .....	1
Abbildungsverzeichnis.....	4
Tabellenverzeichnis.....	6
<b>1. Zusammenfassung.....</b>	<b>7</b>
<b>Summary .....</b>	<b>10</b>
<b>2. Einleitung.....</b>	<b>12</b>
2.1. Repetitive DNA und Transposition.....	12
2.2. Humane endogene Retroviren (HERVs) .....	15
2.2.1. Struktur humaner endogener Retroviren .....	18
2.2.2. Promotoraktivität retroviraler LTRs und Regulation benachbarter Sequenzen .....	19
2.2.3. Transkription humaner endogener Retroviren.....	21
2.3. Die mögliche Involvierung von HERVs in der Krankheitspathogenese .....	23
2.3.1. Die HERV-K-Supergruppe humaner endogener Retroviren.....	24
2.3.1.1. Die Gruppe HERV-K(HML-2) .....	25
2.3.1.2. Die Rolle von Retroviren in der Pathogenese von Krebs .....	28
2.3.1.3. Die potentielle Rolle von HERVs in der Krebs-Pathogenese .....	30
2.3.1.4. Die mögliche Rolle der HERV-K(HML-2)-Gruppe in der Tumorphagenese .....	31
2.3.1.5. Die potentielle Rolle der HML-2-Proteine Rec und Np9 in der Tumorgenese.....	32
2.3.2. Die HERV-W-Gruppe humaner endogener Retroviren .....	33
2.3.2.1. Die potentielle Rolle von HERV-W in der Pathogenese von Multipler Sklerose.....	34
2.4. Zielsetzungen der Arbeit.....	36
<b>3. Material und Methoden .....</b>	<b>38</b>
3.1. Material .....	38
3.1.1. Chemikalien .....	38
3.1.2. Lösungen und Puffer .....	40
3.1.3. Kulturmedien .....	42
3.1.4. Reaktions-Kits .....	43
3.1.5. Enzyme.....	44
3.1.6. Antikörper .....	44
3.1.7. Größenstandards für die Agarose- und Polyacrylamid-Gelelektrophorese .....	44
3.1.8. Gewebeproben .....	45
3.1.9. RNA-Proben .....	46
3.1.10. Zellen.....	47

3.1.11. Vektoren .....	48
3.1.12. Oligonukleotid- <i>Primer</i> .....	51
3.1.13. Geräte und Verbrauchsmaterialien.....	55
3.1.14. <i>Software</i> , <i>Online-Tools</i> und Datenbanken.....	57
3.2. Methoden.....	59
3.2.1. Molekularbiologische Methoden .....	59
3.2.1.1. Isolation von Nucleinsäuren .....	59
3.2.1.2. Konzentrationsbestimmung von Nucleinsäuren und Bakterienkulturen .....	61
3.2.1.3. DNase-Behandlung und Synthese von cDNA.....	61
3.2.1.4. Polymerase-Kettenreaktion (PCR) .....	63
3.2.1.4.1. PCR-Amplifikation aus genomischer DNA .....	63
3.2.1.4.2. PCR-Amplifikation aus cDNA .....	66
3.2.1.4.3. Kolonie-PCR.....	67
3.2.1.4.4. Kontroll-PCRs.....	67
3.2.1.4.5. Quantitative <i>Real-Time</i> -PCR.....	68
3.2.1.5. Auftrennung und Aufreinigung von Nucleinsäuren .....	70
3.2.1.6. Ligation und Transformation.....	71
3.2.1.7. Restriktionsverdau von Plasmid-DNA .....	73
3.2.1.8. DNA-Sequenzierung.....	73
3.2.1.8.1. Sanger-Sequenzierung von DNA.....	73
3.2.1.8.2. <i>Next Generation Sequencing</i> (NGS) von DNA.....	74
3.2.1.9. Zellkultur .....	77
3.2.1.10. Transfektion eukaryotischer Zellen.....	78
3.2.1.11. Reporter-Gen-Assays.....	80
3.2.2. Proteinbiochemische Methoden .....	81
3.2.2.1. Proteinisolation und -konzentrationsbestimmung.....	81
3.2.2.2. SDS-Polyacrylamid-Gelelektrophorese (SDS-PAGE).....	82
3.2.2.3. <i>Western Blot</i> .....	83
3.2.2.4. Immunodetektion von auf eine PVDF-Membran übertragenen Proteinen.....	83
3.2.2.5. Fluoreszenzmikroskopie.....	84
3.2.3. Bioinformatische Arbeiten.....	85
3.2.3.1. Sequenzzusammenstellung und <i>Primer</i> -Optimierung.....	85
3.2.3.2. Sequenzrückkartierung und Auszählung generierter HERV cDNA-Sequenzen.....	86
3.2.3.2.1. Sequenzrückkartierung nach Sanger-Sequenzierung.....	86
3.2.3.2.2. Rückkartierung von Sequenzen aus NGS-Experimenten .....	87
3.2.3.2.3. Locus-Auszählung und Berechnung relativer Klonierungsfrequenzen.....	88
3.2.4. Statistik .....	89

<b>4. Ergebnisse</b> .....	<b>90</b>
4.1. Charakterisierung der Transkription von HERV-K(HML-2)-Loci in Tumor- und Normalgeweben .....	90
4.1.1. Arbeitsablauf zur Identifizierung transkribierter HERV-K(HML-2)-Loci, deren näherer Charakterisierung und Quantifizierung ihrer Transkriptmengen.....	91
4.1.2. Optimierung von PCR- <i>Primern</i> und Amplifikationsstrategie zur Identifizierung transkribierter HERV-K(HML-2)-Loci.....	93
4.1.3. Test der PCR- <i>Primer</i> zur Amplifikation von HERV-K(HML-2)-Loci sowie Strategie zur Rückkartierung generierter Sequenzen.....	96
4.1.4. Identifizierung transkribierter HERV-K(HML-2)-Loci in Melanom-Proben und Melanozyten-Zelllinien .....	98
4.1.5. Reproduzierbarkeit der cDNA-Transkriptionsprofile in Melanom-Proben.....	103
4.1.6. Struktur und Kodierungskapazität der in Melanomen und Melanozyten-Zelllinien als transkribiert identifizierten HML-2-Loci.....	104
4.1.7. Einfluss von UVB-Strahlung auf die HERV-K(HML-2) <i>gag</i> -Transkription in Melanom- und Melanozyten-Zelllinien.....	107
4.1.8. Unterschiedliche HERV-K(HML-2)-Transkriptionsprofile sind nicht durch Unterschiede im Allelstatus polymorpher HERV-K(HML-2)-Loci begründet .....	109
4.1.9. Eine Quantifizierung mittels <i>Real-Time</i> -PCR zeigte erhöhte HERV-K(HML-2) <i>gag</i> -Transkriptmengen in einigen Melanom- und Keimzelltumor-Proben .....	111
4.1.10. Identifizierung <i>rec</i> und <i>np9</i> mRNA-produzierender HML-2-Loci in Melanomen und Melanozyten-Zelllinien .....	113
4.1.11. Charakterisierung der Struktur zweier durch reverse Transkription und Reintegration von gespleißter mRNA entstandener HML-2-Loci .....	115
4.1.12. Transkriptionsanalyse der durch reverse Transkription und Reintegration gespleißter mRNA entstandenen HML-2-Loci auf den Chromosomen 2q32.1 und 10q24.2 .....	120
4.1.13. Nachweis von GFP-Fusionsproteinen der von den HML-2-Loci auf den Chromosomen 2q32.1 und 10q24.2 kodierten Proteine in transient transfizierten Zellen .....	121
4.1.14. Identifizierung transkribierter HERV-K(HML-2)-Loci in Keimzelltumor- und Prostatakarzinom-Proben .....	125
4.1.15. HERV-K(HML-2) <i>rec/np9</i> -Transkription in Normalgeweben und Identifizierung transkribierter, potentiell bisher unbekannter HML-2-Loci .....	127
4.1.16. Charakterisierung eines weiteren durch Retrotransposition gespleißter mRNA entstandenen HML-2-Locus auf Chromosom 5q15 .....	131
4.1.17. HML-2-Sequenzvarianten konnten auch auf genomischer Ebene verifiziert werden .....	133
4.2. Charakterisierung der HERV-W-Transkription im Kontext von Multipler Sklerose .....	134
4.2.1. Arbeitsablauf zur Identifizierung transkribierter HERV-W-Loci, deren näherer Charakterisierung und Quantifizierung ihrer Transkriptmengen .....	134

4.2.2. Amplifikationsstrategie zur Identifizierung transkribierter HERV-W-Loci sowie Optimierung und Test der zur Amplifikation gewählten PCR-Primer .....	137
4.2.3. Generierung von HERV-W cDNA-Sequenzen mittels 454/FLX-Sequenzierung und deren Rückkartierung im menschlichen Referenzgenom .....	139
4.2.4. Identifizierung transkribierter HERV-W-Loci in Hirnproben von MS-Patienten und gesunden Kontrollen mittels 454/FLX-Sequenzierung .....	141
4.2.5. Die HERV-W-Transkriptionsmuster in Hirnproben von MS-Patienten und gesunden Kontrollen unterscheiden sich nicht signifikant .....	144
4.2.6. Durch <i>MiSeq/Illumina</i> -Sequenzierung konnten HERV-W-Transkriptionsmuster in MS-Proben und gesunden Kontrollen bestätigt und weitere transkribierte Loci identifiziert werden .....	147
4.2.7. Struktur und Genomlokalisierung der transkribierten HERV-W-Loci .....	149
4.2.8. Identifizierung transkribierter HERV-W-Loci in <i>ENCODE RNA-Seq</i> -Datensätzen .....	152
4.2.9. Die Quantifizierung der HERV-W-Transkriptmengen lieferte keine spezifischen Transkriptionsunterschiede zwischen MS-Proben und gesunden Kontrollen .....	153
4.2.10. Untersuchung der Promotoraktivität von LTR-Elementen diverser HERV-W-Loci .....	155
<b>5. Diskussion .....</b>	<b>158</b>
5.1. Etablierung einer optimierten Amplifikationsstrategie zur Identifizierung transkribierter HERV-Loci .....	160
5.1.1. Rückkartierung generierter cDNA-Sequenzen anhand charakteristischer Nukleotidunterschiede zwischen genomischen HERV-Loci .....	162
5.1.2. Die mögliche Bedeutung von <i>Antisense</i> -Transkripten .....	165
5.2. Transkriptionsprofile der HERV-K(HML-2)-Gruppe und weitere Charakterisierung transkribierter HML-2-Loci in Krebs und Normalgeweben .....	166
5.2.1. Melanom-Proben und Melanozyten-Zelllinien zeigen interindividuell unterschiedliche Muster transkribierter HML-2-Loci .....	166
5.2.2. Unter den in Melanom-Proben und Melanozyten-Zelllinien transkribierten HML-2-Loci finden sich strukturell intakte Proviren mit vollständigen LTRs sowie ORFs für retrovirale Gag- und Env-Proteine .....	168
5.2.3. Die HML-2 <i>gag</i> -Transkription unterscheidet sich zwischen Keimzelltumor- und Prostatakarzinom-Zelllinien sowie von der Transkription in Melanom-Proben und Melanozyten-Zelllinien .....	170
5.2.4. Einige Melanom-Proben und beide Keimzelltumor-Zelllinien zeigen, verglichen mit Melanozyten, eine erhöhte HML-2 <i>gag</i> -Transkription .....	172
5.2.5. Die Bestrahlung von Melanom- und Melanozyten-Zelllinien mit UVB resultierte in veränderten Transkriptionsprofilen sowie verringerten HML-2 <i>gag</i> -Transkriptmengen .....	172
5.2.6. Transkripte <i>rec</i> und <i>np9</i> mRNA-produzierender HML-2-Loci finden sich in Melanom- und Melanozyten-Zelllinien sowie menschlichen Normalgeweben .....	173
5.2.7. Identifizierung von Transkripten potentieller bisher unbekannter HML-2-Loci .....	174

5.2.8. Unter den transkribierten HML-2-Loci finden sich stark verkürzte retrovirale Loci, die durch reverse Transkription und Reintegration gespleißter mRNA entstanden sind und potentiell für Rec- und Env-ähnliche Proteine kodieren.....	176
5.3. Transkriptionsprofile der HERV-W-Gruppe und Charakterisierung transkribierter Loci in Hirnproben von Multiple Sklerose Patienten und gesunden Kontrollen .....	178
5.3.1. Die cDNA-Sequenzierung mittels <i>Next Generation Sequencing</i> erlaubt die Generierung umfassender Datensätze und die Detektion einer großen Anzahl auch schwach transkribierter Loci .....	179
5.3.2. In Hirnproben von MS-Patienten und gesunden Kontrollen findet sich eine hohen Anzahl transkribierter HERV-W-Loci .....	180
5.3.3. Die HERV-W <i>env</i> -Transkriptionsmuster in MS und gesunden Vergleichsgeweben unterscheiden sich nicht signifikant .....	182
5.3.4. Die Quantifizierung der HERV-W-Transkriptmengen liefert kein einheitliches Bild für MS- und Kontroll-Proben und deutet auf interindividuelle Unterschiede hin .....	183
5.3.5. HERV-W-LTRs zeigen trotz zahlreicher Deletionen oft noch Promotoraktivität und könnten die Transkription der entsprechenden Loci induzieren .....	185
5.3.6. Zuvor publizierte MSRV-Sequenzen können auf Rekombinationen transkribierter HERV-W-Loci zurückzuführen sein.....	186
<b>6. Literaturverzeichnis .....</b>	<b>188</b>
<b>7. Anhang.....</b>	<b>200</b>
<b>8. Publikationen.....</b>	<b>219</b>
<b>9. Danksagung.....</b>	<b>220</b>

## Abkürzungsverzeichnis

°C	Grad Celsius
A	Ampere
as	Aminosäuren
ATL	adulte T-Zell-Leukämie
bp	Basenpaare
BLAST	<i>Basic Local Alignment Search Tool</i>
BLAT	<i>BLAST-like Alignment Tool</i>
bzw.	beziehungsweise
c	Zenti ( $10^{-2}$ )
CCD	<i>charge-coupled device</i>
cDNA	<i>complementary DNA</i>
CMV	Cytomegalovirus
Da	Dalton
DEPC	Diethyldicarbonat
d.h.	das heißt
DMEM	<i>Dulbecco's Modified Eagle Medium</i>
DNA	Desoxyribonukleinsäure ( <i>deoxyribonucleic acid</i> )
dNTP	Desoxynukleosidtriphosphat
ddNTP	Didesoxynukleosidtriphosphat
dUTP	Desoxyuridintriphosphat
DPBS	<i>Dulbecco's Phosphate-Buffered Saline</i>
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylendiamintetraacetat
ENCODE	<i>Encyclopedia of DNA Elements</i>
<i>env</i>	<i>envelope</i>
ERV	endogenes Retrovirus
<i>et al.</i>	und andere ( <i>et alii</i> )
FCS	fetales Kälberserum ( <i>fetal calf serum</i> )
FeLV	Felines Leukosevirus ( <i>feline leukemia virus</i> )
g	Erdbeschleunigung
g	Gramm
<i>gag</i>	Gruppenspezifisches Antigen ( <i>group specific antigen</i> )
GCT	Keimzellmor ( <i>germ cell tumour</i> )
GFP	Grün fluoreszierendes Protein ( <i>green fluorescent protein</i> )
ggf.	gegebenenfalls
GOGE	<i>Genetics of Gene Expression</i>
GREM	<i>Genomic Repeat Expression Monitor</i>
h	Stunde(n)
H <sub>2</sub> O <sub>dd</sub>	zweifach deionisiertes Wasser
HDTV	<i>human teratocarcinoma-derived virus</i>
HERV	humanes endogenes Retrovirus
HGNC	<i>HUGO Gene Nomenclature Committee</i>
HIV-1	Humanes Immundefizienz-Virus 1 ( <i>human immunodeficiency virus 1</i> )
HML	<i>human MMTV-like</i>
HRP	Meerrettich-Peroxidase ( <i>horseradish peroxidase</i> )
HTLV-1	Humanes T-lymphotropes Virus 1 ( <i>human T cell leukemia virus 1</i> )
IgG	Immunglobulin G
ISD	immunsuppressive Domäne
ITR	<i>inverted terminal repeat</i>

JSRV	Jaagsiekte-Schaf-Retrovirus ( <i>Jaagsiekte sheep retrovirus</i> )
k	Kilo ( $10^3$ )
kb	Kilobase(n)
l	Liter
L1	LINE-1
LB	<i>Lysogeny Broth</i>
LINE	<i>long interspersed element</i>
LTR	<i>long terminal repeat</i>
$\mu$	Mikro ( $10^{-6}$ )
m	Milli ( $10^{-3}$ )
m	Meter
M	molar (mol/l)
MCS	<i>multiple cloning site</i>
min	Minute(n)
Mio.	Millionen
MLV	Murines Leukämievirus ( <i>murine leukemia virus</i> )
MMTV	Maus-Mammatumovirus ( <i>mouse mammary tumor virus</i> )
mol	Stoffmenge
MOPS	3-(N-Morpholino)-Propansulfonsäure
mRNA	<i>messenger-RNA</i>
MS	Multiple Sklerose
MSRV	<i>MS associated retrovirus</i>
n	Nano ( $10^{-9}$ )
NGS	<i>Next Generation Sequencing</i>
nt	Nukleotid(e)
OD	optische Dichte
ORF	offener Leserahmen ( <i>open reading frame</i> )
PAGE	Polyacrylamid-Gelelektrophorese
PBMC	<i>peripheral blood mononuclear cell</i>
PBS	Phosphatgepufferte Salzlösung ( <i>phosphate buffered saline</i> )
PBS	<i>primer binding site</i>
PCR	Polymerase-Kettenreaktion ( <i>polymerase chain reaction</i> )
pH	potential Hydrogenii
<i>pol</i>	Polymerase
<i>pro</i>	Protease
PVDF	Polyvinylidenfluorid
RIPA	<i>radioimmunoprecipitation assay</i>
RNA	Ribonukleinsäure ( <i>ribonucleic acid</i> )
rpm	Umdrehungen pro Minute ( <i>revolutions per minute</i> )
RT	Reverse Transkriptase
RVLPs	retrovirusähnliche Partikel ( <i>retrovirus-like particles</i> )
s.u.	siehe unten
s	Sekunde(n)
SA	<i>Splice Acceptor</i>
SD	<i>Splice Donor</i>
SDS	Natriumdodecylsulfat
SINE	<i>short interspersed element</i>
SLE	Lupus erythematodes ( <i>systemic lupus erythematosus</i> )
SOB	<i>Super Optimal Broth</i>
SPRI	<i>Solid Phase Reversible Immobilization</i>
SVA	SINE-VNTR-Alu
TAE	TRIS-Acetat-EDTA

Taq	<i>Thermus aquaticus</i>
TBS	<i>Tris-Buffered Saline</i>
TE	TRIS-EDTA
TEMED	N,N,N',N'-Tetramethylethyldiamin
TfB	Transformations-Puffer ( <i>transformation buffer</i> )
TPRT	<i>target primed reverse transcription</i>
TRIS	Tris(hydroxymethyl)-aminomethan
tRNA	<i>transfer-RNA</i>
TSD	<i>target site duplication</i>
UCSC	University of California, Santa Cruz
u.a.	unter anderem
ü.N.	über Nacht
u.U.	unter Umständen
U	Unit
UTR	untranslatierte Region
UV	Ultraviolett
V	Volt
VNTR	<i>variable number of tandem repeats</i>
v/v	Volumenprozent ( <i>volume per volume</i> )
w/v	Gewichtsprozent ( <i>weight per volume</i> )
X-Gal	5-Brom-4-chlor-3-indoxyl- $\beta$ -D-galactopyranosid
z.B.	zum Beispiel
ZNS	zentrales Nervensystem

### Nukleinbasen

A	Adenin
C	Cytosin
G	Guanin
T	Thymin
U	Uracil

### Einbuchstabencode der Aminosäuren

A	Alanin	I	Isoleucin	R	Arginin
C	Cystein	K	Lysin	S	Serin
D	Asparaginsäure	L	Leucin	T	Threonin
E	Glutaminsäure	M	Methionin	V	Valin
F	Phenylalanin	N	Asparagin	W	Tryptophan
G	Glycin	P	Prolin	Y	Tyrosin
H	Histidin	Q	Glutamin		

## Abbildungsverzeichnis

<b>Abbildung 1:</b>	Klassifizierung transposabler Elemente und deren allgemeine Struktur .....	13
<b>Abbildung 2:</b>	Prinzip der <i>target primed reverse transcription</i> (TPRT) am Beispiel einer L1-Sequenz ...	15
<b>Abbildung 3:</b>	Struktur eines HERV-Provirus am Beispiel von HERV-K(HML-2) .....	18
<b>Abbildung 4:</b>	Struktur und Spleißvarianten verschiedener Provirus-Typen der HERV-K(HML-2)-Gruppe .....	26
<b>Abbildung 5:</b>	Für die Gelelektrophorese verwendete Größenstandards .....	45
<b>Abbildung 6:</b>	Für die SDS-Polyacrylamid-Gelelektrophorese verwendeter Größenstandard .....	45
<b>Abbildung 7:</b>	Karte des <i>pGEM<sup>®</sup>-T Easy</i> -Vektors der Firma Promega .....	48
<b>Abbildung 8:</b>	Karten der Vektoren <i>pGLuc-Basic</i> und <i>pCMV-GLuc</i> der Firma NEB .....	49
<b>Abbildung 9:</b>	Karte des <i>pEGFP-N1</i> -Vektors der Firma Clontech Laboratories/Takara Bio .....	50
<b>Abbildung 10:</b>	Karte des <i>pCMV<math>\beta</math></i> -Vektors der Firma Clontech Laboratories/Takara Bio .....	51
<b>Abbildung 11:</b>	Prinzip der <i>MiSeq/Illumina</i> -Sequenzierung .....	76
<b>Abbildung 12:</b>	Arbeitsablauf zur Identifizierung transkribierter HERV-K(HML-2)-Loci in Melanom-Proben, Melanozyten und Normalgeweben sowie zur Quantifizierung der HML-2-Transkriptmengen .....	92
<b>Abbildung 13:</b>	Arbeitsablauf zu Untersuchung der Translation von Proteinen, die von zwei durch reverse Transkription und Reintegration gespleißter mRNA entstandenen HML-2-Loci auf den Chromosomen 2q32.1 und 10q24.2 kodiert werden.....	93
<b>Abbildung 14:</b>	Position des untersuchten Amplikons im HERV-K(HML-2) <i>gag</i> -ORF .....	94
<b>Abbildung 15:</b>	Graphische Darstellung eines Sequenzvergleiches des untersuchten HERV-K(HML-2) <i>gag</i> -Amplikons in allen theoretisch amplifizierbaren HML-2-Loci .....	95
<b>Abbildung 16:</b>	Relative Klonierungsfrequenzen von aus genomischer DNA amplifizierten HERV-K(HML-2)-Loci .....	98
<b>Abbildung 17:</b>	Sequenzunterschiede der aus Melanom- und Melanozyten-cDNA generierten HERV-K(HML-2) <i>gag</i> cDNA-Sequenzen zu ihrem jeweils besten <i>BLAT-Match</i> .....	99
<b>Abbildung 18:</b>	Transkriptionsprofile verschiedener HERV-K(HML-2)-Loci in Melanomen und Melanozyten-Zelllinien .....	102
<b>Abbildung 19:</b>	Relative Klonierungsfrequenzen in unterschiedlichen Passagen zweier Melanom-Zelllinien als transkribiert identifizierter HERV-K(HML-2)-Loci .....	104
<b>Abbildung 20:</b>	Anzahl an Nukleotidunterschieden der nach UVB-Bestrahlung aus Melanom- und Melanozyten-Zelllinien generierten HERV-K(HML-2) <i>gag</i> cDNA-Sequenzen zu ihrem jeweils besten <i>BLAT-Match</i> .....	108
<b>Abbildung 21:</b>	Einfluss von UVB-Strahlung auf die relativen Transkriptionsfrequenzen von HERV-K(HML-2)-Loci .....	109
<b>Abbildung 22:</b>	Quantifizierung der HERV-K(HML-2) <i>gag</i> -Transkriptmengen mittels <i>Real-Time-PCR</i> ...	112
<b>Abbildung 23:</b>	Position des untersuchten Amplikons zur Identifizierung transkribierter <i>rec/np9</i> mRNA-produzierender HERV-K(HML-2)-Loci .....	113
<b>Abbildung 24:</b>	Struktur eines durch L1-Retrotransposition entstandenen HERV-K(HML-2)-Locus auf Chromosom 2q32.1., der in der Melanozyten-Zelllinie Benno als transkribiert identifiziert wurde .....	116

<b>Abbildung 25:</b> Multiples <i>Alignment</i> von Aminosäuresequenzen der hypothetischen Proteine, die von den beiden durch reverse Transkription und Reintegration gespleißter mRNA entstandenen HML-2-Loci auf den Chromosomen 2q32.1 und 10q24.2 kodiert werden .....	117
<b>Abbildung 26:</b> Struktur eines durch reverse Transkription und Reintegration gespleißter mRNA entstandenen HML-2-Locus auf Chromosom 10q24.2, der in der Melanozyten-Zelllinie Benno als transkribiert identifiziert wurde .....	119
<b>Abbildung 27:</b> Ergebnis der gelelektrophoretischen Auftrennung spezifischer RT-PCR-Produkte zweier durch reverse Transkription und Reintegration gespleißter mRNA gebildeter HML-2-Loci auf den Chromosomen 2q32.1 und 10q24.2 .....	121
<b>Abbildung 28:</b> Nachweis von HERV-K(HML-2) GFP-Fusionsproteinen in Zelllysaten transient transfizierter Zellen mittels <i>Western Blot</i> .....	123
<b>Abbildung 29:</b> Fluoreszenzmikroskopischer Nachweis der Expression von HERV-K(HML-2) GFP-Fusionsproteinen .....	124
<b>Abbildung 30:</b> Transkriptionsprofile der in Keimzelltumor- und Prostatakarzinom-Zelllinien als transkribiert identifizierten HERV-K(HML-2)-Loci .....	126
<b>Abbildung 31:</b> Aus cDNA gesunder Gewebe generierte, deutlich variante <i>rec/np9</i> -Sequenzen zweier potentieller neuer HML-2-Loci .....	129
<b>Abbildung 32:</b> Struktur eines weiteren durch Retrotransposition entstandenen HML-2-Locus auf Chromosom 5q15, der in verschiedenen menschlichen Normalgeweben als transkribiert identifiziert wurde .....	132
<b>Abbildung 33:</b> Arbeitsablauf zur Identifizierung transkribierter HERV-W-Loci in Hirnproben von MS-Patienten und gesunden Kontrollen sowie zur Quantifizierung der HERV-W-Transkriptmengen .....	135
<b>Abbildung 34:</b> Arbeitsablauf zur Untersuchung der potentiellen Promotoraktivität spezifischer HERV-W 5'- und 3'-LTR-Elemente .....	136
<b>Abbildung 35:</b> Relative Klonierungsfrequenzen aus genomischer DNA amplifizierter HERV-W-Loci ...	139
<b>Abbildung 36:</b> Sequenzunterschiede der aus Hirnproben von MS-Patienten und gesunden Kontrollen generierten HERV-W 5' <i>env</i> und 3' <i>env</i> cDNA-Sequenzen zu ihrem jeweils besten <i>BLAT-Match</i> .....	140
<b>Abbildung 37:</b> Relative cDNA-Frequenzen der mit dem 5' <i>env</i> -Amplikon in Hirnproben von MS-Patienten und gesunden Kontrollen als transkribiert identifizierten HERV-W-Loci .....	145
<b>Abbildung 38:</b> Relative cDNA-Frequenzen der mit dem 3' <i>env</i> -Amplikon in Hirnproben von MS-Patienten und gesunden Kontrollen als transkribiert identifizierten HERV-W-Loci .....	146
<b>Abbildung 39:</b> Relative cDNA-Frequenzen des aus Hirnproben von MS-Patienten und gesunden Kontrollen amplifizierten HERV-W 5' <i>env</i> -Amplikons nach 454/ <i>FLX</i> - sowie <i>MiSeq/Illumina</i> -Sequenzierung .....	149
<b>Abbildung 40:</b> Struktur der in Hirnproben von MS-Patienten und gesunden Kontrollen als transkribiert identifizierten HERV-W-Loci .....	151
<b>Abbildung 41:</b> Quantifizierung der HERV-W-Transkriptmengen mittels <i>Real-Time-PCR</i> .....	154
<b>Abbildung 42:</b> Promotoraktivität von 5'- und 3'-LTRs von als transkribiert identifizierten HERV-W-Loci .....	156

## Tabellenverzeichnis

<b>Tabelle 1:</b>	Zur Analyse der HERV-W-Transkription im Kontext von Multipler Sklerose verwendete <i>post mortem</i> Hirnproben .....	46
<b>Tabelle 2:</b>	Verwendete eukaryotische Zelllinien mit ihren Kulturmedien und Zusätzen .....	47
<b>Tabelle 3:</b>	In <i>pGLuc-Basic</i> klonierte HERV-W LTR-Sequenzen, deren potentielle Promotoraktivität mittels Luciferase-Assay bestimmt wurde .....	50
<b>Tabelle 4:</b>	<i>Primer</i> zur Amplifikation von HERV-K(HML-2)-Loci aus genomischer DNA und cDNA .....	52
<b>Tabelle 5:</b>	<i>Primer</i> zur Amplifikation von HERV-W-Loci aus genomischer DNA sowie aus cDNA zur anschließenden Sequenzierung mittels <i>MiSeq/Illumina</i> -Technologie .....	52
<b>Tabelle 6:</b>	<i>Adaptor</i> - und <i>Key</i> -Sequenzen zur Amplifikation von HERV-W-Loci aus cDNA zur anschließenden Sequenzierung mittels <i>454/FLX</i> -Technologie .....	53
<b>Tabelle 7:</b>	<i>Primer</i> zur Bestimmung des Allelstatus polymorpher HERV-K(HML-2)-Proviren in genomischer DNA .....	53
<b>Tabelle 8:</b>	<i>Primer</i> zur spezifischen Amplifikation zweier durch reverse Transkription und Reintegration gespleißter mRNA entstandener HERV-K(HML-2)-Loci .....	54
<b>Tabelle 9:</b>	<i>Primer</i> zur spezifischen Amplifikation der ORFs zweier durch reverse Transkription und Reintegration gespleißter mRNA entstandener HERV-K(HML-2)-Loci .....	54
<b>Tabelle 10:</b>	<i>Nested-PCR-Primer</i> zur spezifischen Amplifikation zweier potentieller neuer HERV-K(HML-2)-Loci .....	54
<b>Tabelle 11:</b>	Primer zur Quantifizierung der HERV-W-Transkriptmengen sowie der Transkriptmengen der Haushaltsgene <i>G6PDH</i> und <i>RPII</i> mittels <i>Real-Time-PCR</i> .....	55
<b>Tabelle 12:</b>	Sonstige <i>Primer</i> .....	55
<b>Tabelle 13:</b>	Erwartete PCR-Produktgrößen im Zuge der Überprüfung des Allelstatus polymorpher HERV-K(HML-2)-Loci in genomischer DNA .....	65
<b>Tabelle 14:</b>	Zur Immunodetektion von Proteinen auf PVDF-Membranen verwendete Primär- und Sekundärantikörper sowie verwendete <i>Blocking</i> - und Waschlösungen .....	84
<b>Tabelle 15:</b>	Relative Klonierungsfrequenzen der in Melanom-Proben und Melanozyten als transkribiert identifizierten HERV-K(HML-2)-Loci .....	100
<b>Tabelle 16:</b>	Lokalisation, Struktur und Kodierungskapazität der in Melanom-Proben und Melanozyten-Zelllinien als transkribiert identifizierten HERV-K(HML-2)-Loci .....	104
<b>Tabelle 17:</b>	Allelstatus polymorpher HERV-K(HML-2)-Proviren in den in der Transkriptionsanalyse untersuchten Melanom- und Melanozyten-Zelllinien .....	111
<b>Tabelle 18:</b>	In Melanom- und Melanozyten-Zelllinien als transkribiert identifizierte <i>rec</i> oder <i>np9</i> mRNA-produzierende HERV-K(HML-2)-Loci .....	114
<b>Tabelle 19:</b>	Absolute Anzahl an cDNA-Sequenzen der in 16 menschlichen Normalgeweben als transkribiert identifizierten <i>rec</i> oder <i>np9</i> mRNA-produzierenden HML-2-Loci .....	130
<b>Tabelle 20:</b>	cDNA-Frequenzen und Struktur der in Hirnproben von MS-Patienten und gesunden Kontrollen mit dem 5' <i>env</i> -Amplikon als transkribiert identifizierten HERV-W-Loci .....	142

## 1. Zusammenfassung

Etwa 8 % des menschlichen Genoms bilden Sequenzen humaner endogener Retroviren (HERVs) und von HERVs abgeleitete Elemente. Zahlreiche Studien beschrieben retrovirale Transkripte und Proteine sowie eine deregulierte Expression unterschiedlicher HERV-Gruppen im Kontext verschiedener Krankheiten und diskutierten daher eine mögliche Involvierung von HERVs. Insbesondere wird hierbei eine potentielle Rolle der HERV-K(HML-2)-Gruppe, kurz HML-2, in verschiedenen Tumorerkrankungen diskutiert sowie eine mögliche Involvierung der HERV-W-Gruppe in der Pathogenese von Multipler Sklerose (MS).

Als Beitrag zum besseren Verständnis ihrer potentiellen Rolle in der Krankheitspathogenese wurden in der vorliegenden Arbeit umfassende Analysen transkribierter HERV-Loci der Gruppen HML-2 und HERV-W in verschiedenen Tumor- und MS-Proben sowie normalen Vergleichsgewebe durchgeführt. Transkribierte Loci wurden durch Generierung HERV-Gruppen-spezifischer RT-PCR-Produkte identifiziert, die sequenziert und deren Sequenzen im Anschluss anhand charakteristischer Nukleotidunterschiede im menschlichen Referenzgenom rückkartiert wurden. Es wurden umfassende Transkriptionsanalysen für alle Proben durchgeführt, potentielle Unterschiede der Transkriptionsprofile zwischen den Proben analysiert und transkribierte Loci wurden hinsichtlich ihrer Struktur, ihrer Kodierungskapazität für retrovirale Proteine sowie ihrer Genomlokalisierung näher charakterisiert.

In verschiedenen Tumor- sowie Normalgeweben wurden insgesamt über 25 verschiedene Loci der HML-2-Gruppe als transkribiert identifiziert. Alle Proben zeigten hierbei ein unterschiedliches, charakteristisches Transkriptionsprofil spezifischer HML-2-Loci. Melanom-, Keimzelltumor- und Prostatakarzinom-Proben unterschieden sich untereinander in der Transkription einiger charakteristischer Loci, die eine alleinige oder dominante Transkription in der entsprechenden Erkrankung zeigten und u.U. spezifische Marker repräsentieren könnten. Durch Quantifizierung der HML-2 *gag*-Transkription wurden in einigen Melanom-Proben sowie Keimzelltumoren erhöhte relative Transkriptmengen detektiert, die im Falle der Melanom-Proben und Melanozyten nach UVB-Bestrahlung verringert wurden. Die UVB-Bestrahlung resultierte generell in veränderten Transkriptionsprofilen aller Zelllinien, wobei kein spezifisches Muster erkennbar war, das Melanome von Melanozyten unterschied. Einige der transkribierten Loci besitzen *open reading frames* (ORFs) für retrovirale Gag- und/oder Env-Proteine, sowie für die akzessorischen Proteine Rec und Np9. Die Transkription *rec*- oder *np9*-mRNA-produzierender Loci wurde aufgrund der potentiellen Rolle der Proteine in der Tumorgenese spezifisch weiter untersucht und es wurden zahlreiche transkribierte Loci beider Provirus-Typen in Melanom-

Proben sowie auch verschiedenen menschlichen Normalgeweben identifiziert. Im Zuge der Analysen wurden auch HML-2-Loci als transkribiert identifiziert, die durch reverse Transkription und Reintegration gespleißter mRNA durch die L1-Maschinerie oder einen retroviralen Mechanismus entstanden sind, wobei zwei dieser Loci potentiell HML-2-ähnliche Proteine kodieren. Da Transkripte dieser Loci sowohl in Tumor- als auch Normalgeweben detektiert wurden, müsste die potentielle Translation der hypothetischen Proteine *in vivo* sowie deren potentielle biologische Funktion in weiterführenden Studien näher analysiert werden. Des Weiteren wurden Sequenzpopulationen generiert, die potentiell bisher unbekannte HML-2-Proviren oder allelische Varianten genomischer Loci repräsentieren.

Die Analysen deuten insgesamt auf eine komplexe Proben-spezifische Transkription von HML-2-Loci in verschiedenen Tumorerkrankungen hin und es wurden zahlreiche transkribierte Loci identifiziert, die für weiterführende Analysen von Interesse sind. Es konnten für Melanome, GCTs und Prostatakarzinome dominant transkribierte Loci identifiziert werden, deren potentielle Rolle als Biomarker in weiterführenden Studien mit größerer Probenanzahl näher untersucht werden müsste. Insbesondere sind auch die von den als transkribiert identifizierten Loci kodierten retroviralen Proteine für weitere Studien von besonderem Interesse, da potentielle Unterschiede auf Proteinebene sowie die biologische Funktion dieser Proteine im Zuge der Krankheitspathogenese von Bedeutung sein könnten.

Durch *Next Generation Sequencing* (NGS) HERV-W *env*-spezifischer RT-PCR-Produkte aus Hirnproben von MS-Patienten und gesunden Kontrollen wurden über 100 Loci der HERV-W-Gruppe als transkribiert identifiziert. Hierbei waren nur wenige Loci, was ihre relativen Transkriptmengen betrifft, dominant vertreten, darunter der *ERVWE1/ERVW-1/syncytin-1*-Locus, für dessen Genprodukt, Syncytin-1, eine mögliche Rolle in der MS-Pathogenese diskutiert wird. Insgesamt zeigten MS- und Kontrollproben sehr ähnliche Muster transkribierter Loci, was gegen eine deregulierte HERV-W *env*-Transkription in MS spricht, die aufgrund der Ergebnisse früherer Studien diskutiert wird. Dies wurde weiter durch die Quantifizierung der HERV-W *env*-Transkriptmengen sowie die spezifische Quantifizierung der Transkriptmengen der Loci *ERVWE1* und *ERVWE2* unterstützt, die keine signifikanten Unterschiede zwischen MS und Kontrollen lieferte, sondern auf interindividuelle Unterschiede in den Transkriptmengen hindeutet. In Reportergen-Assays zeigten viele HERV-W *long terminal repeats* (LTRs) Promotoraktivität, auch wenn diese, was ihre Struktur betrifft, zum Teil große Deletionen aufweisen. Dies deutet auf eine mögliche Transkriptionsinitiation vieler Loci durch ihre eigenen, wenn auch unvollständigen, LTRs hin, wobei solche LTRs auch benachbarte Genomregionen beeinflussen könnten.

Die umfassenden NGS-Analysen deuten insgesamt auf komplexe, interindividuell verschiedene Muster einer großen Zahl transkribierter HERV-W-Loci hin. Auf RNA-Ebene zeigten sich keine signifikanten Unterschiede zwischen der HERV-W *env*-Transkription in Hirnproben aus MS-Patienten und gesunden Kontrollen, die sich beide durch Transkription der gleichen dominant vertretenen Loci auszeichneten. Ähnliche Muster auf RNA-Ebene schließen jedoch potentielle Unterschiede auf Protein-Ebene sowie eine mögliche biologische Funktion der retroviralen Proteine nicht aus, was für weiterführende Studien von Interesse wäre.

## Summary

### **Identification and characterisation of transcribed loci of the human endogenous retrovirus groups HERV-K(HML-2) and HERV-W in tumours and multiple sclerosis**

Human endogenous retroviruses (HERVs) and HERV-derived sequences comprise about 8 % of the human genome. Numerous studies described retroviral transcripts and proteins, as well as a deregulated expression of certain HERV groups in various diseases and thus suggested the potential involvement of HERVs in pathogenesis. In particular, the HERV-K(HML-2) group, short HML-2, has been implicated in the pathogenesis of cancer, and a potential involvement of HERV-W in multiple sclerosis (MS) has been suggested.

To contribute to a better understanding of the potential involvement of HERVs in tumours and MS pathogenesis, extensive analyses of transcribed HML-2 and HERV-W loci in various tumour and MS samples have been performed. Transcribed loci were identified by amplification and sequencing of HERV group-specific RT-PCR products and assignment of generated cDNA sequences to genomic loci based on characteristic nucleotide differences between these loci. Comprehensive transcription analyses have been performed, potential differences between samples were analysed and transcribed loci were characterised in regard to their structure, coding-capacity for retroviral proteins as well as their genomic localisation.

More than 25 different HML-2 loci were identified to be transcribed in tumour and normal tissue. All samples were characterised by a different, distinctive transcription pattern of specific HML-2 loci. The transcription patterns in melanoma, germ cell tumours and prostate carcinoma differed, and in each disease, certain HML-2 loci showed exclusive or dominant transcription and might represent specific biomarkers. Quantification of HML-2 *gag* transcription revealed increased relative transcription levels in certain melanoma samples and germ cell tumours. Relative transcription levels in melanoma and melanocyte cell lines decreased after irradiation with UVB, which also caused alterations in HML-2 transcription patterns in all cell lines. However, no characteristic differences were observed to distinguish melanoma from melanocytes. Some transcribed HML-2 loci harbour open reading frames (ORFs) for retroviral Gag and/or Env proteins, as well as for accessory proteins Rec and Np9. Due to the possible involvement of Rec and Np9 proteins in tumorigenesis, transcription of *rec* or *np9* mRNA producing loci has been analysed. Numerous transcribed loci of both provirus types were identified in melanoma as well as various normal tissues. In the course of the transcription analyses, HML-2 loci have been identified to be transcribed that were generated by reverse transcription and reintegration of

spliced mRNA by L1-machinery or in a retroviral fashion, with two of these loci potentially encoding HML-2-like proteins. As transcripts of both loci have been detected in tumour as well as normal tissue, potential *in vivo* translation of the hypothetical proteins as well as their potential biological function will have to be analysed in further studies. In addition, sequence populations were generated that represent as yet unknown HML-2 loci or allelic variants of genomic loci.

Taken together, transcriptional analyses suggest complex, sample-specific transcription of HML-2 loci in certain tumours and numerous transcribed loci were identified which are of interest for further studies. Dominantly transcribed loci could be identified in melanoma, germ cell tumours as well as prostate carcinoma, whose potential roles as biomarkers will have to be elucidated in further studies with larger sample numbers. In particular, translation of retroviral proteins encoded by transcribed HML-2 loci are of interest, as potential differences in protein expression as well as the biological function of these proteins might be important for pathogenesis.

By *next generation sequencing* (NGS) of HERV-W *env* specific RT-PCR products from brain samples of MS patients and healthy controls, more than 100 HERV-W loci could be identified as transcribed. In terms of relative transcription frequencies, only few loci were dominantly transcribed. Amongst them was the *ERVWE1/ERVW-1/syncytin-1* locus, whose gene product, Syncytin-1, has been implicated in the pathogenesis of MS. On the whole, HERV-W transcription patterns were very similar between MS and control samples, arguing against a deregulated HERV-W *env* transcription discussed in previous studies. This was further supported by quantification of HERV-W *env* transcription, as well as the locus-specific transcription of *ERVWE1* and *ERVWE2*, which did not show significant differences in transcription levels between MS and controls, but rather suggested interindividual differences. Reporter gene assays indicate promoter activity of many HERV-W long terminal repeats (LTRs), even if structurally incomplete. This suggests a possible transcription initiation of many HERV-W loci from their own LTRs, which might also affect transcription of neighbouring genomic regions.

On the whole, extensive NGS-analyses indicate complex, interindividually different patterns of a large number of transcribed HERV-W loci. On the RNA level, no significant differences were observed in HERV-W *env* transcription between MS and control samples, which were both characterised by the same dominantly transcribed loci. Still, similar patterns on the RNA level do not exclude possible differences in protein expression, which will be of interest for further studies.

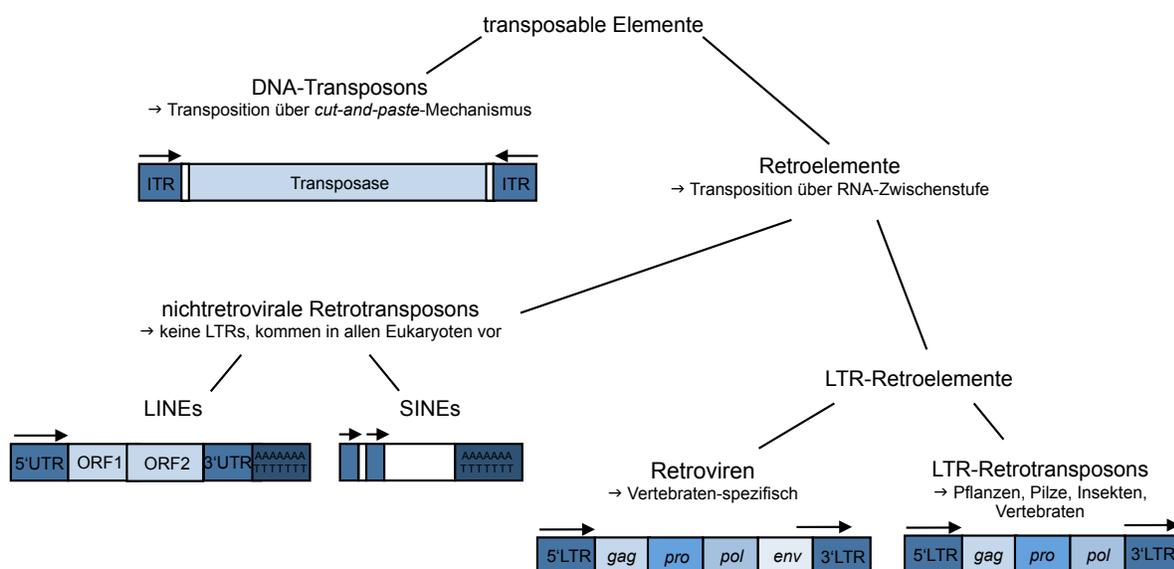
## 2. Einleitung

### 2.1. Repetitive DNA und Transposition

Das menschliche Genom besteht nur zu einem geringen Teil aus Genen, die für Proteine oder untranslatierte RNAs kodieren. Nur etwa 5 % des Genoms, darunter die etwa 1,5 % der kodierenden DNA sowie weitere untranslatierte konservierte Sequenzen und regulatorische Elemente, zeigen eine starke Konservierung. Knapp die Hälfte des menschlichen Genoms besteht hingegen aus repetitiver, nichtkodierender DNA. Diese Bereiche sind zum großen Teil auf transposable Elemente zurückzuführen, mobile DNA-Elemente, die sich nach Integration in das Genom durch Kopiermechanismen und Reintegration an anderer Stelle weiter ausbreiteten. Es wird angenommen, dass sich auch ein großer Teil der restlichen DNA von Kopien transposabler Elemente ableitet, die jedoch mittlerweile zu sehr von diesen abweichen, um noch als solche erkannt zu werden (International Human Genome Sequencing Consortium, 2001). Wurden nichtkodierende repetitive Sequenzen zunächst noch als „*junk DNA*“ abgetan, gewannen sie in den letzten Jahren immer mehr an Bedeutung für medizinische und populationsgenetische Untersuchungen. Repetitive Sequenzen liefern Hinweise auf evolutive Prozesse und dienen als Marker, um Mutations- und Selektionsprozesse zu untersuchen. Zudem haben repetitive Sequenzen durch Rekombination nicht-allelischer Sequenzen und daraus resultierender Translokationen und Umgestaltungen (Hartl, 1996), Bildung neuer Gene sowie Veränderung und Umstrukturierung existierender Gene (Takahara *et al.*, 1996) die Gestalt letztendlich des menschlichen Genoms tiefgehend geprägt (Yoder *et al.*, 1997; Makalowski *et al.*, 2007; Jurka *et al.*, 2007).

Der große Teil der repetitiven Sequenzen, der sich von transposablen Elementen ableitet, lässt sich aufgrund von Strukturähnlichkeiten in verschiedene Gruppen einteilen (vgl. **Abb. 1**). Neben den DNA-Transposons, die oft für das Enzym Transposase kodieren und die direkt oder nach Generierung einer DNA-Kopie durch einen *cut-and-paste*-Mechanismus transponiert werden, unterscheidet man zwei Gruppen sogenannter Retroelemente, bei denen die Transposition jeweils über eine RNA-Zwischenstufe vermittelt wird (Kazazian, 2004). Retroelemente werden, abhängig vom Vorhandensein flankierender *long terminal repeats* (LTRs), in die LTR-losen nichtretroviralen Retrotransposons sowie die retrovirusähnlichen Retrotransposons oder LTR-Retroelemente unterteilt. Zu den nichtretroviralen Retrotransposons zählen die Klassen der *long interspersed elements* (LINEs) und der *short interspersed elements* (SINEs). Ihre Transkription wird durch interne Promotoren stromab der Transkriptions-Startstelle reguliert. Die reverse Transkription von LINE- und SINE-Sequenzen findet im Zellkern statt, nachdem in der DNA-

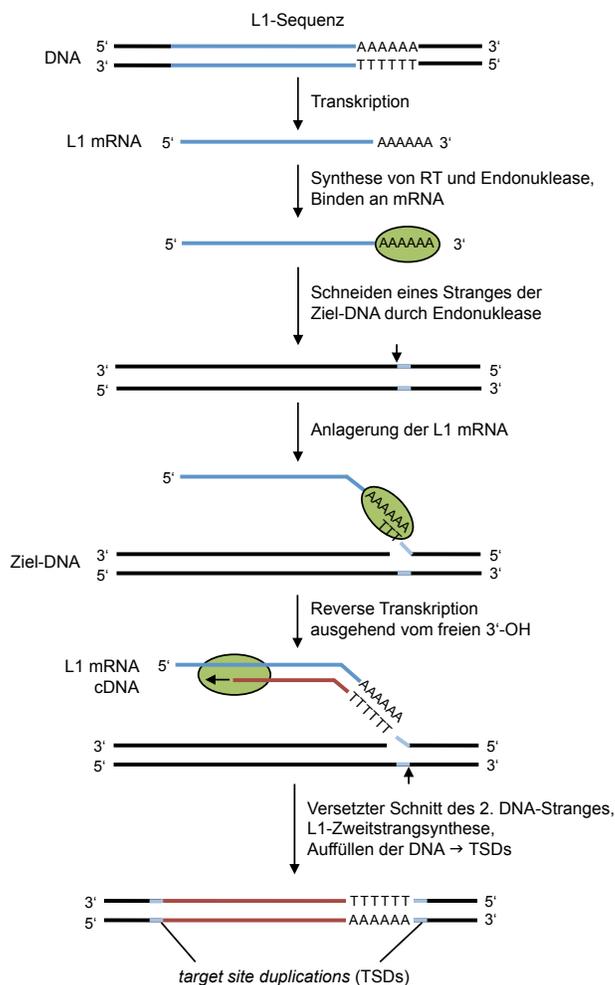
Zielsequenz durch Schneiden ein freies 3'-OH erzeugt wurde. Im Gegensatz hierzu wird die Transkription der LTR-Retroelemente durch Promotor-Sequenzen und Initiations- und Terminationssignale in den flankierenden LTRs reguliert. LTR-Retroelemente werden in einem zytosolischen virusähnlichen Partikel revers transkribiert, wobei der Prozess durch eine tRNA initiiert wird, die komplementär zur Region der *primer binding site* (PBS) des RNA-Transkriptes ist. LTR-Retroelemente lassen sich weiter in Retrotransposons, denen in den meisten Fällen das *env*-Gen fehlt, und die Vertebraten-spezifischen Retroviren unterteilen, zu denen u.a. die endogenen Retroviren (ERVs) gezählt werden (de Parseval und Heidmann, 2005; Han und Boeke, 2005) (s. Abschnitt 2.2.).



**Abbildung 1:** Klassifizierung transposabler Elemente und deren allgemeine Struktur. Abhängig vom Transpositionsmechanismus werden allgemein DNA-Transposons und Retroelemente unterschieden. DNA-Transposons transponieren hierbei über eine DNA-Zwischenstufe durch einen *cut-and-paste*-Mechanismus und werden von kurzen Sequenzwiederholungen (*inverted terminal repeats*, ITRs) flankiert. Die Transposition der Retroelemente verläuft hingegen über eine RNA-Zwischenstufe. Retroelemente werden weiter in die nicht-retroviralen Retrotransposons, denen flankierende *long terminal repeats* (LTRs) fehlen, und LTR-Retroelemente unterteilt. Zu den nichtretroviralen Retrotransposons zählen hierbei die *long interspersed elements* (LINEs), die zwei offene Leserahmen (*open reading frames*, ORFs) für ein DNA-Bindeprotein und die Enzyme Reverse Transkriptase und Endonuklease besitzen, sowie die *short interspersed elements* (SINEs), die selbst für kein Protein kodieren. Die LTR-Retroelemente können weiter in Retrotransposons und die Vertebraten-spezifischen Retroviren unterteilt werden, die ORFs für die retroviralen Gene *gag*, *pro*, *pol* und *env* besitzen (Abbildung modifiziert nach Alberts *et al.*, 2002; de Parseval und Heidmann, 2005).

Alle Retroelemente zeichnen sich dadurch aus, dass sie durch den Prozess der reversen Transkription transponiert werden, wobei zunächst eine RNA als Zwischenstufe transkribiert wird, die im Anschluss durch das Enzym Reverse Transkriptase (RT) in DNA umgeschrieben und an anderer Stelle im Genom integriert wird. Die meisten Retroelemente haben die Fähigkeit, selbst zu transponieren, verloren. Man nimmt an, dass die meisten Prozesse reverser Transkription durch LINE-1 (L1)-Sequenzen vermittelt werden, die den größten Teil

transposabler Elemente im menschlichen Genom bilden. Darunter fällt auch die Transposition der nicht-autonomen SINEs sowie die Bildung prozessierter Pseudogene (Okada *et al.*, 1997; Esnault *et al.*, 2000). Nach Initiation der L1-Transkription von einem Promotor im Bereich der 5'-untranslatierten Region (UTR) (Swergold, 1990; Ostertag und Kazazian, 2001; Lavie *et al.*, 2004) werden das von der bicistronischen mRNA kodierte RNA-Bindeprotein ORF1 sowie das ORF2-Protein mit Endonuklease- und Reverse Transkriptase-Aktivität synthetisiert (vgl. **Abb. 2**). Nach Anlagerung der Proteine an die mRNA wird diese über den Prozess der *target primed reverse transcription* (TPRT) im Genom integriert (Luan *et al.*, 1993). Die Endonuklease-Domäne des ORF2-Proteins bewirkt zunächst den Schnitt eines Stranges der Ziel-DNA, wobei ein freies 3'-OH-Ende entsteht (Moran *et al.*, 1996; Martin *et al.*, 2005). Der Schnitt erfolgt hierbei oft spezifisch in Bereichen, die der Hexanukleotid-Sequenz 5'-TTAAAA-3' ähneln, wobei jeweils der *Antisense*-Strang 3'-AATTTT-5' zwischen den Konsensus-Basen A und T geschnitten wird (Jurka, 1997). Aufgrund von Mikrohomologien lagert sich der Poly(A)-Schwanz der revers zu transkribierenden mRNA an einen komplementären kurzen Poly(T)-Bereich des geschnittenen Stranges der Ziel-DNA an. Das freie 3'-OH der Ziel-DNA wird als *Primer* für die reverse Transkription genutzt, wobei eine einzelsträngige DNA-Kopie der LINE-mRNA synthetisiert wird, die kovalent mit der Ziel-DNA verbunden ist. Nach Schneiden des zweiten Stranges der Ziel-DNA durch die L1-Endonuklease wird dieser zur Zweitstrangsynthese des neu integrierten Elementes genutzt. Da der zweite Strang der Ziel-DNA in unterschiedlicher Entfernung zur 5'-TTAAAA-3'-Konsensussequenz geschnitten wird, entstehen kurze einzelsträngige DNA-Bereiche, die im Anschluss durch zelluläre Reparaturprozesse aufgefüllt werden, was in kurzen Sequenzwiederholungen, sog. *target site duplications* (TSDs), beiderseits des integrierten Retroelementes resultiert (Jensen *et al.*, 1994; Jurka und Klonowski, 1996; Boeke und Stoye 1997; Moran und Gilbert, 2002). Die reverse Transkription der LINE-Sequenz schreitet, möglicherweise aufgrund einer niedrigen Prozessivität der Reversen Transkriptase oder aufgrund der Aktivität der zellulären RNase H, oft nicht bis zum 5'-Ende der mRNA voran, weswegen ein Großteil der neuen Insertionen im 5'-Bereich verkürzt ist (Boissinot *et al.*, 2000; Ostertag und Kazazian, 2001).



**Abbildung 2:** Prinzip der *target primed reverse transcription* (TPRT) am Beispiel einer L1-Sequenz. Nach Transkription der L1-Sequenz werden die Enzyme Endonuklease und Reverse Transkriptase (RT) synthetisiert, die sich an die mRNA anlagern. Durch die Endonuklease wird ein Strang der Ziel-DNA geschnitten und die mRNA lagert sich über ihren Poly(A)-Schwanz an einen kurzen komplementären Poly(T)-Bereich der Ziel-DNA an. Das freie 3'-OH-Ende der DNA wird im Anschluss als Ausgangspunkt für die cDNA-Synthese genutzt. Nach einem versetzten Schnitt des zweiten Stranges der DNA wird der komplementäre Strang der L1-cDNA synthetisiert. Die durch das Schneiden entstandenen kurzen Lücken in der DNA werden durch zelluläre Reparaturprozesse aufgefüllt, wobei zwei kurze, sequenzgleiche Abschnitte beiderseits des neu inserierten Elementes entstehen, die sog. *target site duplications* (TSDs) (Abbildung modifiziert nach Alberts *et al.*, 2002).

## 2.2. Humane endogene Retroviren (HERVs)

Etwa 8 % des menschlichen Genoms sind retroviralen Ursprungs. Diese Sequenzen umfassen sog. humane endogene Retroviren (HERVs), die als DNA im Wirtsgenom integrierte Retroviren, sog. Proviren, repräsentieren, sowie Sequenzfamilien, die lediglich von HERVs abgeleitete Sequenzabschnitte besitzen. Darunter fallen SINE-R, die aus einer SINE-Sequenz mit längerem HERV-Sequenzabschnitt bestehen (Ono *et al.*, 1987), sowie SVA-Elemente, die sich aus

SINE-R-Bereichen, CG-reichen VNTRs (*variable number of tandem repeats*) und einer Alu-ähnlichen Sequenz zusammensetzen (Shen *et al.*, 1994; Ostertag *et al.*, 2003).

HERV-Sequenzen sind auf Infektionen der Keimbahn menschlicher Vorfahren durch exogene Retroviren zurückzuführen, deren einzelsträngiges RNA-Genom nach reverser Transkription in das Genom der infizierten Keimbahnzellen integriert wurde. Diese Proviren, die aus internen Sequenzen mit Homologie zu retroviralen Genen sowie flankierenden LTRs bestehen, wurden anschließend nach Mendel'schen Regeln weitervererbt und, sofern kein negativer Selektionsdruck bestand, in der Population fixiert (Bannert und Kurth 2004, 2006). Im Laufe der Zeit kam es wiederholt zu (Re-)Infektionen durch verschiedene exogene Retroviren und die retroviralen Sequenzen breiteten sich durch intrazelluläre Amplifikation in einer retroviralen Weise weiter aus. Analog des retroviralen Replikationszyklus wurde hierbei die retrovirale RNA des Provirus, ausgehend von Promotorsequenzen im LTR, transkribiert. Im Fall von exogenen Retroviren werden die mRNA-Transkripte im Anschluss in die retroviralen (Vorläufer-)Proteine translatiert, die zusammen mit der RNA in retrovirale Partikel verpackt werden. Diese können nach Abschnüren von der Wirtszelle und weiterer Prozessierung der Proteine weitere Zellen infizieren. Die in die infizierte Zelle eingeschleuste RNA wird dabei erneut revers transkribiert und in das Wirtsgenom integriert. Im Falle von endogenen Retroviren fehlt dieses extrazelluläre Stadium vermutlich, weswegen die Integration hier im gleichen Genom erfolgte, was für die meisten HERV-Gruppen in der Bildung zahlreicher Kopien des ursprünglichen Provirus resultierte (Mayer und Meese, 2005). Zudem finden sich zahlreiche Vertreter der HERV-W-Gruppe, die durch L1-vermittelte Retrotransposition der ursprünglichen proviralen Sequenz generiert wurden (Pavlíček *et al.*, 2002; Costas, 2002).

Infektionen durch verschiedene exogene Retroviren resultierten in der Bildung unterschiedlicher HERV-Gruppen, von denen heute, je nach Definition, etwa 30 bis 50 unterschieden werden (Gifford und Tristem, 2003; Sperber *et al.*, 2007; Blomberg *et al.*, 2009). In der *Repbase*-Datenbank für repetitive Sequenzen sind sogar einige hundert verschiedene HERV- und LTR-Gruppen aufgeführt (Jurka *et al.*, 2005). Wurden HERV-Sequenzen zunächst in retroviralen Partikeln sowie durch Hybridisierungsexperimente mit markierten Sonden bekannter exogener Retroviren aus anderen Spezies detektiert (Callahan *et al.*, 1982; Boller *et al.*, 1983; Repaske *et al.*, 1983; Keydar *et al.*, 1984; Löwer *et al.*, 1984; Westley und May, 1984; Ono *et al.*, 1986), erhöhte sich die Anzahl bekannter Proviren durch computerbasierte Techniken nach Sequenzierung des menschlichen Genoms enorm. So sind zum heutigen Zeitpunkt mehr als eine halbe Million HERV-Loci bekannt (Sperber *et al.*, 2007).

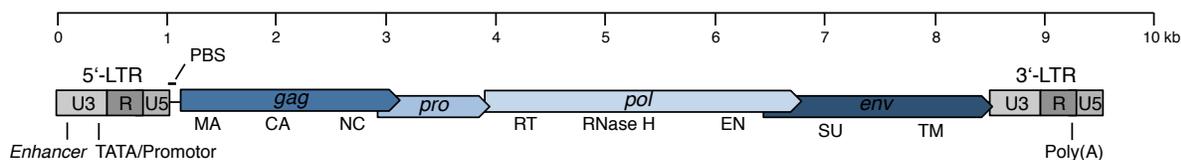
Das evolutionäre Alter verschiedener HERV-Gruppen lässt sich durch Vergleiche mit anderen Primaten-Spezies abschätzen. Geht man davon aus, dass jede retrovirale Gruppe nur einmal in das Genom eingebracht wurde und kein horizontaler DNA-Transfer zwischen Spezies stattfand, lässt sich unter Berücksichtigung der bekannten Phylogenie ein ungefährer Zeitpunkt für die Keimbahnintegration bzw. Fixierung in der Population abschätzen. Das Vorhandensein bestimmter Proviren in verschiedenen Spezies deutet hierbei auf eine Infektion vor der evolutiven Aufspaltung dieser Spezies hin (Voisset *et al.*, 1999; Goodchild *et al.*, 1993). Aufgrund solcher Studien schätzt man, dass die meisten HERV-Gruppen bereits vor über 30 Millionen Jahren in das menschliche Genom eingebracht wurden. Die HERV-K(HML-2)-Gruppe ist hierbei ein Beispiel dafür, dass Reintegrationen in einigen Fällen über einen längeren Zeitpunkt stattfanden, da einige Proviren dieser Gruppe erst nach der evolutiven Trennung von Mensch und Schimpanse entstanden, die vor etwa 6 Millionen Jahren erfolgte (Barbulescu *et al.*, 1999). Ein weiterer Ansatz zur Abschätzung des evolutiven Alters beruht auf Sequenzvergleichen verschiedener Proviren mit einer Konsensussequenz der jeweiligen Gruppe (Costas und Naveira, 2000) oder auf Sequenzvergleichen der flankierenden LTRs. In den 5'- und 3'-LTRs, die zum Zeitpunkt der Provirus-Integration identisch sind, sammeln sich im Laufe der Zeit Mutationen an, anhand derer sich, unter Berücksichtigung der Mutationsrate nicht-kodierender DNA, der ungefähre Integrationszeitpunkt abschätzen lässt (Dangel *et al.*, 1995).

Die Klassifizierung verschiedener HERV-Gruppen in drei der bekannten retroviralen Genera beruhte auf Sequenzvergleichen der retroviralen Gene. Aufgrund von Vergleichen der *pol*-RT-Region finden sich HERV-Gruppen mit Ähnlichkeit zu Gamma- und Beta-Retroviren sowie zu Spuma-ähnlichen Retroviren. Zur Klasse I, den Gamma-Retroviren, zählen hierbei u.a. die Gruppen HERV-W, HERV-E, HERV-F und HERV-I, die Klasse II der Beta-Retroviren umfasst die HERV-K Gruppen HML-1 bis HML-10 und zur Klasse III, den Spuma-ähnlichen Retroviren, zählen die Gruppen HERV-S und HERV-L (Mager und Medstrand, 2003; Jern *et al.*, 2005). Die Nomenklatur verschiedener HERV-Proviren ist oft nicht eindeutig und eine standardisierte Nomenklatur der verschiedenen HERV-Gruppen gibt es bis heute nicht, insbesondere da neu identifizierte HERV-Sequenzen oft laborinterne Bezeichnungen erhielten. Eine oft verwendete Bezeichnung der verschiedenen HERV-Gruppen richtete sich nach der *primer binding site* (PBS), an der die tRNA bindet, die die reverse Transkription initiierte, wobei der Einbuchstaben-Code der Aminosäure der jeweiligen tRNA dem HERV-Namen angehängt wird (Mayer und Meese, 2005). So nutzten Mitglieder der HERV-W-Gruppe eine Tryptophan-tRNA zur Initiation der reversen Transkription und Mitglieder der HERV-K-Gruppe eine Lysin-tRNA. Zu Problemen kommt es hier jedoch, wenn die PBS nicht mehr vorhanden ist, stark

mutiert vorliegt, oder wenn deutlich voneinander unterscheidbare Sequenzen die gleiche PBS nutzen. Ein Beispiel hierfür ist die HERV-K-Supergruppe, die sich wiederum aus zehn Gruppen zusammensetzt, wobei Mitglieder der einzelnen Untergruppen mehr als 80 % Sequenzidentität zueinander aufweisen (Andersson *et al.*, 1999). Für HERV-Loci, die als transkribiert beschrieben wurden, wurde eine einheitliche Nomenklatur vorgeschlagen, die sich auch auf neu identifizierte Loci anwenden lässt. Die Loci werden hierbei einheitlich als „ERV“ benannt, wobei diesem Kürzel das Gruppensymbol, wie beispielsweise W oder K, angehängt wird und neue Loci nach Identifizierung fortlaufend durchnummeriert werden (Mayer *et al.*, 2011).

### 2.2.1. Struktur humaner endogener Retroviren

Die provirale Struktur humaner endogener Retroviren entspricht der proviralen Struktur exogener Retroviren. So bestehen vollständige HERV-Proviren aus einer internen proviralen Sequenz, die für retrovirale Proteine kodiert, sowie flankierenden *long terminal repeats* (vgl. **Abb. 3**). Der retrovirale *gag* (*group-specific antigen*)-ORF kodiert hierbei die Matrix-, Kapsid- und Nukleokapsidproteine, *pro* kodiert die Protease, *pol* die Enzyme Reverse Transkriptase (RT) und Integrase, und *env* die Hüllproteine. Das Env-Protein kann nach Translation weiter in seine Oberflächen- und Transmembran-Untereinheiten prozessiert werden. Die flankierenden LTRs repräsentieren autonome Einheiten, die die Transkription retroviraler RNA regulieren und initiieren. In den LTRs, die, von 5' nach 3', weiter in eine U3-, R- und U5-Region eingeteilt werden, findet sich der ursprüngliche retrovirale Promotor mit einer TATA-Box am 3'-Ende der U3-Region und ein Polyadenylierungs-Signal (5'-AATAAA-3') im 3'-Bereich der R-Region sowie ggf. *Enhancer*-Elemente oder Bindestellen für Transkriptionsfaktoren (de Parseval und Heidmann, 2005; Jern und Coffin; 2008).



**Abbildung 3:** Struktur eines HERV-Provirus am Beispiel von HERV-K(HML-2). Ein knapp 10 kb langes Vollängen-Provirus besteht aus internen proviralen Sequenzen, die für die retroviralen Proteine Gag, Pro, Pol und Env kodieren, sowie flankierenden *long terminal repeats* (LTRs), die sich in die Domänen U3, R und U5 einteilen lassen. Die Position des Promotors im 3'-Bereich der U3 Region des 5'-LTRs sowie des Polyadenylierungs-Signals im 3'-Bereich der R-Region des 3'-LTRs sind eingezeichnet. MA = Matrix, CA = Kapsid (*capsid*), NC = Nukleokapsid (*nucleocapsid*), PBS = *primer binding site*, RT = Reverse Transkriptase, EN = Endonuklease, SU = Oberflächenprotein (*surface protein*), TM = Transmembran-Protein.

Insgesamt finden sich etwa 3.000 relativ komplette provirale Strukturen im menschlichen Genom (Sperber *et al.*, 2007). Ein Großteil der HERV-Loci ist aufgrund der langen Präsenz im Genom stark mutiert und die proviralen Sequenzen weisen zahlreiche Deletionen, Insertionen oder *Nonsense*-Mutationen auf. Es finden sich jedoch, insbesondere bei Vertretern der Gruppe HERV-K(HML-2), noch Proviren mit mehr oder weniger intakten ORFs für retrovirale Proteine (Barbulescu *et al.*, 1999; Tönjes *et al.*, 1999). Beispiele aus dieser Gruppe sind das HERV-K(HML-2.HOM)-Provirus auf Chromosom 7p22.1, das intakte *gag*- und *env*-ORFs besitzt (Mayer *et al.*, 1999), sowie das HERV-K113-Provirus auf Chromosom 19p12, das intakte ORFs für alle retroviralen Gene aufweist (Turner *et al.*, 2001). Ein Beispiel aus der HERV-W-Gruppe stellt der *syncytin-1/ERVWE1/ERVW-1*-Locus auf Chromosom 7q21.2 dar, dessen Env-Protein, Syncytin-1, eine biologische Funktion während der menschlichen Plazentaentwicklung besitzt (Blond *et al.*, 2000), wobei die restlichen ORFs des Locus jedoch nicht mehr intakt sind.

Neben solchen mehr oder weniger vollständigen Proviren mit internen proviralen Sequenzen liegt ein großer Teil der genomischen HERV-Loci in Form sog. solitärer LTRs vor, die die Zahl an Proviren deutlich übersteigen. So beinhaltet beispielsweise die HERV-E-Gruppe zwischen 85 und 250 Proviren oder die HERV-P-Gruppe zwischen 70 und 200. Die Anzahl solitärer LTRs beider Gruppen wird jedoch auf jeweils etwa 1.000 geschätzt (Gifford und Tristem, 2003; Mager und Medstrand, 2003). Solche solitären LTRs resultieren aus homologer Rekombination der 5'- und 3'-LTRs eines Provirus mit daraus resultierendem Verlust der dazwischenliegenden internen Sequenz (Mager und Medstrand, 2003; Bannert und Kurth, 2006; Stoye, 2012).

### 2.2.2. Promotoraktivität retroviraler LTRs und Regulation benachbarter Sequenzen

Auch wenn die Mehrheit, wenn nicht sogar alle HERVs die Fähigkeit verloren haben, selbst neue Kopien zu bilden, besitzen ihre LTRs dennoch oft funktionelle regulatorische Sequenzen zur Transkriptions-Initiation und -Termination. Im Falle von HERV-K(HML-2) ist hierbei insbesondere die U3-Region von Bedeutung, da sich in dieser alle zur Transkriptions-Initiation der retroviralen Gene nötigen Sequenzen befinden, darunter der retrovirale Promotor, *Enhancer* sowie Bindestellen für zahlreiche Transkriptionsfaktoren (Manghera und Douville, 2013). Es wurden weitere alternative Transkriptions-Startpunkte beschrieben, die ggf. unter verschiedenen physiologischen Bedingungen genutzt werden könnten (Kovalskaya *et al.*, 2006; Fuchs *et al.*, 2011; Kato *et al.*, 2011). Mittels GREM (*Genomic Repeat Expression Monitor*)-Technologie konnte gezeigt werden, dass neben 5'-LTRs auch 3'- sowie solitäre LTRs der HERV-K(HML-2)-Gruppe *in vivo* als aktive Promotoren fungieren können (Buzdin *et al.*, 2006). Im Falle der

HERV-W-Gruppe wurde eine Transkriptions-Initiation ausgehend von der U3- und R-Region (Gimenez *et al.*, 2010) sowie von der R/U5-Region beschrieben (Lee *et al.*, 2003). Die Transkription retroviraler Sequenzen wird hierbei epigenetisch durch DNA-Methylierung der LTRs reguliert (Lavie *et al.*, 2005; Gimenez *et al.*, 2009; Trejbalová *et al.*, 2011).

Des Weiteren können retrovirale LTRs auch als (alternative) Promotoren für benachbarte Gene oder für nichtkodierende RNAs dienen (Cohen *et al.*, 2009). So werden beispielsweise die Gene *ApoC-I* (*apolipoprotein C1*) und *EBR* (*endothelin B receptor*) (Medstrand *et al.*, 2001) sowie das mit dem Opitz-Syndrom assoziierte Gen *MID1* (Landry *et al.*, 2002) durch LTRs der HERV-E-Gruppe transkribiert. Im Falle des *beta1,3-galactosyltransferase 5*-Gens wird ein HERV-LTR als einer von mindestens zwei alternativen Promotoren genutzt, wobei der LTR-Promotor hierbei durch einen gewebespezifischen Transkriptionsfaktor aktiviert wird (Dunn *et al.*, 2003). Eine genomweite HERV-K(HML-2)-Expressions-Analyse in gesundem und kanzerogenem Keimzellgewebe ergab, dass mindestens 50 % der untersuchten LTRs Promotoraktivität besitzen, wobei 5'-LTRs generell eine höhere Aktivität als solitäre und 3'-LTRs zeigten (Buzdin *et al.*, 2006). Auch eine Funktion von HERV-LTRs in der Regulation benachbarter Gene durch Generierung von *Antisense*-Transkripten wurde beschrieben. So dienen zwei HERV-LTRs, die intronisch in *Antisense*-Orientierung in den Genen *SLC4A8* (*sodium bicarbonate cotransporter*) und *IFT172* (*intraflagellar transport protein 172*) lokalisiert sind, *in vivo* als Promotoren für RNAs, die als komplementäre Transkripte der benachbarten Exons die mRNA-Mengen der entsprechenden Gene reprimieren (Gogvadze *et al.*, 2009). LTRs können als bidirektionale Promotoren in *Sense*- und *Antisense*-Orientierung fungieren, wie beispielsweise im Falle eines ERV1-LTRs gezeigt, der die Transkription der Gene *DSCR4* (*human down syndrome critical region 4*) und *DSCR8* reguliert, die sich den dazwischen liegenden LTR-Promotor teilen (Dunn *et al.*, 2006). Auch an der Regulation der Transkription von Zielgenen des Tumorsuppressors p53 könnten HERV-LTRs beteiligt sein, da gezeigt werden konnte, dass die LTRs etwa ein Drittel aller p53-Bindungsstellen im menschlichen Genom ausmachen (Wang *et al.*, 2007).

Neben der Promotoraktivität finden sich Beispiele für LTRs, die die Transkription benachbarter Gene beeinflussen, indem sie als *Enhancer* fungieren, wie im Fall des Leptin-kodierenden *LEP*-Gens, für das ein LTR-*Enhancer* eine Plazenta-spezifische Expression bewirkt (Bi *et al.*, 1997). Des Weiteren finden sich LTRs, die alternative Spleiß- und Polyadenylierungs-Signale für zelluläre Gene liefern (Cohen *et al.*, 2009). So wird im Falle des *ORB* (*human leptin receptor*)-Gens das kürzere *OBRa*-Transkript durch ein zusätzliches Spleißen innerhalb eines LTRs der HERV-K-Gruppe generiert, der zudem die 67 terminalen Aminosäuren des ORBa-Proteins kodiert (Kapitonov und Jurka, 1999). Auch für das *VEGFR-3/FLT4* (*vascular endothelial growth*

*factor receptor 3*)-Gen wird eine kürzere Isoform durch Spleißen in einen HERV-K-LTR generiert, der zwischen den letzten beiden Exons lokalisiert ist (Hughes, 2001). Ein Beispiel für Transkripte, die in einem LTR der HERV-H-Gruppe polyadenyliert werden, stellen die beiden Gene *HHLA2* und *HHLA3* dar, die durch Analyse von *expressed sequence tag* (EST)-Daten identifiziert wurden (Mager *et al.*, 1999).

### 2.2.3. Transkription humaner endogener Retroviren

Auch wenn die Mehrheit der HERV-Proviren nicht mehr für intakte retrovirale Proteine kodiert, werden viele HERV-Loci dennoch transkribiert. Im Gegensatz zu früheren Annahmen, dass HERVs, wie viele andere Retroelemente, durch epigenetische Mechanismen stillgelegt sind und es nur in Keimzellen und Plazentagewebe zur Expression kommt (Reiss und Mager, 2007), konnten unter Verwendung unterschiedlicher Techniken Transkripte verschiedener HERV-Gruppen sowohl in malignen als auch in gesunden Geweben und Zelllinien identifiziert werden. So wurden beispielsweise durch Hybridisierungs-Experimente HERV-K-Transkripte in Brustkrebs und gesundem Brustgewebe, PBMCs (*peripheral blood mononuclear cells*) sowie Plazentagewebe detektiert (Yin *et al.*, 1997). Mittels *Real-Time*-PCR wurden Transkripte verschiedener HERV-Gruppen, darunter HERV-E und HERV-W, in Endometriose und Ovarialkarzinomen und gesunden Vergleichsgeweben (Hu *et al.*, 2006) sowie HERV-K-Transkripte in Hirnproben von Multiple Sklerose-Patienten sowie gesunden menschlichen Geweben nachgewiesen (Muradrasoli *et al.*, 2006). Auch mithilfe *Microarray*-basierender Experimente konnte die Transkription unterschiedlicher HERV-Gruppen in zahlreichen untersuchten Geweben detektiert werden. So wurde in zwei Studien die Transkription diverser HERV-Gruppen in Hirnproben aus Personen mit Schizophrenie, bipolaren Störungen und gesunden Kontrollen sowie in nicht-malignem und Tumor-Brustgewebe nachgewiesen (Frank *et al.*, 2005, 2008). In einer weiteren *Microarray*-Studie wurden Transkripte von 20 verschiedenen HERV-Gruppen in 19 verschiedenen gesunden menschlichen Geweben detektiert, darunter Haut, Schilddrüse, Plazenta und Fortpflanzungsorgane (Seifarth *et al.*, 2005). Zudem wurden mit dem gleichen HERV-spezifischen *Microarray* HERV-Transkripte in malignem und gesundem Nierengewebe und Zelllinien nachgewiesen (Haupt *et al.*, 2011).

Das Vorkommen von HERV-Transkripten in allen Normalgeweben, die in diesen Studien getestet wurden, deutet darauf hin, dass HERVs einen ständigen Bestandteil des menschlichen Transkriptom darstellen. Da sich die HERV-Transkription hierbei in verschiedenen Geweben

und Zelllinien sowie zwischen Individuen unterscheidet (Andersson *et al.*, 1998), muss sie einer bestimmten Regulation unterliegen.

Im Gegensatz zu den oben genannten Studien, in denen die Transkription verschiedener HERV-Gruppen insgesamt untersucht wurde, gibt es jedoch noch vergleichsweise wenig Information darüber, von welchen genomischen HERV-Loci die in diesen Studien detektierten Transkripte stammen und welche spezifischen Loci generell in verschiedenen Geweben und Zelltypen aktiv sind. Zur spezifischen Identifizierung von transkribierten HERV-Loci wurden bisher bioinformatische und molekulargenetische Ansätze verfolgt. In *in silico*-Experimenten wurden transkribierte HERV-Proviren durch Analyse von Sequenzinformationen in der *dbEST*-Datenbank für *expressed sequence tags* (ESTs) identifiziert (Stauffer *et al.*, 2004; Oja *et al.*, 2007). Es konnte jedoch in einer späteren Studie gezeigt werden, dass die dort hinterlegten Daten für eine detaillierte Beschreibung der HERV-Transkriptionsmuster in verschiedenen Zelltypen und Geweben nicht ausreichend sind und weit mehr HERV-Loci transkribiert sind als solche Studien vermuten lassen (Flockerzi *et al.*, 2008). Darauf deuten auch Analysen der im Rahmen des ENCODE (*Encyclopedia of DNA Elements*)-Projektes generierten Daten hin, die darauf schließen lassen, dass das menschliche Genom evtl. nahezu vollständig transkribiert sein könnte (ENCODE Project Consortium, 2007; Gerstein *et al.*, 2007; Trinklein *et al.*, 2007; siehe aber auch van Bakel *et al.*, 2010; Clark *et al.*, 2011; Mercer *et al.*, 2012). Aufgrund ihrer im vorangehenden Abschnitt beschriebenen Promotoraktivität könnten hierbei auch zahlreiche Transkripte auf HERV-Sequenzen zurückzuführen sein.

In molekulargenetischen Studien wurden HERV-spezifische cDNA-Sequenzen generiert und genomischen Loci anhand charakteristischer Nukleotidunterschiede zugeordnet (Sugimoto *et al.*, 2001; Mayer *et al.*, 2004; Buzdin *et al.*, 2006; Ruprecht *et al.*, 2008a, Frank *et al.*, 2008). So konnten beispielsweise durch Sequenzierung Gruppen-spezifischer RT-PCR-Produkte transkribierte Loci der HERV-W-Gruppe in PBMCs aus MS-Patienten und gesunden Kontrollen identifiziert werden (Laufer *et al.*, 2009). Eine weitere Studie identifizierte transkribierte HERV-K(HML-2)-Loci in verschiedenen menschlichen Geweben, darunter Hirnproben aus Patienten mit bipolaren Störungen und Schizophrenie, gesunde Hirnproben sowie in verschiedenen Hirn- und Keimzelltumoren (Flockerzi *et al.*, 2008). Die Identifizierung zahlreicher transkribierter HERV-Proviren in einer hohen Anzahl an untersuchten Proben wurde durch die kürzliche Entwicklung von HERV-Locus-spezifischen *Microarrays* angestrebt. So wurde in einer Analyse von 40 gesunden und 39 Tumor-Proben berichtet, dass fast ein Drittel der etwa 10.000 dort untersuchten HERV-Sequenzen tatsächlich transkribiert werden (Pérot *et al.*, 2012).

### 2.3. Die mögliche Involvierung von HERVs in der Krankheitspathogenese

Neben der Bedeutung von HERV-Transkripten im Rahmen allgemeiner Untersuchungen der Transkriptionsaktivität des menschlichen Genoms ist die Transkription verschiedener HERV-Loci insbesondere auch im Falle der Pathogenese von Krankheiten von großem Interesse, für die eine mögliche Involvierung humaner endogener Retroviren oder retroviraler Proteine diskutiert wird. Im Falle der direkten oder indirekten Bedeutung retroviraler Transkripte und/oder Proteine in der Krankheitspathogenese sind sicherlich ausschließlich die transkribierten Loci von Bedeutung, auf die in zahlreichen Studien detektierte retrovirale Transkripte oder Proteine zurückzuführen sind.

Eine direkte Involvierung von HERVs in der Pathogenese einer Krankheit konnte bisher nicht gezeigt werden. Generell würde man annehmen, dass offensichtlich pathogene HERVs, wie Gene, einem negativen Selektionsdruck unterliegen, der die Fixierung solcher Proviren in der Population verhindert, und sich daher nur neutrale Nachfahren einst aktiver Elemente im menschlichen Genom finden (Belshaw *et al.*, 2005). Die Fixierung solcher Proviren wäre jedoch wahrscheinlicher, wenn sich ihre pathogene Wirkung beispielsweise erst nach der reproduktiven Phase zeigen würde, oder wenn die Integration einiger Proviren in evolutiven Vorteilen resultieren würde, wie durch die bereits diskutierten Funktionen der retroviralen LTRs als alternative Promotoren, Spleiß- und Polyadenylierungssignale, oder auch durch vorteilhafte chromosomale Um- und Neuordnungen durch Rekombination retroviraler Sequenzen (Blomberg *et al.*, 2005). Auch von den Proviren kodierte Proteine könnten vom Wirtsorganismus über die Zeit zu seinem Vorteil genutzt werden. Ein Beispiel hierfür ist die *env*-Sequenz eines HERV-W-Locus auf Chromosom 7q21.2, die sich u.a. im Menschen zu einem funktionellen Gen entwickelt hat und dessen Proteinprodukt, Syncytin-1, während der Plazentaentwicklung exprimiert ist und dort vermutlich an der Fusion von Trophoblasten zu Syncytiotrophoblasten beteiligt ist (Mallet *et al.*, 2004; vgl. Abschnitt 2.3.2.). Zudem könnten endogene Retroviren durch den Mechanismus der *receptor interference* einen Schutzmechanismus vor Infektion durch verwandte exogene Retroviren darstellen. So wird beispielsweise durch eine endogene Expression des *Jaagsiekte sheep retrovirus* (JSRV) in Schafen eine Infektion durch das exogene JSRV verhindert (Spencer *et al.*, 2003). In Mäusen bewirkt die Expression des vom *Fv-4*-Gen kodierten Env-Proteins eine Resistenz gegen Infektionen durch ecotrope murine Leukämieviren (MuLVs) (Taylor *et al.*, 2001). Solche positiven Effekte könnten somit dazu beigetragen haben, dass sich endogene Retroviren, trotz potentieller negativer Auswirkungen im Falle der Fehlregulation ihrer Expression in menschlichen Krankheiten, in der Evolution

durchgesetzt haben. In diesem Fall würden HERV-Gene zellulären krankheitsrelevanten Genen gleichen, die ebenfalls erst nach Fehlregulation die Krankheitspathogenese begünstigen.

Geht man davon aus, dass eine potentielle pathogene Funktion humaner endogener Retroviren auf der Kompetenz beruht, intakte und funktionale Vollängen-Proteine zu kodieren, wäre die Pathogenität auf relativ wenige kodierungs-kompetente Proviren beschränkt. Es wäre jedoch auch eine pathogene Funktion retroviraler mRNA-Transkripte, beispielsweise durch RNA-Interferenz (Fire *et al.*, 1998) und ähnliche Mechanismen, nicht auszuschließen (Li *et al.*, 2013).

Diskutiert wurde die Beteiligung diverser HERV-Gruppen in einer Vielzahl verschiedener Krankheiten, darunter verschiedene Krebsarten (Ruprecht *et al.*, 2008b), Autoimmunerkrankungen, wie Gelenkrheumatismus (Nakagawa *et al.*, 1997; Sicat *et al.*, 2005; Ehlhardt *et al.*, 2006; Reynier *et al.*, 2009), Psoriasis (Bessis *et al.*, 2004; Moles *et al.*, 2005) oder Lupus erythematodes (*systemic lupus erythematosus*, SLE) (Ogasawara *et al.*, 2001; Adelman und Marchalonis 2002; Piotrowski *et al.*, 2005) sowie neuropathogene Krankheiten, wie Multiple Sklerose (MS) (Christensen, 2005), Schizophrenie (Yolken *et al.*, 2000) und bipolare Störungen (Weis *et al.*, 2007). Zahlreiche Studien befassten sich hierbei insbesondere mit der HERV-K(HML-2)-Gruppe, für die eine mögliche Involvierung in diversen Tumorerkrankungen diskutiert wird, sowie mit der HERV-W-Gruppe und deren möglicher Rolle in der Pathogenese von Multipler Sklerose.

### 2.3.1. Die HERV-K-Supergruppe humaner endogener Retroviren

Die ersten HERV-K-Sequenzen wurde aufgrund ihrer Ähnlichkeit zur proviralen DNA des *mouse mammary tumor virus* (MMTV), das bei Mäusen die Bildung von Tumoren der Brustdrüsen bewirkt (Callahan und Smith, 2000), aus dem menschlichen Genom isoliert (Callahan *et al.*, 1982). Die HERV-K-Supergruppe wurde zunächst, basierend auf der Sequenz ihrer stark konservierten RT-Region, in sechs verschiedene Gruppen eingeteilt (Medstrand und Blomberg, 1993), die als HERV-K(HML-1) bis HERV-K(HML-6) benannt wurden. Die Abkürzung „HML“ steht hierbei, aufgrund ihrer Sequenzähnlichkeit im *env*-ORF zu MMTV, für „*human MMTV-like*“. Mittlerweile sind, auch in der Datenbank *Repbase*, zehn unterschiedliche HML-Gruppen definiert (Andersson *et al.*, 1999; Jurka, 2000). Schätzungen zufolge gibt es im menschlichen Genom etwa 5.550 HERV-K-Proviren und über 6.000 solitäre LTRs (Mager und Medstrand, 2003).

Die Sequenzierung des menschlichen Genoms, die nahezu alle HERV-K-Sequenzen zugänglich machte, ermöglichte tiefgreifende Analysen der Sequenzen und evolutiven Beziehungen der

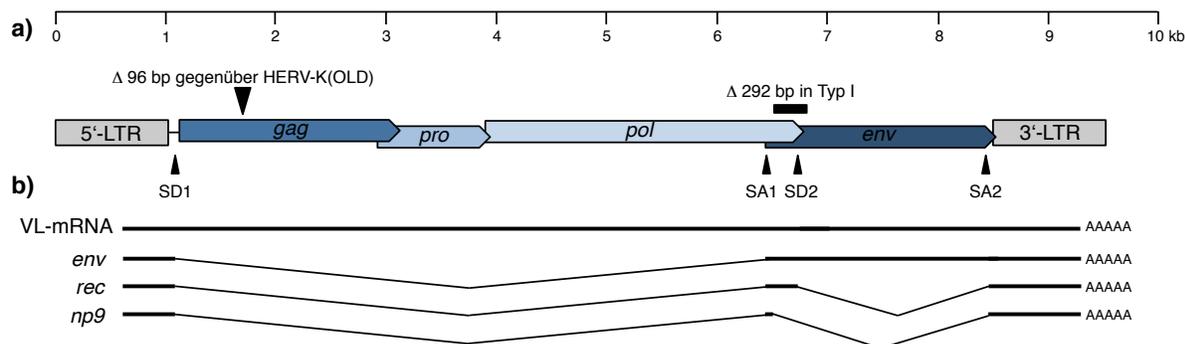
verschiedenen Gruppen. Die Generierung einer Konsensussequenz der HERV-K(HML-5)-Gruppe aus genomischen HML-5-Sequenzen ergab beispielsweise eine PBS mit Ähnlichkeit zur Methionin-tRNA, weswegen im Falle dieser Untergruppe die Bezeichnung HERV-M angebracht wäre (Lavie *et al.*, 2004). Die größten Sequenzähnlichkeiten der verschiedenen HERV-K-Gruppen finden sich in der *pol*-Region, wohingegen die restlichen retroviralen Gene und LTRs stärker voneinander abweichen. Phylogenetische Analysen der HERV-K(HML-3)-Gruppe deuten auf zwei unterscheidbare Sub-Populationen hin, die während einer kurzen Zeitperiode vor etwa 36 Millionen Jahren im Genom fixiert wurden. Durch multiple *Alignments* von 73 genomischen HERV-K(HML-3)-Loci konnte eine Konsensussequenz mit ORFs aller retroviralen Gene generiert werden, die der ursprünglichen Sequenz des exogenen Retrovirus vermutlich sehr nahe kommt (Mayer und Meese, 2002). In dieser Konsensussequenz fand sich zudem eine intakte Domäne des dUTPase-Enzyms, welches durch Verminderung der intrazellulären dUTP-Konzentration den Einbau von Uracil in die DNA verhindert. Weitere Sequenzvergleiche ergaben, dass auch Vorfahren weiterer HERV-K-Gruppen dUTPase-Domänen aufwiesen und deren exogenen Vorfahren vermutlich für aktive Enzyme kodierten (Mayer und Meese, 2003).

### 2.3.1.1. Die Gruppe HERV-K(HML-2)

Zu den am besten untersuchten HERV-Gruppen gehört HERV-K(HML-2), kurz HML-2. HML-2-Sequenzen wurden erstmals im menschlichen Genom durch Hybridisierungsexperimente mit markierter proviraler MMTV-DNA detektiert, wobei spezifische Fragmente Sequenzähnlichkeiten zu den *gag*-, *pol*- und *env*-Bereichen des MMTV-Genoms aufwiesen (Callahan *et al.*, 1982). Das erste, knapp 9,2 kb lange Vollängen-Provirus, HERV-K10, wurde aus dem menschlichen Genom als homologe Sequenz des *Syrian hamster intracisternal A-particle*-Gens isoliert und zeigte lediglich wenige *Nonsense*-Mutationen (Ono *et al.*, 1986).

HML-2-Sequenzen finden sich in Altweltaffen, zu denen auch die Hominoidea zählen, die die Familien der Gibbons sowie der Menschenaffen umfasst, in Neuweltaffen fehlen homologe Sequenzen jedoch. Die ursprüngliche Insertion in das Genom fand somit vermutlich vor etwa 35 Millionen Jahren, nach Divergenz der Alt- und Neuweltaffen, statt. Nach der initialen Integration in die Keimbahn kam es zur Amplifikation der retroviralen Sequenzen, so dass sich im menschlichen Genom heute schätzungsweise 60 HML-2-Proviren und 2.500 solitäre LTRs finden (Mager und Medstrand 2003). Vor Divergenz der Hominoidea, die vor etwa 20 Millionen Jahren stattfand, kam es zu zwei Deletions-Ereignissen innerhalb der proviralen Sequenz (vgl. **Abb. 4a**). Zum einen kam es zu einer 96 bp großen Deletion im *gag*-ORF, wobei sich während

der Evolution der Hominoidea jedoch nur die Proviren mit dieser Deletion weiter ausbreiteten. Die sog. HERV-K(OLD)-Proviren, die die ursprüngliche längere Version des *gag*-ORFs aufweisen, sind im menschlichen Genom lediglich als fossile Sequenzen vorhanden (Reus *et al.*, 2001b). Hierbei deutet das Alter von in diesen Sequenzen integrierten Alu-Elementen sowie der Vergleich der LTRs auf ein evolutives Alter der HERV-K(OLD)-Proviren von etwa 28 Millionen Jahren hin, was die Integration der proviralen Sequenzen vor der evolutiven Trennung der Hominoidea von den restlichen Altweltaffen unterstützt (Mayer *et al.*, 1998; Reus *et al.*, 2001b). Des Weiteren traten HML-2-Proviren auf, die eine 292 bp lange Deletion an der Grenze von *pol* zu *env* aufwiesen. Die Kopienzahl dieser als Typ I bezeichneten Proviren erhöhte sich, ebenso wie die der ursprünglichen Typ II Proviren ohne Deletion (Ono *et al.*, 1986; Mayer und Meese, 2005).



**Abbildung 4:** Struktur und Spleißvarianten verschiedener Provirus-Typen der HERV-K(HML-2)-Gruppe. **a)** Allgemeine Struktur eines HML-2-Volllängen-Provirus am Beispiel von HERV-K(HML-2.HOM) (*Genbank acc. no.* AF074086.2) mit ORFs der retroviralen Gene *gag*, *pro*, *pol* und *env* sowie den flankierenden LTRs. Die Position einer 96 bp großen Deletion ( $\Delta$ ) im *gag*-ORF, die HML-2 von den älteren HERV-K(OLD) Proviren (vgl. Text) unterscheidet, ist über der Povirus-Struktur eingezeichnet. Basierend auf einer weiteren 292 bp langen Deletion an der Grenze von *pol* zu *env* (schwarzer Balken), ist eine weitere Unterteilung der HML-2-Proviren in Typ I und II möglich. Bei HERV-K(HML-2.HOM) handelt es sich um ein Typ II-Provirus, das den 292 bp langen Sequenzabschnitt besitzt, im Gegensatz zu Proviren des Typs I, wo der Bereich deletiert ist. **b)** Spleißvarianten von mRNA der HML-2-Proviren des Typs I und II. Die Position von je zwei *splice donor* (SD) und *splice acceptor* (SA) sites sowie die verschiedenen mRNA-Spleißvarianten sind eingezeichnet, wobei gespleißte Bereiche durch dünnere schwarze Linien verdeutlicht sind. Ein Volllängen-Transkript (VL-mRNA) erstreckt sich vom Promotorbereich im 5'-LTR bis zum Polyadenylierungssignal im 3'-LTR. Im Falle der *env*-mRNA wird ein etwa 5,5 kb langer Bereich zwischen SD1 und SA1 gespleißt. Proviren des Typs I und II kodieren für die beiden zusätzlichen Proteine Np9 (Typ I) und Rec (Typ II), deren mRNAs durch ein weiteres Spleißen der *env*-mRNA entstehen. Die *splice donor site* für *rec*-mRNA (SD2) liegt hierbei innerhalb der 292 bp Deletionsregion, wobei ein Bereich bis zur *splice acceptor site* SA2 etwa 110 bp stromauf des 3'-LTR gespleißt wird. In Proviren des Typs I, wo die SD2 aufgrund der Deletion fehlt, wird eine alternative *splice donor site* unmittelbar stromauf des Deletionsbereiches genutzt.

Neben den üblichen (mehr oder weniger vollständigen) retroviralen Genen kodieren Proviren des Typs I und II jeweils ein zusätzliches Protein im *env*-Bereich des Provirus. Dies ist im Falle von Typ I Proviren das etwa 9 kDa große Np9-Protein und bei Proviren des Typs II das etwa 14 kDa große Rec/cORF-Protein. Bei Rec handelt es sich hierbei um ein Protein mit funktioneller Homologie zum HIV-1 (*human immunodeficiency virus 1*) Rev-Protein, das den Export unge-

spleißter mRNA aus dem Zellkern über den Export-Rezeptor Crm1/exportin1 vermittelt (Magin *et al.*, 1999; Boese *et al.*, 2000). Die mRNA der beiden akzessorischen Proteine Rec und Np9 entsteht jeweils durch ein zusätzliches Spleißen der *env*-mRNA (vgl. **Abb. 4b**). Die in Typ II Proviren zum Spleißen der *rec*-mRNA genutzte *splice donor site* liegt hierbei im 3'-Bereich der 292 bp Deletions-Region und die *splice acceptor site* findet sich etwa 110 bp stromauf des 3'-LTRs (Löwer *et al.*, 1995). In Proviren des Typs I wird zum Spleißen der *np9*-mRNA eine alternative *splice donor site* unmittelbar stromauf der Deletions-Region genutzt (Armbruster *et al.*, 2002).

Neben dem großen Teil der evolutiv alten HML-2-Proviren wurden auch jüngere Vollängen-Proviren beschrieben, von denen einige spezifisch im Menschen vorkommen und somit erst nach Aufspaltung der evolutiven Linien von Mensch und Schimpanse, die vor etwa 6 Millionen Jahren stattfand, im menschlichen Genom integriert wurden (Barbulescu *et al.*, 1999). Auch wenn die Mehrheit der HML-2-Proviren in zwei allelischen Kopien im menschlichen Genom vorhanden ist, wurden einige polymorphe HML-2-Loci beschrieben, die noch nicht in der Population fixiert sind und von denen sich einige erst vor weniger als 200.000 Jahren in die menschliche Keimbahn integriert haben (Turner *et al.*, 2001, Hughes und Coffin, 2004). Hierunter finden sich u.a. die Proviren HERV-K113 auf Chromosom 19p12, das intakte ORFs aller retroviralen Gene aufweist (Turner *et al.*, 2001; Beimforde *et al.*, 2008), sowie das HERV-K115-Provirus auf Chromosom 8p23.1. HERV-K115 kodiert für ein im Carboxy-Terminus verändertes Gag-Protein sowie Vollängen Rec- und Env-Proteine (Turner *et al.*, 2001). Polymorphe Loci kommen in Teilen der Bevölkerung als Vollängen-Provirus oder als solitärer LTR vor, bzw. findet sich in einigen Fällen lediglich eine leere Präintegrationsstelle, sofern das Provirus noch nicht in der Population fixiert ist (Burmeister *et al.*, 2004; Macfarlane und Simmonds 2004; Moyes *et al.*, 2005; Belshaw *et al.*, 2005). Auch das bereits erwähnte HERV-K(HML-2.HOM)-Provirus stellt ein Beispiel für einen polymorphen HML-2-Locus dar, da das Provirus in Teilen der Bevölkerung in Form eines Tandem-Provirus vorliegt, wobei sich zwei nahezu identische Vollängen-Proviren einen dazwischenliegenden LTR teilen (Reus *et al.*, 2001a).

Im Menschen finden sich einige HML-2-Proviren, die noch intakte ORFs für die retroviralen Proteine Gag, Pro, Pol und Env aufweisen (Loewer *et al.*, 1993; Mayer *et al.*, 1997a, 1997b; Tönjes *et al.*, 1999; Barbulescu *et al.*, 1999). So kodiert beispielsweise das HERV-K10-Provirus, das auf Chromosom 5q33.3 lokalisiert ist (Mayer *et al.*, 2011), ein Vollängen Gag-Protein und eine funktionelle Protease (Mueller-Lantzsch *et al.*, 1993). Zudem konnte für das HERV-K10-Provirus in prokaryotischen Zellen eine enzymatische Aktivität der Integrase (Kitamura *et al.*,

1996) sowie der Reversen Transkriptase gezeigt werden (Berkhout *et al.*, 1999). Ein Beispiel für ein nahezu intaktes Volllängen-Provirus mit ORFs der retroviralen Proteine, das lediglich eine Mutation in einem konservierten RT-Bereich aufweist, stellt das bereits erwähnte HERV-K(HML-2.HOM)-Provirus auf Chromosom 7p22.1 dar (Barbulescu *et al.*, 1999; Mayer *et al.*, 1999; Tönjes *et al.*, 1999). Auch wenn zahlreiche HML-2-Proviren mit intakten ORFs existieren und in früheren Experimenten retrovirusähnliche Partikel (*retrovirus-like particles*, RVLPS), die HML-2-Sequenzen enthielten, detektiert wurden (Boller *et al.*, 1993), konnte bisher keine Infektiosität von HML-2 nachgewiesen werden. Dies könnte darauf zurückzuführen sein, dass das etwa 80 - 90 kDa große Env-Volllängen-Protein nicht in die Oberflächen- und Transmembran-Glykoproteine prozessiert und somit auch nicht in die Zellmembran eingebaut wird (Tönjes *et al.*, 1997). In einer späteren Studie wurde zwar *in vitro* die Prozessierung und Expression eines HML-2 Env-Proteins beschrieben (Dewannieux *et al.*, 2005), jedoch ist zur Prozessierung des Proteins *in vivo* bisher nichts bekannt. Sollte die Spaltung des Env-Proteins *in vivo* nicht stattfinden, wäre eine Infektion weiterer Zellen, die über Erkennung von Rezeptormolekülen durch die viralen Oberflächenproteine vermittelt wird, nicht möglich. Zudem könnten frühere Rezeptormoleküle an der Oberfläche der Ziel-Zellen mittlerweile stark verändert vorliegen, was eine Erkennung durch die viralen Partikel verhindern würde (Mayer und Meese, 2005).

### 2.3.1.2. Die Rolle von Retroviren in der Pathogenese von Krebs

Der Entstehung von Krebs liegen komplexe mehrstufige Prozesse zugrunde, die das natürliche Gleichgewicht zwischen Zellproliferation und Apoptose stören. Durch Akkumulation verschiedenster Mutationen in Proto-Onkogenen und Tumorsuppressor-Genen kommt es zur malignen Transformation von Zellen, deren ungehinderte Proliferation sonst durch Kontrollmechanismen unterbunden wird (Wood *et al.*, 2007). Zu den Umweltfaktoren, die mit der Tumorgenese in Verbindung stehen, gehören, neben Strahlung und chemischen Kanzerogenen, auch Viren.

Die Identifizierung von ERV-Sequenzen im Menschen wurde erst durch vorangehende Studien exogener Retroviren ermöglicht, die in Tieren Krebs induzieren können (Martin *et al.*, 1981; Callahan *et al.*, 1982). In Tieren werden, Viren betreffend, zwei Haupt-Mechanismen der Tumorgenese unterschieden, die sog. *oncogene capture* sowie Insertionsmutagenese. Im Falle der *oncogene capture* wird das virale Onkogen *v-onc* von akut transformierenden Retroviren in die infizierten Zellen eingebracht, was in einem vermehrten Zellwachstum resultiert. Die *v-onc*-

Gene stammen hierbei ursprünglich von homologen zellulären *c-onc*-Proto-Onkogenen ab, die die Zellproliferation und Zelldifferenzierung regulieren, und wurden durch Rekombinationsereignisse während der reversen Transkription in das retrovirale Genom integriert (Sinkovics, 1988; Vogt, 1997). Im Falle von nicht-akuten Retroviren, die keine *v-onc* Gene besitzen, werden Tumoren durch Insertionsmutagenese verursacht. Hierbei wird entweder die Funktion von Tumorsuppressor-Genen durch Integration der retroviralen Sequenzen in diese gestört, oder es kommt zur Expression zellulärer Proto-Onkogene durch Integration retroviraler Sequenzen in benachbarte Bereiche, wobei die Expression durch regulatorische Sequenzen innerhalb der LTRs aktiviert wird (Ruprecht *et al.*, 2008).

Das erste infektiöse menschliche Retrovirus, HTLV-1 (*human T cell leukemia virus 1*), Verursacher der adulten T-Zell-Leukämie (ATL), wurde erstmals 1980 beschrieben (Poiesz *et al.*, 1980). Die maligne Zelltransformation resultiert in diesem Fall aus der onkogenen Wirkung des regulatorischen Proteins Tax (Curren *et al.*, 2012). Auch eine Infektion mit HIV-1 (*human immunodeficiency virus type 1*) ist indirekt, aufgrund fehlender immunologischer Kontrolle neoplastischer Zellen, mit einem erhöhten Tumoraufkommen assoziiert. Daneben wurde diskutiert, dass HIV-1 durch Insertionsmutagenese die Aktivierung von Onkogenen in Non-Hodgkin Lymphomen begünstigt (Shiramizu *et al.*, 1994). HTLV-1 ist jedoch bis heute das einzige Beispiel für ein exogenes Retrovirus, das im Menschen erwiesenermaßen direkt an der Krebsentstehung beteiligt ist, wobei es jedoch nur in einem geringen Anteil HTLV-1 infizierter Personen und oft erst nach langer Zeit zur Ausbildung von ATL kommt (Vogt, 1997). Für die Krankheitsentstehung werden somit vermutlich weitere Cofaktoren benötigt, da eine retrovirale Replikation in den meisten Fällen aktiv durch Faktoren wie die APOBEC (*apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like*) Cytidin-Deaminase-Familie und das TRIM (*tripartite motif*) 5 $\alpha$ -Protein unterbunden werden kann (Goff, 2004). Beide Faktoren interagieren hierbei auch mit murinen und humanen ERVs und könnten somit auch einen potentiell pathogenen Einfluss von ERVs kontrollieren. So wird die Retrotransposition muriner Retroelemente durch APOBEC 3G inhibiert (Esnault *et al.*, 2005). Auch für ein rekonstruiertes HERV-K(HML-2)-Provirus, dessen Sequenz vermutlich der des HML-2-Vorläufers entspricht, der sich ursprünglich in das menschliche Genom integriert hatte, konnte eine Inhibition durch APOBEC 3F gezeigt werden (Lee und Bieniasz, 2007). Solche zellulären Verteidigungsmechanismen könnten somit zur Kontrolle der potentiellen Pathogenität von ERVs im Menschen beitragen (Ruprecht *et al.*, 2008).

### 2.3.1.3. Die potentielle Rolle von HERVs in der Krebs-Pathogenese

Aufgrund der Ähnlichkeit endogener Retroviren zu infektiösen exogenen Retroviren wird seit vielen Jahren eine mögliche Involvierung in der Pathogenese diverser Krebsarten diskutiert. Eine direkte Beteiligung von ERVs an der Tumorgenese konnte bisher in Mäusen und Schafen nachgewiesen werden. In Mäusen verursacht beispielsweise das endogene *murine leukemia virus* (MLV) Leukämie und das bereits erwähnte *mouse mammary tumor virus* (MMTV), das strukturelle Ähnlichkeiten zu HERV-K aufweist, die Entstehung von Tumoren der Brustdrüsen (Stocking und Kozak, 2008). In Schafen sind endogene Vertreter des *Jaagsiekte sheep retrovirus* (JSRV) an der Entstehung von Lungenkarzinomen beteiligt (Spencer und Palmarini, 2012). In allen Fällen existieren hier jedoch auch exogene Retrovirus-Varianten, die ebenfalls die Entstehung von Tumoren verursachen können, welche im Falle von HERVs bisher nicht nachgewiesen werden konnten.

Es wurden verschiedene Mechanismen diskutiert, durch die HERVs aktiv an der Tumorgenese beteiligt sein könnten. Wie bereits beschrieben, beruht die kanzerogene Eigenschaft von ERVs in Tieren oft auf Insertionsmutagenese. Dies setzt jedoch voraus, dass sich neue retrovirale Kopien in den somatischen oder Keimzellen des Wirtsorganismus bilden (Boeke und Stoye, 1997). Bis heute wurde kein replikationskompetentes HERV-Provirus identifiziert, das durch einen solchen Mechanismus eine Rolle in der Krebs-Pathogenese spielen könnte. Aufgrund der mehr oder weniger vollständigen Struktur einiger Proviren und der Tatsache, dass einige Proviren erst nach evolutiver Aufspaltung der Linien von Mensch und Schimpanse integriert wurden (Medstrand und Mager, 1998; Barbulescu und Turner, 1999), stellt die HML-2-Gruppe jedoch den wahrscheinlichsten Vertreter replikationsfähiger Elemente dar (Bannert und Kurth, 2004). Zumal gezeigt werden konnte, dass es sich bei HML-2 um die einzige HERV-Gruppe handelt, die retrovirus-ähnliche Partikel (*retrovirus-like particles*, RVLPs) produzieren kann (Boller *et al.*, 2008). Die Tatsache, dass in der menschlichen Genomsequenz bisher keine replikationskompetenten Proviren identifiziert wurden, schließt nicht aus, dass solche Sequenzen tatsächlich existieren. Es wurde diskutiert, dass immer noch einige aktive und infektiöse HML-2-Proviren in der Keimbahn existieren könnten und in den Individuen, die sie tragen, eine pathogene Wirkung zeigen. Aufgrund der Pathogenität erreichen diese jedoch keine hohen Allel-Frequenzen in der Population und werden erst nach Akkumulation inaktivierender Mutationen fixiert (Belshaw *et al.*, 2005). Neben diesen pathogenen Mechanismen, die eine Bewegung von HERVs voraussetzen, wäre eine direkte onkogene Wirkung retroviraler Proteine, wie im Falle des bereits im Zusammenhang mit HTLV-1 erwähnten Tax-Proteins, denkbar. Hierbei waren in zahlreichen Studien v.a. die akzessorischen Proteine Rec und Np9 von

Interesse, auf die weiter unten im Detail eingegangen wird. Außerdem könnten Env-Proteine mit fusionierenden Eigenschaften, wie das von der HERV-W-Gruppe kodierte Syncytin-1-Protein, auch in Tumoren an der Zellfusion beteiligt sein (Strick *et al.*, 2007). Eine weitere mögliche Rolle an der Tumorgenese könnte HML-2 durch immunsuppressive Eigenschaften der retroviralen Env-Proteine zukommen, die die normale Kontrolle des Zellwachstums durch das Immunsystem unterbinden. Solche immunsupprimierenden Eigenschaften wurden beispielsweise für die TM-Proteindomänen und insbesondere eine darin lokalisierte hoch konservierte immunsuppressive Domäne (ISD) des katzenspezifischen FeLV (*feline leukemia virus*) gezeigt (Haraguchi *et al.*, 1997, Cianciolo *et al.*, 1985). Auch in zehn von 16 HERV-Elementen, die komplette *env*-ORFs aufweisen, konnte eine solche Domäne nachgewiesen werden, wobei sich jedoch in HML-2-Proviren keine ISD fand (de Parseval *et al.*, 2003).

#### **2.3.1.4. Die mögliche Rolle der HERV-K(HML-2)-Gruppe in der Tumorphagenese**

Auch wenn eine direkte oder indirekte Beteiligung humaner endogener Retroviren an der Krebsentstehung im Menschen bisher nicht gezeigt werden konnte, wurden in zahlreichen untersuchten Tumorproben verschiedener Krebsarten und verschiedenen Krebs-Zelllinien HML-2-Transkripte und retrovirale Proteine nachgewiesen, darunter Keimzelltumoren und Melanome.

Bei Keimzelltumoren (*germ cell tumours*, GCTs) handelt es sich um die häufigste Tumorerkrankung junger Männer in westlichen Industrienationen (Adami *et al.*, 1994). GCTs entwickeln sich aus entarteten Zellen der Keimbahn und es werden allgemein zwei Typen, Seminome und Nichtseminome, unterschieden (Horwich *et al.*, 2006; Vasdev *et al.*, 2013). Schon früh wurden retrovirus-ähnliche Partikel in verschiedenen GCT-Zelllinien detektiert (Bronson *et al.*, 1979; Löwer *et al.*, 1981), die daraufhin als HDTV (*human teratocarcinoma-derived virus*)-Partikel benannt wurden. Nachdem sich herausstellte, dass diese Partikel von HML-2-Loci kodiert werden, wurde ihr Name zu HDTV/HERV-K geändert (Boller *et al.*, 1993; Löwer *et al.*, 1993). Alle bisher analysierten HDTV/HERV-K-Partikel stellten sich jedoch als morphologisch defekt und nicht infektiös heraus (Löwer *et al.*, 1984; Bieda *et al.*, 2001; Boller *et al.*, 2008). Auch wenn bisher keine infektiösen Partikel detektiert werden konnten, wurde eine mögliche Involvierung von HERV-K(HML-2) in der GCT-Pathogenese durch Detektion von HML-2-RNA und retroviralen Proteinen weiter unterstützt (Bieda *et al.*, 2001). So wurden beispielsweise durch *in situ*-Hybridisierung HML-2 *gag*- und *env*-Transkripte in bestimmten GCT-Proben detektiert (Herbst *et al.*, 1996) und HML-2 Gag-Proteine wurden durch

Immunfluoreszenzfärbung in Seminom-Patienten (Sauter *et al.*, 1995) und in GCT-Zelllinien nachgewiesen (Bieda *et al.*, 2001). Zudem zeigten einige Keimzelltumor-Patienten einen hohen Antikörper-Titer gegen HML-2 Gag- und Env-Proteine, was auf eine möglicherweise spezifische Expression und potentielle Involvierung retroviraler Proteine in der Krankheitspathogenese hindeuten könnte (Sauter *et al.*, 1995, 1996; Boller *et al.*, 1997; Kleiman *et al.*, 2004).

Eine weitere Krebsart, bei der eine mögliche Pathogenität humaner ERVs diskutiert wird, stellen Melanome dar, die bösartigste Form des Hautkrebses mit jährlich steigender Häufigkeit (Gordon, 2013). Melanome entstehen durch unkontrollierte Proliferation Pigment-produzierender Melanozyten der Haut, der Schleimhäute sowie der Uvea (Dennis, 1999). In einigen Melanom-Zelllinien wurde die Produktion retrovirus-ähnlicher Partikel mit Reverser Transkriptase-Aktivität nachgewiesen, in denen sich HML-2-ähnliche Sequenzen sowie Gag- und Env-Proteine fanden (Muster *et al.*, 2003). Es stellte sich jedoch heraus, dass die RVLPs defekt und nicht infektiös sind (Büscher *et al.*, 2005). Eine in Melanomen erhöhte HML-2-Expression könnte auf eine durch Demethylierung bedingte erhöhte Promotoraktivität spezifischer HML-2 5'-LTRs zurückzuführen sein (Stengel *et al.*, 2010). Wie schon im Falle der GCTs wurden auch in Melanom-Geweben und -Zelllinien HML-2-Transkripte sowie retrovirale Gag-, Env-, Rec- und Np9-Proteine nachgewiesen (Muster *et al.*, 2003; Büscher *et al.*, 2005, 2006; Reiche *et al.*, 2010). In Melanom-Patienten wurden HERV-K-spezifische Antikörper detektiert, wobei die Immunantwort gegen HML-2 Gag- und Env-Proteine schwach mit einer geringeren Lebenserwartung korrelierte (Hahn *et al.*, 2008). In *in vitro*-Experimenten bewirkte eine Stress-induzierte HML-2-Expression in adhären Melanom-Zelllinien den Übergang zu einem nicht-adhären, maligneren Phänotyp (Serafino *et al.*, 2009). Des Weiteren wurde eine spezifische Aktivierung der HML-2-Expression infolge von UV-Bestrahlung berichtet, dem bekanntesten Risikofaktor, der die Melanomentstehung begünstigt (Schanab *et al.*, 2011).

### **2.3.1.5. Die potentielle Rolle der HML-2-Proteine Rec und Np9 in der Tumorgenese**

Von besonderem Interesse in der Krebs-Pathogenese sind auch die von HML-2-Proviren des Typs I und II kodierten akzessorischen Proteine Np9 und Rec. Das Np9-Protein wurde zunächst in zahlreichen Tumor-Zelllinien sowie transformierten Zelllinien detektiert, wo es überwiegend im Zellkern lokalisiert ist. Eine Expression in gesunden Zellen konnte nicht nachgewiesen werden (Armbruster *et al.*, 2002). Eine mögliche Rolle in der Tumorgenese könnte Np9 hierbei aufgrund seiner Interaktion mit dem LNX (*ligand of Numb protein X*)-Protein zukommen. Bei LNX handelt es sich um eine Ubiquitin-Ligase, die den proteasomalen Abbau des Numb-

Proteins bewirkt, das sonst als Antagonist die Expression des Notch-Proteins supprimiert und so die Zelldifferenzierung und Proliferation reguliert (Armbruster *et al.*, 2004). In einer Studie konnte in etwa der Hälfte der untersuchten menschlichen Brustkarzinome eine Störung des Numb/Notch-Signalweges aufgrund Ubiquitin-vermittelter Degradation des Numb-Proteins nachgewiesen werden (Pece *et al.*, 2004). Sowohl für Rec als auch für Np9 konnte eine Interaktion mit dem PLZF (*promyelocytic zinc finger*)-Protein gezeigt werden, das als transkriptioneller Repressor an der Kontrolle des Zellzyklus beteiligt ist. Eine Coexpression von Np9 und Rec inhibiert hierbei das PLZF-Protein, was in einer Überexpression des Proto-Onkogens *c-myc* resultiert, das sonst durch PLZF reprimiert wird (Denne *et al.*, 2007). In transgenen Mäusen bewirkte die induzierte Expression von Rec eine gestörte Keimzellentwicklung sowie die Entwicklung von testikulärem *Carcinoma in situ* (Galli *et al.*, 2005).

Es bleibt jedoch vorerst unklar, ob das Vorhandensein retroviraler Proteine in Tumorgewebe auf eine funktionelle Rolle und direkte Beteiligung an der Tumorgenese schließen lässt, oder ob eine solche Expression als eher zufälliges Ereignis anzusehen ist, das auf einer deregulierten Genexpression beruht. Eine wichtige Rolle könnte hierbei auch polymorphen HML-2-Proviren zukommen, da diese nur in einem Teil der Bevölkerung vorkommen und von ihnen ausgehende Transkripte und Proteine, im Gegensatz zu denen anderer Proviren, nicht in allen Individuen exprimiert würden (Moyes *et al.*, 2007). Des Weiteren könnten spezifische Polymorphismen innerhalb proteinkodierender HERV-Gene in der Expression potentiell pathogener Proteine resultieren (de Parseval und Heidmann, 2005). Auch aufgrund epigenetischer Veränderungen, wie der Hypomethylierung von HERV-Sequenzen in Tumorzellen, könnte es zur Expression retroviraler Proteine oder Epitope kommen, die in gesunden Geweben nicht exprimiert werden (Götzinger *et al.*, 1996; Lavie *et al.*, 2005).

### **2.3.2. Die HERV-W-Gruppe humaner endogener Retroviren**

Die HERV-W-Gruppe humaner endogener Retroviren wurde vermutlich nach der evolutiven Trennung von Alt- und Neuweltaffen im Genom integriert (Mayer und Meese, 2005). Strukturell unterscheidet man zwei Hauptgruppen. Zum einen findet man etwa 77 HERV-W-Proviren mit mehr oder weniger vollständigen internen Sequenzen und kompletten flankierenden LTRs sowie etwa 343 komplette solitäre LTRs. Daneben gibt es etwa 176 Sequenzen, die durch unvollständige LTRs gekennzeichnet sind. Diesen fehlt der U3-Bereich des 5'-LTRs sowie der U5-Bereich des 3'-LTRs und sie werden oft durch einen Poly(A)-Bereich am 3'-Ende flankiert. Ein großer Teil dieser zweiten Gruppe ist zudem durch unterschiedlich große Deletionen im

5'-Bereich gekennzeichnet, was insgesamt auf eine Entstehung durch einen L1-vermittelten Retrotranspositionsmechanismus hindeutet (Costas, 2002; Pavlicek *et al.*, 2002).

Ein Großteil der HERV-W-Loci kodiert aufgrund des hohen evolutiven Alters nicht mehr für funktionelle Proteine (Voisset *et al.*, 1999; Costas, 2002). Eine Ausnahme stellt hierbei der bereits erwähnte *syncytin-1/ERVWE1/ERVW-1*-Locus auf Chromosom 7q21.2 dar. Dieser kodiert ein 80 kDa großes Volllängen Env-Protein namens Syncytin-1, welches das erste Beispiel für ein HERV-kodiertes Protein mit potentieller biologischer Funktion im Menschen darstellt. Während der Plazentaentwicklung vermittelt Syncytin-1, dessen Expression im Trophoblasten stark erhöht ist, wahrscheinlich die Membranfusion während der Bildung des Syncytiotrophoblasten (Mi *et al.*, 2000; Frendo *et al.*, 2003). Auch *in vitro* konnte eine Syncytin-1 abhängige Membranfusion verschiedener Zelllinien zu Synzytien gezeigt werden (Blond *et al.*, 2000). Ein weiterer HERV-W-Locus auf Chromosom Xq22.3, *ERVWE2/ERVW-2*, kodiert ein Amino-terminal verkürztes Env-Protein, N-Trenv, dessen Expression *ex vivo* in menschlichen PBMCs nachgewiesen werden konnte. Eine Expression *in vivo* sowie die mögliche Funktion des Proteins sind jedoch noch unklar (Roebke *et al.*, 2010).

Die ersten retroviralen Sequenzen der HERV-W-Gruppe wurden aus Plasma von Multiple Sklerose Patienten isoliert (Perron *et al.*, 1991) und zahlreiche Studien diskutierten seither eine mögliche Rolle von HERV-W in der Pathogenese von Multipler Sklerose (Antony *et al.*, 2011).

### **2.3.2.1. Die potentielle Rolle von HERV-W in der Pathogenese von Multipler Sklerose**

Bei Multipler Sklerose (MS) handelt es sich um eine Neuro-Immunerkrankung, die durch chronische Entzündungen des zentralen Nervensystems (ZNS) gekennzeichnet ist (Sospedra und Martin, 2005). Weltweit sind über eine Million Individuen betroffen, wobei die Krankheit oft im Alter von 20 bis 40 Jahren auftritt und mit zahlreichen physischen und kognitiven Störungen sowie einer verkürzten Lebenserwartung assoziiert ist (Paty, 1997). Im Verlauf der Krankheit kommt es, vermutlich durch körpereigene Immunreaktionen auf die neuronale Myelinscheide und Zerstörung dieser, in der weißen Substanz von Gehirn und Rückenmark zur Entstehung entzündlicher Entmarkungsherde, sog. *Plaques*. Da es im gesamten ZNS zur Entstehung solcher *Plaques* kommen kann, werden zahlreiche verschiedene neurologische Symptome bei MS-Patienten beobachtet. Die genaue Krankheits-Ursache ist bis heute nicht geklärt, es werden jedoch sowohl genetische Ursachen als auch Umwelteinflüsse diskutiert (Marrie 2004; De Jager *et al.*, 2009). Aktuelle Studien deuten darauf hin, dass infektiöse Agenzien, die an der

MS-Entwicklung beteiligt sind, in der Population vermutlich weit verbreitet sind und bestimmte genetische Eigenschaften benötigt werden, um zur Krankheitsentwicklung zu führen. Auch eine Rolle infektiöser viraler Pathogene wird diskutiert. Es wurden verschiedene Mechanismen vorgeschlagen, durch die Viren eine autoreaktive Immunantwort auslösen könnten. Neben der generellen Virus-induzierten Aktivierung des Immunsystems könnte es durch virale Genprodukte zur Anregung spezifischer Immunantworten kommen, die Kreuzreaktionen mit Autoantigenen bewirken (Kakalacheva *et al.*, 2011).

Aufgrund der Detektion retroviraler Sequenzen in MS-Patienten wurde in mehreren Studien das potentielle Vorhandensein eines MS-assoziierten exogenen replikationsfähigen HERV-W-Retrovirus diskutiert, das als MSR (*MS associated retrovirus*) benannt wurde (Blond *et al.*, 1999; Mameli *et al.*, 2007, 2009; Garcia-Montojo *et al.*, 2013). Spätere genauere Untersuchungen der retroviralen Sequenzen ergaben jedoch, dass diese eher Transkripte endogener HERV-W-Sequenzen darstellen oder *in vitro* während der cDNA-Synthese durch Springen der reversen Transkriptase zwischen RNA-Molekülen verschiedener HERV-W-Loci generiert wurden (Laufer *et al.*, 2009).

In weiteren Studien wurden HERV-W-Transkripte sowie Env-Proteine in akut demyelinisierenden Läsionen in MS-Patienten berichtet (Antony *et al.* 2004, 2007b; Roebke *et al.*, 2010). Die Oberflächendomäne eines weiteren Env-Proteins löste eine Immunantwort aufgrund der Aktivierung der CD14- und TLR4-Rezeptoren aus (Rolland *et al.*, 2006). Für das gleiche Env-Protein wurde die Inhibition der Differenzierung von Oligodendroglia-Vorläuferzellen in der direkten Umgebung von MS-Läsionen berichtet (Kremer *et al.*, 2013). Für das Syncytin-1-Protein wurde die Überexpression in Gliazellen in demyelinisierenden Läsionen von MS-Patienten berichtet. Des Weiteren bewirkte die Syncytin-1-Expression in Astrozyten die Freisetzung von reaktiven Sauerstoffspezies, die zytotoxisch auf Oligodendrozyten wirkten, und im Mausmodell resultierte die Implantation von Syncytin-1 in das *Corpus callosum* in einer Demyelinisierung (Antony *et al.*, 2004). In transgenen Mäusen, die Syncytin-1 unter Kontrolle des GFAP (*glial fibrillary acidic protein*)-Promotors exprimierten, wurden neuronale Entzündungen und reduzierte Mengen an Myelin-Protein im *Corpus callosum* berichtet (Antony *et al.*, 2004, 2007a).

RNA-Transkripte der HERV-W-Gruppe konnten in verschiedenen Studien nachgewiesen werden, wobei diese teils unterschiedliche Ergebnisse lieferten. So wurde eine erhöhte Transkription von HERV-W *env*-RNA in Gehirnproben (Antony *et al.*, 2006, 2007b) und PBMCs (Mameli *et al.*, 2007) von MS-Patienten gegenüber gesunden Kontrollen beschrieben. Die

Erhöhte *env*-Transkription in PBMCs konnte in einer anderen Studie jedoch nicht bestätigt werden (Antony *et al.*, 2006). Es wurden jedoch kürzlich potentielle technische Fehler in zwei dieser Studien diskutiert (Garson *et al.*, 2009). Durch Sequenzierung HERV-W *env*-spezifischer RT-PCR-Produkte konnten sieben transkribierte HERV-W-Loci in PBMCs aus MS-Patienten und gesunden Kontrollen identifiziert werden, wobei sich kein signifikanter Unterschied zwischen MS und gesunden Proben ergab (Laufer *et al.*, 2009).

## 2.4. Zielsetzungen der Arbeit

Humane endogene Retroviren stellen einen wichtigen Bestandteil des menschlichen Transkriptoms dar und beeinflussen durch regulatorische Sequenzen sowie die Promotoraktivität ihrer LTRs auch die Transkription benachbarter Gene. In allen bisher untersuchten Geweben und Zelllinien wurden HERV-Transkripte verschiedener Gruppen detektiert, wobei sich die Transkription jedoch zwischen den Geweben und Zelllinien unterschied. Die Transkription muss somit einer spezifischen Kontrolle unterliegen.

Im Gegensatz zu Studien, in denen die Transkription verschiedener HERV-Gruppen insgesamt untersucht wurde, gibt es jedoch noch vergleichsweise wenig Information darüber, von welchen genomischen HERV-Loci diese Transkripte stammen und welche spezifischen Loci generell in verschiedenen Geweben und Zelltypen aktiv sind. Als Beitrag zum besseren Verständnis des Beitrages von HERVs zum menschlichen Transkriptom sowie insbesondere ihrer möglichen Rolle in der Pathogenese verschiedener Krankheiten, sollten in der vorliegenden Arbeit daher transkribierte HERV-Loci in verschiedenen kranken und Normalgeweben identifiziert werden.

Im Speziellen sollte hierbei die Transkription von Loci der HERV-K(HML-2)-Gruppe in verschiedenen Tumorerkrankungen untersucht werden. Hierfür sollten HML-2-Transkriptionsprofile im Kontext von Melanomen analysiert werden, wobei hier zusätzlich potentiell veränderte Transkriptionsmuster nach einer UVB-Bestrahlung bestimmt werden sollten, sowie in Keimzelltumoren und Prostatakarzinomen und verschiedenen Normalgeweben. Des Weiteren sollte die Transkription von Loci der HERV-W-Gruppe im Kontext von Multipler Sklerose untersucht und umfassende Transkriptionsprofile von HERV-W-Loci erstellt werden.

Generell sollten hierbei die Fragen geklärt werden, wie sich HERV-Transkriptionsprofile in verschiedenen Proben unterscheiden. Insbesondere war hierbei von Interesse, ob sich kranke und gesunde Vergleichsgewebe durch die Transkription charakteristischer HERV-Loci auszeichnen, die u.U. auch ORFs für retrovirale Proteine besitzen, die in der jeweiligen

Krankheit von Bedeutung sein könnten. Aus diesem Grund sollte die Kodierungskapazität transkribierte Loci näher untersucht werden, ebenso wie ihre Struktur und Genomlokalisierung, da die Transkription spezifischer Loci im Kontext einer Krankheit auch Hinweise auf relevante Veränderungen in für die Pathogenese relevanten Genomregionen liefern könnte.

In vorangehenden Studien (Flockerzi *et al.*, 2008; Laufer *et al.*, 2009) wurden bereits im kleineren Maßstab transkribierte HERV-Loci durch Generierung von HERV-Gruppen-spezifischen RT-PCR-Produkten und deren anschließende Sequenzierung identifiziert. In der vorliegenden Arbeit wurde die gleiche Methode zur Identifizierung transkribierter Loci gewählt, wobei hierbei zunächst gegenüber diesen Studien optimierte Amplifikationsstrategien etabliert werden sollten, um möglichst umfassende Transkriptionsanalysen zu ermöglichen. Es sollten zuverlässige Methoden etabliert werden, um aus cDNA-Proben in einer PCR genomweit möglichst viele Loci gleichzeitig zu amplifizieren. Hierdurch sollte, im Falle ihrer Transkription, die Detektion der größtmöglichen Anzahl verschiedener HERV-Loci ermöglicht werden, was zudem durch Generierung einer großen Zahl an cDNA-Sequenzen weiter unterstützt werden sollte. Die cDNA-Amplikons sollten, im Fall von HML-2 mittels Sanger-Sequenzierung sowie im Fall von HERV-W durch *Next Generation Sequencing*-Techniken, sequenziert und generierte Sequenzen aufgrund charakteristischer Nukleotidunterschiede spezifischen HERV-Loci im menschlichen Referenzgenom zugeordnet werden, was diese als in der Ausgangsprobe transkribiert identifiziert.

Zur Detektion potentiell veränderter HERV-Transkriptmengen im Kontext von Tumorerkrankungen und MS, sollten die Gruppen-spezifischen Transkriptmengen von HML-2 und HERV-W in kranken und gesunden Vergleichsgeweben mittels *Real-Time*-PCR quantifiziert werden. Des Weiteren sollten zur Klärung des Ursprungs detektierter Transkripte ausgewählte LTR-Elemente transkribierter Loci in Reporter-Gen-Assays auf ihre potentielle Promotoraktivität hin untersucht werden.

### 3. Material und Methoden

#### 3.1. Material

##### 3.1.1. Chemikalien

<b>Chemikalie</b>	<b>Hersteller</b>
Acrylamid/Bisacrylamid (30 %) (37,5 : 1)	Bio-Rad (Hercules, USA)
Agar, <i>Bacto</i> <sup>TM</sup>	Becton, Dickinson and Company (Le Pont de Claix Cedex, Frankreich)
Agarose, <i>Ultra Pure</i> <sup>TM</sup>	Invitrogen/LifeTechnologies (CarlsGenbank)
Ammoniumperoxodisulfat (APS)	Carl Roth (Karlsruhe, Deutschland)
Ampicillin	AppliChem (Darmstadt, Deutschland)
5-Brom-4-chlor-3-indoxyl-β-D-galactopyranosid (X-Gal)	AppliChem (Darmstadt, Deutschland)
Bromphenolblau	Carl Roth (Karlsruhe, Deutschland)
Calciumchlorid (CaCl <sub>2</sub> )	Serva (Heidelberg, Deutschland)
Chloroform	Sigma-Aldrich (St. Louis, USA)
Desoxynukleosid-Triphosphat Set	Roche (Basel, Schweiz)
Diethyldicarbonat (DEPC)	Carl Roth (Karlsruhe, Deutschland)
Dimethylformamid	Carl Roth (Karlsruhe, Deutschland)
Ethylendiamintetraessigsäure (EDTA) Dinatriumsalz Dihydrat	Carl Roth (Karlsruhe, Deutschland)
Eisessig	VWR (Radnor, USA)
Ethanol (C <sub>2</sub> H <sub>5</sub> OH)	Sigma-Aldrich (St. Louis, USA)
Ethidiumbromid	Carl Roth (Karlsruhe, Deutschland)
FCS ( <i>fetal calf serum</i> ) (Charge 1243-T)	Biochrom (Berlin, Deutschland)
FCS (für Tu2 % Medium)	PAA (Pasching, Österreich)
Fixierer	Kodak (Rochester, USA)
Formaldehyd	Carl Roth (Karlsruhe, Deutschland)
Formamid	Carl Roth (Karlsruhe, Deutschland)
Fotoentwickler	Kodak (Rochester, USA)
<i>GelStar</i> <sup>TM</sup> <i>Nucleic Acid Gel Stain</i>	Lonza (Rockland, USA)
Glucose	Carl Roth (Karlsruhe, Deutschland)
Glycerin	VWR (Radnor, USA)

Glycin	Carl Roth (Karlsruhe, Deutschland)
Hefeextrakt ( <i>Bacto</i> <sup>TM</sup> <i>Yeast Extract</i> )	Becton, Dickinson and Company (Le Pont de Claix Cedex, Frankreich)
<i>Hoechst Stain Solution 33258</i>	Sigma-Aldrich (St. Louis, USA)
<i>Insulin Solution (human)</i>	Sigma-Aldrich (St. Louis, USA)
Isopropanol (C <sub>3</sub> H <sub>7</sub> OH)	VWR (Radnor, USA)
Kaliumacetat (CH <sub>3</sub> CO <sub>2</sub> K)	VWR (Radnor, USA)
Kaliumchlorid (KCl)	Carl Roth (Karlsruhe, Deutschland)
Kaliumdihydrogenphosphat (KH <sub>2</sub> PO <sub>4</sub> )	Carl Roth (Karlsruhe, Deutschland)
Kanamycin	Sigma-Aldrich (St. Louis, USA)
Magnesiumchlorid (MgCl <sub>2</sub> )	Carl Roth (Karlsruhe, Deutschland)
Mangan(II)-chlorid (MnCl <sub>2</sub> )	Merck (Whitehouse Station, USA)
Methanol	BCD Chemie (Hamburg, Deutschland)
Milchpulver	Carl Roth (Karlsruhe, Deutschland)
3-(N-Morpholino)-Propansulfonsäure (MOPS)	Carl Roth (Karlsruhe, Deutschland)
MOPS-Natriumsalz	AppliChem (Darmstadt, Deutschland)
Nagellack, klar	p2 Cosmetics (Wiener Neudorf, Österreich)
Natriumacetat (CH <sub>3</sub> COONa)	Sigma-Aldrich (St. Louis, USA)
Natriumchlorid (NaCl)	VWR (Radnor, USA)
Natriumdesoxycholat	Sigma-Aldrich (St. Louis, USA)
Natriumdihydrogenphosphat (NaH <sub>2</sub> PO <sub>4</sub> )	Merck (Whitehouse Station, USA)
Natriumdodecylsulfat (SDS)	Carl Roth (Karlsruhe, Deutschland)
<i>Nonidet</i> <sup>®</sup> <i>P40</i>	Fluka Chemie AG (Buchs, Schweiz)
Paraformaldehyd	Merck (Whitehouse Station, USA)
Penicillin	Gibco/Life Technologies (Carlsbad, USA)
<i>ProLong</i> <sup>®</sup> <i>Gold Antifade</i>	Invitrogen/Life Technologies (Carlsbad, USA)
Streptomycin	Gibco/Life Technologies (Carlsbad, USA)
N,N,N',N'-Tetramethylethylendiamin (TEMED)	Sigma-Aldrich (St. Louis, USA)
<i>Tween</i> <sup>®</sup> <i>20</i>	Carl Roth (Karlsruhe, Deutschland)
TRIS-Base	Carl Roth (Karlsruhe, Deutschland)
Tris-HCl	Carl Roth (Karlsruhe, Deutschland)
Triton X-100	Carl Roth (Karlsruhe, Deutschland)
Trizol	Invitrogen/Life Technologies (Carlsbad, USA)

Trypton, <i>Bacto</i> <sup>TM</sup>	Becton, Dickinson and Company (Le Pont de Claix Cedex, Frankreich)
Xylencyanol	Carl Roth (Karlsruhe, Deutschland)

### 3.1.2. Lösungen und Puffer

<b>Puffer</b>	<b>Ansatz bzw. Hersteller</b>
DEPC-H <sub>2</sub> O	0,1 % [v/v] DEPC H <sub>2</sub> O <sub>dd</sub> ü. N. inkubiert bei 37 °C, autoklaviert
<i>Dulbecco's Phosphate-Buffered Saline</i> (DPBS)	Gibco/Life Technologies (Carlsbad, USA)
Laufpuffer SDS-PAGE (1x)	25 mM Tris 190 mM Glycin 0,1 % [w/v] SDS
<i>Luciferase Cell Lysis Buffer</i> (5x)	NEB (Ipswich, USA)
MOPS-Puffer (10x)	2 M MOPS 0,05 M Natriumacetat 0,01 M EDTA
PBS (1x)	100 mM NaH <sub>2</sub> PO <sub>4</sub> 130 mM NaCl eingestellt auf pH 7, autoklaviert
PBS-T	1x PBS 0,2 % [v/v] <i>Tween</i> <sup>®</sup> 20
RIPA-Puffer (1x)	25 mM Tris-HCl, pH 7,6 150 mM NaCl 1 % [v/v] <i>Nonidet</i> <sup>®</sup> P40 1 % [v/v] Natriumdesoxycholat 0,1 % [w/v] SDS
RNA-Ladepuffer (5x)	55 % [v/v] Formamid, deionisiert 1x MOPS 0,77 µg/µl Ethidiumbromid 0,06 % [v/v] Glycerin 0,06 % gesättigte Bromphenolblau-Lösung

	alles verdünnt in DEPC-H <sub>2</sub> O <sub>dd</sub>
SE-Puffer	75 mM NaCl 25 mM EDTA eingestellt auf pH 8, autoklaviert
Stopp-Ladepuffer (Agarosegele)	1x TAE-Puffer 70 % [v/v] Glycerin 0,04 M EDTA Bromphenolblau (~1 Spatelspitze pro 5 ml) Xylencyanol (~1 Spatelspitze pro 5 ml)
<i>Stripping Buffer</i> (1x)	200 mM Glycin 3,467 mM SDS 1 % [v/v] <i>Tween</i> <sup>®</sup> 20 eingestellt auf pH 2,2
TAE (50x)	2 M TRIS-Base 1 M Eisessig 0,05 M EDTA (pH 8)
TBS ( <i>Tris-Buffered Saline</i> ) (10x)	1,5 M NaCl 500 mM Tris-HCl, pH 7,5
TBS-T	1x TBS 0,05 % [v/v] <i>Tween</i> <sup>®</sup> 20
TE-Puffer (1x)	0,01 M TRIS 0,001 M EDTA eingestellt auf pH 8, autoklaviert
TfBI	30 mM CH <sub>3</sub> CO <sub>2</sub> K 50 mM MnCl <sub>2</sub> 100 mM KCl 10 mM CaCl <sub>2</sub> 15 % [v/v] Glycerol sterilfiltriert (0,2 µm Filter)
TfBII	10 mM MOPS-Na, pH7 75 mM CaCl <sub>2</sub> 10 mM KCl 15 % [v/v] Glycerol sterilfiltriert (0,2 µm Filter)

Transferpuffer <i>Western Blot</i> (1x)	25 mM TRIS 190 mM Glycin 0,1 % [w/v] SDS 20 % [v/v] Methanol
X-Gal	30 mg/ml X-Gal in Dimethylformamid

### 3.1.3. Kulturmedien

#### Kulturmedien für prokaryotische Zellen

Medium	Ansatz
LB-Medium	1 % [w/v] Trypton 0,5 % [w/v] Hefeextrakt 170 mM NaCl autoklaviert
LB-Agar-Platten	LB-Medium 1,5 % [w/v] Agar autoklaviert

Nach Abkühlen auf 55 °C wurde das Medium mit 100 µg/ml Ampicillin und 50 µg/ml X-Gal oder mit 30 µg/ml Kanamycin versetzt und in 55 cm<sup>2</sup> Petrischalen gegossen.

SOB-Medium	2 % [w/v] Trypton 0,5 % [w/v] Hefeextrakt 10 mM MgCl <sub>2</sub> 10 mM NaCl 2,5 mM KCl autoklaviert
SOC-Medium	SOB-Medium, autoklaviert 20 mM Glucose sterilfiltriert (0,2 µm Filter)

#### Kulturmedien für eukaryotische Zellen

Medium	Ansatz bzw. Hersteller
<i>DMEM</i>	Gibco/Life Technologies (Carlsbad, USA)
<i>Ham's F-12</i>	Biochrom (Berlin, Deutschland)

<i>L-15 Leibovitz Medium</i>	Biochrom (Berlin, Deutschland)
<i>McCoy's 5A (1x) + GlutaMAX™-I</i>	Gibco/Life Technologies (Carlsbad, USA)
<i>MCDB 153 Basal Medium</i>	Biochrom (Berlin, Deutschland)
<i>Melanocyte Growth Medium</i>	PromoCell (Heidelberg, Deutschland)
<i>Opti-MEM® I (1x) + GlutaMAX™-I</i>	Gibco/Life Technologies (Carlsbad, USA)
<i>RPMI-1640</i>	Gibco/Life Technologies (Carlsbad, USA)
<i>Tu2 %-Medium</i>	80 % [v/v] <i>MCDB 153 Basal Medium</i>
	20 % [v/v] <i>L-15 Leibovitz Medium</i>
	2 % [v/v] FCS
	1,68 mM CaCl <sub>2</sub>
	2,5 ng/ml Insulin

### 3.1.4. Reaktions-Kits

#### **Kit**

*Agencourt CosMCPrep*  
*β-Galactosidase Enzyme Assay*  
*System with Reporter Lysis Buffer*  
*BioLux® Gaussia Luciferase Assay Kit*  
*Amersham ECL Prime Western Blotting*  
*Detection Reagent*  
*Omniscript RT Kit*  
*peqGOLD Plasmid Miniprep Kit I*  
*QIAGEN Plasmid Midi Kit*  
*QIAprep Spin Miniprep Kit*  
*QIAquick Gel Extraction Kit*  
*QIAquick PCR Purification Kit*  
*Quant-iT™ dsDNA HS Assay Kit*  
*(0,2 - 100 ng)*  
*RNaseOUT™ Ribonuclease Inhibitor*  
*RNeasy Mini Kit*  
*TURBO DNA-free™ Kit*  
*TurboFect™ Transfection Reagent*

#### **Hersteller**

Beckman Coulter Genomics (Danvers, USA)  
 Promega (Fitchburg, USA)  
 NEB (Ipswich, USA)  
 GE Healthcare (Uppsala, Schweden)  
 QIAGEN (Hilden, Deutschland)  
 PEQLAB (Erlangen, Deutschland)  
 QIAGEN (Hilden, Deutschland)  
 QIAGEN (Hilden, Deutschland)  
 QIAGEN (Hilden, Deutschland)  
 QIAGEN (Hilden, Deutschland)  
 Invitrogen/Life Technologies (Carlsbad, USA)  
 Invitrogen/Life Technologies (Carlsbad, USA)  
 QIAGEN (Hilden, Deutschland)  
 Applied Biosystems (California, USA)  
 Thermo Fisher Scientific (Waltham, USA)

### 3.1.5. Enzyme

<b>Enzym</b>	<b>Hersteller</b>
<i>Bam</i> HI-HF™	NEB (Ipswich, USA)
<i>HOT FIREPo</i> ® DNA Polymerase	Solis Biodyne (Tartu, Estland)
<i>Phusion</i> ® <i>High-Fidelity</i> DNA Polymerase	Finnzymes/Thermo Fisher Scientific (Waltham, USA)
Proteinase K	Carl Roth (Karlsruhe, Deutschland)
<i>Taq</i> DNA Polymerase, <i>recombinant</i>	Invitrogen/Life Technologies (Carlsbad, USA)
T4 DNA Ligase	Promega (Fitchburg, USA)
0,05 % Trypsin-EDTA	Gibco/Life Technologies (Carlsbad, USA)

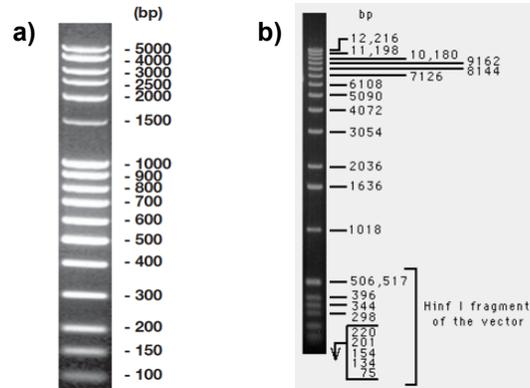
### 3.1.6. Antikörper

<b>Antikörper</b>	<b>Hersteller</b>
$\alpha$ -GFP	AG Schlenstedt, Medizinische Biochemie und Molekularbiologie, Universitätsklinikum Homburg, polyklonaler Kaninchen-Antikörper, affinitäts-gereinigt zur Verfügung gestellt
$\alpha$ - $\beta$ -Actin (monoklonal, Maus)	Sigma-Aldrich (St. Louis, USA)
<i>HRP-conjugated Goat-<math>\alpha</math>-Rabbit IgG</i> (H+L)	Dianova (Hamburg, Deutschland)
<i>HRP-conjugated Sheep-<math>\alpha</math>-Mouse IgG</i> (H+L)	Dianova (Hamburg, Deutschland)

### 3.1.7. Größenstandards für die Agarose- und Polyacrylamid-Gelelektrophorese

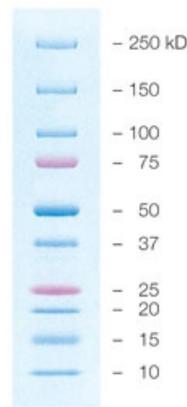
Für die Agarose-Gelelektrophorese wurden folgende Größenstandards verwendet (vgl. **Abb. 5**).

- 100 bp DNA-Leiter, *extended*, 100 - 5.000 bp, 0,1  $\mu$ g/ $\mu$ l (Carl Roth, Karlsruhe, Deutschland).
- 1 kb DNA-Leiter, 1  $\mu$ g/ $\mu$ l (Invitrogen/Life Technologies, Carlsbad, USA).



**Abbildung 5:** Für die Gelelektrophorese verwendete Größenstandards. **a)** 100 bp DNA-Leiter, *extended*, 100 - 5.000 bp. **b)** 1 kb DNA-Leiter µg/µl (Quelle: *Technical Manuals* Carl Roth (Karlsruhe, Deutschland) und Invitrogen/Life Technologies (Carlsbad, USA)).

Für die SDS-Polyacrylamid-Gelelektrophorese wurde der Größenstandard *Precision Plus Protein™ Standard Dual Color* (Bio-Rad, Hercules, USA) verwendet (vgl. **Abb. 6**).



**Abbildung 6:** Für die SDS-Polyacrylamid-Gelelektrophorese verwendeter Größenstandard *Precision Plus Protein™ Standard Dual Color* (Quelle: [www.bio-rad.com](http://www.bio-rad.com)).

### 3.1.8. Gewebeproben

Zwei Lymphknoten-Metastasen von Melanom-Patienten wurden von Herrn Prof. Dr. med. Jörg Reichrath und Herrn PD Dr. med. Alexander Rösch aus der Klinik für Dermatologie, Venerologie und Allergologie des Universitätsklinikums des Saarlandes (Homburg, Deutschland) zur Verfügung gestellt. Die Proben wurden nach der operativen Entfernung bei -80 °C gelagert.

*Post mortem* Hirnproben aus *Plaques* von Multiple Sklerose-Patienten sowie aus gesunden Kontrollen wurden von der *Netherlands Brain Bank* (Netherlands Institute for Neuroscience, Amsterdam, Niederlande) bezogen. Es handelte sich um sieben Proben aus neuropathologisch bestätigten Läsionen der weißen Substanz von sechs MS-Patienten sowie sieben Proben der

weißen Substanz von Kontrollpersonen ohne zerebrale Anomalien und neurologischen Erkrankungen während ihrer Lebenszeit und mit nicht-neurologischer Todesursache, wie Herzinfarkt, Herzversagen, postoperative retroperitoneale Blutungen, Multi-Organversagen oder Lungeninfektion. Die Patientendetails sind in **Tabelle 1** zusammengefasst.

**Tabelle 1:** Zur Analyse der HERV-W-Transkription im Kontext von Multipler Sklerose verwendete *post mortem* Hirnproben.\*

Probe	NBB	Autopsie	Geschlecht	Alter	PM	Diagnose	Hirnregion
MS1	2001-130	S01/298	weiblich	53	10:45	Multiple Sklerose	Plaque
MS2	2001-135	S01/316	männlich	43	08:30	Multiple Sklerose	Plaque
MS3	2002-055	S02/156	weiblich	48	04:50	Multiple Sklerose	Plaque
MS4	2006-045	S06/139	männlich	56	08:00	Multiple Sklerose	Plaque
MS5	2006-045	S06/139	männlich	56	08:00	Multiple Sklerose	Plaque
MS6	2007-010	S07/051	männlich	47	07:15	Multiple Sklerose	Plaque
MS7	2009-067	S09/219	männlich	44	12:00	Multiple Sklerose	Plaque
K1	1991-124	91/225.4	männlich	38	7:00	Kontrolle	weiße Substanz
K2	1996-057	S96/163	weiblich	69	08:30	Kontrolle	weiße Substanz
K3	2010-070	S10/196	weiblich	60	07:30	Kontrolle	weiße Substanz
K4	1991-125	*91/230	männlich	61	05:40	Kontrolle	weiße Substanz
K5	1997-043	S97/133	männlich	68	10:10	Kontrolle	weiße Substanz
K6	1998-127	S98/235	männlich	56	05:25	Kontrolle	weiße Substanz
K7	2009-003	S09/007	männlich	62	07:20	Kontrolle	weiße Substanz

\*Es standen Proben von sechs Multiple Sklerose-Patienten (MS) und sieben nicht-dementen Kontrollen (K) zur Verfügung, wobei die Proben MS4 und MS5 vom selben Patienten stammten. Angegeben sind die *Netherlands Brain Bank*- (NBB), sowie Autopsie-Nummer, Geschlecht und Alter in Jahren der Spenderpersonen, die *post mortem* Zeit bis zur Probenentnahme (PM) in Stunden und Minuten, die Diagnose sowie die Hirnregion, aus der die Probe entnommen wurde.

### 3.1.9. RNA-Proben

Drei RNA-Proben aus malignen Melanom-Metastasen, in der weiteren Arbeit als RNA1, RNA2 und RNA3 bezeichnet, wurden von *Cambridge Bioscience* (Cambridge, England) bezogen. Die Metastasen enthielten nach Angaben des Lieferanten mindestens 90 % Tumorgewebe und stammten aus Jejunum (RNA1, *cat#* CR562868), Lunge (RNA2, *cat#* CR562901) und Lymphknoten (RNA3, *cat#* CR561386).

RNA aus den Prostatakarzinom-Zelllinien PTu 230, PTu 232, MDA PCa 2B, VCaP sowie PC-3 wurde von Herrn Dr. rer. nat. Wolfgang Göring (AG Wolfgang Schulz, Urologische Klinik, Universitätsklinikum Düsseldorf, Deutschland) zur Verfügung gestellt. Für MDA PCa 2B, VCaP und PC-3 stand RNA aus zwei unterschiedlichen Passagen der Zelllinien zur Verfügung, in der weiteren Arbeit mit I und II gekennzeichnet.

### 3.1.10. Zellen

#### Prokaryotische Zellen

Zur Transformation von Plasmiden mittels Hitzeschock wurden chemokompetente Zellen des *E. coli* Stammes DH5 $\alpha$  (*Library Efficiency*<sup>®</sup> DH5 $\alpha$ <sup>TM</sup> *Competent Cells*, Genotyp F-  $\Phi$ 80*lacZ* $\Delta$ M15  $\Delta$ (*lacZYA-argF*)U169 *recA1 endA1 hsdR17*( $r_k^-$ ,  $m_k^+$ ) *phoA supE44 thi-1 gyrA96 relA1*  $\lambda^-$ ) der Firma Invitrogen (Karlsruhe, Deutschland) verwendet. Die DH5 $\alpha$ -Zellen zeichnen sich durch Sensitivität gegenüber den Antibiotika Ampicillin und Kanamycin aus, was als Selektionskriterium bei der Transformation ausgenutzt wurde.

#### Eukaryotische Zellen

Die Melanom-Zelllinien SK-Mel-25, SK-Mel-28, MEWO und WM3734a sowie die Melanozyten-Zelllinien Benno und Oskar wurden von Herrn PD Dr. med. Alexander Rösch (Klinik für Dermatologie, Venerologie und Allergologie, Universitätsklinikums des Saarlandes, Homburg, Deutschland) zur Verfügung gestellt. Der Genotyp der verwendeten Zelllinien wurde zuvor im Auftrag der Klinik für Dermatologie durch DNA-Fingerprinting im Institut für Humangenetik der Universitätsklinik des Saarlandes (Homburg, Deutschland) bestätigt.

Die menschliche Chorionkarzinom-Zelllinie JEG-3 wurde von Frau Dr. Sindy Sohr (Molekulare Onkologie, Universitätsfrauenklinik Leipzig, Deutschland) zur Verfügung gestellt.

Die menschlichen Keimzelltumor-Zelllinien Tera-1 und NCCIT wurden von Frau Dr. rer. nat. Marlies Sauter (Institut für Virologie, Universitätsklinikum Homburg, Deutschland) und Herrn Dr. rer. nat. Wolfgang Göring (AG Wolfgang Schulz, Urologische Klinik, Universitätsklinikum Düsseldorf, Deutschland) zur Verfügung gestellt.

Alle Zellen wurden bei 37 °C und 5 % [v/v] CO<sub>2</sub> kultiviert. Die zur Kultur der verschiedenen Zelllinien verwendeten Medien sind in **Tabelle 2** aufgelistet.

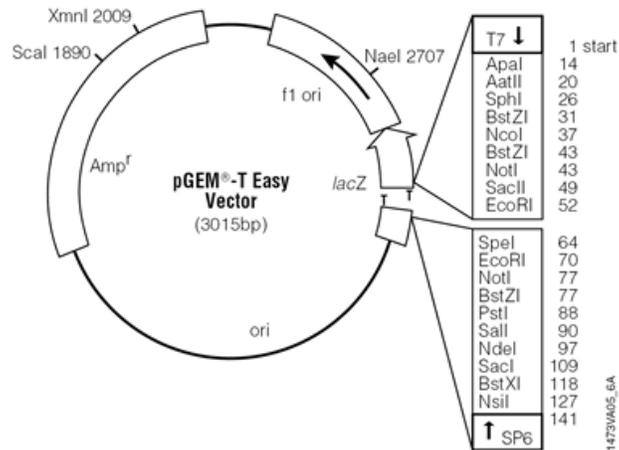
**Tabelle 2:** Verwendete eukaryotische Zelllinien mit ihren Kulturmedien und Zusätzen.

Zelllinie	Kulturmedium
JEG-3	<i>Ham's F-12</i> (10 % [v/v] FCS, 10.000 U/ml Penicillin, 10 mg/ml Streptomycin)
Benno, Oscar	<i>Melanocyte Growth Medium</i>
MEWO	<i>RPMI 1640</i> (10 % [v/v] FCS, 10.000 U/ml Penicillin, 10 mg/ml Streptomycin)
NCCIT	<i>RPMI 1640</i> (10 % [v/v] FCS, 10.000 U/ml Penicillin, 10 mg/ml Streptomycin)
SK-Mel-25	<i>RPMI 1640</i> (10 % [v/v] FCS, 10.000 U/ml Penicillin, 10 mg/ml Streptomycin)
SK-Mel-28	<i>RPMI 1640</i> (10 % [v/v] FCS, 10.000 U/ml Penicillin, 10 mg/ml Streptomycin)
Tera-1	<i>McCoy's 5A (1x) + GlutaMAX<sup>TM</sup>-1</i> (10 % [v/v] FCS, 10.000 U/ml Penicillin, 10 mg/ml Streptomycin)
WM3734a	<i>Tu2</i> %-Medium

### 3.1.11. Vektoren

#### *pGEM<sup>®</sup>-T Easy*

Für die Klonierung von PCR-Produkten zur anschließenden Sanger-Sequenzierung oder zum Restriktionsverdau wurde der *pGEM<sup>®</sup>-T Easy*-Vektor der Firma Promega (Fitchburg, USA) verwendet (vgl. **Abb. 7**).

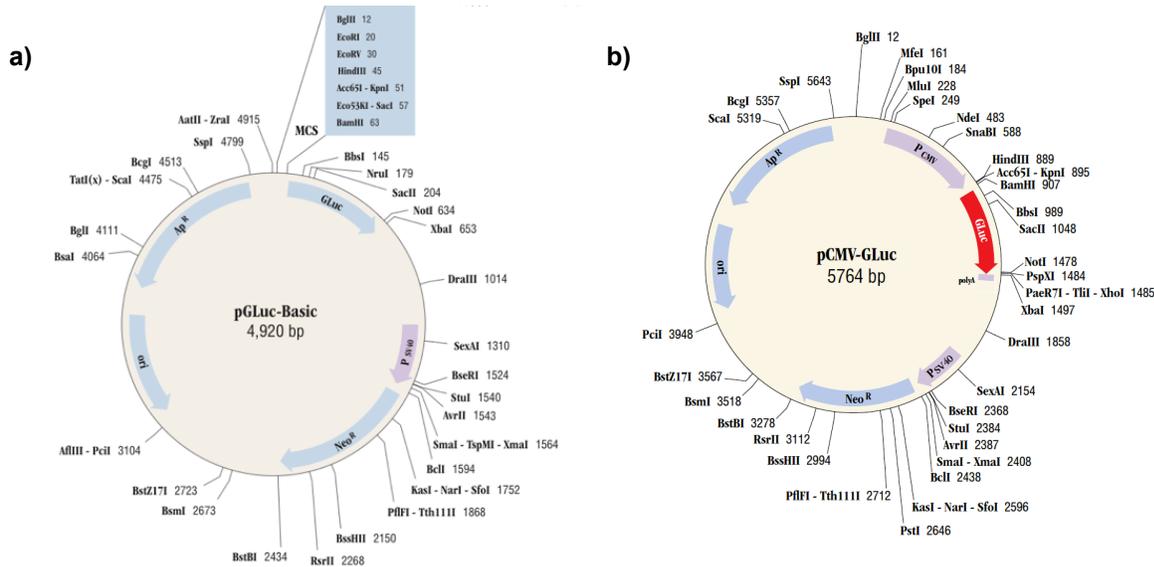


**Abbildung 7:** Karte des *pGEM<sup>®</sup>-T Easy*-Vektors der Firma Promega (Fitchburg, USA). Amp<sup>r</sup> = Gen für Ampicillin-resistenz, f1 ori = Phage f1 Region, *lacZ* =  $\beta$ -Galactosidase-Gen ori = Replikationsursprung (Quelle: *Technical Manual*, Promega).

Die für die Amplifikation der zu klonierenden Produkte verwendete thermostabile *Taq*-Polymerase hängt nach der Elongationsphase der PCR ein einzelnes Deoxyadenosin am 3'-Ende des neu synthetisierten Stranges an. Der im *lacZ*-Gen linearisierte *pGEM<sup>®</sup>-T Easy*-Vektor besitzt an beiden 3'-Enden ein überhängendes Thymidin, was die Ligation der PCR-Produkte ermöglicht. Zur Selektion von Bakterien mit aufgenommenem Plasmid enthält der Vektor das Gen für Ampicillin-Resistenz. Transformierte Bakterien mit in den Vektor ligiertem PCR-Produkt werden durch Blau/Weiß-Selektion identifiziert. Der Vektor enthält das *lacZ*-Gen, dessen Produkt,  $\beta$ -Galactosidase, das den LB-Agarplatten zugegebene Substrat X-Gal zu Galactose und 5-Brom-4-chlor-indoxyl hydrolysiert, welches anschließend durch Luftsauerstoff zum dunkelblauen 5,5'-Dibrom-4,4'-dichlor-indigo oxidiert wird. Bei erfolgreicher Ligation eines PCR-Produktes in den Vektor wird das *lacZ*-Gen unterbrochen und das Genprodukt nicht mehr gebildet. Die Kolonien mit in den Vektor ligiertem PCR-Produkt bleiben weiß, im Gegensatz zu blauen Kolonien mit intaktem *lacZ*-Gen.

**pGLuc-Basic**

Zur Ermittlung einer möglichen Promotoraktivität ausgesuchter 5'- und 3'-LTR-Elemente der HERV-W-Gruppe wurden diese in den *pGLuc-Basic*-Vektor der Firma NEB (Ipswich, USA) kloniert (vgl. **Abb. 8a**).



**Abbildung 8:** Karten der Vektoren *pGLuc-Basic* (a) und *pCMV-GLuc* (b) der Firma NEB (Ipswich, USA). MCS = multiple cloning site, GLuc = *Gaussia princeps* Luciferase-Gen, P<sub>SV40</sub> = SV40 Promotorregion, Neo<sup>R</sup>/Ap<sup>R</sup> = Gen für Neomycin- bzw. Ampicillin-Resistenz, ori = Replikationsursprung (Quelle: *Technical Manual*, NEB).

Der *pGLuc-Basic* Vektor erlaubt die Klonierung potentieller Promotorsequenzen in die *multiple cloning site* (MCS) unmittelbar stromauf des promotorlosen Reportergens *GLuc*, das für die *Gaussia princeps*-Luciferase codiert. Die anschließende Bestimmung der Luciferaseaktivität in Säugerzellen gibt Aufschluss über die Promotoraktivität der einklonierten Sequenzen.

Die hier getesteten HERV-W 5'- und 3'-LTR-Konstrukte wurden im Rahmen von Richter (2012) generiert. Die teilweise stark verkürzten LTRs (vgl. **Anhang 17**) wurden hierfür aus genomischer DNA amplifiziert, in *pGEM<sup>®</sup>-T Easy* zwischenkloniert und über *EcoRI*-Verdau in Promotorposition in *pGLuc-Basic* einkloniert. Die in der vorliegenden Arbeit verwendeten LTR-Konstrukte, deren Orientierung zuvor erneut durch Sanger-Sequenzierung bestätigt wurde, sind in **Tabelle 3** zusammengefasst.

Im Zuge der Luciferase-Assays wurde als Positivkontrolle das *pCMV-GLuc*-Kontrollplasmid (NEB, Ipswich, USA) in JEG-3-Zellen transfiziert und die Luciferaseaktivität bestimmt. Der Vektor enthält die Kodierungssequenz der *Gaussia princeps*-Luciferase unter Kontrolle des konstitutiven Cytomegalovirus (CMV)-Promotors, wobei die Sequenz ansonsten der von *pGLuc-Basic* entspricht (vgl. **Abb. 8b**).

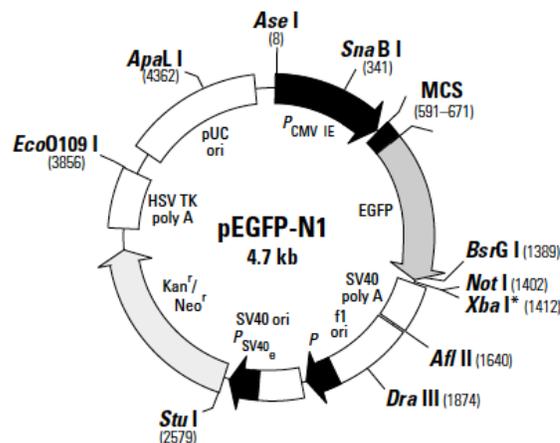
**Tabelle 3:** In *pGLuc-Basic* klonierte HERV-W LTR-Sequenzen, deren potentielle Promotoraktivität mittels Luciferase-Assay bestimmt wurde.\*

HGNC: ERVW-	HERV-W-Locus incl. LTRs			LTR-Amplikon		Orientierung	
	Chromosomenposition	Bande	LTR	Chromosomenposition	[bp]	sense	antisense
-	chr1:55149270-55157786	1p32.3	3'	chr1:55149175-55150115	941	+	+
12	chr2:53837284-53842066	2p16.2	5'	chr2:53837211-53837765	555	+	+
13	chr2:112513394-112519037	2q13	3'	chr2:112513280-112513751	472	+	+
-	chr6:1275282-1276849	6p25.3	3'	chr6:1275212-1275640	429		+
17	chr6:106782705-106790382	6q21	3'	chr6:106790020-106790515	496	+	+
1	chr7:91935249-91945442	7q21.2	5'	chr7:91944555-91945473	919	+	
25	chr13:36428787-36431118	13q13.3	5'	chr13:36428734-36428936	203	+	+
27	chr14:90761762-90762965	14q32.12	3'	chr14:90761700-90762099	400	+	+
4	chr15:53384372-53391866	15q21.3	3'	chr15:53384237-53384727	491	+	+
2	chrX:106182017-106184757	Xq22.3	3'	chrX:106181886-106182366	481		+

\*Angegeben ist der HGNC-Name des kompletten Locus (vgl. Mayer *et al.*, 2011) sowie dessen Chromosomenposition und -bande im menschlichen hg18 Referenzgenom, der jeweils untersuchte LTR (5' oder 3'), die Chromosomenposition des klonierten Amplikons, das die komplette noch vorhandene LTR-Sequenz mit flankierender Sequenz umfasste, und dessen Größe in [bp], sowie die Orientierung des LTR-Amplikons (*sense* und/oder *antisense*) im *pGLuc-Basic*-Vektor.

### **pEGFP-N1**

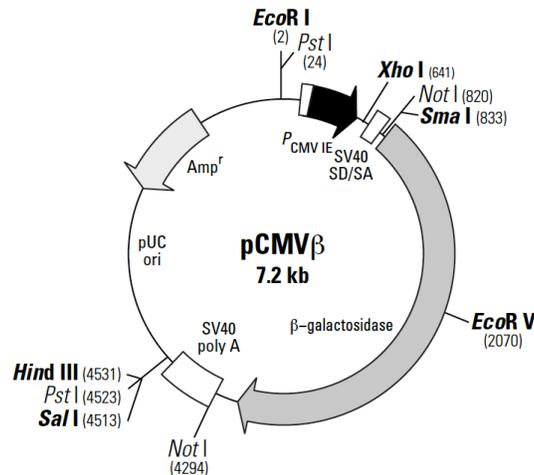
Der *pEGFP-N1*-Vektor (Clontech Laboratories/Takara Bio, Otsu, Japan) kodiert eine Variante des Wildtyp-GFP (*green fluorescent protein*), die hinsichtlich hellerer Fluoreszenz und höherer Expression in Säugerzellen optimiert wurde. Um die Translationseffizienz in eukaryotischen Zellen zu erhöhen, wurden die EGFP (*enhanced green fluorescent protein*) flankierenden Sequenzen zu einer Kozak-Sequenz zur Translations-Initiation umgewandelt. Die MCS befindet sich zwischen dem CMV-Promotor und der EGFP-kodierenden Sequenz. Im richtigen Leserahmen in die MCS einklonierte Gene werden als Fusionsproteine exprimiert, an deren C-Terminus sich das grünfluoreszierende Protein befindet (vgl. **Abb. 9**).



**Abbildung 9:** Karte des *pEGFP-N1*-Vektors der Firma Clontech Laboratories/Takara Bio (Otsu, Japan).  $P_{CMV IE}$  = *immediate early* CMV-Promotor, EGFP = *enhanced green fluorescent protein*, SV40 polyA = SV40 Polyadenylierungssignal, f1 ori = *f1 single-strand DNA-origin*, SV40<sub>ori</sub> = SV40 *origin of replication*,  $P_{SV40_e}$  = SV40 *early* Promotor, Kan<sup>r</sup>/Neo<sup>r</sup> = Gen für Kanamycin/Neomycin-Resistenz, HSV TK poly A = Herpes-simplex-Virus Thymidin-Kinase Polyadenylierungssignal, pUCori = pUC Plasmid Replikationsursprung (Quelle: *Technical Manual*, Clontech Laboratories/Takara Bio).

**pCMV $\beta$** 

Der pCMV $\beta$ -Reportervektor der Firma Clontech Laboratories/Takara Bio (Otsu, Japan) enthält das Gen für  $\beta$ -Galactosidase mit eukaryotischen Translations-Initiationssignalen unter Kontrolle des konstitutiven Cytomegalovirus (CMV)-Promotors (vgl. **Abb. 10**). Der Vektor exprimiert hohe Mengen an  $\beta$ -Galactosidase und wurde mit den Luciferase-exprimierenden pGLuc-Vektorkonstrukten in Reporter-Gen-Assays cotransfiziert. Die  $\beta$ -Galactosidase-Aktivität wurde zur Normalisierung der Luciferasewerte herangezogen.



**Abbildung 10:** Karte des pCMV $\beta$  Vektors der Firma Clontech Laboratories/Takara Bio (Otsu, Japan). P<sub>CMV IE</sub> = *immediate early* CMV-Promotor, SV40 SD/SA = SV40 *Splice Donor/Acceptor*,  $\beta$ -Galactosidase = *E. coli*  $\beta$ -Galactosidase-Gen mit eukaryotischen Translations-Initiationssignalen, SV40 polyA = SV40 Polyadenylierungssignal, pUC ori = pUC Replikationsursprung, Amp<sup>R</sup> = Gen für Ampicillin-Resistenz (Quelle: *Technical Manual*, Clontech Laboratories/Takara Bio).

**3.1.12. Oligonukleotid-Primer**

Alle in dieser Arbeit verwendeten Oligonukleotid-Primer wurden von Eurofins MWG Operon (Ebersberg, Deutschland) bezogen.

Die für die HERV-Gruppen spezifischen PCRs verwendeten Oligonukleotid-Primer sind in den **Tabellen 4** und **5** aufgelistet. Im Rahmen der HERV-K(HML-2)-Studien wurde ein zentraler Bereich im *gag*-ORF amplifiziert sowie ein Bereich in der *rec/np9*-Region. Für die HERV-W-Studien wurden zwei Amplikons im 5'- und 3'-Bereich des *env*-ORFs, in der Arbeit als 5'*env* und 3'*env* benannt, untersucht. Um in einer PCR genomweit möglichst viele Loci einer HERV-Gruppe zu amplifizieren, wurden jeweils mehrere *Forward*- und *Reverse-Primer*-Varianten gemischt, die Nukleotidunterschiede der verschiedenen Loci in den von den *Primern* gebundenen Bereichen berücksichtigten. Das Mischverhältnis entsprach hierbei in etwa der theoretischen Anzahl an Loci, die von den jeweiligen *Primer*-Varianten erkannt werden.

**Tabelle 4:** Primer zur Amplifikation von HERV-K(HML-2)-Loci aus genomischer DNA und cDNA.\*

amplifizierter Bereich	Primer-Name	Sequenz in 5'-3'-Orientierung	Verhältnis
HERV-K(HML-2) gag, zentrale Region	HML-2_1778_for1	CCCCAGAAAGTCAGTATGGA	10:3:2:1
	HML-2_1778_for2	TCTCCAGAGGTTTCAGTATGGA	
	HML-2_1778_for3	CCCCAGAAAATCAGTATGGA	
	HML-2_1778_for4	TCTCCAGAGGTCAGTATAGA	
	HML-2_2396_rev1	TTTCCCAGGCTCTAAGGCAG	12:6:1
	HML-2_2396_rev2	TTCCAGGCCCTGAGGCAA	
HML-2_2396_rev3	TTTCTAGGCTCTAAGGCAG		
HERV-K(HML-2) np9/rec	np9-FOR-1 <sup>a</sup>	ATGAACCCATCAGAGATGCAA	17:1:1:1
	np9-FOR-2	ATGAATCCATCAGAGATGCAA	
	np9-FOR-3	GCGAACCCCTTCAGAGATGCAA	
	np9-FOR-4	ATGAACCCATCGGAGATGAAA	
	np9-REV-1	AGCATCTGTTTAAACAAAGCA	19:1
	np9-REV-2	AGCATGTTTAAACAAAGCA	

\*Angegeben sind der jeweils zu amplifizierende HERV-Bereich sowie Name und Sequenz der *Forward*- (for) und *Reverse-Primer* (rev). Die Zahl im *Primer*-Namen bezieht sich auf die Position des 5'-Endes des *Primers* in der Sequenz des HERV-K(HML-2.HOM)-Provirus (*Genbank acc. no.* AF074086.2). Zusätzlich ist das Verhältnis angegeben, in dem die *Primer*-Varianten gemischt wurden.

<sup>a</sup>aus Armbruster *et al.* (2002)

**Tabelle 5:** Primer zur Amplifikation von HERV-W-Loci aus genomischer DNA sowie aus cDNA zur anschließenden Sequenzierung mittels *MiSeq/Illumina*-Technologie.\*

amplifizierter Bereich	Primer-Name	Sequenz in 5'-3'-Orientierung	Verhältnis
HERV-W 5'env	HERV17_7548_for	GAACAATGGAACAACCTTCAGCAC	1:1
	HERV17_7568_for	GCACAGAAATAAACACCACCTTCC	
	HERV17_7821_rev1	CACTAAGAATGAGAGGAAGCAC	1:1
	HERV17_7821_rev2	CACTAAGAATGACAGGAAGCAC	
HERV-W 3'env	HERV17_8349_for1	CCTCCTTGTTAAGTTTGTCTC	for2-for4 zusammen 10 %
	HERV17_8349_for2	CCTCCTTGTTAACTTTGTCTC	
	HERV17_8349_for3	CCTCCTTGTTAAGTTTGTCTT	
	HERV17_8349_for4	CCTCCTTATTAATTTGGTCTC	
	HERV17_8587_rev1	AACCCAAGTGCTGTTGGGGA	rev2-rev7 zusammen 10 %
	HERV17_8587_rev2	AACCTAAGTGCTGTTGGGGA	
	HERV17_8587_rev3	AACCCAAGTGCTGCTGGGGA	
	HERV17_8587_rev4	AACCCAAGTGCTGTTGGGGA	
	HERV17_8587_rev5	AACCCAAGTGCTTTGGGGA	
	HERV17_8587_rev6	AACCAAAGTGCTGTTGGGGA	
HERV17_8587_rev7	AACTCAAGTGCTGTTGGGTT		

\*Angegeben sind der jeweils zu amplifizierende HERV-Bereich sowie Name und Sequenz der *Forward*- (for) und *Reverse-Primer* (rev). Die Zahl im *Primer*-Namen bezieht sich auf die Position des 5'-Endes des *Primers* in der *Repbse* Konsensussequenz von HERV17. Zusätzlich ist das Verhältnis angegeben, in dem die *Primer*-Varianten gemischt wurden. Für die Amplifikation aus cDNA zur anschließenden Sequenzierung mittels *454/FLX*-Technologie wurden *Primer* mit entsprechenden *Adaptor*-Sequenzen am 5'-Ende verwendet (vgl. Text und **Tab. 6**).

Für die Amplifikation von HERV-W-Loci aus cDNA zur anschließenden Sequenzierung mittels *454/FLX*-Technologie wurden *Primer* verwendet, deren Sequenz denen aus **Tabelle 5** entspricht, wobei am jeweiligen 5'-Ende die in **Tabelle 6** aufgelisteten *Adaptor*- und *Key*-Sequenzen ligiert waren.

**Tabelle 6:** Adaptor- und Key-Sequenzen zur Amplifikation von HERV-W-Loci aus cDNA zur anschließenden Sequenzierung mittels 454/FLX-Technologie.\*

FLX-Lauf	Proben	Primer	Adaptor- und Key-Sequenz (5'-3')
1	MS1, MS2, MS3, K1, K2, K3	Forward-Primer	CGTATCGCCTCCCTCGGCCATCAG
		Reverse-Primer	CTATGCGCCTTGCCAGCCGCTCAG
2	MS4, MS5, MS6, MS7, K4, K5, K6, K7	Forward-Primer	CCATCTCATCCCTGCGTGTCTCCGACGACT
		Reverse-Primer	CCTATCCCCTGTGTGCCTTGGCAGTCGACT

\*Angegeben sind die Sequenzen, die am 5'-Ende der jeweiligen *Forward*- und *Reverse-Primer* (vgl. **Tab. 5**) ligiert waren (*Key*-Sequenz unterstrichen). Die jeweiligen Proben MS1-3 und K1-3, sowie MS4-7 und K4-7 wurden in zwei 454/FLX-Läufen mit unterschiedlichen *Adaptor*- und *Key*-Sequenzen sequenziert.

Die *Primer* zur Bestimmung des Allelstatus polymorpher HERV-K(HML-2)-Proviren in genomischer DNA sind in **Tabelle 7** zusammengefasst.

**Tabelle 7:** *Primer* zur Bestimmung des Allelstatus polymorpher HERV-K(HML-2)-Proviren in genomischer DNA.\*

Locus	Position	Primer	Sequenz
1p31.1	chr1:75615359-75621731	1p31.1_flank_for	AATGGAACAAAGCTGGATGG
		1p31.1_LTR	AGCCCTAAGGCAGTTTTTCC
		1p31.1_int	TCGTCGACTTGTCCTCAATG
3p25.3	chr3:9864346-9871236	3p25.3_flank_for	AGCAACTCGCACCTTTTGT
		3p25.3_flank_rev	ATGCCAGCTGATCTTGGACT
		3p25.3_LTR	GCTGCCTTCAAGCATCTGTT
3q13.2 (K106)	chr3:114225814-114234972	3q13.2_flank_for <sup>a</sup>	GGTGTGCTGTGAAGGTATTC
		3q13.2_flank_rev <sup>a</sup>	TCCATGGCTATCCACGAGA
		3q13.2_LTR	AATGGAGTCTCCTATGTCTACT
6q14.1 (K109)	chr6:78483381-78492802	6q14.1_flank_for <sup>a</sup>	GTCTTTAATGTCTCCCCTC
		6q14.1_flank_rev <sup>a</sup>	CAGATGAGATGTCAAGCAAGGT
		3q13.2_LTR	AATGGAGTCTCCTATGTCTACT
7p22.1 (K108, HOM)	chr7:4588583-4606557	5'flankFOR <sup>b</sup>	TACCCACAGCACCCAAGAGATG
		3'flankREV <sup>b</sup>	GCTTTCGGGACTTGAACATTGG
		LTRgagREV2 <sup>b</sup>	TTTGCCCCATTATCACCCCTA
8p23.1 (K115)	chr8:7342807-7352269	K115_flank_for <sup>c</sup>	CCGCACCTAGTCAACTTAGC
		K115_flank_rev <sup>c</sup>	TCAGTTCCCGATTTTCTGCC
		K115_LTR	TTGTGTCCCTGGGTACTTGAG
19p12 (K113)	chr19:21633376-21633382	K113_flank_for <sup>c</sup>	GCATGGGGAGATTCAGAACC
		K113_flank_rev <sup>c</sup>	CATGTTTCCTGTCTGCCAC
		K113_LTR	TTCTCAAAGAGGGGATGTG

\*Angegeben sind die Chromosomenbande und genomische Position des Locus im hg18 Referenzgenom (Alternativnamen sind in Klammern ergänzt) sowie die zur Amplifikation verwendeten *Primer* jeweils mit Sequenz in 5'-3' Orientierung. HOM = HERV-K(HML-2.HOM) (*Genbank acc. no.* AF074086.2).

<sup>a</sup> aus Barbulescu *et al.* (1999)

<sup>b</sup> aus Mayer *et al.* (2005)

<sup>c</sup> aus Turner *et al.* (2011)

Die *Primer* zur spezifischen Amplifikation der durch reverse Transkription und Reintegration gespleißter mRNA entstandenen HERV-K(HML-2)-Loci auf den Chromosomen 2q32.1 und 10q24.2 sind in **Tabelle 8** zusammengefasst.

**Tabelle 8:** *Primer* zur spezifischen Amplifikation zweier durch reverse Transkription und Reintegration gespleißter mRNA entstandener HERV-K(HML-2)-Loci.\*

Locus	Primer	Sequenz in 5'-3'-Orientierung	[bp]
chr2:187093878-187095344 (2q32.1)	chr2_specific_for	CACAGTTAGCTAAAAAAAAGC	363
	chr2_specific_rev	AAATCCATTAAATCTTGACACA	
chr10:101570559-101577735 (10q24.2)	chr10_specific_for	TTATCACATGTCTGCAGGTGTA	230
	chr10_specific_rev	GGGGTAAGGTCACAGAATCTTA	

\*Angegeben sind der *Primer*-Namen, seine Sequenz sowie die Größe des zu erwartenden PCR-Produktes in [bp].

In **Tabelle 9** finden sich die *Primer*, die zur spezifischen Amplifikation der ORFs der durch reverse Transkription und Reintegration gespleißter mRNA entstandenen HERV-K(HML-2)-Loci auf den Chromosomen 2q32.1 und 10q24.2 verwendet wurden. Am 5'-Ende der *Primer* war jeweils eine *Bam*HI-Schnittstelle angehängt, was die spätere Klonierung in *pEGFP-N1* erlaubte.

**Tabelle 9:** *Primer* zur spezifischen Amplifikation der ORFs zweier durch reverse Transkription und Reintegration gespleißter mRNA entstandener HERV-K(HML-2)-Loci.\*

Locus	Primer	Sequenz in 5'-3'-Orientierung	[bp]
chr2:187093878-187095344 (2q32.1)	c2_ORF_F_Bam	<u>GGATCCGAAGGTATGCTCGAGCATGAT</u>	506
	c2_ORF_R_Bam	<u>GGATCCTTCTTTCTACATAGACACAGTAAC</u>	
chr10:101570559-101577735 (10q24.2)	c10_ORF_F_Bam	<u>GGATCCTGCTTGTGTTTCACCAGGAG</u>	724
	c10_ORF_R_Bam	<u>GGATCCCATGGCCCCTTCTCGATGGTC</u>	

\*Angegeben sind die Chromosomenposition des Locus im hg18 Referenzgenom, die *Primer*-Namen, die Sequenz der *Forward*- (F) und *Reverse-Primer* (R) sowie die Größe des zu erwartenden PCR-Produktes in [bp]. Die jeweils am 5'-Ende der *Primers* befindliche *Bam*HI-Schnittstelle ist unterstrichen.

Die *nested-PCR-Primer* zur spezifischen Amplifikation zweier im Rahmen der HML-2 *rec/np9*-Analysen in Normalgeweben identifizierten potentiellen neuen HERV-K(HML-2)-Loci mit Sequenzähnlichkeit zu *ERVK-5* und *ERVK-18* sind in **Tabelle 10** zusammengefasst. Die *Primer* liegen hierbei mit ihrem 3'-Ende jeweils auf charakteristischen Nukleotidunterschieden der potentiellen neuen Loci verglichen mit *ERVK-5* und *ERVK-18*.

**Tabelle 10:** *Nested-PCR-Primer* zur spezifischen Amplifikation zweier potentieller neuer HERV-K(HML-2)-Loci.\*

Locus	Primer	Sequenz in 5'-3'-Orientierung	[bp]
ähnlich <i>ERVK-5</i>	Rec_Np9_nested_F	<u>AGAAAAGCGCCTCCACG</u>	~2000
	Rec_Np9_nested_R	<u>CCCTGAGTTTGACACAGCACA</u>	
ähnlich <i>ERVK-18</i>	Rec_Np9_2_nested_F	<u>AAAAGCGCCTCCACGGAGAC</u>	~270
	Rec_Np9_2_nested_R	<u>AGTGAGTCTACTGTTAGGACCGG</u>	

\*Die beiden potentiellen neuen HML-2-Loci ähneln *ERVK-5* und *ERVK-18*. Angegeben sind jeweils die *Primer*-Namen, die Sequenzen der *Forward*- (F) und *Reverse-Primer* (R) sowie die potentielle Produktgröße in [bp]. Die *Primer* liegen hierbei auf charakteristischen Nukleotidunterschieden der potentiellen neuen Loci gegenüber *ERVK-5* und *ERVK-18*, die hier unterstrichen sind.

Die für die quantitativen *Real-Time*-PCRs verwendeten *Primer* sind in **Tabelle 11** aufgelistet. Zur Amplifikation und Quantifizierung der Transkriptmengen von HERV-K(HML-2) *gag* sowie HERV-W 5'*env* wurden die jeweiligen *Primer*-Gemische aus den **Tabellen 4** und **5** eingesetzt.

**Tabelle 11:** Primer zur Quantifizierung der HERV-W-Transkriptmengen sowie der Transkriptmengen der Haushaltsgene *G6PDH* und *RPII* mittels *Real-Time-PCR*.\*

Amplikon	Primer	Sequenz (5'-3')	Referenz
HERV-W <i>Syncytin-1</i> ( <i>ERVWE1/ERVW-1</i> )	<i>syncytin-1_for</i>	TTCACCTGCCACACCCAT	Lauer <i>et al.</i> , 2009
	<i>syncytin-1_rev</i>	CCCCATCAGACATACCAGTT	de Parseval <i>et al.</i> , 2003
HERV-W Xq22.3 ( <i>ERVWE2/ERVW-2</i> )	<i>Xq22.3_for</i>	GCTGCTGTACAACCAGTAGCTC	Roebke <i>et al.</i> , 2010
	<i>Xq22.3_rev</i>	TTCTCTTGCTGACCTTGAAT	Roebke <i>et al.</i> , 2010
G6PDH	<i>G6PDH_for</i>	ATCGACCACTACCTGGGCAA	Radonić <i>et al.</i> , 2004
	<i>G6PDH_rev</i>	TTCTGCATCACGTCCCGGA	Radonić <i>et al.</i> , 2004
RPII	<i>RPII_for</i>	GCACCACGTCCAATGACAT	Radonić <i>et al.</i> , 2004
	<i>RPII_rev</i>	GTGCGGCTGCTTCCATAA	Radonić <i>et al.</i> , 2004

\*Angegeben sind die spezifischen *Forward*- (*for*) und *Reverse-Primer* (*rev*) der verschiedenen Amplikons, die *Primer*-Sequenzen in 5'-3'-Orientierung sowie deren Referenzen.

Alle weiteren verwendeten *Primer* finden sich in **Tabelle 12**.

**Tabelle 12:** Sonstige *Primer*.

Primer	Sequenz in 5'-3'-Orientierung	Verwendung
Alu A1S	TCATGTCGACGCGAGACTCCATCTCAA	Alu-Kontroll-PCR*
$\beta$ -Actin_for	CACCTCTCCAGCCTTCCTTC	$\beta$ -Actin-Kontroll-PCR
$\beta$ -Actin_rev	GTCACCTTCACCGTTCAGT	
M13_for	CGCCAGGGTTTTCCAGTCACGAC	Kolonie-PCR <i>Primer</i> für <i>pGEM<sup>®</sup>-T Easy</i> und <i>pCR<sup>®</sup>-2.1-TOPO</i>
M13_rev	TCACACAGGAAACAGCTATGAC	
T7	TAATACGACTCACTATAGGG	Sequenzier- <i>Primer</i> für <i>pGEM<sup>®</sup>-T Easy</i> und <i>pCR<sup>®</sup>-2.1-TOPO</i>
SP6	ATTTAGGTGACACTATAGAA	
pEGFP-N1_for	TTTAGTGAACCGTCAGATC	Kolonie-PCR <i>Primer</i> für <i>pEGFP-N1</i>
pEGFP-N1_rev	CGTCGCCGTCCAGCTCGACCA	
pEGFP-N1_seq	CGTCGCCGTCCAGCTCGACCA	Sequenzier- <i>Primer</i> <i>pEGFP-N1</i>
pGLuc_col_for2	ACATTTCCCCGAAAAGTGC	Sequenzier- <i>Primer</i> <i>pGLuc-Basic</i>

\*aus Brooks-Wilson *et al.* (1990)

### 3.1.13. Geräte und Verbrauchsmaterialien

Gerät, Verbrauchsmaterial	Hersteller
3730 DNA-Analyzer	Applied Biosystems (Foster City, USA)
Amersham Hyperfilm <sup>TM</sup> ECL	GE Healthcare (Uppsala, Schweden)
8-Kanalpipette <i>Proline</i> 50 - 200 $\mu$ l	Biohit (Helsinki, Finnland)
8-Kanalpipette <i>peqPETTE</i> 10M	PEQLAB (Erlangen, Deutschland)
Autoklav VE-150	System (Wettenberg, Deutschland)
Blaulicht-Transilluminator <i>Flu-O-blu</i>	Biozym Scientific GmbH (Hessisch Oldendorf, Deutschland)
Brutschrank <i>Jouan</i>	Thermo Fisher Scientific (Waltham, USA)
CCD-Kamera LU105M	Lumenera (Ottawa, Kanada)
CCD-Kamera DP71	Olympus (Tokyo, Japan)
CO <sub>2</sub> -Inkubator	Sanyo (Osaka, Japan)

Einkanalpipetten <i>Pipetman</i> P10, P20, P200 und P1000	Gilson Inc. (Middleton, USA)
Einkanalpipetten <i>Discovery Comfort</i> 0,1 - 2 µl und 0,5 - 10 µl	Abimed GmbH (Langenfeld, Deutschland)
Eismaschine	Tritec (Hannover, Deutschland)
Elektrophoresekammern	Renner GmbH (Dannstadt, Deutschland)
Elektrophoresekammer 40-2314-N	PEQLAB (Erlangen, Deutschland)
Feinwaage <i>Basic</i>	Satorius (Göttingen, Deutschland)
Feinwaage <i>Europe 60</i>	Gibertini Elettronica (Novate Milanese, Italien)
<i>Genome Sequencer FLX™ System</i>	GS FLX/Roche (Basel, Schweiz)
Fluoreszenzmikroskop <i>AX70</i>	Olympus (Tokio, Japan)
Geldokumentation <i>DeVision DBox</i>	DC Science Tec (Hohengandern, Deutschland)
Kühlzentrifuge 2-16K, ausgerüstet mit	
• Falconrotor 12151	
• Plattenrotor 11123	Sigma (Osterode am Harz, Deutschland)
Lichtmikroskop <i>Axiovert 25</i>	Zeiss (Jena, Deutschland)
Luminometer <i>Lumat LB 9507</i>	Berthold Technologis (Bad Wildbad, Deutschland)
Magnetplatte, <i>96R Ring</i>	Alpaqua (Beverly, USA)
Magnetrührer <i>IKAMAG RCT</i>	IKA Labortechnik (Staufen, Deutschland)
Minishaker <i>MS2</i>	IKA Labortechnik (Staufen, Deutschland)
Multipette <i>HandyStep</i>	Brand (Wertheim, Deutschland)
<i>NanoDrop™ 2000</i>	
Spektralphotometer	Thermo Fisher Scientific (Waltham, USA)
Neubauer Zählkammer	Brand (Wertheim, Deutschland)
pH-Meter <i>SevenEasy</i>	Mettler-Toledo (Gießen, Deutschland)
Photometer <i>BioPhotometer</i>	Eppendorf (Hamburg, Deutschland)
<i>Powersupply 200/2.0</i>	Bio-Rad (Hercules, USA)
PVDF-Membran	GE Healthcare (Uppsala, Schweden)
<i>QIAshredder</i> Säule	QIAGEN (Hilden, Deutschland)
<i>Qubit™</i> Fluorometer	Invitrogen/Life Technologies (Carlsbad, USA)
<i>Real-Time-PCR System</i>	
<i>StepOnePlus™</i>	Applied Biosystems (Foster City, USA)
Rollinkubator <i>RM5</i>	Ingenieurbüro CAT (Staufen, Deutschland)
Schüttelinkubator <i>Multitron II</i>	Infors (Basel, Schweiz)

Schüttelwasserbad 1086	GFL (Burgwedel, Deutschland)
Schüttler KS 125 Basic	IKA Labortechnik (Staufen, Deutschland)
Sterilbank BioHit Polaris	Satorius (Göttingen, Deutschland)
Test Tube Thermostat TCR 100	Carl Roth (Karlsruhe, Deutschland)
Thermal Controller PTC-100™	MJ Research, Inc. (St. Bruno, Kanada)
Thermocycler T-Gradient	Biometra (Göttingen, Deutschland)
Tischzentrifuge Mikro 20	Hettich Zentrifugen (Tuttlingen, Deutschland)
Tischzentrifuge Biofuge pico	Heraeus (Hanau, Deutschland)
Überkopfschüttler CMV	Fröbel Labortechnik GmbH (Lindau, Deutschland)
Ultrazentrifuge, ausgerüstet mit	
• Rotor JLA-8.100	Beckman Coulter (Brea, USA)
UV-Lampe UV409T	Waldmann (Villingen-Schwenningen, Deutschland)
UV-Meter Vario Control	Waldmann (Villingen-Schwenningen, Deutschland)
UV Transilluminator	Wealtec (Sparks, USA)
Wasseranlage Milli-Q	Millipore (Bedford, USA)
Wasserbad Thermomix®BU	B. Braun (Melsungen, Deutschland)
Whatman™ Papier	GE Healthcare (Uppsala, Schweden)
XCell II Blotting Kammer	Invitrogen/Life Technologies (Carlsbad, USA)
XCell SureLock™ Kammer	Invitrogen/Life Technologies (Carlsbad, USA)
XCell SureLock™ Mini-Kassette	Invitrogen/Life Technologies (Carlsbad, USA)

### 3.1.14 Software, Online-Tools und Datenbanken

- *BiSearch: in silico* PCR  
<http://bisearch.enzim.hu/>
- *cell^F*: Bearbeitung fluoreszenzmikroskopischer Aufnahmen
- *DeVision G Version 2.0* (DC Science Tec, Hohengandern, Deutschland): Dokumentation von Agarosegelen
- *ExpASy - Translate Tool*: Translation von Nukleinsäure- in Aminosäuresequenzen  
<http://web.expasy.org/translate/>
- *FinchTV* (Geospiza Inc., PerkinElmer, Seattle, WA): Qualitätsbeurteilung und Bearbeitung von generierten Sanger-Sequenzen  
[www.geospiza.com/Products/finchtv.shtml](http://www.geospiza.com/Products/finchtv.shtml)

- *Galaxy* (Giardine *et al.*, 2005; Blankenberg *et al.*, 2010; Goecks *et al.*, 2010): Bearbeitung von *MiSeq/Illumina*-Sequenzen sowie der *BLAT*-Ergebnisse  
<https://main.g2.bx.psu.edu/>
- *Geneious* (Biomatters Ltd., Auckland, Neuseeland): verschiedene Sequenzanalysen und Sequenzbearbeitungen, *Primer*-Suche, Entfernung von *Primer*-Sequenzen, ORF-Vorhersage, Erstellung von Abbildungen
- *MacVector v10*: Erstellung von *Pustell Matrices*; die *Dot-Matrix*-Vergleiche in **Abb. 24** und **26** wurde von Herrn apl.-Prof. Dr. rer. nat. Jens Mayer erstellt
- *MAFFT Version 6* und *7* (Kato *et al.*, 2002; Kato und Toh, 2008; Kato und Standley, 2013): multiples Sequenz-*Alignment*  
<http://mafft.cbrc.jp/alignment/server/>
- *MEGA Version 4* (Tamura *et al.*, 2007): phylogenetische Analyse; der *Neighbour-joining Tree* in **Abb. 15** wurde von Herrn apl.-Prof. Dr. rer. nat. Jens Mayer erstellt
- *Microsoft Office 2011*: Text- und Tabellenverarbeitung, Erstellung von Abbildungen und Diagrammen
- *NEBcutter V2.0*: Suche von Restriktionsenzymen, *in silico* Verdaue von DNA-Sequenzen  
<http://tools.neb.com/NEBcutter2/>
- *Primer3 v.0.4.0* (Rozen und Skaletsky, 2000): Auswahl von *Primer*-Sequenzen  
<http://bioinfo.ut.ee/primer3-0.4.0/>
- *Protein Molecular Weight Calculator*: Berechnung der Molekulargewichte von Proteinen  
<http://www.sciencegateway.org/tools/proteinmw.htm>
- *pslReps*: Filterung bester *BLAT-Matches* der mittels *454/FLX* generierten und im Genom rückkartierten Sequenzen; die Filterungen wurden von Frau Dr. rer. nat. Christina Backes durchgeführt  
<http://www.epigenomebrowser.org/goldenPath/help/blatSpec.html>
- *Repbase* (Jurka, 2000; Jurka *et al.*, 2005): Quelle von HERV- und LTR-Konsensussequenzen  
<http://www.girinst.org/>
- *Reverse Complement*: Generierung revers komplementärer DNA-Sequenzen  
[http://www.bioinformatics.org/sms/rev\\_comp.html](http://www.bioinformatics.org/sms/rev_comp.html)
- *Sequence Format Converter*: Konvertierung verschiedener Sequenzformate  
<http://genome.ncbi.nlm.nih.gov/tools/reformat.html>
- *SIAS*: Berechnung von Aminosäure-Sequenzidentitäten und -ähnlichkeiten  
<http://imed.med.ucm.es/Tools/sias.html>

- *StepOne™ Software Version 2.2.2* (Applied Biosystems, Foster City, USA): Auswertung von *Real-Time-PCR* Quantifizierungen
- *UCSC Genome Browser* (Kent *et al.* 2002): Genomsequenzdatenbank; für alle Analysen in der vorliegenden Arbeit wurde die menschliche Referenzgenomsequenz NCBI36/hg18 verwendet  
<http://genome.ucsc.edu/>
  - *BLAT* (Kent, 2002): Rückkartierung von Sanger-Sequenzen
  - *Table Browser* (Karolchik *et al.*, 2004): *Download* von Sequenzen definierter Genomabschnitte, Chromosomenkoordinaten, *Repeat*-Annotationen, *RNA-Seq*-Daten
- *UniProt*: Proteindatenbank  
<http://www.uniprot.org/>

## 3.2. Methoden

### 3.2.1. Molekularbiologische Methoden

#### 3.2.1.1. Isolation von Nukleinsäuren

##### DNA-Isolation

Zur Isolation genomischer DNA aus eukaryotischen Zellen, die in 75 cm<sup>2</sup> Zellkulturflaschen oder 55 cm<sup>2</sup> Zellkulturschalen kultiviert wurden, wurde zunächst das Zellkulturmedium abgenommen und die Zellen wurden mit 1,5 ml 1x DPBS gewaschen. Die Zellen wurden anschließend durch Zugabe von 1,5 ml 0,05 % Trypsin-EDTA vom Boden der Zellkulturflasche oder -schale gelöst, in 8,5 ml Zellkulturmedium aufgenommen und durch 5-minütige Zentrifugation bei 300 g pelletiert. Der Überstand wurde abgenommen und das Pellet in 5 ml SE-Puffer und 0,25 ml 20 % [v/v] SDS resuspendiert. Nach Zugabe einer Spatelspitze Proteinase K wurde der Ansatz ü. N. bei 35 rpm in einem Wasserbad bei 56 °C inkubiert. Anschließend wurden 2 ml 5 M NaCl und 8 ml Chloroform zugegeben, der Ansatz wurde für 1 - 2 h auf dem Überkopfschüttler inkubiert und 15 min bei 1.300 g zentrifugiert. Die obere wässrige Phase wurde in ein 15 ml Falcon-Gefäß überführt und die darin enthaltene DNA durch Zugabe von 1 Vol. Isopropanol (vorgekühlt auf -20 °C) gefällt. Die gefällte DNA wurde in 2 ml 70 % [v/v] Ethanol überführt und ü. N. bei -20 °C inkubiert. Bei geringen DNA-Konzentrationen, wo es zu keinem sichtbaren Ausfällen von DNA kam, wurde die DNA in Isopropanol ü. N. bei

-20 °C inkubiert und anschließend für 5 min bei 8.600 g pelletiert. Das Isopropanol wurde verworfen und die DNA in 2 ml 70 % [v/v] Ethanol überführt und ü. N. bei -20 °C inkubiert. Das Ethanol wurde abgenommen, das DNA-Pellet für 10 - 20 min bei Raumtemperatur getrocknet und abschließend in 200 - 500 µl TE-Puffer gelöst.

### RNA-Isolation

Zur Isolation von RNA aus Zelllinien und Gewebeproben wurde das *RNeasy Mini Kit* verwendet. Während der Lyse der biologischen Proben in einem stark denaturierenden Guanidin-Thiocyanat-haltigen Puffer werden RNasen inaktiviert und die RNA im Anschluss nach Zugabe von Ethanol und Waschpuffer über Bindung an eine, in einer zentrifugierbaren Säule befindliche, Silika-Membran aufgereinigt. Es werden hierbei selektiv RNA-Moleküle >200 nt gebunden und aufgereinigt.

Kultivierte Zellen einer 75 cm<sup>2</sup> Zellkulturflasche bzw. transfizierte Zellen aus zwei bis drei *Wells* einer 12-*Well*-Platte wurden nach Abnahme des Mediums mit 1,5 ml/Flasche bzw. 0,5 ml/*Well* 1x DPBS gewaschen, durch Zugabe von 1,5 ml/Flasche bzw. 0,5 ml/*Well* 0,05 % Trypsin-EDTA vom Boden gelöst, in 8,5 ml Medium aufgenommen und durch 5-minütige Zentrifugation bei 300 g pelletiert. Der Überstand wurde abgenommen und die Zellen wurden bei -70 °C gelagert oder direkt weiterverwendet. Nach Resuspension der Zellen in 600 µl des im *Kit* enthaltenen *RLT*-Puffer wurden diese auf eine *QIAshredder* Säule gegeben, für 2 min bei 15.000 g zentrifugiert und die RNA wurde aus dem Durchfluss anschließend über die *RNeasy Mini Spin* Säule nach Herstellerprotokoll aufgereinigt.

Bei -70 °C gelagerte Gewebeproben (~0,25 cm<sup>3</sup>) wurden zunächst mit einem Skalpell im noch gefrorenen Zustand zerkleinert, in 1,5 ml kaltem Trizol homogenisiert, auf eine *QIAshredder* Säule gegeben und für 2 min bei 15.000 g zentrifugiert. Der Durchfluss wurde in ein 15 ml Falcon-Gefäß überführt und für 5 min in einem Wasserbad bei 30 °C inkubiert, danach mit 0,4 ml Chloroform versetzt, durch 40-sekündiges Vortexen gemischt, 3 min bei 30 °C inkubiert und für 20 min bei 4 °C und 8.600 g zentrifugiert. Die obere, wässrige Phase wurde in ein neues Gefäß überführt, die darin befindliche RNA durch Zugabe von 1 Vol. 70 % [v/v] Ethanol gefällt und anschließend über eine *RNeasy Mini Spin* Säule nach Herstellerprotokoll aufgereinigt.

Zur Überprüfung der Qualität der aufgereinigten RNA wurde 1 µg RNA in einem 1%-igen Agarose-Gel (1x MOPS mit 1 % [v/v] Formaldehyd) aufgetrennt.

### 3.2.1.2. Konzentrationsbestimmung von Nukleinsäuren und Bakterienkulturen

#### Spektralphotometrische DNA- und RNA-Konzentrationsbestimmung

Die Konzentration isolierter DNA und RNA wurde mit einem *NanoDrop<sup>TM</sup> 2000* Spektralphotometer durch Messung der optischen Dichte bei 260 nm ( $OD_{260}$ ) bestimmt. Zur Abschätzung von Proteinkontaminationen wurde hierbei zusätzlich die  $OD_{280}$  gemessen, wobei der Quotient  $OD_{260}/OD_{280}$  Aufschluss über die Reinheit der Probe gibt und bei etwa 1,8 - 2 liegen sollte.

#### Fluorometrische DNA-Konzentrationsbestimmung

Die Konzentration von PCR-Produkten zur anschließenden *454/FLX*-Sequenzierung wurde mit dem *Quant-iT<sup>TM</sup> dsDNA HS Assay Kit (0,2 - 100 ng)* mit einem *Qubit<sup>TM</sup>* Fluorometer bestimmt. Die Messung erfolgte nach Herstellerprotokoll. Es wurden jeweils 2  $\mu$ l der aus dem Agarosegel aufgereinigten DNA eingesetzt. Die Messungen wurden im Institut für Genetik/Epigenetik der Universität des Saarlandes (Saarbrücken, Deutschland) durchgeführt.

#### Photometrische Messung von Bakterienkulturen

Das Wachstum von Bakterienkulturen wurde photometrisch mit einem *BioPhotometer* durch Messung der optischen Dichte bei 600 nm ( $OD_{600}$ ) bestimmt.

### 3.2.1.3. DNase-Behandlung und Synthese von cDNA

#### DNase-Behandlung

Um Reste genomischer DNA aus aufgereinigten RNA-Proben zu entfernen, wurden diese mit dem *Turbo DNA-free<sup>TM</sup> Kit* einem rigorosen DNase-Verdau unterzogen. Es wurden je 10  $\mu$ g RNA in einer 50  $\mu$ l Reaktion nach Herstellerprotokoll behandelt. Im Falle der Wiederholung des Experimentes mit einer weiteren Passage der Prostatakarzinom-Zelllinien PC-3, MDA PCA 2B sowie VCaP standen lediglich 5 ng RNA zur Verfügung, die vollständig in die Reaktion eingesetzt wurden. Ein Ansatz setzte sich wie folgt zusammen:

10 $\mu$ g	RNA
1x	<i>Turbo DNase<sup>TM</sup> Buffer</i>
2 U	<i>Turbo DNase<sup>TM</sup></i>
ad 50 $\mu$ l	DEPC- $H_2O_{dd}$

Der Ansatz wurde für 30 min bei 37 °C inkubiert. Das Enzym wurde durch Zugabe von 5 µl *DNase Inactivation Reagent* und 5 min Inkubation bei Raumtemperatur inaktiviert. Der Ansatz wurde während dieser Zeit 2 - 3x gemischt und das *Inactivation Reagent* anschließend für 1,5 min bei 2.100 g abzentrifugiert. Der RNA-enhaltende Überstand wurde in ein frisches 200 µl Reaktionsgefäß überführt und bei -70 °C gelagert.

Der erfolgreiche Verdau wurde durch eine Alu-Element-spezifische PCR (s.u.) sowie weitere Kontrollen in nachfolgenden Experimenten überprüft. Es wurden nur RNA-Proben weiterverwendet, die negativ auf Reste genomischer DNA getestet wurden.

### **cDNA-Erststrangsynthese**

cDNA wurde mit dem *Omniscript RT Kit* in einer 30 µl Reaktion nach Herstellerangaben generiert. RNA wird hierbei, ausgehend von einem der Reaktion zugegebenen Oligonukleotid-*Primer*, vom Enzym Reverse Transkriptase in eine komplementäre einzelsträngige cDNA umgeschrieben und die im Ansatz enthaltene RNA anschließend durch die RNase H Aktivität des Enzyms degradiert.

RNA und *Random-Hexaoligonukleotide* wurden gemischt, für 5 min bei 70 °C denaturiert und auf Raumtemperatur abgekühlt, um den *Primern* die Anlagerung an die RNA zu ermöglichen. Für jede RNA-Probe wurden zwei Reaktionen angesetzt, wobei eine Reaktion ohne Enzym als RT(-) Negativkontrolle diente (eine Ausnahme bildete die RNA aus SK-Mel-25, SK-Mel-28 und MEWO, hier wurde nur eine RT(-) Kontrolle für einen gemischten Ansatz dieser drei RNAs pipettiert). Es wurde ein *Mastermix* aus Puffer, dNTPs und RNase-Inhibitor hergestellt und auf die Reaktionen verteilt. Im Anschluss wurde das Enzym zu den RT(+) Reaktionen, bzw. eine entsprechende Menge DEPC-H<sub>2</sub>O<sub>dd</sub> zu den RT(-) Negativkontrollen gegeben.

Ein Ansatz setzte sich wie folgt zusammen:

6 µl	DNase-behandelte RNA
1x	<i>Buffer RT</i>
je 0,5 mM	dNTPs
10 µM	<i>Random-Hexaoligonukleotid-Primer</i>
15 U	<i>RNaseOUT™ Ribonuclease Inhibitor</i>
6 U	Reverse Transkriptase
ad 30 µl	DEPC-H <sub>2</sub> O <sub>dd</sub>

Der Ansatz wurde für 60 min bei 37 °C inkubiert und das Enzym anschließend für 5 min bei 93 °C hitzeinaktiviert.

Der Erfolg der Reaktion wurde durch eine  $\beta$ -Actin-spezifische PCR (s.u.) überprüft. Es wurden ausschließlich Ansätze weiterverwendet, die negativ auf DNA in der RT(-) Negativkontrolle getestet wurden.

### 3.2.1.4. Polymerase-Kettenreaktion (PCR)

Die Polymerase-Kettenreaktion (*polymerase chain reaction*, PCR) ist ein Verfahren zur *in vitro* Amplifikation eines gewünschten DNA-Abschnittes. Die DNA wird zunächst durch Erhitzen denaturiert. Nach Abkühlen der Reaktion auf die *Annealing*-Temperatur der einzelsträngigen Oligonukleotid-*Primer*, die den gewünschten DNA-Bereich flankieren, lagern sich diese an die komplementären Stränge der DNA an. Der Ansatz wird im Anschluss auf das Temperaturoptimum der verwendeten DNA-Polymerase erhitzt, woraufhin diese, ausgehend vom 3'-Ende der *Primer*, mit der Neusynthese des jeweils komplementären DNA-Stranges beginnt. Durch zyklische Wiederholung dieser drei Schritte kommt es zu einer exponentiellen Amplifikation des gewünschten DNA-Abschnittes (Mullis, 1990).

In allen im Rahmen dieser Arbeit durchgeführten PCRs diente ein Ansatz mit H<sub>2</sub>O<sub>dd</sub> statt DNA bzw. cDNA als Kontaminationskontrolle.

#### 3.2.1.4.1. PCR-Amplifikation aus genomischer DNA

Amplikons in den *gag*- und *rec/np9*-Bereichen von HERV-K(HML-2)-Loci sowie in zwei Bereichen im *env*-ORF von HERV-W-Loci, bezeichnet als 5'*env* und 3'*env*, wurden nach folgendem Ansatz aus genomischer DNA amplifiziert. Die verwendeten *Primer* sind in den **Tabellen 4** und **5** zusammengefasst. Nach dem gleichen Ansatz wurden zudem die beiden potentiellen neuen HML-2-Proviren mit Ähnlichkeit zu *ERVK-5* und *ERVK-18* mit den *Primern* aus **Tabelle 10** aus genomischer DNA amplifiziert.

PCR-Ansatz:	PCR-Programm:
1x <i>PCR Buffer minus Mg<sup>++</sup></i>	95 °C 3 min
je 0,2 mM dNTPs	95 °C 50 s
1,5 mM MgCl <sub>2</sub>	53 °C* 50 s
je 0,5 µM <i>Forward-/Reverse-Primer(-Mix)</i>	72 °C X min <sup>#</sup>
1 U <i>Taq</i> DNA-Polymerase	72 °C 10 min
100 ng genomische DNA	10 °C ∞
ad 20 µl H <sub>2</sub> O <sub>dd</sub>	

} x 40 - 45

\*Für den *Annealing*-Schritt wurde eine relativ niedrige Temperatur von 53 °C gewählt, die jedoch noch eine effiziente Amplifikation der PCR-Produkte erlaubt. Eine geeignete Temperatur wurde zuvor in einer Gradienten-PCR ermittelt.

#Die Elongationzeit betrug für die HML-2 Amplikons *gag* und *rec/np9* sowie die HERV-W-Amplikons *5'env* und *3'env* je 1 min. Für die Amplifikation der potentiellen neuen Proviren mit Ähnlichkeit zu *ERVK-5* bzw. *ERVK-18* wurden Elongationszeiten von 140 s bzw. 15 s gewählt.

Die 506 bzw. 724 bp großen ORF-Amplikons der beiden durch reverse Transkription und Reintegration gespleißter mRNA entstandenen HERV-K(HML-2)-Loci auf den Chromosomen 2q32.1 und 10q24.2 wurden mit einem leicht modifizierten Ansatz amplifiziert. Die verwendeten *Primer* besaßen an ihrem 5'-Ende jeweils Schnittstellen für das Restriktionsenzym *Bam*HI (vgl. **Tab. 9**). Als *Template* dienten etwa 100 ng eines *pGEM*<sup>®</sup>-*T Easy*-Vektorkonstruktes, in das der komplette 2q32.1-Locus bzw. der ORF- und 3'-LTR-Bereich des 10q24.2-Locus ligiert war. Die *pGEM*<sup>®</sup>-*T Easy*-Konstrukte wurden im Rahmen von Biermann (2013) generiert. Nach dem gleichen Ansatz wurde ein 363 bp bzw. 230 bp großer Abschnitt des *rec/np9*-Bereiches der beiden Loci aus 50 ng genomischer DNA amplifiziert. Die verwendeten *Primer* sind in **Tabelle 8** zusammengefasst.

PCR-Ansatz:	PCR-Programm:	
1x <i>PCR Buffer minus Mg</i> <sup>++</sup>	95 °C	3 min
je 0,2 mM dNTPs	95 °C	50 s
1,5 mM <i>MgCl</i> <sub>2</sub>	X °C*	50 s
je 0,5 µM <i>Forward-/Reverse-Primer</i>	72 °C	1 min
1 U <i>Taq DNA-Polymerase</i>	72 °C	10 min
100/50 ng genomische DNA/ <i>pGEM</i> <sup>®</sup> - <i>T Easy</i> -Konstrukt	10 °C	∞
ad 50 µl <i>H</i> <sub>2</sub> O <sub>dd</sub>		

} x 40

\*Die *Annealing*-Temperatur der ORF-Amplikon-*Primer* lag bei 55 °C und die der 2q32.1 bzw. 10q24.2 spezifischen *Primer* bei 50 bzw. 62 °C.

### **Long-PCR Amplifikation polymorpher HML-2-Proviren aus genomischer DNA**

Der Allelstatus polymorpher HERV-K(HML-2)-Loci in genomischer DNA der in der vorliegenden Arbeit verwendeten Melanom- und Melanozyten-Zelllinien wurde mit *long*-PCRs getestet. Hierfür wurden jeweils den Locus flankierende *Primer* verwendet, wobei die Größe des entstandenen PCR-Produktes Aufschluss darüber lieferte, ob in der getesteten DNA ein komplettes Provirus,

ein durch homologe Rekombination entstandener solitärer LTR oder eine leere Präintegrationsstelle vorhanden ist, wenn eines der möglichen Allele noch nicht in der Population fixiert ist. In einer zweiten PCR wurde das Vorhandensein eines Provirus oder solitären LTRs durch Verwendung eines flankierenden *Primers* sowie eines Primers im LTR bestätigt. Die verwendeten *Primer* sind in **Tabelle 7** zusammengefasst. Im Falle des 1p31.1 Locus, der von LINE-Sequenzen verschiedener Subfamilien (L1HS und L1PH2) flankiert wird, war eine Amplifikation mit flankierenden *Primern* nicht möglich. Stattdessen wurde hier ein flankierender *Primer* sowie ein zweiter *Primer* stromab des 5'-LTRs im Provirus verwendet.

PCR-Ansatz:		PCR-Programm:	
1x	<i>Phusion HF Buffer</i>	98 °C	30 s min
je 200 µM	dNTPs	98 °C	10 s
1,5 mM	MgCl <sub>2</sub>	62 °C	30 s
0,5 µM	<i>Forward-Primer</i>	72 °C	5 min/45 s*
0,5 µM	<i>Reverse-Primer</i>	72 °C	10 min
0,4 U	<i>Phusion® High-Fidelity DNA Polymerase</i>	10 °C	∞
100 ng	genomische DNA		
ad 20 µl	H <sub>2</sub> O <sub>dd</sub>		

\*Die Elongationszeit betrug bei Amplifikation mit zwei flankierenden *Primern* 5 min, bei Amplifikation mit einem flankierenden *Primer* und einem *Primer* im LTR 45 s.

Alle erwarteten Produktgrößen sind in **Tabelle 13** zusammengefasst.

**Tabelle 13:** Erwartete PCR-Produktgrößen im Zuge der Überprüfung des Allelstatus polymorpher HERV-K(HML-2)-Loci in genomischer DNA.\*

Locus	Primer-Kombination	erwartete Produktgrößen [bp]
1p31.1	1p31.1_flank_for + 1p31.1_int	1323 (PV)
	1p31.1_flank_for + 1p31.1_LTR	867
3p25.3	3p25.3_flank_for + 3p25.3_flank_rev	7291 (PV), 1318 (sLTR), 400 (PI)
	3p25.3_flank_for + 3p25.3_LTR	320
3q13.2 (K106)	3q13.2_flank_for + 3q13.2_flank_rev	9704 (PV), 1505 (sLTR)
	3q13.2_flank_for + 3q13.2_LTR	382
6q14.1 (K109)	6q14.1_flank_for + 6q14.1_flank_rev	9794 (PV), 1332 (sLTR)
	6q14.1_flank_for + 3q13.2_LTR	346
7p22.1 (K108, HOM)	5'flankFOR + 3'flankREV	9834 (PV), 18338 (TPV)
	5'flankFOR + LTRgagREV2	1361
8p23.1 (K115)	K115_flank_for + K115_flank_rev	10025 (PV), 1521 (sLTR), 561 (PI)
	K115_flank_for + K115_LTR	549
19p12 (K113)	K113_flank_for + K113_flank_rev	9781 (PV), 1277 (sLTR), 303 (PI)
	K113_flank_for + K113_LTR	976

\*Angegeben sind der jeweilige Locus, Alternativnamen (HOM = HERV-K(HML-2.HOM)) sowie die bei den unterschiedlichen *Primer*-Kombinationen erwarteten Produktgrößen. Die Position des jeweiligen *Primers* geht aus dem Namen hervor. Es wurden den Locus flankierende *Primer* (*flank*), interne *Primer* im Provirus (*int* bzw. *gag*), sowie *Primer* im 5'-LTR (*LTR*) verwendet. *for* = *Forward-Primer*, *rev* = *Reverse-Primer*, PV = Provirus, TPV = Tandem-*Provirus*, sLTR = solitärer LTR, PI = Präintegrationsstelle.

### 3.2.1.4.2. PCR-Amplifikation aus cDNA

Amplikons im *gag*- sowie *rec/np9*-Bereich von HERV-K(HML-2)-Loci sowie die spezifischen *rec/np9*-Produkte der beiden durch reverse Transkription und Reintegration gespleißter mRNA entstandenen HML-2-Loci auf den Chromosomen 2q32.1 und 10q24.2 wurden nach folgendem Ansatz aus cDNA der untersuchten Zelllinien und Gewebe amplifiziert. Die verwendeten *Primer* sind in den **Tabellen 4** und **8** zusammengefasst. Als Negativkontrolle diente für jede untersuchte Probe ein Ansatz mit 1 µl der RT(-) Kontrolle als *Template*.

PCR-Ansatz:	PCR-Programm:
1x <i>PCR Buffer minus Mg<sup>++</sup></i>	95 °C 5 min
je 0,2 mM dNTPs	95 °C 50 s
1,5 mM MgCl <sub>2</sub>	53 °C* 50 s
0,5 µM <i>Forward-Primer(-Mix)</i>	72 °C 1 min
0,5 µM <i>Reverse-Primer(-Mix)</i>	72 °C 10 min
2,5 U <i>Taq DNA-Polymerase</i>	10 °C ∞
1 µl cDNA	
ad 50 µl H <sub>2</sub> O <sub>dd</sub>	

} x 40 - 45

Die Amplikons 5'*env* und 3'*env* im *env*-ORF von HERV-W-Loci, die im Anschluss mittels 454/FLX-System und *MiSeq/Illumina* sequenziert wurden, wurden nach folgendem Ansatz aus cDNA amplifiziert. Als Negativkontrolle diente für jede Probe ein Ansatz mit 1 µl der Negativkontrolle RT(-) als *Template*. Die verwendeten *Primer* finden sich in **Tabelle 5**.

PCR-Ansatz:	PCR-Programm:
1x <i>Buffer B</i>	95 °C 15 min
10 mM dNTPs	95 °C 1 min
2,5 mM MgCl <sub>2</sub>	53 °C* 1 min
0,6 µM <i>Forward-Primer-Mix*</i>	72 °C 45 min
0,6 µM <i>Reverse-Primer-Mix*</i>	72 °C 10 min
2,5 U <i>HOT FIREPoI DNA-Polymerase</i>	10 °C ∞
1 µl cDNA	
ad 30 µl H <sub>2</sub> O <sub>dd</sub>	

} x 32

\*Für die Amplifikation zur anschließenden 454/FLX-Sequenzierung wurden *Primer* verwendet, die an ihrem 5'-Ende zur Sequenzierung benötigte *Adaptor*- und *Key*-Sequenzen aufwiesen. Diese sind in **Tabelle 6** zusammengefasst. Im Falle der *MiSeq/Illumina*-Sequenzierung erfolgte

die Ligation von entsprechenden Adaptoren erst nachträglich im Zuge der *Library-Preparation* durch die Seq-IT GmbH (Kaiserslautern, Deutschland).

### 3.2.1.4.3. Kolonie-PCR

Die erfolgreiche Ligation amplifizierter PCR-Produkte in den jeweiligen Zielvektor wurde mittels Kolonie-PCR überprüft. Hierfür wurden im Falle von *pGEM<sup>®</sup>-T Easy* und *pCR<sup>®</sup>-2.1-TOPO* die vektorständigen *Primer* M13\_for/rev verwendet, im Falle von *pEGFP-N1* die *Primer* pEGFP-N1\_for/rev (vgl. **Tab. 12**). Diese flankieren den jeweiligen Insertionsbereich und liefern für den Leervektor ein 290 bp (*pGEM<sup>®</sup>-T Easy*), 245 bp (*pCR<sup>®</sup>-2.1-TOPO*) oder 175 bp (*pEGFP-N1*) großes, bzw. im Falle einer erfolgreichen Ligation ein der *Insert*-Größe entsprechend größeres Produkt. Dem PCR-Ansatz wurde ein Bakterienklon als Ausgangsmaterial zugegeben. Während des ersten Denaturierungsschrittes zerplatzen die Bakterienzellen und geben die Vektor-DNA frei, die dann als *Template* der PCR-Amplifikation dient.

PCR-Ansatz:		PCR-Programm:		
1	Bakterienklon	95 °C	5 min	
1x	<i>PCR Buffer minus Mg<sup>++</sup></i>	95 °C	30 s	} x 30
je 0,2 mM	dNTPs	50 °C*	30 s	
1,5 mM	MgCl <sub>2</sub>	72 °C	1 min	
0,5 µM	<i>Forward-Primer-Mix</i>	72 °C	10 min	
0,5 µM	<i>Reverse-Primer-Mix</i>	10 °C	∞	
1 U	<i>Taq DNA-Polymerase</i>			
ad 20 µl	H <sub>2</sub> O <sub>dd</sub>			

### 3.2.1.4.4. Kontroll-PCRs

#### Alu-PCR

Nach der DNase Behandlung wurden die RNA-Proben zur Ermittlung von DNA-Kontaminationen einer Kontroll-PCR unterzogen. In dieser werden durch Verwendung eines einzelnen Alu-Element spezifischen *Primers* genomische Bereiche zwischen zwei Alu-Elementen amplifiziert, die sich in gegensätzlicher Orientierung zueinander befinden und maximal so weit voneinander entfernt sind, dass eine PCR-Amplifikation noch möglich ist (vgl. Brooks-Wilson *et al.*, 1990; die *Primer*-Sequenz findet sich in **Tabelle 12**). Nur Proben, in denen es zu keiner nachweisbaren

DNA-Amplifikation kam, wurden für die weiteren Experimente verwendet. Als Positivkontrolle diente ein Ansatz mit 50 ng humaner genomischer DNA als *Template*.

PCR-Ansatz:	PCR-Programm:
1x <i>PCR Buffer minus Mg<sup>++</sup></i>	94 °C 5 min
je 0,2 mM dNTPs	94 °C 1 min
1,5 mM MgCl <sub>2</sub>	56 °C* 1 min
0,5 µM <i>Alu A1S-Primer</i>	72 °C 2 min
2,5 U <i>Taq DNA-Polymerase</i>	72 °C 10 min
1 µl DNase-behandelte RNA	10 °C ∞
ad 50 µl H <sub>2</sub> O <sub>dd</sub>	

} x 30

### β-Actin PCR

Zur Überprüfung des Erfolges der cDNA-Erststrangsynthese sowie zum Ausschluss einer DNA-Kontamination in den RT(-) Negativkontrollen, wurden die RT(+) und RT(-) Reaktionen nach der reversen Transkription einer Kontroll-PCR unterzogen, in der ein 722 bp großer Abschnitt des β-Actin-Haushaltsgens *ACTB* amplifiziert wurde. Die verwendeten *Primer* sind in **Tabelle 12** aufgelistet. Nur Proben, in denen es in der RT(-) Kontrolle zu keiner nachweisbaren DNA-Amplifikation kam, wurden für die weiteren Experimente verwendet.

PCR-Ansatz:	PCR-Programm:
1x <i>PCR Buffer minus Mg<sup>++</sup></i>	95 °C 5 min
je 0,2 mM dNTPs	95 °C 1 min
1,5 mM MgCl <sub>2</sub>	57 °C* 45 s
0,5 µM <i>Forward-Primer</i>	72 °C 1 min
0,5 µM <i>Reverse-Primer</i>	72 °C 10 min
2,5 U <i>Taq DNA-Polymerase</i>	10 °C ∞
1 µl RT(+) Ansatz bzw. RT(-) Kontrolle	
ad 50 µl H <sub>2</sub> O <sub>dd</sub>	

} x 25

#### 3.2.1.4.5. Quantitative *Real-Time*-PCR

Die relativen Transkriptmengen der HERV-Gruppen HERV-W und HERV-K(HML-2) wurden mittels *Real-Time*-PCR quantifiziert. Die DNA wird hierbei nach dem Verfahren einer PCR amplifiziert, wobei die Quantifizierung am Ende jeder Elongationsphase durch Messung der Fluoreszenzintensität des interkalierenden Farbstoffes *SYBR Green* erfolgt. Dieser lagert sich in

die neu synthetisierte doppelsträngige DNA ein, wobei die Zunahme der Fluoreszenz mit jedem Zyklus der PCR mit der DNA-Menge korreliert. Für die *Real-Time-PCR* wurde das *KAPA™ SYBR® FAST qPCR MasterMix Universal Kit* und der *StepOnePlus™ Real-Time-PCR Cycler* verwendet.

Ansatz:

1x	<i>KAPA™ SYBR® FAST qPCR MasterMix Universal</i>
10 µM	<i>Forward-Primer*</i>
10 µM	<i>Reverse-Primer*</i>
1x	<i>ROX Reference Dye High</i>
1 µl	cDNA oder RT(-) Negativkontrolle (1:10 verdünnt)
ad 10 µl	H <sub>2</sub> O <sub>dd</sub>

\*Zur Quantifizierung der Transkriptmengen von HERV-K(HML-2) *gag*, sowie HERV-W 5'*env* wurden die in den **Tabellen 4** und **5** aufgelisteten PCR-*Primer*-Mischungen verwendet. Im Falle von HERV-W wurden zusätzlich die Transkriptmengen der Loci *syncytin-1/ERVWE1/ERVW-1* sowie *ERVWE2/ERVW-2* quantifiziert. Die *Primer* finden sich in **Tabelle 11**.

Programm:

*Cycling Stage:*

95 °C	1 min	
95 °C	3 s	} x 40
55 °C	3 s	
72 °C	30 s	

*Melt Curve Stage:*

95 °C	15 s
60 °C	1 min
95 °C	15 s

Die Messung des Reportersignals erfolgte alle 0,3 °C während der zweiten 95 °C Phase.

Die gemessenen HERV-Transkriptmengen wurde gegen die Transkriptmengen der Haushaltsgene *G6PDH* und *RPII* normalisiert (die *Primer* finden sich in **Tabelle 11**). Die RT(+) cDNA-Proben jedes Amplikons wurden in Triplikaten gemessen, die RT(-) Negativkontrollen je einmal. Es wurde zudem für jedes Amplikon ein Kontrollansatz mit 1 µl H<sub>2</sub>O<sub>dd</sub> gemessen. Für die Normalisierung und weitere Analyse der relativen Transkriptmengen wurde die *StepOne™ Software Version 2.2.2* verwendet.

### 3.2.1.5. Auftrennung und Aufreinigung von Nukleinsäuren

#### Agarose-Gelelektrophorese

Nukleinsäurefragmente unterschiedlicher Größe wurden mittels Agarose-Gelelektrophorese aufgetrennt. Die, bedingt durch das Phosphatrückgrat, negativ geladenen DNA- und RNA-Fragmente wandern hierbei nach Anlegen eines elektrischen Feldes durch ein poröses Agarosegel zur Anode und werden hierbei der Größe nach aufgetrennt.

Die Auftrennung von DNA erfolgte in 1%-igen [w/v] Agarosegelen (1x TAE-Puffer) mit 0,2 µg/ml Ethidiumbromid, die Auftrennung von RNA in 1%-igen [w/v] Agarosegelen (1x MOPS-Puffer mit 2 % Formaldehyd) bei Spannungen von 70 - 110 V. Die Dokumentation erfolgte mit einem UV-Transilluminator und der CCD-Kamera *LU105M* sowie der *DeVision G Software Version 2.0*. Bei aufgetrennten DNA-Fragmenten aus Restriktionsverdauen wurde zur Schonung der DNA auf eine Färbung mit Ethidiumbromid und anschließende UV-Detektion verzichtet. Stattdessen enthielt das 1%-ige [w/v] Agarosegel (1x TAE-Puffer) den Fluoreszenzfarbstoff *GelStar™ Nucleic Acid Gel Strain* in einfacher Endkonzentration. Die Detektion der Banden erfolgte im Anschluss mit dem Blaulicht-Transilluminator *Flu-O-blu* bei 470 nm.

Aus Gründen der Übersichtlichkeit wurden bei den in der vorliegenden Arbeit gezeigten Gelbildern irrelevante Bereiche, wie leere Spuren oder Spuren, in denen DNA-Fragmente aus anderen Projekten aufgetrennt wurden, ausgeschnitten. Die restlichen Teile wurden maßstabsgetreu zusammengefügt, was jeweils durch einen weißen Spalt zwischen den Bildbereichen deutlich gemacht wurde.

#### Extraktion von Nukleinsäuren aus Agarosegelen

Zur Aufreinigung von DNA aus Agarosegelen wurde das *QIAquick Gel Extraction Kit* verwendet. DNA wird hierbei unter hoher Salzkonzentration an eine in einer zentrifugierbaren Säule befindliche Silika-Membran gebunden, gewaschen und anschließend unter niedriger Salzkonzentration eluiert. Die Aufreinigung erfolgte nach Herstellerprotokoll, wobei in 30 µl H<sub>2</sub>O<sub>dd</sub> eluiert wurde.

#### Isolation von Plasmid-DNA

Zur Isolation von 15 - 25 µg Plasmid-DNA aus Bakterienzellen wurde das *peqGOLD Plasmid Miniprep Kit I* verwendet. Die Bakterienzellen werden hierbei durch alkalische Lyse aufgeschlossen und die Plasmid-DNA an eine in einer zentrifugierbaren Säule befindliche Silika-

Membran gebunden, durch Waschpuffer von Kontaminationen befreit und in  $\text{H}_2\text{O}_{\text{dd}}$  eluiert. Ein das Plasmid von Interesse enthaltender Bakterienklon wurde in 3 ml LB-Medium mit 100  $\mu\text{g}/\text{ml}$  Ampicillin oder 30  $\mu\text{g}/\text{ml}$  Kanamycin angeimpft, ü. N. bei 37 °C und 230 rpm inkubiert und die Plasmid-DNA im Anschluss nach Herstellerprotokoll aufgereinigt.

Zur Isolation größerer Plasmidmengen (max. 100  $\mu\text{g}$ ) aus Bakterienzellen diente das *QIAGEN Plasmid Midi Kit*. Hierbei werden die Bakterienzellen zunächst alkalisch lysiert und die Plasmid-DNA anschließend spezifisch unter niedriger Salzkonzentration und geeigneten pH-Bedingungen an eine Anionen-Austausch-Säule gebunden, gewaschen und unter hoher Salzkonzentration eluiert. Die Plasmid-DNA wird im Anschluss durch Isopropanolfällung aufkonzentriert und entsalzt. Es wurde ein Bakterienklon in 2 ml LB-Medium mit 100  $\mu\text{g}/\text{ml}$  Ampicillin oder 30  $\mu\text{g}/\text{ml}$  Kanamycin angeimpft und für 8 h bei 37 °C und 230 rpm inkubiert. Von dieser Starterkultur wurden im Anschluss 25  $\mu\text{l}$  in 25 ml LB-Medium übergeimpft, ü. N. bei 37 °C und 230 rpm inkubiert und die Plasmid-DNA im Anschluss nach Herstellerprotokoll aufgereinigt.

Zur Plasmid-Aufreinigung im 96-Well-Maßstab wurden das *Agencourt CosMCPrep Kit* und eine *96R Ring* Magnetplatte verwendet. Die Aufreinigung basiert auf der *Solid Phase Reversible Immobilization (SPRI<sup>®</sup>)*-Technik. Die Plasmid-DNA bindet hierbei unter spezifischen Pufferbedingungen an paramagnetische Mikropartikel, die sich nach Platzierung der 96-Well-Platte in der Magnetplatte an der Wand jedes *Wells* sammeln und so Waschschriffe sowie das Eluieren der Plasmid-DNA ohne Zentrifugation oder Filtration ermöglichen. Es wurde pro *Well* einer 96 *Deep-Well*-Platte je ein Bakterienklon in 800  $\mu\text{l}$  LB-Medium mit 100  $\mu\text{g}/\text{ml}$  Ampicillin angeimpft und ü. N. bei 37 °C und 230 rpm inkubiert. Die Aufreinigung erfolgte im Anschluss nach Herstellerprotokoll.

### 3.2.1.6. Ligation und Transformation

#### Ligation von DNA-Fragmenten

Die Ligation von PCR-Produkten in den *pGEM<sup>®</sup>-T Easy*-Vektor beruhte auf dem Prinzip der TA-Klonierung. Die zur PCR verwendete thermostabile *Taq*-Polymerase erzeugt am 3'-Ende neu synthetisierter Fragmente einen Adenosin-Überhang. Der linearisierte *pGEM<sup>®</sup>-T Easy*-Vektor besitzt an jedem freien 3'-Ende ein überhängendes Thymidin, was die einfache Ligation der PCR-Produkte ermöglicht. Aus genomischer DNA oder cDNA amplifizierte PCR-Produkte

wurden nach folgendem Ansatz in den *pGEM<sup>®</sup>-T Easy*-Vektor ligiert. Der Ansatz wurde über Nacht bei 4 °C inkubiert.

1x            *Ligation Buffer*  
25 - 50 ng *pGEM-T<sup>®</sup> Easy*-Vektor  
1,5 - 3 U    T4 DNA Ligase  
2 - 3,5 µl   PCR-Produkt  
ad 10 µl    H<sub>2</sub>O<sub>dd</sub>

Die amplifizierten ORFs der durch reverse Transkription und Reintegration gespleißter mRNA entstandenen HERV-K(HML-2)-Loci auf den Chromosomen 2q32.1 und 10q24.2 wurden nach *Bam*HI-Verdau nach folgendem Ansatz in den linearisierten und dephosphorylierten *pEGFP-N1*-Vektor ligiert.

1x            *Ligation Buffer*  
0,5 µl       linearisierter und dephosphorylierter *pEGFP-N1*-Vektor\*  
3 U          T4 DNA Ligase  
1 - 3 µl      *Bam*HI-verdautes, aufgereinigtes PCR-Produkt  
ad 10 µl     H<sub>2</sub>O<sub>dd</sub>

\*Für die Ligation wurde *pEGFP-N1*-Vektor verwendet, der im Rahmen von Biermann (2013) mit *Bam*HI geschnitten und dephosphoryliert worden war.

### Herstellung chemokompetenter Bakterien

Zur Transformation verwendbare, chemokompetente Bakterienzellen wurden mittels der Calciumchloridmethode hergestellt. Die Behandlung mit Calciumchlorid reduziert die abstoßenden Kräfte, die zwischen negativ geladener DNA und der ebenfalls negativ geladenen Zellmembran der Bakterien bestehen.

Als Vorkultur wurden 10 µl *E. coli Library Efficiency<sup>®</sup> DH5α<sup>TM</sup> Competent Cells* (Invitrogen/Life Technologies, Carlsbad, USA) in 3 ml SOB-Medium angeimpft und ü. N. bei 37 °C und 225 rpm inkubiert. Die Übernachtskultur wurde im Anschluss vollständig in 500 ml SOB-Medium überimpft und bei 37 °C und 225 rpm bis zu einer Dichte von OD<sub>600</sub> = 0,6 inkubiert. Die Kultur wurde für 15 min bei 4.000 g zentrifugiert, der Überstand verworfen und das Bakterienpellet in 100 ml kaltem TfbI-Puffer auf Eis resuspendiert. Die Bakterien wurden erneut 8 min bei 4.000 g zentrifugiert, der Überstand wurde verworfen und das Bakterienpellet in 20 ml kaltem TfbII auf

Eis resuspendiert. Die Bakterien wurden zu je 550 µl aliquotiert, in flüssigem Stickstoff weggefroren und bei -70 °C gelagert.

### Transformation von Plasmid-DNA in Bakterienzellen

Zur Transformation wurden chemokompetente Bakterien des *E. coli* Stammes DH5α verwendet. 50 µl kompetente Zellen wurden mit 5 µl des Ligationsansatzes versetzt und für 30 min auf Eis inkubiert, währenddessen sich die Plasmid-DNA an die Bakterienwände anlagert. Der Hitzeschock, bei dem sich Poren in der Zellmembran öffnen, wodurch Plasmid-DNA in die Bakterienzellen gelangen kann, erfolgte für 45 s bei 42 °C. Die Bakterien wurden im Anschluss für 3 min auf Eis inkubiert, wobei sich die Poren der Zellmembran wieder schließen, danach mit 300 µl SOC-Medium versetzt und für eine Stunde bei 37 °C inkubiert. Der Ansatz wurde auf LB-Agarplatten mit 100 µg/ml Ampicillin und 50 µg/ml X-Gal (*pGEM<sup>®</sup>-T Easy*) oder 30 µg/ml Kanamycin (*pEGFP-N1*) ausplattiert und ü. N. bei 37 °C inkubiert.

#### 3.2.1.7. Restriktionsverdau von Plasmid-DNA

Die aus genomischer DNA amplifizierten und in *pGEM<sup>®</sup>-T Easy* klonierten DNA-Fragmente der ORFs der durch reverse Transkription und Reintegration gespleißter mRNA entstandenen HERV-K(HML-2)-Loci auf den Chromosomen 2q32.1 und 10q24.2 wurden nach folgendem Ansatz mit dem Restriktionsenzym *Bam*HI aus *pGEM<sup>®</sup>-T Easy* geschnitten. Der Restriktionsansatz wurde für 3 h bei 37 °C inkubiert.

1x	Puffer
1-3 µg	<i>pGEM<sup>®</sup>-T Easy</i> mit ORF-Insert
100 U	<i>Bam</i> HI-HF™

#### 3.2.1.8. DNA-Sequenzierung

##### 3.2.1.8.1. Sanger-Sequenzierung von DNA

Sequenzen der aus genomischer DNA amplifizierten und in *pGEM<sup>®</sup>-T Easy* klonierten HERV-W 5'*env*- und 3'*env*-Amplikons, der HERV-K(HML-2) *gag*-, *rec/np9*- und chr2q32.1- und 10q24.2-Amplikons sowie der aus cDNA amplifizierten HERV-K(HML-2) *gag*- und *rec/np9*-Amplikons wurden mittels Sanger-Sequenzierung generiert. Die Orientierung der in *pGLuc-Basic* klonierten

HERV-W und HERV-K(HML-2) LTR-Elemente sowie die Sequenzen der in *pEGFP-N1* klonierten ORFs der HML-2-Loci auf den Chromosomen 2q32.1 und 10q24.2 wurde ebenfalls mittels Sanger-Sequenzierung überprüft. DNA-Sequenzen wurden von der Seq-IT GmbH (Kaiserslautern, Deutschland) mit einem *3730 DNA-Analyzer* generiert. Für die Sequenzierungen wurden die für die Vektoren *pGEM<sup>®</sup>-T Easy*, *pGLuc-Basic* und *pEGFP-N1* spezifischen Sequenzier-Primer T7, pGLuc\_col\_for2 und pEGFP-N1\_seq (vgl. **Tab. 12**) verwendet.

Die Sequenzierung beruht auf der Didesoxymethode nach Sanger (Sanger *et al.*, 1977). Hierbei lagert sich in einer Sequenzierreaktion nach Denaturierung der DNA ein dem Bereich von Interesse benachbarter Sequenzier-Primer an, der von einer DNA-Polymerase verlängert wird. Der Ansatz enthält hierbei, neben Desoxynukleosidtriphosphaten, die vier Basen als Didesoxynukleosidtriphosphate (ddNTPs). Die vier ddNTPs tragen eine unterschiedliche Fluoreszenzmarkierung und werden zufällig in den neu synthetisierten Strang eingebaut, was aufgrund der fehlenden 3'-Hydroxylgruppe zum Abbruch der Reaktion führt. Es entstehen hierbei unterschiedlich lange Fragmente mit jeweils unterschiedlich markiertem ddNTP an ihrem 3'-Ende, die im Anschluss kapillarelektrophoretisch nach ihrer Länge aufgetrennt werden, wobei die Fluoreszenzsignale erfasst und die DNA-Sequenzen mit geeigneten Algorithmen hergeleitet werden.

#### **3.2.1.8.1. Next Generation Sequencing (NGS) von DNA**

Die aus cDNA amplifizierten HERV-W 5'*env*- und 3'*env*-Amplikons wurden mittels *454/FLX*-sowie *MiSeq/Illumina*-Technologie sequenziert.

##### **454/FLX-Sequenzierung**

Das *Genome Sequencer FLX<sup>TM</sup> System* ermöglicht mit einer hohen Genauigkeit die Hochdurchsatz Pyrosequenzierung bis 400 bp langer DNA-Sequenzen (Droege und Hill, 2008).

Die zu sequenzierenden Amplikons tragen zu beiden Seiten entsprechende *Adaptor*-Sequenzen, die während der initialen PCR-Amplifikation über entsprechende *Primer* generiert oder nachträglich ligiert werden können. Über diese Adaptoren erfolgt eine Bindung an spezifische *Capture Beads* aus Sepharose, die zu den Adaptoren komplementäre Oligonukleotide tragen, wobei durch die geeignete Wahl des Mischverhältnisses pro *Capture Bead* jeweils nur ein einziges DNA-Molekül bindet, das im Anschluss in einer *emulsion PCR* (emPCR) am *Bead* amplifiziert wird. Hierbei werden nach Zugabe der Amplifikationsreagenzien

DNA-Moleküle und Sepharose-*Beads* in einem Wasser-Öl-Gemisch emulgiert, so dass sich jeder *Bead* in seinem eigenen Mikroreaktor befindet. Während des ersten Denaturierungsschrittes der PCR bindet ein einzelsträngiges DNA-Molekül über seinen Adaptor an einen *Capture Bead* und wird im Anschluss amplifiziert, wobei jeder *Bead* am Ende der Reaktion ca. 10 Mio. identische DNA-Moleküle trägt. Nach Aufreinigung der *Beads*, die im Anschluss, bedingt durch das Aufreinigungsverfahren, erneut einzelsträngige DNA tragen, werden diese mit dem, die DNA-Polymerase enthaltenden, *DNA Bead Incubation Mix* vermischt und auf eine 7 cm x 7,5 cm große faseroptische *454 PicoTiterPlate™* geladen. Diese Platte enthält 1,6 Mio. *Wells*, in denen mit ihrem Durchmesser von 44 µm jeweils ein *Bead* Platz findet. Die *Capture Beads* werden in den *Wells* mit Enzym-*Beads* überschichtet, die die Enzyme Sulfurylase und Luciferase tragen. In der eigentlichen Sequenzierreaktion, die für alle *Beads*, die jeweils Mio. Kopien eines einzelsträngigen DNA-Moleküls tragen, parallel abläuft, wird von einem Sequenzier-*Primer* aus der jeweils komplementäre Strang der einzelsträngigen DNA-Moleküle synthetisiert. Hierbei werden in festgelegter Reihenfolge die Sequenzierreagenzien, die Puffer und je ein bestimmtes Nukleotid enthalten, über die *454 PicoTiterPlate™* gespült. Wird hierbei ein Nukleotid in ein *Well* gespült, in dem es komplementär zum wachsenden DNA-Strang ist, wird es von der Polymerase eingebaut. Bei dieser Reaktion wird ein inorganisches Pyrophosphat freigesetzt, das eine chemische Kaskade in Gang setzt, die ein Lichtsignal generiert, welches von einer CCD-Kamera erfasst wird, wobei die Stärke des Lichtsignals proportional zur Anzahl der im jeweiligen Nukleotid-Durchlauf eingebauten Nukleotide ist.

Zur initialen PCR-Amplifikation der hier sequenzierten *454/FLX*-Amplikons wurden *Primer* mit entsprechenden *Adaptor*- und *Key*-Sequenzen an ihrem 5'-Ende verwendet (vgl. **Tabelle 6**). Die Sequenzierung erfolgte in zwei separaten Sequenzier-Läufen, wobei Amplikons der Proben MS1 - MS3 und K1 - K3 im ersten und Amplikons der Proben MS4 - MS7 und K4 - K7 in einem zweiten Lauf sequenziert wurden. Die jeweils sechs bzw. acht Proben pro Lauf wurden in unterschiedliche *Lanes* einer *8-Lane PicoTiter*-Platte geladen und sequenziert. Die Sequenzierungen wurden von Frau Jasmin Gries im Institut für Genetik/Epigenetik der Universität des Saarlandes (Saarbrücken, Deutschland) durchgeführt.

### ***MiSeq/Illumina*-Sequenzierung**

Die *MiSeq/Illumina* NGS-Technologie ermöglicht die schnelle Sequenzierung von 2x 300 bp langen *paired-end* Reads und die Generierung großer Datenmengen von bis zu 15 GB innerhalb wenigen Stunden. Das *MiSeq*-System nutzt hierbei, wie alle *Illumina*-Sequenzierplattformen, die *Sequencing by Synthesis* (SBS) Technologie. Das Prinzip ist in **Abbildung 11** gezeigt.

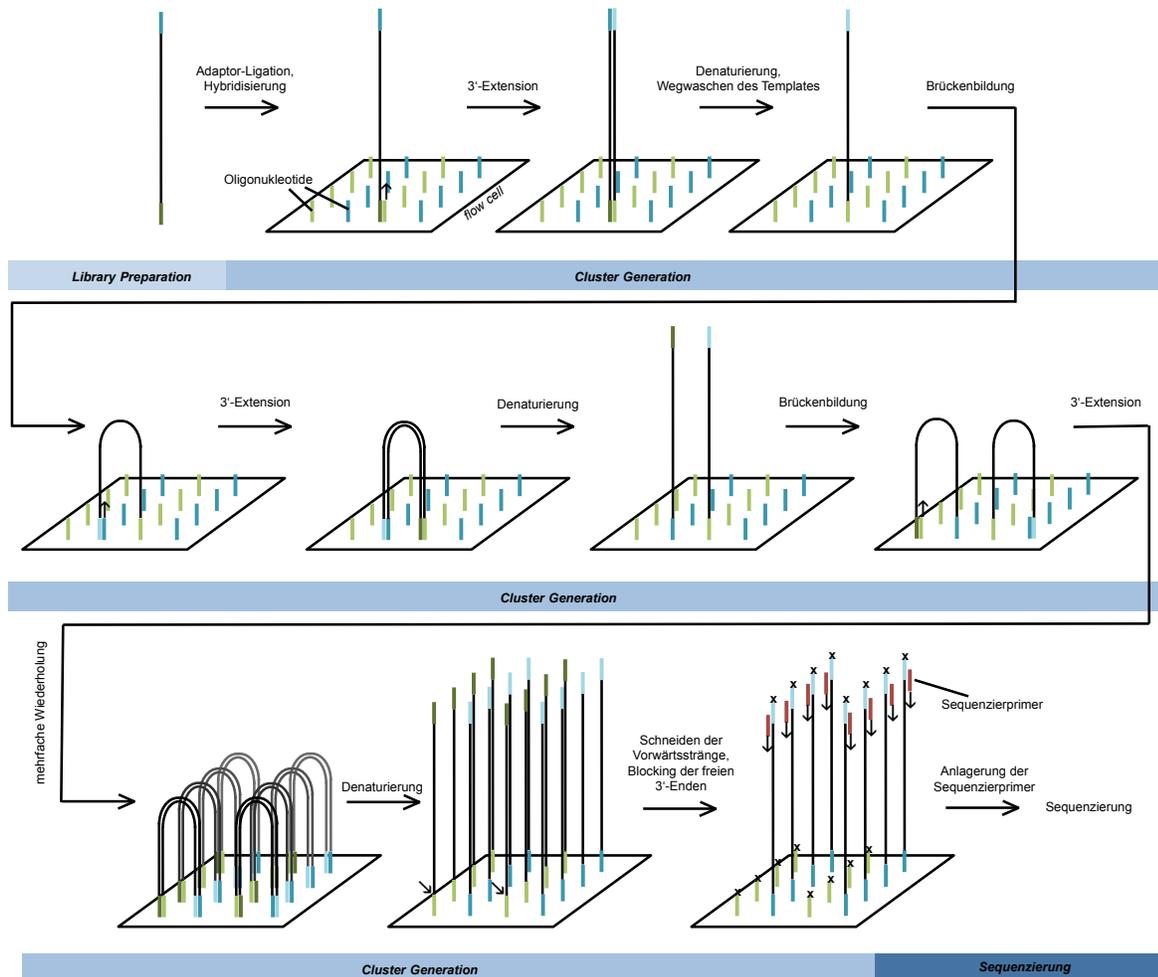


Abbildung 11: Prinzip der MiSeq/Illumina-Sequenzierung. Zur Erklärung, s. Text.

Im Zuge der *Library Preparation* werden die zu sequenzierenden DNA-Moleküle zunächst an beiden Enden mit Adaptoren versehen. Einzelsträngige DNA-Moleküle binden dann über einen dieser Adaptoren an komplementäre Oligonukleotide, die auf einem Glas-Objektträger (*Flow Cell*) immobilisiert sind und werden über die sog. *Bridge Amplification*-Technik amplifiziert. Ausgehend von den Oligos wird der komplementäre Strang des DNA-Moleküls synthetisiert, das doppelsträngige Molekül anschließend denaturiert und der ursprüngliche DNA-Strang gewaschen. Der neu synthetisierte Einzelstrang ist an einem Ende kovalent mit der Oberfläche der *Flow Cell* verbunden, klappt um und hybridisiert mit dem am anderen Ende des Moleküls befindlichen Adaptor an ein anderes, benachbartes auf der *Flow Cell* hybridisiertes komplementäres Oligonukleotid, von dem ausgehend erneut der komplementäre Strang synthetisiert wird. Hierbei entsteht eine doppelsträngige Brückenstruktur, die im Anschluss denaturiert wird, wobei zwei Kopien mit unterschiedlichem Ende kovalent gebundener einzelsträngiger *Templates* entstehen. Umklappen, Neusynthese des Doppelstranges und

Denaturierung wiederholen sich, bis die *Flow Cell* mit einzelsträngigen DNA-Molekülen bedeckt ist. Die Reversstränge werden anschließend oberhalb der *Flow Cell* Oligonukleotide geschnitten und hinterlassen einen *Cluster* an Vorwärtssträngen, deren freie 3'-Enden, ebenso wie die der *Flow Cell*-Oligos, geblockt werden, um unerwünschte Reaktionen zu verhindern. Die Sequenzier-*Primer* binden nun am 3'-Ende der Vorwärtsstränge und die Sequenzierung erfolgt nach einer auf einem reversiblen Terminator basierenden Methode. Hierbei werden alle vier dNTPs in einer Reaktion zugegeben, wobei diese mit unterschiedlich fluoreszenzmarkierten Terminatoren blockiert sind, die beim Einbau der dNTPs abgespalten und detektiert werden. Bei der *paired-end*-Sequenzierung wird der während der Sequenzierung neu synthetisierte Strang und die Blockierung der 3'-Enden entfernt. Der Einzelstrang klappt erneut um und hybridisiert mit einem benachbarten komplementären *Flow Cell*-Oligonukleotid. Von diesem aus wird der komplementäre Strang synthetisiert, der Doppelstrang denaturiert und dieses Mal der Vorwärtsstrang oberhalb des *Flow Cell*-Oligonukleotids geschnitten, wobei ein *Cluster* an Reverssträngen entsteht, die nach Blockierung der freien 3'-Enden analog der Vorwärtsstränge sequenziert werden.

Die *MiSeq/Illumina*-Sequenzierung wurde von der Seq-IT GmbH (Kaiserslautern, Deutschland) durchgeführt. Die *Adaptor*-Ligation erfolgte hierbei im Anschluss an die Amplifikation im Zuge der *Library Preparation* durch die Seq-IT GmbH.

### 3.2.1.9. Zellkultur

#### Kultivieren und Splitten eukaryotischer Zellen

Alle hier verwendeten adhärennten eukaryotischen Zelllinien wurden bei 37 °C und 5 % [v/v] CO<sub>2</sub> in 75 cm<sup>2</sup> Zellkulturflaschen oder 55 cm<sup>2</sup> Zellkulturschalen kultiviert und, je nach Wachstum und Zellkonfluenz, ein- bis dreimal wöchentlich gesplittet. Hierfür wurde das Zellkulturmedium abgenommen und die Zellen wurden mit 1,5 ml 1x DPBS gewaschen. Das 1x DBPS wurde abgenommen, 1,5 ml 0,05 % Trypsin-EDTA wurden zugegeben und so lange inkubiert, bis sich die Zellen vom Boden gelöst hatten. Die Zellen wurden in 5 - 10 ml frischem Medium mit 10 % [v/v] FCS aufgenommen, durch vorsichtiges Auf- und Abpipettieren vereinzelt und ein Teil der Suspension wurde in eine frische Zellkulturflasche bzw. -schale überführt, in die bereits frisches Medium vorgelegt worden war.

### Zellzahlbestimmung eukaryotischer Zellen

Die Bestimmung der Zellzahl eukaryotischer Zellen erfolgte mit Hilfe eines Hämocytometers (Neubauer Zählkammer). Die Zellen einer 75 cm<sup>2</sup> Zellkulturflasche wurden zunächst, wie im voranstehenden Abschnitt beschrieben, gewaschen, trypsiniert und in Medium aufgenommen. 10 µl der Suspension wurden seitlich unter das Deckglas in den 0,1 mm hohen Zwischenraum der Kammer pipettiert, von wo sie durch Kapillarkräfte passiv in die Kammer gesaugt wurden. Mit einem Auflichtmikroskop wurden im Anschluss die Zellen in vier Großquadranten (1 mm x 1 mm), über denen sich jeweils ein Volumen von 0,1 µl befindet, einzeln ausgezählt und die Werte gemittelt. Mit dem Mittelwert  $\bar{x}$  wurde die Gesamtzahl der Zellen im Anschluss nach folgender Formel ermittelt, wobei im Falle einer vorherigen Verdünnung der entsprechende Verdünnungsfaktor mit eingerechnet wurde:

$$\frac{\bar{x}}{10^{-4} \text{ ml}} = \bar{x} * 10^4 \frac{\text{Zellen}}{\text{ml Zellsuspension}}$$

### UV-Bestrahlung kultivierter Melanom- und Melanozyten-Zelllinien

Melanom- und Melanozyten-Zelllinien wurden in 55 cm<sup>2</sup> Zellkulturschalen bis zu einer Konfluenz von ~70 - 80 % kultiviert. Das Zellkulturmedium wurde abgenommen und die Zellen wurden mit 1x DPBS gewaschen, wobei ein Rest 1x DPBS in der Kulturschale über den Zellen belassen wurde, um das Austrocknen der Zellen während der Bestrahlung zu verhindern. Die Zellen wurden unter Verwendung einer *UV409T* UV-Lampe und eines *Vario Control* UV-Meters mit 200 mJ/cm<sup>2</sup> UVB bestrahlt, im Anschluss in frischem Medium bei 37 °C und 5% CO<sub>2</sub> weiter kultiviert und nach 24 h geerntet, um die RNA zu isolieren. Die Bestrahlung der Zellen wurde in der Klinik für Dermatologie, Venerologie und Allergologie des Universitätsklinikums des Saarlandes (Homburg, Deutschland) durchgeführt.

#### 3.2.1.10. Transfektion eukaryotischer Zellen

##### Transfektion von HERV-K(HML-2) ORF-Vektorkonstrukten in MEWO-Zellen

Zur Detektion der Expression möglicher GFP-Fusionsproteine mittels Immunfluoreszenz und *Western Blot* wurden MEWO-Zellen mit *pEGFP-N1*-Vektorkonstrukten transfiziert, die stromauf des GFP-Gens einen der ORFs der beiden durch reverse Transkription und Reintegration gespleißter mRNA entstandenen HERV-K(HML-2)-Loci auf den Chromosomen 2q32.1 und 10q24.2 enthielten.

MEWO-Zellen wurden in 6-Well Platten zu je  $2 \times 10^5$  Zellen pro Well ausgesät und bei 37 °C und 5% [v/v] CO<sub>2</sub> kultiviert. Wells, in die die Zellen ausgesät wurden, die später zur Fluoreszenzmikroskopie verwendet wurden, enthielten je ein 18 mm x 18 mm großes Deckglas, auf dem die Zellen anwuchsen. Die Zellen wurden nach 24 h pro Well mittels 6 µl TurboFect™ Transfection Reagent nach Herstellerprotokoll mit je 3 µg des jeweiligen pEGFP-N1-ORF-Konstruktes in 100 µl Opti-MEM® I (1x) + GlutaMAX™-I Medium transfiziert und etwa 48 h bei 37 °C und 5 % [v/v] CO<sub>2</sub> kultiviert. Die auf Deckgläsern ausgesäten Zellen wurden anschließend für die Immunfluoreszenz fixiert und aus den Zellen der restlichen Wells wurde Gesamtprotein isoliert. Zur späteren Proteinisolation wurden je sechs Wells mit dem gleichen Konstrukt transfiziert, für die Fluoreszenzmikroskopie je drei Wells.

### Transfektion von HERV-W LTR-Vektorkonstrukten in JEG-3 Zellen

Zur Ermittlung von möglichen Promotoraktivitäten ausgesuchter HERV-W 5'- und 3'LTR-Elemente, wurden diese aus genomischer DNA amplifiziert und die PCR-Produkte in Sense- und/oder Antisense-Orientierung in Promotorposition in den Luciferase-exprimierenden pGLuc-Basic-Vektor kloniert und die Luciferaseaktivität in transfizierten JEG-3-Zellen ermittelt. Zur Normalisierung der erhaltenen Werte wurde der pCMVβ-Vektor cotransfiziert und die Aktivität der von diesem Vektor exprimierten β-Galactosidase bestimmt. Die getesteten pGLuc LTR-Konstrukte wurden im Rahmen von Richter (2012) generiert und sind in **Tabelle 3** zusammengefasst. Es wurden jeweils die LTRs, die im Falle einiger Loci jedoch stark verkürzt sind, mit flankierender Sequenz kloniert (vgl. **Anhang 17** für die Sequenzen der LTRs und der zur Amplifikation verwendeten Primer).

JEG-3 Zellen wurden in 12-Well Platten zu je  $6 \times 10^4$  Zellen pro Well ausgesät und bei 37 °C und 5 % [v/v] CO<sub>2</sub> kultiviert. Die Zellen wurden nach 24 h mittels TurboFect™ Transfection Reagent nach Herstellerprotokoll mit je 1 µg des pGLuc-LTR-Konstruktes und 0,2 µg des pCMVβ-Vektors in 100 µl Opti-MEM® I (1x) + GlutaMAX™-I Medium cotransfiziert und etwa 48 h bei 37 °C und 5 % [v/v] CO<sub>2</sub> kultiviert. Als Kontrolle wurden Zellen mit dem promotorlosen pGLuc-Basic-Vektor bzw. dem pCMV-GLuc-Vektor, der die Gaussia Luciferase unter Kontrolle des konstitutiven CMV-Promotors exprimiert, transfiziert. Es wurden je drei Wells mit dem gleichen Konstrukt transfiziert und das Experiment wurde zweimal wiederholt. Die Transfektionen der JEG-3 Zellen wurden von Herrn apl.-Prof Dr. rer. nat. Jens Mayer durchgeführt.

### 3.2.1.11. Reporteragen-Assays

Die Luciferase- und  $\beta$ -Galactosidase-Aktivitäten in Lysaten von JEG-3-Zellen wurden 48 h nach Transfektion mit den HERV-W LTR-Konstrukten bestimmt. Das Zellkulturmedium wurde hierfür abgenommen und die Zellen wurden mit je 1 ml 1x DPBS gewaschen. Die Zellen wurden durch Zugabe von 300  $\mu$ l 1x *Luciferase Cell Lysis Buffer* pro Well und 15 min Inkubation unter gelegentlichem leichtem Klopfen lysiert. Lysate wurden in 1,5 ml Reaktionsgefäße überführt, 15 s gevortext und Zellreste durch 1-minütige Zentrifugation bei 16.000 g pelletiert. Der Überstand wurde für die Luciferase- und  $\beta$ -Galactosidase-Assays eingesetzt.

#### Luciferase-Assay

Die Luciferaseaktivität in Lysaten von JEG-3-Zellen wurde mit dem *BioLux<sup>®</sup> Gaussia Luciferase Assay Kit* bestimmt. Das dem Puffer zugegebene Substrat Coelenterazin wird hierbei von der Luciferase zu Coelenteramide oxidiert, wobei Licht der Wellenlänge 475 nm emittiert wird, welches mittels Luminometer quantifiziert wird.

Es wurden je 5  $\mu$ l des Zellysates mit 30  $\mu$ l *BioLux GLuc Assay Buffer* und 0,3  $\mu$ l *BioLux GLuc Substrat* gemischt und die relative Luciferaseaktivität in einem *Lumat LB 9507* Luminometer gemessen. Die Luminometer-Messungen wurden gemeinsam mit Herrn apl.-Prof. Dr. rer. nat. Jens Mayer durchgeführt

#### $\beta$ -Galactosidase-Assay

Die  $\beta$ -Galactosidase-Aktivität in JEG-3-Zellysaten wurde mit dem  *$\beta$ -Galactosidase Enzyme Assay System with Reporter Lysis Buffer* bestimmt. Das im Reaktionspuffer enthaltene farblose Substrat *o*-Nitrophenyl- $\beta$ -D-galactopyranosid wird hierbei von  $\beta$ -Galactosidase zum gelben *o*-Nitrophenol hydrolysiert, dessen Absorption bei 420 nm spektralphotometrisch bestimmt wird.

Es wurden je 100  $\mu$ l Zellysate mit 100  $\mu$ l *Assay 2x Buffer* versetzt und für 30 min bei 37 °C inkubiert. Die Reaktion wurde durch Zugabe von 333  $\mu$ l Natriumcarbonat gestoppt und die Absorption bei 420 nm in einem *NanoDrop<sup>TM</sup> 2000* Spektralphotometer bestimmt. Parallel wurden, wie im Herstellerprotokoll beschrieben, sieben Standards mit bekannten  $\beta$ -Galactosidase-Konzentrationen von 0 bis 6 mU in je 100  $\mu$ l Gesamtvolumen in die Reaktion mit *Assay 2x Buffer* eingesetzt. Anhand der erhaltenen Absorptionen wurde eine Standardgerade generiert und die  $\beta$ -Galactosidase Konzentrationen in mU der JEG-3 Zellysate

berechnet, indem die Absorption einer Probe durch den Wert der Steigung der Geraden dividiert wurde.

Zur Bestimmung der relativen Luciferaseaktivität wurde die Luciferaseaktivität durch die  $\beta$ -Galactosidase-Menge [mU] dividiert.

### **3.2.2. Proteinbiochemische Methoden**

#### **3.2.2.1. Proteinisolation und -konzentrationsbestimmung**

##### **Ernte von eukaryotischen Zellen**

Das Medium von in 6-Well Platten ausgesäten Zellen wurde 24 h nach Transfektion abgenommen und die Zellen wurden mit 1,5 ml 1x DPBS pro Well gewaschen. Das 1x DPBS wurde abgenommen, die Zellen wurden durch Zugabe von 1,5 ml 0,05 % Trypsin-EDTA pro Well vom Boden gelöst und die mit dem gleichen Konstrukt transfizierten Zellen aus sechs Wells zusammen in insgesamt 6 ml 1x DPBS aufgenommen. Die Zellen wurden für 5 min bei 300 g pelletiert, der Überstand wurde abgenommen und das Zellpellet bis zur Weiterverwendung bei -70°C gelagert.

##### **Herstellung von Lysaten eukaryotischer Zellen**

Pellets von transfizierten Zellen aus sechs Wells einer 6-Well Schale wurden in 300  $\mu$ l 1x RIPA Puffer resuspendiert und 20 min auf Eis inkubiert. Zelldebris wurde im Anschluss für 10 min bei 8.000 g abzentrifugiert, der klare Überstand in ein frisches Reaktionsgefäß überführt und bei -20 °C gelagert.

##### **Spektralphotometrische Protein-Konzentrationsbestimmung**

Die Protein-Konzentration in Zelllysaten transfizierter Zellen wurde mit einem *NanoDrop*<sup>TM</sup> 2000 Spektralphotometer über die Messung der optischen Dichte bei 280 nm ( $OD_{280}$ ) bestimmt.

### 3.2.2.2. SDS-Polyacrylamid-Gelelektrophorese (SDS-PAGE)

Die Auftrennung von Proteinen erfolgte mittels SDS-Polyacrylamid-Gelelektrophorese (SDS-PAGE). Proteine werden hierbei zunächst hitzedenaturiert. Das zugegebene Detergenz SDS bindet an die Proteine, überdeckt deren Eigenladung und verleiht ihnen eine negative Gesamtladung, so dass sie im elektrischen Feld zur Anode wandern und dabei ihrer Molekülmasse nach in einem Polyacrylamidgel aufgetrennt werden. Die Proteine wandern hierbei zunächst durch ein Sammelgel mit neutralem pH-Wert, in dem sie aufkonzentriert werden, was später in einer besser definierten Bandenschärfe resultiert. Im basischen Trenngel erfolgt dann die Auftrennung nach Molekülmasse. Das Trenngel besteht hierbei aus Acrylamid und dem quervernetzenden N,N'-Methylenbisacrylamid, die nach Radikalisierung durch Ammoniumperoxidsulfat (APS) und Katalyse von Tetramethylethyldiamin (TEMED) polymerisieren und je nach Konzentration ein Gel bestimmter Porengröße bilden.

In der vorliegenden Arbeit wurden Polyacrylamid-Gele mit 4 %-igem Sammelgel und 12 %-igem Trenngel verwendet, die nach folgendem Ansatz hergestellt wurden.

	Trenngel	Sammelgel
Tris-HCl, pH 6,8	-	125 mM
Tris-HCl, pH 8,8	375 mM	-
Acrylamid/Bisacrylamid (37,5:1)	12 % [v/v]	4 % [v/v]
SDS	0,1 % [w/v]	0,1 % [w/v]
APS	0,1 % [w/v]	0,1 % [w/v]
TEMED	0,1 % [v/v]	0,25 % [v/v]
H <sub>2</sub> O <sub>dd</sub>	ad 5 ml	ad 2 ml

Der Ansatz für das Trenngel wurde in eine *XCell SureLock™* Mini-Kassette gegossen und mit Isopropanol überschichtet. Nach Polymerisierung wurde das Isopropanol abgegossen und über dem Trenngel das Sammelgel gegossen, in welches im noch nicht polymerisierten Zustand ein Kamm gesteckt wurde. Die Gelkassetten wurden feucht bei 4 °C gelagert und zur SDS-PAGE in eine mit Laufpuffer gefüllte *XCell Sure Lock™* Elektrophoresekammer eingebaut.

Proteinproben (je 2 µg) wurden mit dem reduzierenden, β-Mercaptoethanol enthaltenden, Probenpuffer *Roti®-Load* in einfacher Endkonzentration gemischt, durch 5-minütige Inkubation bei 95 °C denaturiert und im Anschluss in eine Tasche im Gel pipettiert. Als Größenstandard dienten 7 µl des *Precision Plus Protein™ Standard Dual Color*. Die Auftrennung erfolgte für etwa 2 h bei einer Spannung von konstant 125 V.

### 3.2.2.3. Western Blot

Mittels SDS-PAGE aufgetrennte Proteine wurden durch ein senkrecht zum Gel gerichtetes elektrisches Feld aus dem Gel auf eine Polyvinylidenfluorid (PVDF)-Membran übertragen. Die durch das angelagerte SDS negativ geladenen Proteine wandern hierbei Richtung Anode und werden auf der PVDF-Membran aufgrund hydrophober und polarer Wechselwirkungen festgehalten, wo die Proteine später mittels spezifischer Antikörper detektiert werden können.

Das Polyacrylamidgel wurde auf eine Schicht aus zwei bis drei Schwämmen und zwei *Whatman*<sup>TM</sup> Papieren in eine *XCell II* Blotting Kammer gelegt. Auf dem Gel wurde eine zuvor in Methanol aktivierte PVDF-Membran platziert, gefolgt von zwei weiteren *Whatman*<sup>TM</sup> Papieren und zwei bis drei Schwämmen. Schwämme und *Whatman*<sup>TM</sup> Papiere wurden zuvor in Transferpuffer getränkt. Der Aufbau erfolgte so, dass nach Einbau in die Kammer das Gel in Richtung Kathode und die PVDF-Membran in Richtung Anode zeigte. Die innere *Blotting*-Kammer wurde mit Transferpuffer gefüllt und die äußere Kammer mit Wasser. Der anschließende Transfer erfolgte für 3 h bei konstant 25 V.

### 3.2.2.4. Immunodetektion von auf eine PVDF-Membran übertragenen Proteinen

Die Immunodetektion von Proteinen nach Transfer auf eine PVDF-Membran fand durch einen GFP-spezifischen Primärantikörper sowie einen gegen diesen gerichteten Meerrettich-Peroxidase (*horseradish peroxidase*, HRP)-gekoppelten Sekundärantikörper statt. Zur Detektion der Proteine mittels Chemilumineszenz diente das *Amersham ECL Prime Western Blotting Detection Reagent*. Die Peroxidase katalysiert in einem mehrstufigen Prozess die Oxidation des darin enthaltenen Luminols zu Luminolendoperoxid, das bei Zerfall Licht emittiert. Die Lichtemission ist hierbei proportional zur Proteinmenge und kann mit einem für Chemilumineszenz optimierten Film detektiert werden.

Nach Transfer der Proteine auf die PVDF-Membran wurde die Membran in ein 50 ml Falcon-Gefäß überführt und nicht besetzte Bindestellen auf der Membran durch Zugabe von 10 ml *Blocking*-Lösung und Inkubation ü. N. oder für 1 h bei 4 °C geblockt. Alle Inkubations- und Waschschrte fanden auf einem Rollinkubator statt. Die verwendeten *Blocking*-, Wasch- und Antikörperlösungen sind in **Tabelle 14** zusammengefasst. Die *Blocking*-Lösung wurde abgegossen und die Membran, abhängig von den verwendeten Antikörpern, 3 x 10 min mit je 10 ml 1x TBS-T oder 10 ml 1x PBS-T gewaschen. Es folgte eine einstündige Inkubation mit der

Primärantikörper-Lösung. Zum Entfernen nicht gebundener Primärantikörper wurde die Membran erneut 3 x 10 min gewaschen. Anschließend wurde die Membran für 1 h mit dem Peroxidase-gekoppelten Sekundärantikörper inkubiert. Nicht gebundene Sekundärantikörper wurden durch 3 x 10 min Waschen entfernt.

**Tabelle 14:** Zur Immunodetektion von Proteinen auf PVDF-Membranen verwendete Primär- und Sekundärantikörper sowie verwendete *Blocking*- und Waschlösungen.

Protein	Blocken	Waschen	Primärantikörper	Sekundärantikörper
GFP	ü. N., 5 % Milchpulver in 1x TBS-T	1x TBS-T	$\alpha$ -GFP, 1:1.000 verdünnt in 5 % Milchpulver in 1x TBS-T	<i>HRP-conjugated Goat-<math>\alpha</math>-Rabbit</i> IgG, 1:10.000 verdünnt in 5 % Milchpulver in 1x TBS-T
$\beta$ -Actin	1 h, 5 % Milchpulver in 1x PBS-T	1x PBS-T	$\alpha$ - $\beta$ -Actin, 1:1.000 verdünnt in 5 % Milchpulver in 1x PBS-T	<i>HRP-conjugated Sheep-<math>\alpha</math>-Mouse</i> IgG, 1:50.000 verdünnt in 5 % Milchpulver in 1x PBS-T

Zur Detektion der Proteine wurden die Luminol- und Peroxid-Lösungen des *Amersham ECL Prime Western Blotting Detection Reagent Kit* zu gleichen Anteilen gemischt, über die PVDF-Membran verteilt und für 5 min inkubiert. Es wurde für 15 s bis 2,5 min ein *Amersham Hyperfilm™ ECL* Chemilumineszenz-Film auf die Membran aufgelegt, der anschließend bis zur gewünschten Bandenintensität in einer Entwicklerlösung inkubiert wurde. Die Reaktion wurde durch Überführen in eine Fixierlösung gestoppt und der Film in Wasser gewaschen und anschließend luftgetrocknet.

Zum Nachweis unterschiedlicher Proteine auf derselben Membran wurden gebundene Primär- und Sekundärantikörper durch 2 x 10 min Waschen mit 1x *Stripping Buffer* von der Membran gelöst. Der *Stripping Buffer* wurde abgegossen und die Membran, abhängig von den im Anschluss verwendeten Antikörpern (vgl. **Tab. 14**), 2 x 10 min mit 1x TBS-T bzw. 1x PBS-T gewaschen, bevor sie erneut, wie zuvor beschrieben, geblockt und mit den spezifischen Antikörperlösungen inkubiert wurde.

### 3.2.2.5. Fluoreszenzmikroskopie

Vor fluoreszenzmikroskopischen Auswertungen wurden transfizierte Zellen mit Paraformaldehyd fixiert, die Zell- und Kernmembran mittels *Triton X-100* permeabilisiert und die DNA in den Zellkernen mit *Hoechst Stain Solution* angefärbt. Die *Hoechst*-Färbung kann bei Anregung mit Licht der Wellenlänge 350 nm detektiert werden. Zur Verhinderung eines schnellen Ausbleichens der Fluoreszenz wurden die Präparate mit *ProLong® Gold Antifade* behandelt, das zudem eine längere Lagerung erlaubt.

24 h nach Transfektion mit den HML-2 ORF-Konstrukten wurde das Medium der auf Deckgläser in 6-Well Platten ausgesäten MEWO-Zellen abgenommen und die Zellen wurden durch Zugabe von 1,5 ml 1x DPBS pro Well gewaschen. Die Zellen wurden durch Zugabe von 2 ml 4 % [v/v] Paraformaldehyd in 1x DPBS pro Well und 20 min Inkubation bei 4 °C fixiert und 3 x 5 min mit je 0,5 ml 1x DPBS gewaschen. Die Deckgläser wurden mit der Zellseite nach oben auf ein Stück Parafilm überführt, mit je 200 µl 0,1 % [v/v] Triton X-100 in 1x DPBS überschichtet und 2 min bei Raumtemperatur inkubiert. Das Triton X-100 wurde abgenommen und die Zellen wurden 3 x 5 min mit 200 µl 1x DPBS gewaschen. Zur Färbung der DNA wurde jedes Deckglas mit 200 µl Hoechst Staining Solution 33258 überschichtet, für 10 min bei Raumtemperatur inkubiert, die Färbelösung im Anschluss abgenommen und die Zellen erneut 3 x 5 min mit 200 µl 1x DPBS gewaschen. Im Anschluss wurden 20 µl ProLong® Gold Antifade auf einen Objektträger gegeben, das Deckglas mit der Zellseite nach unten darauf gelegt und die Ecken des Deckglases mit klarem Nagellack fixiert.

Zur fluoreszenzmikroskopischen Auswertung diente ein AX70 Mikroskop und die CCD-Kamera DP71. Fluoreszenzmikroskopische Bilder wurde mit der Software cell<sup>^</sup>F aufgenommen und bearbeitet.

### 3.2.3. Bioinformatische Arbeiten

#### 3.2.3.1. Sequenzzusammenstellung und Primer-Optimierung

Für HERV-K(HML-2) und HERV-W wurde jeweils eine Übersicht der genomischen Loci zusammengestellt. Hierfür wurde im UCSC Genome Browser (Kent *et al.*, 2002) eine BLAT-Suche (Kent 2002) mit der bei Repbase (Jurka *et al.*, 2005) hinterlegten HERVK- und HERV17-Konsensussequenz im menschlichen NCBI36/hg18 Referenzgenom durchgeführt. Im Falle von HERV17 (entspricht HERV-W) wurde lediglich der *env*-ORF der Konsensussequenz als Suchsequenz in der BLAT-Suche verwendet, da vielen HERV-W-Loci, die durch L1-vermittelte Retrotransposition entstanden sind, unterschiedlich große Abschnitte im 5'-Bereich der Proviren fehlen. Für die Chromosomenpositionen der gefundenen HML-2- und HERV-W-Loci wurden mit dem UCSC Table Browser die entsprechenden DNA-Sequenzen zusammengestellt und mittels MAFFT (Kato *et al.*, 2002) ein multiples Sequenz-Alignment aller Loci generiert. Multiple Alignments wurden in Geneious importiert und geeignete Amplikonbereiche ausgewählt, die in möglichst vielen Loci vorkommen, jedoch noch genügend Sequenzunterschiede aufweisen, um eine eindeutige Zuordnung der generierten Sequenzen zu spezifischen Loci zu gewährleisten.

Im Falle von HERV-K(HML-2) wurde ein zentraler Bereich von etwa 620 bp im *gag*-ORF gewählt, der nt 1.778 - 2.396 in der HERV-K(HML-2.HOM)-Sequenz (*Genbank acc. no.* AF074086.2) entspricht sowie ein weiterer ~370 bp (*np9*) bzw. ~580 bp (*rec*) großer Bereich in der *np9/rec* Region, der nt 6.451 - 8.728 der *Rec*-kodierenden HERV-K(HML-2.HOM)-Sequenz entspricht. Für HERV-W wurden zwei Bereiche von ~280 bp und ~330 bp im 5'- und 3'-Bereich des *env*-ORFs gewählt, 5'*env* und 3'*env*, die den Bereichen von nt 7.548/7.568 - 7.821 bzw. nt 8.349 - 8.587 der HERV17 *Repb*ase-Konsensussequenz entsprechen.

Um die Zahl der in einer PCR grundsätzlich amplifizierbaren Loci zu erhöhen, wurden für jedes Amplikon mehrere an der gleichen Stelle bindende *Forward*- und *Reverse-Primer*-Varianten gewählt, die Nukleotidunterschiede in den *Primer*-Binderegionen zwischen den verschiedenen Loci berücksichtigen. Das Mischverhältnis der *Primer*-Varianten in späteren PCRs entsprach dabei in etwa der Anzahl an potentiellen Zielsequenzen der einzelnen Varianten.

### **3.2.3.2. Sequenzrückkartierung und Auszählung generierter HERV cDNA-Sequenzen**

#### **3.2.3.2.1. Sequenzrückkartierung nach Sanger-Sequenzierung**

Die Qualität von Sanger-Sequenzen wurde in *FinchTV* überprüft und Sequenzen mit Chromatogrammen schlechter Qualität verworfen. Sequenzen mit ausreichender Qualität wurden mit *MAFFT* (Kato et al., 2002) aliniert und in *Geneious* importiert, wo Vektor- sowie *PCR-Primer*-Sequenzen identifiziert und manuell entfernt wurden. Die bearbeiteten Sequenzen wurden im *UCSC Genome Browser* (Kent et al., 2002) mittels *BLAT* (Kent 2002) im menschlichen NCBI36/hg18 Referenzgenom rückkartiert. Es ist zu beachten, dass das polymorphe HERV-K113-Provirus (Chromosom 19p12; *Genbank acc. no.* AY037928.1) nicht im hier verwendeten *Genome Browser Assembly NCBI36/hg18* enthalten ist. Es hätte somit, sofern transkribiert, mittels *BLAT*-Rückkartierung nicht identifiziert werden können. Da die hier untersuchte *gag*-Amplikonregion in HERV-K113 mit Ausnahme eines Nukleotidunterschiedes der eines HERV-K(HML-2)-Locus auf Chromosom 1q22 entspricht, wurde die entsprechende Nukleotidposition in allen Sequenzen überprüft, die 1q22 zugeordnet wurden, um auszuschließen, dass diese Sequenzen eigentlich vom K113-Provirus stammen.

### 3.2.3.2.2. Rückkartierung von Sequenzen aus NGS-Experimenten

#### Rückkartierung von 454/FLX-Sequenzen

Die Filterung der durch 454/FLX-Sequenzierung generierten Sequenzen aus dem gesamten Datensatz des 454/FLX-Laufes wurde am Institut für Genetik/Epigenetik der Universität des Saarlandes (Saarbrücken, Deutschland) von Herrn M.Sc. Pavlo Lutsik durchgeführt. Aufgrund der Sequenzierung jeder Probe in unterschiedlichen *Lanes* der *PicoTiter*-Platte waren die Sequenzen bereits automatisch nach Proben sortiert und die kurzen *Key*-Sequenzen an deren 5'-Ende entfernt. Die Zuordnung zu den Amplikons 5'*env* und 3'*env* erfolgte anhand der *Primer*-Sequenzen. Hierbei wurden die ersten neun Nukleotide aller *Forward-Primer* als Erkennungssequenz definiert und die Sequenzen entsprechend dieser Nukleotide den Amplikons 5'*env* und 3'*env* zugeordnet, wobei bis zu vier Nukleotidunterschiede toleriert wurden.

Für jede Probe und jedes Amplikon wurde im Anschluss mittels *MAFFT* (Kato et al., 2002) ein eigenes multiples Sequenz-*Alignment* generiert. Die multiplen *Alignments* wurden in *Geneious* importiert und schlechte und zu kurze Sequenzen sowie *Primer*-Dimere manuell gelöscht. In den verbleibenden Sequenzen wurden die *PCR-Primer*-Bereiche identifiziert und die entsprechenden Sequenzabschnitte gelöscht. Die bearbeiteten Sequenzen wurden im Anschluss in einer lokalen *BLAT*-Suche von Frau Dr. rer. nat. Christina Backes im menschlichen Referenzgenom (NCBI36/hg18) rückkartiert und die besten *BLAT-Matches* mit dem beim *UCSC Genome Browser* zur Verfügung gestellten *pslReps*-Skript gefiltert.

#### *MiSeq/Illumina*-Auswertung

Die Auswertung der mittels *MiSeq/Illumina* generierten Sequenzen erfolgte mit diversen Werkzeugen des *Galaxy*-Servers (Giardine et al., 2005; Blankenberg et al., 2010; Goecks et al., 2010), die im Folgenden jeweils in Klammern genannt sind. Die Sequenzen wurden zunächst in das *FASTQ*-Format umgewandelt (*NGS: QC and manipulation, FASTQ Groomer*) und die 19 bis 23 nt langen *PCR-Primer*-Sequenzen entfernt (*NGS: QC and Manipulation, Trim Sequences*). Eine genaue Suche nach *Primer*-Sequenzen war hierbei nicht möglich, weswegen stattdessen von allen Sequenzen die ersten und letzten 25 Basen entfernt wurden. Nach Umwandlung der Sequenzen in das *FASTA*-Format (*NGS: QC and Manipulation, FASTQ to FASTA converter*) wurden die Sequenzen in einer lokalen *BLAT*-Suche im menschlichen Referenzgenom (NCBI36/hg18) rückkartiert. Die *BLAT*-Rückkartierung wurde von Dr. rer. nat. Christina Backes durchgeführt. Zur Ermittlung des besten *BLAT-Matches* wurden die erhaltenen *psl*-Dateien

erneut in *Galaxy* geladen und zunächst alle kurzen Sequenzen mit einem *Match* <185 nt gelöscht (*Filter and Sort, Filter data*). Nach Sortieren der *BLAT*-Ergebnisse nach Name und *Match*-Länge (*Filter and Sort, Sort data*) wurden die *BLAT*-Ergebnisse in Einzeldateien zu je 1 Mio. Zeilen gesplittet (*Text Manipulation, Select first lines* und *Remove beginning*), heruntergeladen und in *Excel* weiterbearbeitet. Alle Ergebnisse mit mehr als einem besten *Match* sowie mehr als drei Nukleotidunterschieden wurden gelöscht und die verschiedenen Loci im Anschluss in *Galaxy* (*Join, Subtract and Group, Group*) und *Excel* ausgezählt. Die Zuordnung der erhaltenen Loci zu den Amplikons 5'*env* und 3'*env* erfolgte anhand der Chromosomenkoordinaten. Für die Chromosomenkoordinaten der *BLAT*-Ergebnisse wurden mittels *UCSC Table Browser* (Karolchik *et al.*, 2004) die *Repeatmasker*-Annotationen der entsprechenden Chromosomenbereiche heruntergeladen (*Track: RepMask 3.2.7*). Die erhaltenen Tabellen enthalten hierbei Informationen zur Position des Amplikons in der HERV17int Konsensussequenz (*repStart, repEnd*), wobei 5'*env* etwa den Bereich von nt 7548 - 7821 und 3'*env* den Bereich von nt 8349 - 8587 umfasst. Sequenzen, die in anderen Bereichen der Konsensussequenz lokalisiert waren oder nicht als HERV17int annotiert waren, wurden verworfen.

### 3.2.3.2.3. Locus-Auszählung und Berechnung der relativen Klonierungsfrequenzen

Für alle Analysen wurden nur eindeutig rückkartierbare Sequenzen verwendet. Als eindeutig rückkartierbar wurden Sequenzen mit nur einem besten Treffer in der *BLAT*-Suche (in der weiteren Arbeit als *BLAT-Match* bezeichnet) definiert, der weniger als drei (HERV-W 5'*env* und 3'*env* und HERV-K(HML-2) *reclnp9*) bzw. sechs (HERV-K(HML-2) *gag*) Nukleotidunterschiede zur entsprechenden Genomsequenz aufwies. Hierdurch wurde etwa 1 % Sequenzunterschied aufgrund von Fehlern während der RT-PCR und Sequenzierung sowie interindividueller Unterschiede toleriert. Die Auszählung der verschiedenen Loci erfolgte für jede untersuchte Probe manuell in *Excel*. Die relativen (Klonierungs-)Frequenzen jedes Locus in jeder Probe in [%] berechneten sich aus dem Quotienten aus der Anzahl an Sequenzen, die eindeutig dem jeweiligen Locus zugeordnet werden konnten und der Gesamtanzahl an eindeutig rückkartierbaren Sequenzen der entsprechenden Probe.

### 3.2.4. Statistik

#### Mittelwert und Standardabweichung

Arithmetische Mittelwerte  $\bar{x}$  und Standardabweichungen  $\sigma$  wurden in *Excel* nach folgenden Formeln berechnet ( $x$  = Messwert,  $n$  = Probenanzahl).

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \qquad \sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{(n - 1)}}$$

#### Wilcoxon-Mann-Whitney-Test

Zum Vergleich der relative Frequenzen der als transkribiert identifizierten HERV-W-Loci in MS und gesunden Kontrollen, sowie der mittels *Real-Time-PCR* bestimmten relativen Transkriptmengen zwischen MS und gesunden Kontrollen wurde ein Wilcoxon-Mann-Whitney-Test durchgeführt. Hierbei wird überprüft, ob zwei unabhängige Verteilungen, hier die Frequenzen bzw. Transkriptmengen in MS und gesunden Kontrollen, zur selben Grundgesamtheit gehören. Es wurden jeweils die Frequenzen bzw. Transkriptmengen zweier Stichproben (MS vs. gesund) verglichen und geprüft, ob diese gleich sind (Nullhypothese). Im Falle eines signifikanten Unterschiedes ergibt sich ein p-Wert  $< 0,05$  ( $< 5\%$ ). Die Berechnungen wurden von Frau Dr. rer. nat. Christina Backes in der *Software-Umgebung R* (*wilcox.test*) durchgeführt.

## 4. Ergebnisse

Einen wichtigen Bestandteil des menschlichen Transkriptoms stellen Transkripte humaner endogener Retroviren (HERVs) dar (Frank *et al.*, 2005, 2008; Seifarth *et al.*, 2005). Als Beitrag zum besseren Verständnis der spezifischen Rolle von HERVs in verschiedensten Geweben ist die Identifizierung transkribierter HERV-Loci von besonderem Interesse. Eine umfassende Analyse der Transkription zahlreicher HERV-Loci sowie potentieller Unterschiede zwischen Geweben sind hierbei ebenso von Interesse, wie eine detaillierte strukturelle Analyse transkribierter HERV-Loci hinsichtlich ihrer Kodierungskapazität für retrovirale Proteine sowie des Vorhandenseins von Promotor- und regulatorischen Sequenzen, die nicht nur die Transkription der HERV-Sequenzen selbst beeinflussen, sondern ggf. auch benachbarte Genomregionen regulieren können (Jern und Coffin, 2008; Cohen *et al.*, 2009).

Für einige HERV-Gruppen wird aufgrund der Ergebnisse früherer Studien, in denen HERV-Transkripte sowie retrovirale Proteine z.T. spezifisch in verschiedenen kranken Geweben detektiert wurden, eine potentielle direkte oder indirekte Rolle in der Pathogenese verschiedener Krankheiten diskutiert. Es wird hierbei u.a. eine potentielle Involvierung der HERV-K(HML-2)-Gruppe in verschiedenen Tumorerkrankungen diskutiert (Ruprecht *et al.*, 2008b) sowie eine mögliche Rolle der HERV-W-Gruppe in der Pathogenese von Multipler Sklerose (Antony *et al.*, 2011). Auch in diesem Zusammenhang ist die Identifizierung transkribierter HERV-Loci von besonderem Interesse. Die Klärung der Fragen, ob sich kranke und gesunde Gewebe durch spezifische Transkriptionsmuster oder die Aktivität spezifischer Loci unterscheiden, sowie eine detaillierte Charakterisierung der Struktur und Genomlokalisierung transkribierter Loci können zu einem besseren Verständnis der potentiellen Rolle von HERVs in der Krankheitspathogenese beitragen.

### 4.1. Charakterisierung der Transkription von HERV-K(HML-2)-Loci in Tumor- und Normalgeweben

Eine potentielle Rolle der HERV-K(HML-2)-Gruppe humaner endogener Retroviren wird insbesondere in verschiedenen Tumorerkrankungen diskutiert, darunter Keimzelltumoren (*germ cell tumours*, GCTs), Prostatakarzinome und Melanome. In zahlreichen Tumor-Proben und Zelllinien wurden HML-2-Transkripte (Herbst *et al.*, 1996; Büscher *et al.*, 2005, 2006; Reiche *et al.*, 2010; Goering *et al.*, 2011; Agoni *et al.*, 2013) sowie retrovirale Gag- und/oder Env-Proteine detektiert (Sauter *et al.*, 1995, Bieda *et al.*, 2001, Muster *et al.*, 2003, Reis *et al.*, 2013), und

GCT- sowie Melanom-Patienten zeigten erhöhte Antikörper-Titer gegen HML-2 Gag- und Env-Proteine (Sauter *et al.*, 1995, 1996; Boller *et al.*, 1997; Hahn *et al.*, 2008). Dies deutet möglicherweise auf eine spezifische Expression sowie eine potentielle Rolle der retroviralen Proteine in der Krankheitspathogenese hin (vgl. Abschnitt 2.3.1.4. der Einleitung). Insbesondere wurde hierbei auch eine potentielle Rolle der akzessorischen HML-2-Proteine Rec und Np9 diskutiert, deren Interaktion mit verschiedenen an der Zellproliferation und Zelldifferenzierung beteiligten Proteinen gezeigt wurde (Armbruster *et al.*, 2002, 2004; Pece *et al.*, 04; Denne *et al.*, 2007) (vgl. Abschnitt 2.3.1.5. der Einleitung).

Als Beitrag zum besseren Verständnis der potentiellen Rolle von HML-2 in der Tumorphagenese wurden umfangreiche Transkriptionsanalysen durchgeführt und transkribierte HML-2-Loci in zahlreichen Melanom-, GCT- und Prostatakarzinom-Proben sowie Normalgeweben identifiziert und hinsichtlich ihrer Struktur, Kodierungskapazität und Genomlokalisierung weiter analysiert.

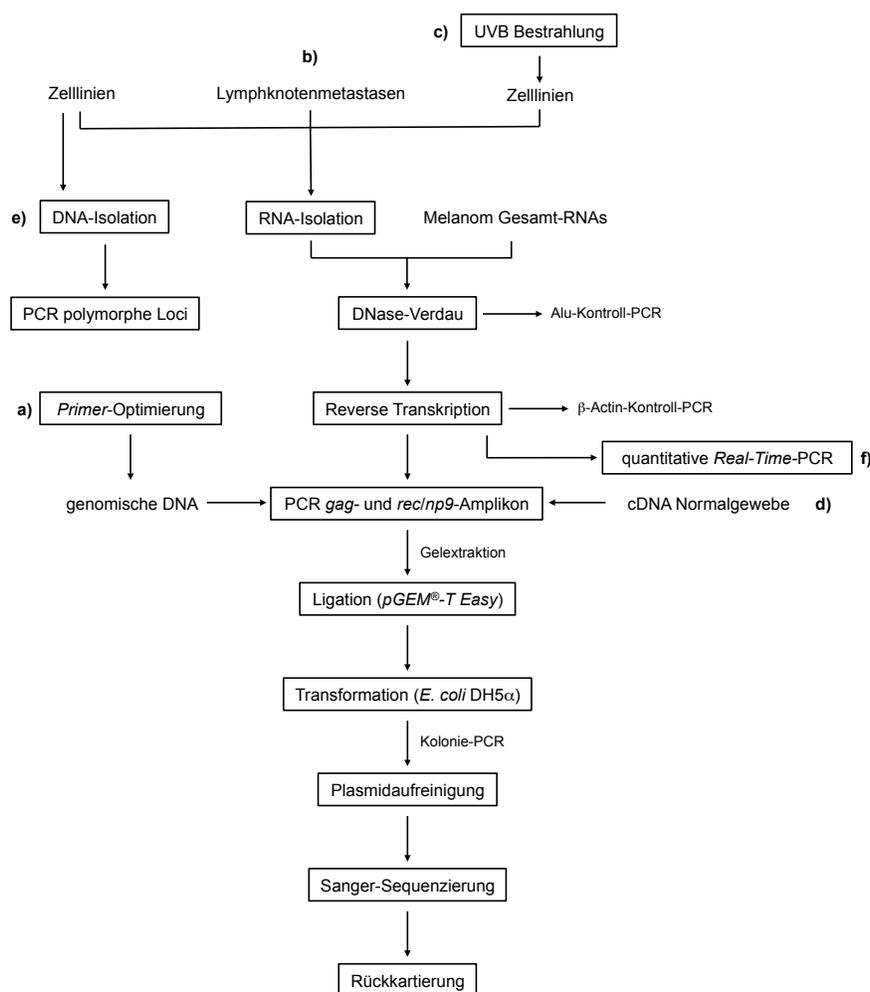
#### **4.1.1. Arbeitsablauf zur Identifizierung transkribierter HERV-K(HML-2)-Loci, deren näherer Charakterisierung und Quantifizierung ihrer Transkriptmengen**

Zur Untersuchung einer möglichen biologischen Involvierung von HERV-K(HML-2) in Melanomen wurden transkribierte HML-2-Loci in vier Melanom-Zelllinien, Lymphknoten-Metastasen aus zwei Melanom-Patienten, Gesamt-RNA aus drei Melanom-Metastasen sowie zwei Melanozyten-Zelllinien identifiziert. Der Arbeitsablauf ist in **Abbildung 12** gezeigt.

Die zur HML-2-Gruppen-spezifischen PCR verwendeten *Primer* wurden hinsichtlich der Amplifizierbarkeit möglichst vieler HML-2-Loci in einer einzelnen PCR optimiert und zur Verifizierung in einer PCR mit genomischer DNA getestet (vgl. **Abb. 12a**). Das untersuchte Amplikon im *gag*-ORF (*open reading frame*) von HML-2 wurde von genomischer DNA amplifiziert, die PCR-Produkte gelelektrophoretisch aufgetrennt, aufgereinigt, in *pGEM<sup>®</sup>-T Easy* kloniert und in *E. coli* DH5a transformiert. Positive Klone wurden mittels Kolonie-PCR identifiziert, die Plasmide aus diesen aufgereinigt und Sequenzen durch Sanger-Sequenzierung generiert und mittels *BLAT* im humanen hg18 Referenzgenom rückkartiert.

Aus den untersuchten Zelllinien und Lymphknoten-Metastasen wurde Gesamt-RNA isoliert (vgl. **Abb. 12b**). Diese RNA-Proben sowie weitere kommerziell erworbene Gesamt-RNA aus drei Melanom-Metastasen wurden einem DNase-Verdau unterzogen, dessen Erfolg im Anschluss mittels Alu-Element-spezifischer PCR überprüft wurde. DNase-behandelte RNA wurde in cDNA

umgeschrieben, wobei für jede RNA-Probe ein Ansatz ohne Enzym als RT(-) Negativkontrolle diente. Der Erfolg der reversen Transkription sowie die Abwesenheit einer DNA-Kontamination in den Negativkontrollen wurde mittels  $\beta$ -Actin-spezifischer PCR überprüft. Es wurden nur Proben weiter verwendet, die sowohl nach der DNase-Behandlung als auch nach der reversen Transkription negativ auf Reste genomischer DNA getestet wurden. Von der generierten cDNA wurden im Anschluss die Amplikons *gag* und *rec/np9* amplifiziert und, wie für genomische DNA beschrieben, durch Klonierung und anschließende Sanger-Sequenzierung Sequenzen generiert und im humanen Referenzgenom rückkartiert.

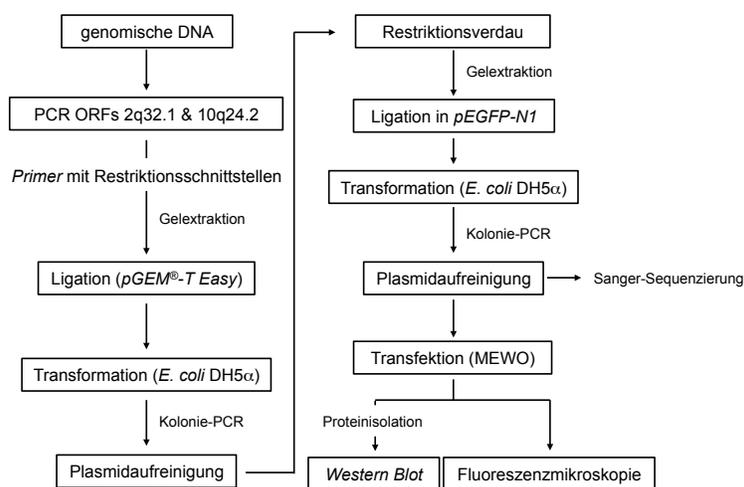


**Abbildung 12:** Arbeitsablauf zur Identifizierung transkribierter HERV-K(HML-2)-Loci in Melanom-Proben, Melanozyten und Normalgeweben sowie zur Quantifizierung der HML-2-Transkriptmengen. Zur Erklärung s. Text.

Zur Untersuchung des Einflusses von UVB-Strahlung auf die HML-2-Transkription wurden die zur Verfügung stehenden Melanom- und Melanozyten-Zelllinien mit  $200 \text{ mJ/cm}^2$  UVB bestrahlt, nach 24 h die Gesamt-RNA isoliert und, wie oben beschrieben, *gag* cDNA-Sequenzen generiert und rückkartiert (vgl. **Abb. 12c**).

In weiteren Experimenten wurde der Allelstatus polymorpher HML-2-Loci in genomischer DNA der verwendeten Zelllinien mittels PCR überprüft (vgl. **Abb. 12e**) sowie die HERV-K(HML-2) *gag*-Transkriptmengen mittels *Real-Time*-PCR quantifiziert (vgl. **Abb. 12f**).

Des Weiteren wurden die potentiellen Proteine, die von zwei durch reverse Transkription und Reintegration gespleißter mRNA entstandenen HML-2-Loci auf den Chromosomen 2q32.1 und 10q24.2 kodiert werden, näher charakterisiert (vgl. **Abb. 13**). Die ORFs beider Loci wurden amplifiziert, jeweils in *pGEM<sup>®</sup>-T Easy* zwischenkloniert und durch *Bam*HI-Verdau in den *pEGFP-N1*-Vektor *in-frame* stromauf des GFP-Gens kloniert. Exprimierte GFP-Fusionsproteine wurden mittels *Western Blot* und Fluoreszenzmikroskopie detektiert.



**Abbildung 13:** Arbeitsablauf zu Untersuchung *der Transkription* von Proteinen, die von zwei durch reverse Transkription und Reintegration gespleißter mRNA entstandenen HML-2-Loci auf den Chromosomen 2q32.1 codiert werden. Zur Erklärung s. Text.

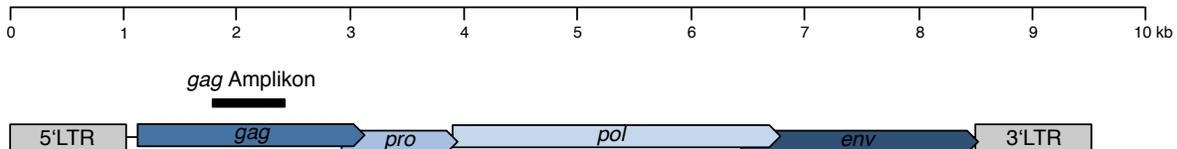
Analog der Melanom-Studie wurden transkribierte HML-2-Loci in zwei Keimzelltumor- sowie sechs Prostatakarzinom-Zelllinien durch Sanger-Sequenzierung von *gag*-spezifischen RT-PCR-Produkten identifiziert. Des Weiteren wurde die HML-2 *rec/np9* Transkription in Normalgeweben untersucht. Hierfür stand cDNA aus 16 menschlichen Normalgeweben zur Verfügung, aus denen das *rec/np9*-Amplikon amplifiziert und, wie in den vorhergehenden Experimenten, cDNA-Sequenzen generiert und rückkartiert wurden (vgl. **Abb. 12d**).

#### 4.1.2. Optimierung von PCR-Primern und Amplifikationsstrategie zur Identifizierung transkribierter HERV-K(HML-2)-Loci

Die zur Amplifikation HML-2-spezifischer cDNA-Sequenzen verwendeten *Primer* wurden zunächst dahingehen optimiert, dass in einer einzelnen PCR möglichst viele der im humanen

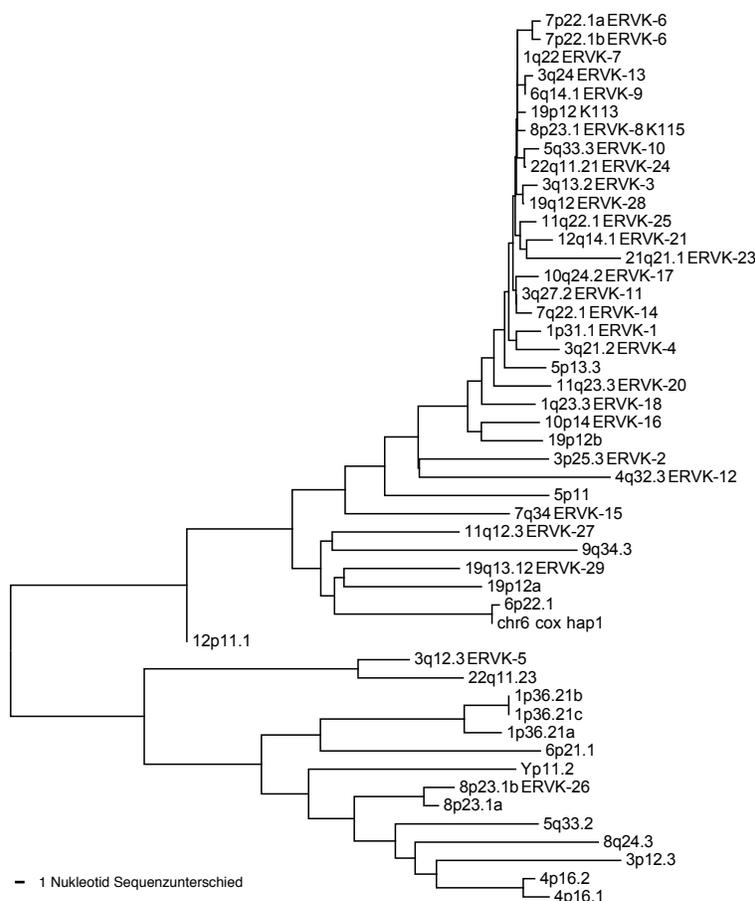
Genom vorhandenen HML-2-Loci gleichzeitig amplifiziert werden können. Hierfür wurde zunächst eine möglichst vollständige Übersicht über die genomischen HML-2-Loci, insbesondere deren Sequenzen, erstellt. Im menschlichen hg18 Referenzgenom wurde eine *BLAT*-Suche mit der bei *Repbase* (Jurka *et al.*, 2005) hinterlegten HML-2-Konsensussequenz durchgeführt. Die DNA-Sequenzen der resultierenden Loci, incl. LTRs, wurden mit dem *UCSC Table Browser* (Karolchik *et al.*, 2004) zusammengestellt und mittels *MAFFT* (Kato *et al.*, 2002) wurde ein multiples *Alignment* dieser Sequenzen generiert. Es ergab sich ein *Alignment* von 80 genomischen HML-2-Loci, inklusive des polymorphen HERV-K113-Locus (vgl. **Anhang 1**). Es ist anzumerken, dass das K113-Provirus nicht im hg18 Referenzgenom annotiert ist und die Sequenz des Locus (*Genbank acc. no.* AY037928.1) daher der Sequenzsammlung nachträglich hinzugefügt wurde. Bei 36 der HML-2-Loci handelt es sich um annähernd vollständige Vollängen-Proviren mit einer internen Sequenz von über 7.000 bp. Die restlichen Sequenzen weisen größere Deletionen auf und besitzen interne Sequenzen zwischen 3.889 und 6.579 bp.

Als Amplikonregion wurde ein zentraler Bereich im *gag*-ORF gewählt, der dem Bereich von nt 1.778 bis 2.396 der Sequenz des HERV-K(HML-2.HOM)-Provirus (*Genbank acc. no.* AF074086.2) entspricht (vgl. **Abb. 14** und **Anhang 1**).



**Abbildung 14:** Position des untersuchten Amplikons im HERV-K(HML-2) *gag*-ORF. Gezeigt ist die Struktur der HERV-K(HML-2.HOM)-Sequenz (*Genbank acc. no.* AF074086.2) mit ORFs der retroviralen Gene *gag*, *pro*, *pol* und *env* sowie den flankierenden LTRs. Der Amplikonbereich ist als schwarzer Balken eingezeichnet. Die Skala entspricht der Größe des Provirus in kb.

Der etwa 620 bp lange Amplikonbereich ist in 35 der 36 (97,22 %) Vollängen-Proviren vorhanden, darunter auch in den polymorphen Loci HERV-K113 und HERV-K115, sowie in 13 der verkürzten Loci. Der Bereich im *gag*-ORF wurde ausgewählt, da viele der Loci hier eine hohe Ähnlichkeit in den *Primer*-bindenden Regionen besitzen, die Amplikonregion insgesamt jedoch noch genügend Nukleotidunterschiede aufweist, um für fast alle Loci eine eindeutige Rückkartierung generierter cDNA-Sequenzen zu gewährleisten (vgl. **Abb. 15**). Alle theoretisch amplifizierbaren Loci, mit Ausnahme von zwei Proviren auf Chromosom 1p36.21, unterscheiden sich innerhalb der Amplikonregion in mindestens einer Nukleotidposition und/oder diversen Insertionen oder Deletionen.



**Abbildung 15:** Graphische Darstellung eines Sequenzvergleiches des untersuchten HERV-K(HML-2) *gag*-Amplikons in allen theoretisch amplifizierbaren HML-2-Loci. Der dargestellte *Neighbour-joining* Baum stellt die absolute Anzahl an Nukleotidunterschieden zwischen den amplifizierbaren HML-2-Loci in der untersuchten *gag*-Amplikonregion dar, wobei Insertionen und Deletionen nicht berücksichtigt wurden. Die Sequenzen der *Primer*-bindenden Regionen der Amplikonbereiche gingen nicht in die Sequenzvergleiche ein. Die einzelnen Loci sind entsprechend ihrer jeweiligen Lage in Chromosomenbanden benannt sowie, falls vorhanden, ihrem offiziellen HGNC-Namen (vgl. Mayer *et al.*, 2011). Bis auf zwei Loci auf Chromosom 1p36.21 unterscheiden sich alle amplifizierbaren Loci in mindestens einer Position. Die Skala gibt die einem Nukleotidunterschied entsprechende Astlänge an. Der *Neighbour-joining* Baum wurde von apl.-Prof. Dr. rer. nat. Jens Mayer erstellt.

In früheren Studien (Flockerzi *et al.*, 2007, 2008; Ruprecht *et al.*, 2008a) wurde bereits eine ~650 bp große Region im *gag*-ORF untersucht, die etwa 150 bp stromauf des in der vorliegenden Arbeit gewählten Amplikons lokalisiert ist und mit diesem in ihrem 3'-Bereich überlappt (vgl. **Anhang 1**). Da es sich bei HERV-K(HML-2) um eine, evolutiv gesehen, relativ junge HERV-Gruppe handelt und sich die Loci untereinander stärker ähneln als die von evolutiv älteren HERV-Gruppen, wurde der Amplikonbereich vergrößert und verschoben, um genügend charakteristische Nukleotidunterschiede der Loci untereinander zu erfassen und so die Rückkartierung von generierten cDNA-Sequenzen grundsätzlich zu verbessern. Die Anzahl an potentiell amplifizierbaren Loci wurde zudem in der vorliegenden Arbeit weiter erhöht, indem verschiedene Sequenzvarianten der *Forward*- und *Reverse-Primer* eingesetzt wurden. Diese *Primer*-Varianten binden zwar jeweils im gleichen Bereich des *gag*-ORFs, berücksichtigen

jedoch Nukleotidunterschiede innerhalb der beiden *Primer*-bindenden Regionen zwischen den verschiedenen HML-2-Loci. Es wurden insgesamt vier *Forward*- und drei *Reverse-Primer*-Varianten eingesetzt, wobei diese zur PCR-Amplifikation in einem Verhältnis gemischt wurden, das in etwa der Anzahl an möglichen Zielsequenzen entspricht. Betrachtet man die Loci, die sowohl zu einer der *Forward*- als auch zu einer der *Reverse-Primer*-Varianten eine 100 %-ige Übereinstimmung aufweisen, ergibt sich eine Anzahl von 28 theoretisch amplifizierbaren Loci, verglichen mit nur 13 Loci in der Studie von Flockerzi *et al.* (2008), die nach Sequenzvergleich diesem Kriterium entsprachen. Weitere elf Loci weisen einen oder zwei Nukleotidunterschiede in den *Primer*-bindenden Regionen zu einer der *Forward*- und/oder *Reverse-Primer*-Varianten auf, die sich jedoch nicht unmittelbar am 3'-Ende des *Primers* befinden und eine Amplifikation dieser Loci unter geeigneten PCR-Bedingungen (s.u.) trotzdem ermöglichen sollten. Weitere sechs HML-2-Loci zeigen mehr als zwei Nukleotidunterschiede und drei weitere Loci weisen einen Nukleotidunterschied am 3'-Ende der *Primer*-bindenden Region auf, was wahrscheinlich in beiden Fällen eine Amplifikation der betreffenden Loci verhindert.

Um auch die Amplifikation von HML-2-Loci zu begünstigen, die Nukleotidunterschiede in den *Primer*-bindenden Regionen aufweisen, wurden PCR-Amplifikationen unter niedrig stringenten Bedingungen bei relativ niedriger *Primer-Annealing*-Temperatur durchgeführt. Tatsächlich wiesen später acht der von genomischer und/oder cDNA amplifizierten Loci einen oder zwei Nukleotidunterschiede zu den *Primer*-Varianten auf (s.u.). Die Anzahl an theoretisch amplifizierbaren Loci sollte durch den Einsatz von *Primer*-Gemischen und die Wahl niedrig stringenter PCR-Bedingungen somit deutlich erhöht werden.

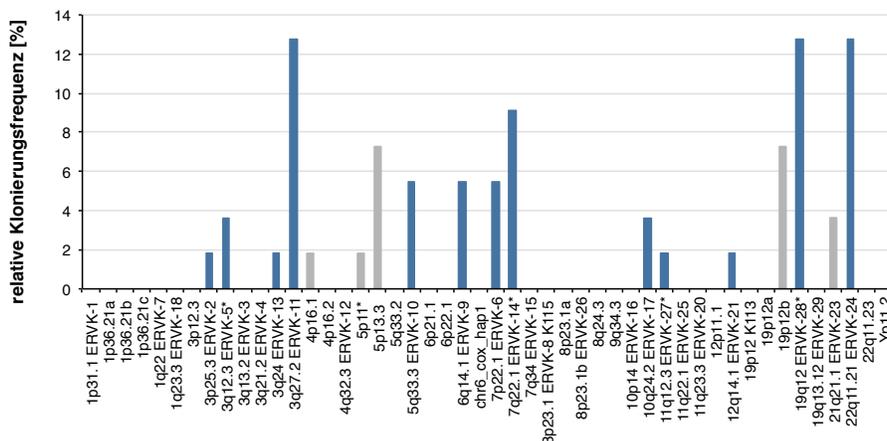
#### **4.1.3. Test der PCR-*Primer* zur Amplifikation von HERV-K(HML-2)-Loci sowie Strategie zur Rückkartierung generierter Sequenzen**

Die ausgewählten *Primer* wurden zunächst in PCRs mit humaner genomischer DNA getestet, um die eindeutige Rückkartierbarkeit der generierten Sequenzen zu verifizieren und eine präferentielle Amplifikation einzelner Loci auszuschließen. Das HML-2 *gag*-Amplikon wurde hierzu von genomischer DNA amplifiziert, in *pGEM<sup>®</sup>-TEasy* kloniert, in *E. coli* DH5a transformiert und Insert-tragende Plasmide mittels Kolonie-PCR identifiziert, aufgereinigt und sequenziert. Die generierten Sanger-Sequenzen wurden mittels *BLAT* (Kent, 2002) im menschlichen hg18 Referenzgenom rückkartiert. Diese PCR wurde, wie auch die späteren PCRs mit cDNA, unter niedrig stringenten Bedingungen durchgeführt, was auch die Amplifikation solcher Loci erlauben sollte, die Sequenzunterschiede zu den verwendeten

*Primern* aufweisen. Aus diesem Grund wurden hier sowie in den nachfolgenden Experimenten mit cDNA die *Primer*-bindenden Regionen der generierten Sequenzen vor der Rückkartierung entfernt, da dies sonst unter Umständen die Rückkartierung durch eine Erhöhung der Sequenzunterschiede zu den genomischen Loci erschwert hätte.

Hier sowie im weiteren Verlauf der Arbeit gingen nur eindeutig rückkartierbare Sequenzen in die weitere Analyse mit ein. Die Rückkartierung zu einem genomischen HML-2-Locus wurde als eindeutig definiert, wenn sich nur ein einziger bester Treffer in der *BLAT*-Suche (in der weiteren Arbeit als „*BLAT-Match*“ bezeichnet) ergab und alle weiteren Treffer mindestens einen weiteren Sequenzunterschied hierzu aufwiesen. Sequenzen mit zwei oder mehr besten *BLAT-Matches* gleicher Größe mit gleicher Anzahl an Nukleotidunterschieden wurden verworfen. Ebenso wurden Sequenzen mit mehr als sechs Nukleotidunterschieden zum besten *BLAT-Match* nicht berücksichtigt. Hierdurch wird ein Sequenzunterschied zwischen generierter Sequenz und Referenzsequenz von bis zu etwa 1 % toleriert, der auf Fehler bei der PCR-Amplifikation und Sequenzierreaktion oder auf interindividuelle Sequenzunterschiede bzw. Einzelnukleotidpolymorphismen (*single nucleotide polymorphisms*, SNPs) zurückzuführen sein kann (vgl. hierzu auch Abschnitt 5.1.1. der Diskussion). Eine Ausnahme bei der Rückkartierung stellte das HERV-K(HML-2.HOM)-Provirus (*ERVK-6*) auf Chromosom 7p22.1 dar, das in Teilen der Bevölkerung in Form von zwei Tandemprovirten vorliegt, die sich einen zentralen LTR teilen (Reus *et al.*, 2001a). Da sich beide Tandemprovirten über ihre gesamte Länge nur in wenigen Nukleotiden unterscheiden, ist hier eine eindeutige Zuordnung zum einen oder anderen Provirus in vielen Fällen nicht möglich, weswegen das HERV-K(HML-2.HOM)-Provirus in den weiteren Analysen als ein Locus gewertet wurde.

Es konnten 55 der von genomischer DNA generierten Sequenzen eindeutig im Genom rückkartiert werden (vgl. **Abb. 16**). Es wurden Sequenzen von 18 der 48 theoretisch amplifizierbaren Loci generiert. Die amplifizierten Loci verteilten sich auf elf verschiedene Chromosomen, wobei vier der Loci auf Chromosom 3 lokalisiert waren und je drei Loci auf den Chromosomen 5 und 7. Fünf der amplifizierten Loci zeigten einen oder zwei Nukleotidunterschiede in der *Primer*-bindenden Region zu den verwendeten *Forward*- und/oder *Reverse-Primer*-Varianten, was die Amplifikation solcher Loci unter niedrig stringenten PCR-Bedingungen bestätigt. Eine explizit präferentielle Amplifikation spezifischer Loci wurde nicht beobachtet. Zwar wurden einige Loci, wie *ERVK-11* (3q27.2), *ERVK-24* (22q11.21) oder *ERVK-28* (19q12), etwas häufiger amplifiziert, diese waren aber bei späterer Amplifikation von cDNA nicht dominant vertreten.



**Abbildung 16:** Relative Klonierungsfrequenzen von aus genomischer DNA amplifizierten HERV-K(HML-2)-Loci. Es konnten 55 generierte Sequenzen des Amplikons im HML-2 *gag*-ORF eindeutig im menschlichen hg18 Referenzgenom rückkartiert werden (vgl. Text). Gezeigt ist die relative Klonierungsfrequenz in [%] aller theoretisch amplifizierbaren Loci. Diese berechnete sich für jeden Locus aus der Anzahl an Sequenzen, die eindeutig diesem Locus zugeordnet werden konnten, dividiert durch die Gesamtzahl an eindeutig rückkartierbaren Sequenzen aller Loci. Die Loci sind entsprechend ihrer Lokalisation in einer Chromosomenbande sowie, falls vorhanden, ihrem offiziellen HGNC-Namen benannt (vgl. Mayer *et al.*, 2011). Mit einem blauen Balken markierte Loci wurden in späteren Experimenten auch von cDNA identifiziert, mit grauem Balken markierte Loci wurden ausschließlich aus genomischer DNA amplifiziert. Mit einem Stern (\*) markierte Loci wiesen einen oder zwei Nukleotidunterschiede zu den zur Amplifikation verwendeten *Forward*- und/oder *Reverse-Primer*-Varianten auf.

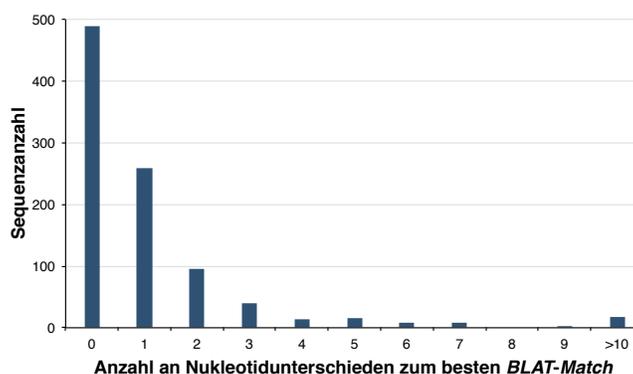
#### 4.1.4. Identifizierung transkribierter HERV-K(HML-2)-Loci in Melanom-Proben und Melanozyten-Zelllinien

Als Beitrag zum besseren Verständnis der biologischen Rolle von HERV-K(HML-2) in Melanomen (vgl. Ruprecht *et al.*, 2008b), wurden transkribierte HML-2-Loci in Melanom-Proben und Melanozyten identifiziert. Hierfür standen die Melanom-Zelllinien SK-Mel-25, SK-Mel-28, MEWO und WM3734a zur Verfügung, zwei Melanom-Lymphknoten-Metastasen (LKM1 und LKM2) sowie die Melanozyten-Zelllinien Benno und Oskar. Aus allen Proben wurde RNA isoliert und drei weitere Gesamt-RNA-Proben (RNA1-3) aus Melanom-Metastasen käuflich erworben.

Die RNA aller Proben wurde, wie in Abschnitt 4.1.1. beschrieben, nach DNase-Verdau in cDNA umgeschrieben und das *gag*-Amplikon daraus amplifiziert (vgl. **Anhang 2a**). Die ~620 bp großen PCR-Produkte, sowie ggf. weitere auftretende Produkte, die durch HML-2-Loci mit größeren Insertionen oder Deletionen begründbar wären, wurden aufgereinigt, in *pGEM<sup>®</sup>-T Easy* kloniert, in *E. coli* DH5a transformiert und Plasmide aus positiven Klonen aufgereinigt und mittels Sanger-Methode sequenziert.

Für jede Probe wurden 96 cDNA-Sequenzen generiert, wobei insgesamt 991 Sanger-Sequenzen guter Qualität in die weitere Analyse eingingen. Durchschnittlich vier Sequenzen pro

Probe (~5 %) wiesen mehr als einen besten *BLAT-Match* in der Rückkartierung auf und wurden verworfen, da eine eindeutige Zuordnung zu einem spezifischen HML-2-Locus in diesem Fall prinzipiell nicht möglich war (vgl. hierzu auch Abschnitt 5.1.1. der Diskussion). Die durchschnittliche Anzahl an Sequenzunterschieden der verbleibenden 945 Sequenzen zu ihrem besten *BLAT-Match* betrug 1,15. Insgesamt zeigten 918 Sequenzen (97,14 %) weniger als sechs Nukleotidunterschiede und wurden für die weitere Analyse verwendet. Von diesen Sequenzen wiesen 489 (51,75 %) keinen und weitere 259 (27,41 %) lediglich einen Sequenzunterschied zu ihrem besten *BLAT-Match* auf sowie weitere 95 (10,05 %), 39 (4,13 %), 13 (1,38 %), 15 (1,59 %) und 8 (0,85 %) Sequenzen, in dieser Reihenfolge, zwischen zwei und sechs Nukleotidunterschiede. Die restlichen 27 Sequenzen (2,86 %) mit mehr als sechs Nukleotidunterschieden wurden verworfen (vgl. **Abb. 17**).



**Abbildung 17:** Sequenzunterschiede der aus Melanom- und Melanozyten-cDNA generierten HERV-K(HML-2) *gag* cDNA-Sequenzen zum ihrem jeweils besten *BLAT-Match*. Gezeigt ist die absolute Anzahl an Nukleotidunterschieden von 945 eindeutig rückkartierbaren Sequenzen. Sequenzen mit mehr als sechs Nukleotidunterschieden wurden verworfen.

Für jede Probe waren im Schnitt 83,45 Sequenzen eindeutig rückkartierbar. Insgesamt konnten 22 verschiedene HML-2-Loci als transkribiert identifiziert werden, die in **Tabelle 15** zusammengefasst sind.

Vier der Loci waren zuvor noch nicht als transkribiert beschrieben worden. Es handelt sich hierbei um Loci auf den Chromosomen 8p23.1 (*ERVK-26*), 11q12.3 (*ERVK-27*), 19q12 (*ERVK-28*) sowie 19q13.12 (*ERVK-29*). Hierbei ist zu beachten, dass es sich bei dem *ERVK-26*-Locus auf Chromosom 8p23.1 nicht um das polymorphe HERV-K115-Provirus handelt, das in derselben Chromosomenbande lokalisiert ist. Im Zuge der eindeutigen Benennung aller transkribierten HERV-Loci (vgl. Mayer *et al.*, 2011), wurden die vier Loci entsprechend der Richtlinien des *HUGO Gene Nomenclature Committee* (HGNC) als *ERVK-26* bis *ERVK-29* benannt, und korrespondierende cDNA-Sequenzen als EST-Sequenzen in *Genbank* hinterlegt (JZ164910, JZ164909, JZ164911 und JZ164912).

**Tabelle 15:** Relative Klonierungsfrequenzen der in Melanom-Proben und Melanozyten als transkribiert identifizierten HERV-K(HML-2)-Loci.\*

Name	HML-2-Locus				Melanom-Zelllinien				Melanom-RNA				Lymphknoten-Met.		Melanozyten			
	Chr	Start	Ende	Bande	ORFs	SK-Mel-25	SK-Mel-28	MEWO	WM3734a	RNA1	RNA2	RNA3	LKM1	LKM2	Benno	Oskar		
ERVK-1	1	75615359	75621731	1p31.1	n	48,08	6,94			2,27	7,06	29,17	9,88	53,66	2,56	1,49		
ERVK-7	1	153863081	153872260	1q22	n					1,14								
ERVK-18	1	158927199	158936430	1q23.3						6,82	12,94	31,94	1,23	1,22	14,10	5,97		
ERVK-5 <sup>b</sup>	3	102893427	102902549	3q12.3	g/n	9,62	6,00	6,00	2,25		1,18							
ERVK-3	3	114225814	114234972	3q13.2	n	10,58	4,00	4,00	3,37		1,18	9,72	4,94	2,44	7,69			
ERVK-4	3	127091992	127101129	3q21.2	r	0,96												
ERVK-13	3	149764167	149768086	3q24														
ERVK-11	3	186763030	186772209	3q27.2	n	1,39	1,00	1,00		3,41	7,06	2,78	4,94		1,49			
ERVK-10	5	156017295	156026474	5q33.3	g/n	5,56	1,00	1,00					2,47	1,22		14,93		
ERVK-9	6	78483381	78492802	6q14.1	g/r													
ERVK-6 <sup>a</sup>	7	4588583	4606557	7p22.1	g/e/r	21,19	74,99	6,00			5,88	1,39	20,99	24,39	69,23	61,19		
ERVK-14 <sup>b</sup>	7	104175604	104180502	7q22.1	g/r	4,17	4,17	79,00	92,13	86,36	61,18	4,17	45,68	1,22	5,13			
ERVK-15 <sup>b</sup>	7	141097395	141102407	7q34		0,96	1,00	1,00	2,25				1,23					
ERVK-26 <sup>b</sup>	8	12360863	12370378	8p23.1														
ERVK-16	10	6906147	6915609	10p14	r	1,92	2,78	1,00					1,23			11,94		
ERVK-17	10	101570559	101577735	10q24.2	r								2,47					
ERVK-27 <sup>ab</sup>	11	61892538	61907139	11q12.3										14,63				
ERVK-21	12	57007509	57016965	12q14.1	e/r													
ERVK-28 <sup>ab</sup>	19	32820338	32829201	19q12	g/e/r								3,70		1,28	1,49		
ERVK-29 <sup>b</sup>	19	42289389	42298906	19q13.12	r	1,92	4,17				1,18	8,33						
ERVK-23	21	18855530	18863833	21q21.1	n	4,81		1,00			2,35		1,23	1,22	1,49			
ERVK-24	22	17306187	17315361	22q11.21	g/n	104	72	100	89	88	85	72	81	82	78	67		
					Gesamtanzahl an Sequenzen												6	8
					Anzahl verschiedener Loci												6	8

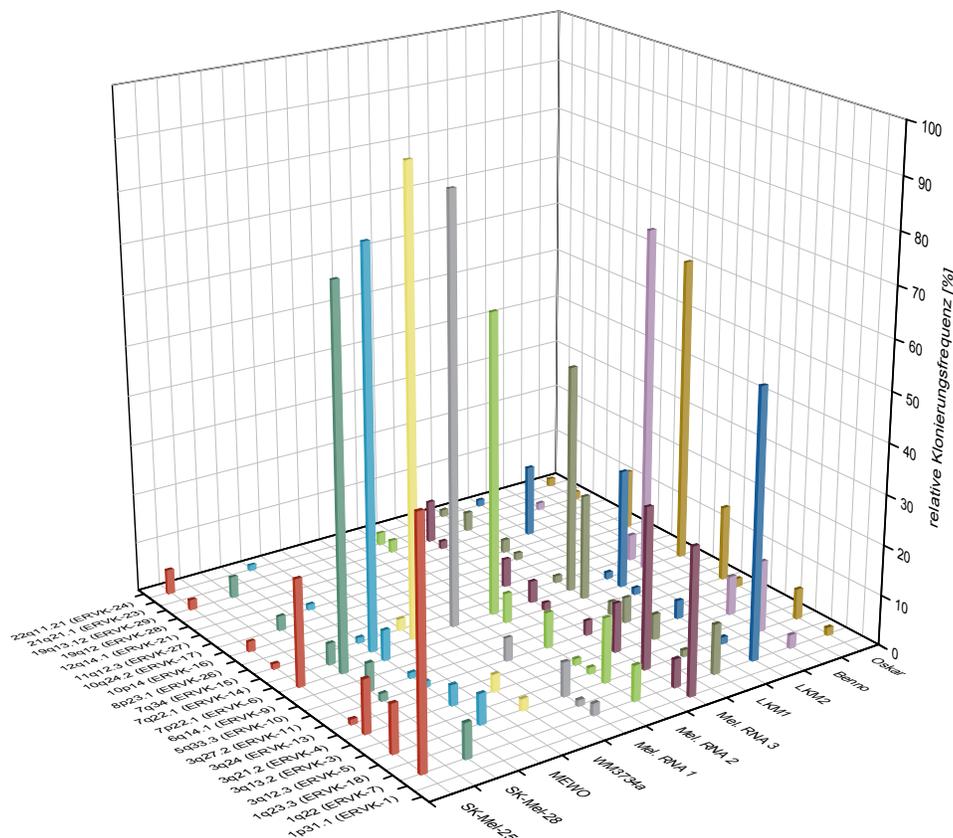
\*Das Amplicon im HERV-K(HML-2) gag-ORF wurde aus cDNA der Melanom-Zelllinien SK-Mel-25, SK-Mel-28, MEWO und WM3734a, drei Melanom-Metastasen (RNA1-3), zwei Melanom Lymphknoten-Metastasen (LKM1 und 2) sowie den beiden Melanozyten-Zelllinien Benno und Oskar amplifiziert. Angegeben sind, neben der Chromosomenposition und -bande des kompletten HML-2-Locus incl. LTRs, dessen offizieller HGNC-Name (vgl. Mayer et al., 2011) sowie seine relative Klonierungsfrequenz in den verschiedenen Proben in [%]. Diese berechnet sich in jeder Probe für jeden Locus aus der Anzahl der cDNA-Sequenzen, die eindeutig dem Locus zugeordnet werden konnten, dividiert durch die Gesamtanzahl an rückkartierbaren Sequenzen, die für die jeweilige Probe generiert wurden. Ebenfalls angegeben ist das Vorhandensein von *open reading frames* (ORFs) der retroviralen Proteine Gag (g), Env (e), Rec (r) und Np9 (n). Zur weiteren Erklärung, vgl. Text.

<sup>a</sup>Loci, die auf genomischer Ebene einen oder zwei Nukleotidunterschiede zu den zur Amplifikation verwendeten *Forward*- und/oder *Reverse-Primer*-Varianten aufweisen.  
<sup>b</sup>In dieser Studie erstmals als transkribiert identifizierte Loci.

In jeder Probe waren im Durchschnitt acht verschiedene HML-2-Loci transkribiert, wobei die Anzahl zwischen vier (WM3734a) und zwölf (LKM1) Loci schwankte. Transkribierte Loci waren auf zwölf verschiedenen Chromosomen lokalisiert, davon fünf Loci auf Chromosom 3 und je drei auf den Chromosomen 1 und 7. Wie schon in den vorangehenden Experimenten mit genomischer DNA, zeigten auch auf cDNA-Ebene sechs der als transkribiert identifizierten Loci auf genomischer Ebene einen oder zwei Sequenzunterschiede zu den zur Amplifikation verwendeten *Primer*-Varianten, was die Amplifikation solcher Loci unter niedrig stringenten PCR-Bedingungen weiter bestätigt.

Die relative Klonierungsfrequenz eines HML-2-Locus berechnete sich für jede Probe aus der Anzahl an generierten cDNA-Sequenzen, die eindeutig dem jeweiligen Locus zugeordnet werden konnten, dividiert durch die Gesamtanzahl an rückkartierbaren Sequenzen, die für die Probe generiert wurden. Eine hohe Klonierungsfrequenz kann hierbei als Zeichen dafür angesehen werden, dass ein Locus in der entsprechenden Probe relativ stark transkribiert war und folglich eine höhere Menge Locus-spezifischer Transkripte in der Ausgangs-RNA vorhanden waren, die anschließend in entsprechend höhere Mengen cDNA umgeschrieben und in der PCR amplifiziert wurden. Hierdurch wäre letztendlich die Wahrscheinlichkeit, dass der betreffende Locus kloniert und nach zufälliger Auswahl positiver, *Insert*-tragender Plasmide sequenziert wird, erhöht. Die relative cDNA-Klonierungsfrequenz eines Locus wurde daher als relatives Maß für die Transkriptionsrate dieses Locus herangezogen.

In den untersuchten Proben zeigten jeweils nur wenige Loci hohe relative Klonierungsfrequenzen von bis zu 92 %, was auf eine relativ starke Transkription dieser Loci hindeutet, wohingegen von den meisten Loci nur relativ wenige cDNA-Sequenzen generiert wurden und die Loci somit schwächer transkribiert zu sein scheinen. Besonders auffällig war hierbei der *ERVK-14*-Locus auf Chromosom 7q22.1, der in fünf der Melanom-Proben und den beiden Melanozyten-Zelllinien hohe relative Klonierungsfrequenzen von 45,68 % (LKM1) bis 92,13 % (WM3734a) zeigte, sowie der *ERVK-6*-Locus auf Chromosom 7p22.1 mit relativen Klonierungsfrequenzen von 21,19 % (SK-Mel-25) bis 74,99 % (SK-Mel-28) in vier der Melanom-Proben. Die Loci *ERVK-1* (1p31.1), *ERVK-7* (1q22) und *ERVK-5* (3q12.3) zeigten ebenfalls hohe relative Klonierungsfrequenzen von 29,17 % (*ERVK-1* in RNA3) bis 53,66 % (*ERVK-7* in LKM2) in jeweils einer oder zwei der Melanom-Proben. Die relativen Klonierungsfrequenzen der transkribierten Loci in den verschiedenen Proben sind in **Abbildung 18** graphisch dargestellt.



**Abbildung 18:** Transkriptionsprofile verschiedener HERV-K(HML-2)-Loci in Melanom-Proben und Melanozyten-Zelllinien. Das HERV-K(HML-2) *gag*-Amplikon wurde aus cDNA der Melanom-Zelllinien SK-Mel-25, SK-Mel-28, MEWO und WM3734a, drei Melanom-Metastasen (Mel. RNA1 - 3), zwei Lymphknoten-Metastasen aus Melanom-Patienten (LKM1 und 2) sowie den Melanozyten-Zelllinien Benno und Oskar amplifiziert und mittels Sanger-Methode sequenziert (vgl. Text). Die als transkribiert identifizierte Loci, die nach ihrer Lokalisation in einer Chromosomenbande und ihrem offiziellen HGNC-Namen (vgl. Mayer *et al.*, 2011) benannt sind, sind auf der x-Achse aufgetragen, die untersuchten Proben auf der y-Achse. Auf der z-Achse sind die relativen Klonierungsfrequenzen in [%] aufgetragen, die sich für jeden Locus in jeder Probe aus der Anzahl an cDNA-Sequenzen berechnete, die eindeutig dem jeweiligen Locus zugeordnet werden konnten, dividiert durch die Gesamtzahl an rückkartierbaren Sequenzen der Probe (die Frequenzen finden sich in **Tabelle 15**).

Alle Proben zeigten unterschiedliche Profile von transkribierten HML-2-Loci, wobei kein signifikantes Muster erkennbar war, das z.B. Melanom-Proben von Melanozyten-Zelllinien unterschied. Keiner der 22 Loci wurde in allen Proben als transkribiert identifiziert. Neun Loci wurden ausschließlich in jeweils einer der Melanom-Proben als transkribiert identifiziert (*ERVK-1*, *ERVK-18*, *ERVK-3*, *ERVK-9*, *ERVK-26*, *ERVK-16*, *ERVK-21*, *ERVK-28* und *ERVK-23*), wobei alle Loci, mit Ausnahme von *ERVK-1* (29,17 % in RNA3) und *ERVK-21* (14,63 % in LKM2) relativ schwach transkribiert zu sein scheinen, mit relativen Klonierungsfrequenzen von max. 5,56 % (*ERVK-26* in RNA3). Die Loci mit höheren relativen Klonierungsfrequenzen, darunter *ERVK-14* (7q22.1, max. 92,13 % in WM3734a), *ERVK-7* (1q22, max. 53,66 % in LKM2) und *ERVK-5* (3q12.3, max. 31,94 % in RNA3) wurden sowohl in Melanom-Proben als auch Melanozyten-Zelllinien detektiert. Eine Ausnahme bildete der bereits

erwähnte *ERVK-1*-Locus, der ausschließlich in einer der Melanom-Proben mit einer relativen Klonierungsfrequenz von 29,17 % (RNA3) als transkribiert identifiziert wurde. Auffällig war zudem der *ERVK-6*-Locus auf Chromosom 7p22.1, der in sieben der Melanom-Proben eine relative Klonierungsfrequenz von bis zu 74,99 % (SK-Mel-28) aufwies und somit relativ stark transkribiert zu sein scheint, der jedoch in den beiden Melanozyten-Zelllinien nicht als transkribiert gefunden wurde.

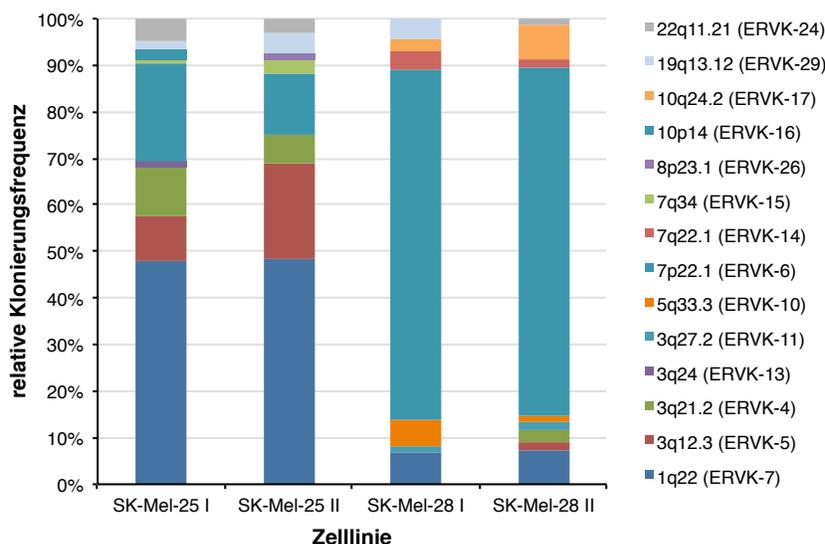
Es wurden somit bis zu zwölf verschiedene HML-2-Loci in Melanomen und Melanozyten-Zelllinien als transkribiert identifiziert, wobei sich die Transkriptionsmuster nicht signifikant zwischen den Melanom-Proben und Melanozyten-Zelllinien unterschieden. Es scheint vielmehr so, als würde sich jede Probe durch ein leicht unterschiedliches, charakteristisches Transkriptionsmuster spezifischer Loci auszeichnen. Einige Loci wurden ausschließlich in verschiedenen Melanom-Proben identifiziert, wobei diese, mit Ausnahme von *ERVK-6*, jedoch nur schwach transkribiert zu sein scheinen.

#### 4.1.5. Reproduzierbarkeit der cDNA-Transkriptionsprofile in Melanom-Proben

Zur Überprüfung der Reproduzierbarkeit bzw. der Stabilität der Transkriptionsprofile wurde das im vorherigen Abschnitt beschriebene Experiment exemplarisch für die Melanom-Zelllinien SK-Mel-25 und SK-Mel-28 wiederholt. Aus Zellpellets einer unterschiedlichen Passage beider Zelllinien wurde Gesamt-RNA isoliert und es wurden, wie beschrieben, *gag* cDNA-Sequenzen generiert und rückkartiert (zur Amplifikation vgl. **Anhang 2e**).

Insgesamt wurden für jede Zelllinie je 68 eindeutig rückkartierbare Sequenzen generiert. Bezüglich der relativen Klonierungsfrequenzen ergab sich für beide Zelllinien im Vergleich der beiden Experimente ein sehr ähnliches Bild (vgl. **Abbildung 19**; die relativen Klonierungsfrequenzen finden sich in **Anhang 3a**). Die in den vorangehenden Experimenten als am stärksten transkribiert identifizierten Loci, *ERVK-7* (1q22), *ERVK-5* (3q12.3) und *ERVK-6* (7p22.1) in SK-Mel-25 sowie *ERVK-6* (7p22.1) in SK-Mel-28, zeigten auch in der Wiederholung die höchsten relativen Klonierungsfrequenzen und scheinen somit dominant transkribiert. Lediglich bei den schwächer transkribierten Loci zeigten sich minimale Unterschiede. So wurden im Fall von SK-Mel-25 die im ersten Experiment mit ein und zwei Sequenzen detektierten Loci *ERVK-13* (3q24) und *ERVK-16* (10p14) im zweiten Experiment nicht detektiert. Für SK-Mel-28 wurden für die im ersten Experiment nicht detektierten Loci *ERVK-5* (3q12.3) und *ERVK-24* (22q11.21) je eine Sequenz und für *ERVK-4* (3q21.2) zwei Sequenzen generiert, wohingegen

der zuvor mit drei Sequenzen detektierte Locus *ERVK-29* (19q13.12) im zweiten Experiment nicht gefunden wurde. Diese minimalen Unterschiede sind sehr wahrscheinlich durch die niedrigen relativen Klonierungsfrequenzen der Loci begründet und würden vermutlich durch Generierung einer größeren Anzahl an cDNA Sequenzen minimiert.



**Abbildung 19:** Relative Klonierungsfrequenzen in unterschiedlichen Passagen zweier Melanom-Zelllinien als transkribiert identifizierter HERV-K(HML-2)-Loci. Das HML-2 *gag*-Amplikon wurde aus cDNA aus je zwei unterschiedlichen Passagen (I und II) der Melanom-Zelllinien SK-Mel-25 und SK-Mel-28 amplifiziert, mittels Sanger-Methode Sequenzen generiert und im menschlichen hg18 Referenzgenom rückkartiert (vgl. Text). Die Ergebnisse für Passage I beider Zelllinien sind bereits in **Abbildung 18** sowie **Tabelle 15** gezeigt. Angegeben sind die relativen Klonierungsfrequenzen jedes Locus in [%], die sich in jeder Zelllinie aus der Anzahl an cDNA-Sequenzen, die eindeutig dem Locus zugeordnet werden konnten, dividiert durch die Gesamtzahl an rückkartierbaren Sequenzen pro Zelllinie berechneten.

Durch Wiederholung des Experimentes konnte somit das spezifische Transkriptionsmuster verschiedener HML-2-Loci für unterschiedliche Passagen zweier Zelllinien exemplarisch bestätigt werden.

#### 4.1.6. Struktur und Kodierungskapazität der in Melanomen und Melanozyten-Zelllinien als transkribiert identifizierten HERV-K(HML-2)-Loci

Aufgrund ihrer potentiellen Kodierungskapazität für retrovirale Proteine sowie potentieller Promotor-Sequenzen und regulatorischer Elemente innerhalb der flankierenden LTRs wurden die als transkribiert identifizierten HML-2-Loci hinsichtlich ihrer Struktur und Kodierungskapazität näher analysiert. Die Ergebnisse sind in **Tabelle 16** zusammengefasst.

**Tabelle 16: Lokalisation, Struktur und Kodierungskapazität der in Melanom-Proben und Melanozyten-Zelllinien als transkribiert identifizierten HERV-K(HML-2)-Loci.\***

Genomlokalisierung					Struktur	5'/3'LTR	ORFs			Typ I/II
Chr	Start	Ende	Strang	Bande			<i>gag</i>	<i>env</i>	Rec/Np9	
1	75615359	75621731	+	1p31.1		+/+		Np9	I	
1	153863081	153872260	-	1q22		+/+		Np9	I	
1	158927199	158936430	+	1q23.3		+/+			I	
3	9864346	9871236	-	3p25.3 <sup>a</sup>		+/+			I	
3	102893427	102902549	+	3q12.3		+/+	+	Np9	I	
3	114225814	114234972	-	3q13.2		+/+		Np9	I	
3	127091992	127101129	+	3q21.2		beide Δ790-914		Rec	II	
3	149764167	149768086	-	3q24		-/-			-	
3	186763030	186772209	-	3q27.2		+/+		Np9	I	
5	156017295	156026474	-	5q33.3		+/+	+	Np9	I	
6	78483381	78492802	-	6q14.1		+/+	+	Rec	II	
7	4588583	4606557	-	7p22.1		+/+	+	Rec	II	
7	104175604	104180502	-	7q22.1		-/+ <sup>c</sup>			-	
7	141097395	141102407	-	7q34		Δ1-968/+ <sup>d</sup>			-	
8	12360863	12370378	-	8p23.1		+/+			II	
10	6906147	6915609	-	10p14		+/+		Rec	II	
10	101570559	101577735	-	10q24.2		Δ1-940/+		Rec	II	
11	61892538	61907139	-	11q12.3 <sup>b</sup>		Δ892-944/+			II	
12	57007509	57016965	-	12q14.1		+/+		Rec	II	
19	32820338	32829201	-	19q12 <sup>c</sup>		-/+	+	Rec	II	
19	42289389	42298906	-	19q13.12		+/+		Rec	II	
21	18865530	18863833	-	21q21.1		+/Δ258-968		Np9	I	
22	17306187	17315361	+	22q11.21		+/+	+	Np9	I	

5'LTR gag pro pol env 3'LTR

\*Die chromosomale Lokalisation der kompletten Loci, incl. LTRs (sofern vorhanden), im hg18 Referenzgenom ist in den ersten fünf Spalten angegeben. Die mittlere Spalte zeigt die graphische Darstellung eines multiplen Sequenz-Alignments der als transkribiert identifizierten HML-2-Loci mit der Sequenz des HERV-K(HML-2.HOM)-Provirus (*Genbank acc. no. AF074086.2*) als Referenz, dessen schematische Struktur unter dem Alignment gezeigt ist. Insertionen von nicht-HML-2 Sequenzen in den Loci auf den Chromosomen 11q12.3 und 19q12 (s.u.) wurden der Übersichtlichkeit halber vor dem Alinieren in den proviralen Sequenzen gelöscht. Alinierbare HERV-Bereiche sind durch graue Rechtecke verdeutlicht, wobei schwarze Bereiche Sequenzunterschiede zur HERV-K(HML-2.HOM)-Referenzsequenz anzeigen. Deletierte Bereiche sind durch horizontale Linien verdeutlicht. Das Vorhandensein (+) oder Fehlen (-) des 5'- und 3'-LTRs ist in der entsprechenden Spalte angezeigt, ebenso wie mögliche Deletionen (Δ) innerhalb der LTRs verglichen mit der bei *RepBase* hinterlegten LTR5-Konsensussequenz. ORFs der retroviralen Proteine Gag, Env und Rec/Np9 wurden vorhergesagt und sind in den entsprechenden Spalten vermerkt. Hierbei ist das Vorkommen eines ORFs, der für ein Volllängen Gag Protein von 666 (± 1) as oder ein Volllängen Env Protein von 699 (± 1) as kodiert (vgl. Text), durch ein „+“ markiert und ein „-“ markiert die Loci, in denen die komplette kodierende Region deletiert ist. In der letzten Spalte findet sich die Zuordnung der Proviren zu Typ I oder II, die sich durch eine 292 bp große Deletion an der Grenze von *pol* zu *env* in Typ I Proviren unterscheiden. Die mit einem „-“ markierten Loci weisen eine Deletion im entscheidenden Bereich auf, was die Zuordnung zum einen oder anderen Typ aufgrund des Kriteriums einer 292 bp Deletion nicht ermöglicht. ORFs für die von Typ I und II Proviren kodierten akzessorischen Proteine Np9 und Rec sind vermerkt (vgl. Mayer *et al.*, 2004 und nicht publizierte Daten).

<sup>a</sup>Der Locus auf Chromosom 3p25.3 wurde nur nach Bestrahlung der Zelllinien SK-Mel-28 und MEWO mit UVB detektiert (vgl. Abschnitt 4.1.7.).

<sup>b</sup>Der 5'-LTR wird zwischen nt 891 und 945 durch eine AluY/AuYb8/AluY Repeat-Sequenz unterbrochen und die interne provirale Sequenz weist eine AluSg-Insertion bei nt 5.717 auf sowie einen (TA)<sub>n</sub>/AluY/(CAG)<sub>n</sub> Repeat und zusätzliche 63 bp von LTR5 nach nt 5.998, die vor dem Alinieren aus den Sequenzen entfernt wurden. Alle Insertionen sind auf dem Gegenstrang annotiert.

<sup>c</sup>Die provirale Sequenz wird bei nt 4.583 von einem AluYa5-Element unterbrochen, das vor dem Alinieren aus der proviralen Sequenz gelöscht wurde.

<sup>d</sup>Die 3'-LTRs der Loci auf Chromosom 7q21.1 und 7q34 sind in *antisense* orientiert.

Die meisten Loci besitzen 5'- und/oder 3'-LTRs sowie interne provirale Sequenzen unterschiedlicher Länge. Fünfzehn der Loci besitzen beide LTRs, wohingegen den restlichen acht Loci mindestens ein LTR fehlt oder die LTRs Deletionen unterschiedlicher Größe

aufweisen. Der interne provirale Bereich ist bei 17 der Loci länger als 7.000 bp und somit mehr oder weniger vollständig im Vergleich zum ~7,5 kb langen internen Bereich des HERV-K(HML-2.HOM)-Provirus (*Genbank acc. no.* AF074086.2). In den restlichen Loci sind die internen Bereiche durch Deletionen unterschiedlicher Größe verkürzt, wobei sich die kürzeste interne Sequenz von 3.900 bp bei *ERVK-15* (7q34) findet.

ORFs für die retroviralen Proteine Gag und Env wurden vorhergesagt und die Loci wurden speziell daraufhin analysiert, ob sie für Vollängen-Proteine kodieren, deren Größen mit den vom HERV-K(HML-2.HOM)-Provirus (*Genbank acc. no.* AF074086.2) kodierten Gag und Env Proteinen vergleichbar sind. Sechs der Loci, die sowohl in Melanom-Proben als auch Melanozyten-Zelllinien als transkribiert identifiziert wurden, besitzen einen ORF, der für ein Gag-Protein der erwarteten Größe von 666 ( $\pm 1$ ) Aminosäuren kodiert (*ERVK-5*, *ERVK-10*, *ERVK-9*, *ERVK-6*, *ERVK-28* und *ERVK-24*). Darunter finden sich Loci mit niedrigen relativen Klonierungsfrequenzen von weniger als 5 % (*ERVK-24*, 22q11.21) sowie Loci mit höheren relativen Klonierungsfrequenzen von bis zu 74,99 % (*ERVK-6*, 7p22.1). Drei der transkribierten Loci besitzen einen ORF, der für ein Env-Protein der erwarteten Größe von 699 ( $\pm 1$ ) Aminosäuren kodiert (*ERVK-6*, *ERVK-21* und *ERVK-28*), wobei diese drei Loci ausschließlich in Melanom-Proben transkribiert waren. *ERVK-21* (12q14.1) und *ERVK-28* (19q12) waren hierbei mit relativ niedrigen relativen Klonierungsfrequenzen von 14,63 % (LKM2) und 1,39 % (RNA3) in je einer Melanom-Probe transkribiert. Der *ERVK-6*-Locus war hingegen in sieben der Melanom-Proben mit niedrigen bis hohen relativen Klonierungsfrequenzen von bis zu 74,99 % (SK-Mel-28) transkribiert. Die für beide Vollängen-Proteine, Gag und Env, kodierenden Loci *ERVK-28* (19q12) und *ERVK-6* (7p22.1), wurden beide nur in den Melanom-Proben als transkribiert identifiziert. Weitere Loci besitzen ORFs, die für hypothetische Proteine kodieren, die stark von den erwarteten Größen abweichen, wobei eine Aussage über die tatsächliche Translation oder biologische Funktion solcher Proteinvarianten schwierig ist und nicht Ziel der vorliegenden Arbeit war.

HML-2-Loci mit ORFs für die akzessorischen retroviralen Proteine Np9 und Rec wurden bereits in früheren Studien untersucht (Mayer *et al.*, 2004, und unveröffentlichte Daten). Np9 wird hierbei von Proviren des Typs I kodiert, die eine 292 bp große Deletion an der Grenze von *pol* zu *env* aufweisen, gegenüber Rec-kodierenden Typ II Proviren, die diesen Bereich besitzen. Die als transkribiert identifizierten Loci wurden mit unveröffentlichten und von Mayer *et al.* (2004) publizierten Daten abgeglichen. Acht der zehn transkribierten Typ I-Loci besitzen einen intakten ORF für ein Np9-Protein und acht der zehn Typ II Proviren weisen einen intakten Rec-ORF auf im Vergleich zur Sequenz publizierter Rec und Np9 Vollängen-Proteine (*Genbank acc. no.*

CAA51305, *UniProt acc. no.* P61577, P61583, P61581, P61582, P61582). Rec- und Np9-kodierende Loci waren sowohl in Melanom-Proben als auch in Melanozyten transkribiert. Bei drei Proviren war die Zuordnung zu Typ I/II aufgrund größerer Deletionen im *pol/env*-Bereich nicht möglich.

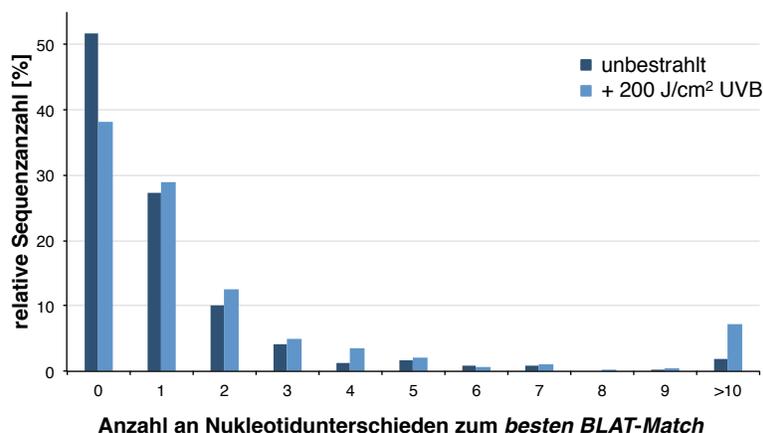
Die meisten HML-2-Loci zeigen somit eine mehr oder weniger vollständige Struktur mit 5'- und 3'-LTRs sowie größtenteils intakten internen Sequenzen. Sechs der Loci weisen ORFs für ein Vollängen Gag-Protein auf und drei Loci kodieren ein Vollängen Env-Protein, wobei die beiden Loci, die sowohl für Gag als auch für Env kodieren, ausschließlich in Melanom-Proben identifiziert wurden. Die jeweils acht Loci, die für Rec- und Np9-Proteine kodieren, waren sowohl in Melanom-Proben als auch Melanozyten-Zelllinien transkribiert.

#### **4.1.7. Einfluss von UVB-Strahlung auf die HERV-K(HML-2) gag-Transkription in Melanom- und Melanozyten-Zelllinien**

In einer früheren Studie von Schanab *et al.* (2011) wurde eine erhöhte HML-2 *pol*-Transkription in Melanom-Zelllinien nach UVB-Bestrahlung beschrieben. Eine solche veränderte Transkription kann in der Veränderung der Transkriptionsstärken spezifischer Loci begründet sein oder in einer generellen Deregulierung der Transkription aller HML-2-Loci. Der Einfluss von UVB-Strahlung auf das HML-2-Transkriptionsmuster wurde in den zur Verfügung stehenden Zelllinien untersucht. Die Melanom-Zelllinien SK-Mel-25, SK-Mel-28- MEWO und WM3734a sowie die Melanozyten-Zelllinien Benno und Oskar wurden mit einer Dosis von 200 mJ/cm<sup>2</sup> UVB bestrahlt, was der Dosis entspricht, die in der Studie von Schanab *et al.* (2011) in der stärksten Deregulierung resultierte. Nach 24 h wurde RNA aus den Zelllinien isoliert und es wurden, wie zuvor beschrieben, cDNA-Sequenzen für das HML-2 *gag*-Amplikon generiert und rückkartiert (zur Amplifikation vgl. **Anhang 2b**).

Insgesamt wurden für alle Zelllinien 528 Sequenzen guter Qualität generiert. 49 dieser Sequenzen (~9,28 %) wurden verworfen, da sie mehr als einen besten *BLAT-Match* aufwiesen und eine eindeutige Rückkartierung somit nicht möglich war. Eine Ausnahme bildete auch hier das HERV-K(HML-2.HOM)-Tandemprovirus, das als ein Locus gezählt wurde (vgl. Abschnitt 4.1.3.). Die restlichen 479 Sequenzen (~90,72 %) konnten eindeutig einem genomischen HML-2-Locus zugeordnet werden, wobei die durchschnittliche Anzahl an Nukleotid-unterschieden zum besten *BLAT-Match* 2,27 betrug. 436 (~91,02 %) der Sequenzen wiesen

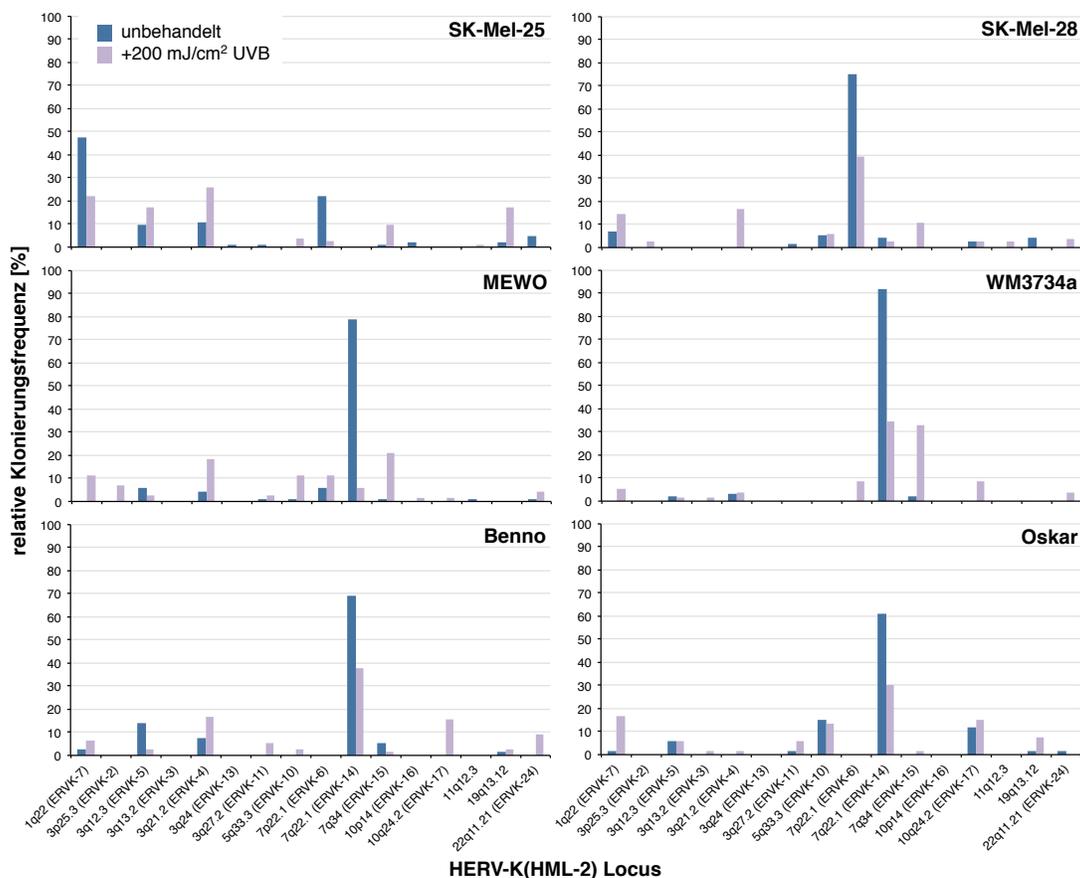
weniger als sechs Nukleotidunterschiede auf und gingen in die weitere Analyse ein (vgl. **Abb. 20**).



**Abbildung 20:** Anzahl an Nukleotidunterschieden der nach UVB-Bestrahlung aus Melanom- und Melanozyten-Zelllinien generierten HERV-K(HML-2) *gag* cDNA-Sequenzen zu ihrem jeweils besten *BLAT-Match*. Die Zellen wurden mit 200 mJ/cm<sup>2</sup> UVB bestrahlt und nach 24 h RNA isoliert und HML-2 *gag*-spezifische RT-PCR-Produkte generiert, sequenziert und rückkartiert (vgl. Text). Gezeigt ist in hellblau die relative Anzahl an eindeutig rückkartierbaren cDNA-Sequenzen nach UVB-Bestrahlung mit einer bestimmten Anzahl an Nukleotidunterschieden zu ihrem besten *BLAT-Match*, die Werte unbehalteter Zellen sind in dunkelblau eingezeichnet. cDNA Sequenzen mit mehr als sechs Nukleotidunterschieden zum besten *BLAT-Match* wurden verworfen.

Die UVB-Bestrahlung resultierte in allen Zelllinien in einem veränderten Transkriptionsprofil (vgl. **Abb. 21**; die relativen Klonierungsfrequenzen finden sich in **Anhang 3b**). Die Gesamtzahl an detektierten transkribierten HML-2-Loci war in allen Zelllinien nach UVB-Bestrahlung leicht erhöht, mit Ausnahme von SK-Mel-25, wo nach Bestrahlung ein Locus weniger detektiert wurde. Im Durchschnitt waren pro Probe zehn Loci transkribiert, verglichen mit sieben Loci in unbehandelten Zellen. Eine spezifische Aktivierung einzelner Loci war nicht zu beobachten. Die relative Klonierungsfrequenzen des *ERVK-14*-Locus (7q22.1), der vor der Bestrahlung Werte von bis zu 92,13 % (WM3734a) aufwies, waren nach Bestrahlung stark reduziert auf Werte von maximal 39,29 % (SK-Mel-28). Der *ERVK-2*-Locus auf Chromosom 3p25.3, der in unbehandelten Zellen nicht als transkribiert identifiziert wurde, wurde nach der UVB-Bestrahlung in den Zelllinien SK-Mel-28 und MEWO als transkribiert detektiert (zur Struktur des Locus vgl. **Tab. 16**). Ein spezifischer Unterschied in der Änderung der Transkriptionsprofile zwischen Melanom- und Melanozyten-Zelllinien war jedoch nicht erkennbar.

Die UVB-Bestrahlung beeinflusste somit die Transkription verschiedener HML-2-Loci, wobei jedoch kein spezifisches Muster erkennbar war, das Melanom- von Melanozyten-Zelllinien unterscheidet.



**Abbildung 21:** Einfluss von UVB-Strahlung auf die relative Transkriptionsfrequenzen von HERV-K(HML-2)-Loci. Die Melanom-Zelllinien SK-Mel-25, SK-Mel-28- MEWO und WM3734a sowie die Melanozyten-Zelllinien Benno und Oskar wurden mit 200 mJ/cm<sup>2</sup> UVB bestrahlt, nach 24 h wurde RNA isoliert und HML-2 *gag*-spezifische RT-PCR-Produkte generiert, sequenziert und im Genom rückkartiert (vgl. Text). Angegeben sind die relativen Klonierungsfrequenzen in [%] der als transkribiert identifizierten HML-2-Loci, berechnet für jede Zelllinie aus der Anzahl an cDNA-Sequenzen, die eindeutig einem bestimmten Locus zugeordnet werden konnten, dividiert durch die Gesamtanzahl an rückkartierbaren Sequenzen für die Zelllinie. Die blauen Balken repräsentieren die relativen Klonierungsfrequenzen in unbehandelten Zellen (vgl. Tab. 15), in violett sind die relativen Frequenzen nach UVB-Bestrahlung gezeigt.

#### 4.1.8. Unterschiedliche HML-2 Transkriptionsprofile sind nicht durch Unterschiede im Allelstatus polymorpher HERV-K(HML-2)-Loci begründet

In früheren Studien wurden mehrere HERV-K(HML-2)-Loci als polymorph beschrieben (Barbulescu *et al.*, 1999; Hughes und Coffin, 2004; Macfarlane und Simmonds 2004; Mayer *et al.*, 2005). Als mögliche Allele existieren hierbei, neben vollständigen Proviren, auch solitäre LTRs, die durch homologe Rekombination der 5'- und 3'-LTRs unter Verlust der dazwischen liegenden proviralen Sequenz entstanden, sowie leere Präintegrationsstellen im Fall von nicht vollständig in der Population fixierten Proviren. In einer Transkriptionsanalyse könnte das Fehlen von Transkripten eines als polymorph beschriebenen Locus somit schlicht auf ein Fehlen des

Provirus in der genomischen DNA der entsprechenden Probe zurückzuführen sein. Die zur Verfügung stehenden DNAs der Melanom- und Melanozyten-Zelllinien wurden daher bzgl. des Allelstatus bekannter polymorpher Loci getestet. Es handelte sich hierbei um die Loci *ERVK-6* (7p22.1, HERV-K(HML-2.HOM)), *ERVK-1* (1p31.1), *ERVK-2* (3p25.3), *ERVK-3* (3q21.2) und *ERVK-9* (6q14.1) sowie die Proviren HERV-K113 (19p12) und HERV-K115 (8p23.1), die beide in der vorangehenden Analyse nicht als transkribiert identifiziert wurden. Der Allelstatus der polymorphen Loci wurde durch spezifische PCRs mit Locus-flankierenden *Primern* überprüft, wobei die Größe des entstehenden PCR-Produktes Aufschluss über das Vorkommen eines Provirus im betreffenden Locus liefert. Das Vorhandensein eines Provirus oder solitären LTRs wurde durch eine zweite PCR mit einem flankierenden *Primer* sowie einem *Primer* im LTR oder Provirus bestätigt.

Mit Ausnahme von HERV-K113 und -K115 (s.u.) konnte für die polymorphen Loci in fast allen Proben ein etwa 10 kb langes Produkt generiert werden, was der Produktgröße im Falle des Vorhandenseins eines Volllängen-Provirus entspricht (vgl. **Tab. 17** und **Anhang 4**). Hierdurch wurde für die untersuchten polymorphen HML-2-Loci jeweils das Vorkommen eines Provirus auf mindestens einem der beiden homologen Chromosomen bestätigt. Die Unterscheidung zwischen einem homozygoten und heterozygoten Zustand bzgl. des Vorhandenseins eines Provirus ist hierbei nicht eindeutig möglich. Es ergaben sich in fast allen Fällen PCR-Produkte, die der Größe eines solitären LTRs entsprechen, was bzgl. des Vorkommens des Provirus auf Heterozygotie hindeuten würde. Das Vorhandensein solitärer LTRs aller polymorpher Loci in nahezu allen untersuchten Proben steht jedoch im Gegensatz zu früher beschriebenen niedrigeren Allelfrequenzen (Hughes und Coffin, 2004) und es handelt sich hierbei vermutlich eher um Produkte, die *in vitro* während der PCR durch Rekombination von LTR-Anteilen von unvollständig synthetisierten PCR-Produkten entstanden sein könnten, zumal noch weitere unspezifische PCR-Produkte im Gel auftraten. Im Gegensatz zu den anderen polymorphen Loci konnte das HERV-K113 Volllängen-Provirus in keiner der Proben detektiert werden und das HERV-K115-Provirus lediglich in der Melanozyten-Zelllinie Benno. Dies deckt sich mit zuvor beschriebenen niedrigen Allelfrequenzen dieser beiden Loci (Burmeister *et al.*, 2004; Macfarlane und Simmonds 2004; Moyes *et al.*, 2005).

**Tabelle 17:** Allelstatus polymorpher HERV-K(HML-2)-Loci in den in der Transkriptionsanalyse untersuchten Melanom- und Melanozyten-Zelllinien.\*

	<i>ERVK-1</i>	<i>ERVK-2</i>	<i>ERVK-3</i>	<i>ERVK-9</i>	<i>ERVK-6</i>	<b>K113</b>	<b>K115</b>
<b>SK-Mel-25</b>	PV	PV	PV	nur sLTR	PV	-	-
<b>SK-Mel-28</b>	PV	PV	nur sLTR	PV	PV	-	-
<b>MEWO</b>	PV	PV	PV	PV	PV	-	-
<b>WM3734a</b>	PV	PV	PV	PV	PV	-	-
<b>Benno</b>	PV	PV	PV	nur sLTR	nur sLTR	nur sLTR	PV
<b>Oskar</b>	nur sLTR	PV	PV	PV	PV	-	-

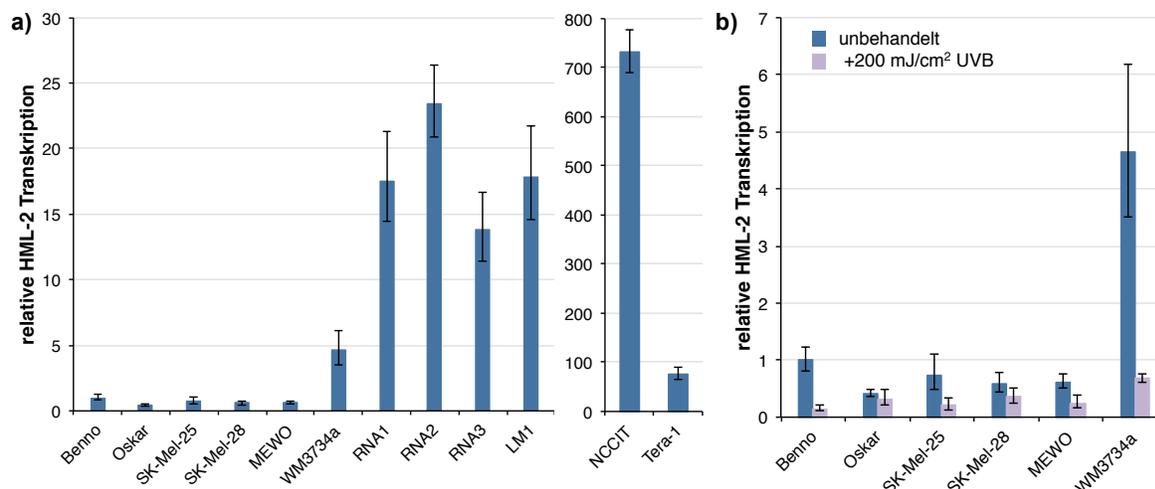
\*Die Melanom-Zelllinien SK-Mel-25, SK-Mel-28, MEWO und WM3734a sowie die Melanozyten-Zelllinien Benno und Oskar wurden bzgl. des Allelstatus bekannter polymorpher Proviren untersucht. Die Proviren wurden mit Locus-flankierenden *Primern* aus DNA der Zelllinien amplifiziert, wobei die Größe des PCR-Produktes Aufschluss über das Vorhandensein eines kompletten Provirus (PV) oder eines solitären LTRs (sLTR) lieferte. Eine Unterscheidung zwischen homo- und heterozygotem Zustand war hierbei nicht eindeutig möglich. Zur weiteren Erklärung, s. Text.

Aufgrund des Vorhandenseins von mindestens einer Provirus-Kopie der meisten polymorphen Loci in fast allen Proben sind zuvor beschriebene Unterschiede in der Transkription somit nicht auf das Fehlen bestimmter Proviren in den spezifischen Proben zurückzuführen.

#### 4.1.9. Eine Quantifizierung mittels *Real-Time-PCR* zeigte erhöhte HERV-K(HML-2) *gag*-Transkriptmengen in einigen Melanom- und Keimzelltumor-Proben

Zur Detektion einer möglicherweise deregulierten HML-2-Expression in Melanomen wurden die HML-2 *gag*-Transkriptmengen in den untersuchten Melanom- und Melanozyten-Proben mittels *Real-Time-PCR* quantifiziert. Zusätzlich wurde die Transkription in den Keimzelltumor (*germ cell tumour*, GCT)-Zelllinien NCCIT und Tera-1 quantifiziert, die in früheren Studien eine relativ starke HML-2-Transkription zeigten (Florl *et al.*, 1999; Ruprecht *et al.*, 2008a). Zur Normalisierung dienten die beiden Haushaltsgene *G6PDH* und *RPII* (Radonić *et al.*, 2004). Als Referenz diente die HML-2 *gag*-Transkriptmenge in der Melanozyten-Zelllinie Benno und die Transkriptmengen in allen weiteren Proben wurden relativ zu dieser berechnet (vgl. **Abb. 22a**).

Die zweite Melanozyten-Zelllinie Oskar sowie die Melanom-Zelllinien SK-Mel-25, SK-Mel-28 und MEWO zeigten mit Benno vergleichbare, niedrige relative HML-2 *gag*-Transkriptmengen. WM3734a, die drei Melanom-Metastasen RNA1-3 und die Lymphknoten-Metastase LKM1 zeigten gegenüber Benno um 5 bis 23-fach erhöhte relative Transkriptmengen. Die höchste Transkriptmenge zeigte sich in den GCT-Zelllinien NCCIT und Tera-1, in denen die relative Transkription 733- bzw. 77-fach höher war als in Benno.



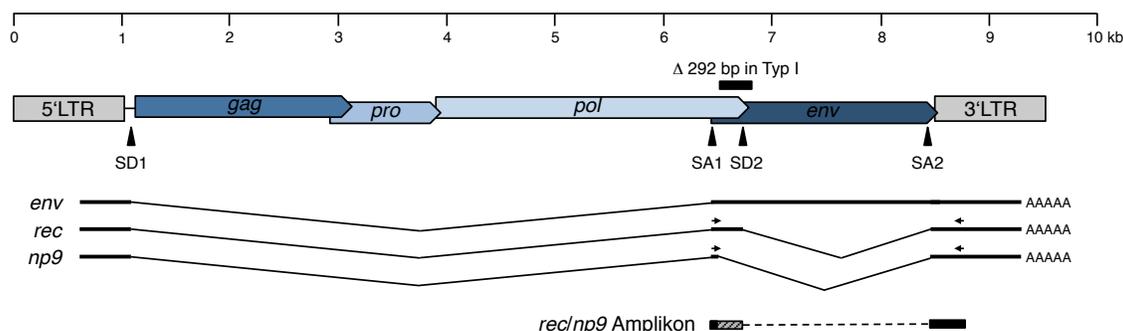
**Abbildung 22:** Quantifizierung der HERV-K(HML-2) *gag*-Transkriptmengen mittels *Real-Time*-PCR. **a)** Relative HML-2 *gag*-Transkriptmengen in den Melanozyten-Zelllinien Benno und Oskar, den Melanom-Zelllinien SK-Mel-25, SK-Mel-28, MEWO, WM3734a, drei Melanom-Proben (RNA1-3), einer Melanom-Lymphknoten-Metastase (LKM1) sowie den GCT-Zelllinien NCCIT und Tera-1 (man beachte die unterschiedliche Skalierung für die GCT-Zelllinien). Zur Normalisierung dienten die Haushaltsgene *G6PDH* und *RPII* und als Referenz die Transkription in der unbehandelten Melanozyten-Zelllinie Benno (vgl. Text). Alle Proben wurden in Triplikaten gemessen, wobei für alle Proben neben den Mittelwerten jeweils auch Minimal- und Maximalwerte durch schwarze Linien dargestellt sind. **b)** Relative HML-2 *gag*-Transkriptmengen 24 h nach Bestrahlung der genannten Zelllinien mit 200 mJ/cm<sup>2</sup> UVB. Unbehandelte Zelllinien sind in blau, bestrahlte Zelllinien in violett gezeigt.

Die relativen HML-2 *gag*-Transkriptmengen in den Melanom- und Melanozyten-Zelllinien wurden zusätzlich 24 h nach Bestrahlung der Zellen mit 200 mJ/cm<sup>2</sup> UVB quantifiziert (vgl. **Abb. 22b** und Abschnitt 4.1.7.). Auch hier diente als Referenz die Transkriptmenge in der unbehandelten Melanozyten-Zelllinie Benno. Es zeigte sich in allen bestrahlten Zelllinien eine verringerte HML-2 *gag*-Transkriptmenge verglichen mit den unbehandelten Zellen. Die relativen Transkriptmengen wurden auf Werte von 77 % (Oskar), 28 % (SK-Mel-25), 60 % (SK-Mel-28) und 40 % (MEWO) der Transkriptmengen in den jeweils unbehandelten Zellen reduziert. Die stärkste Verminderung wurde in der Melanozyten-Zelllinie Benno und der Melanom-Zelllinie WM3734a beobachtet, wo die Transkriptmenge nach Bestrahlung jeweils auf etwa 15 % des Wertes der unbehandelten Zellen reduziert wurde.

Fünf der unbehandelten Melanom-Proben sowie beide GCT-Zelllinien zeigten somit gegenüber der Melanozyten-Zelllinie Benno erhöhte relative HML-2 *gag*-Transkriptmengen. Eine Bestrahlung mit UVB resultierte in allen untersuchten Melanom- und Melanozyten-Zelllinien in verringerten relativen HML-2 *gag*-Transkriptmengen.

#### 4.1.10. Identifizierung *rec* und *np9* mRNA-produzierender HML-2-Loci in Melanomen und Melanozyten-Zelllinien

Frühere Studien diskutierten eine potentielle Rolle der akzessorischen HML-2-Proteine Rec und Np9 in der Tumorphagenese (Armbruster *et al.*, 2002, 2004; Pece *et al.*, 2004; Denne *et al.*, 2007). Analog der bereits beschriebenen Identifizierung transkribierter HML-2-Loci durch Verwendung eines Amplikons im HML-2 *gag*-ORF wurden im Weiteren speziell solche transkribierten HML-2-Loci in cDNA identifiziert, die *rec*- oder *np9*-mRNA produzieren. Hierfür wurde ein Amplikon gewählt, das die in *rec*- und *np9*-mRNA gespleißten Bereiche berücksichtigt (Löwer *et al.*, 1995; Armbruster *et al.*, 2004). Analog dem HML-2 *gag*-Amplikon wurden auch die *Primer* zur Amplifikation des *rec/np9*-Amplikons gegenüber denen einer früheren Studie optimiert (Mayer *et al.*, 2004) und zusätzliche *Primer*-Varianten eingesetzt, die Nukleotid-unterschiede in den *Primer*-bindenden Regionen der verschiedenen HML-2-Loci berücksichtigen. Der *Forward-Primer* lag, wie bereits in einer früheren Studie von Armbruster *et al.* (2004), mit seinem 5'-Ende auf dem ATG-Start-Codon des *rec*- und *np9*-ORFs, der *Reverse-Primer* war etwa 180 bp stromab des Stopp-Codons der *rec*- und *np9*-ORFs im 3'-LTR lokalisiert. Die Größe der RT-PCR-Produkte gespleißter Transkripte betrug ~580 bp für *rec* mRNA-produzierende Loci und ~370 bp für *np9* mRNA-produzierende Loci (vgl. **Abb. 23**).



**Abbildung 23:** Position des untersuchten Amplikons zur Identifizierung transkribierter *rec/np9* mRNA-produzierender HERV-K(HML-2)-Loci. Gezeigt ist die Struktur des HERV-K(HML-2.HOM)-Provirus (*Genbank acc. no.* AF074086.2) mit den retroviralen ORFs für *gag*, *pro*, *pol* und *env*. Unter dem Provirus sind die unterschiedlichen Spleißvarianten für *env*-, *rec*- und *np9*-mRNA verdeutlicht (SD = *splice donor site*, SA = *splice acceptor site*), wobei sowohl *rec*- als auch *np9*-mRNA gegenüber *env*-mRNA durch einen weiteren Spleißprozess im *env*-ORF gekennzeichnet ist. Bei HERV-K(HML-2.HOM) handelt es sich um ein Typ II-Provirus, das einen 292 bp langen Sequenzabschnitt an der *pol/env*-Grenze besitzt, der Typ I Proviren fehlt (vgl. schwarzen Balken über der Provirus-Struktur). In diesem Typ I Deletionsbereich liegt die *splice donor site 2* (SD2), die für das Spleißen der *rec*-mRNA in Typ II Proviren genutzt wird. Typ I Proviren nutzen eine alternative *splice donor site* unmittelbar stromauf der Deletionsregion. Die Positionen der zur Amplifikation verwendeten PCR-*Primer* sind durch Pfeile gekennzeichnet. Das 5'-Ende des *Forward-Primers* lag hierbei auf dem ATG-Start-Codon von *rec/np9*, der *Reverse-Primer* lag ~180 bp stromab des *rec/np9* Stopp-Codons im 3'-LTR.

Das *rec/np9*-Amplikon wurde, analog dem *gag*-Amplikon, aus cDNA der Melanom-Zelllinien SK-Mel-25, SK-Mel-28, MEWO und WM3734a sowie aus den Melanozyten-Zelllinien Benno und Oskar amplifiziert (vgl. **Anhang 5**). Die ~580 bp bzw. ~370 bp großen PCR-Produkte sowie ggf. weitere auftretende Nebenprodukte wurden aus dem Gel ausgeschnitten und zusammen aufgereinigt, in *pGEM<sup>®</sup>-T Easy* kloniert, in *E. coli* DH5a transformiert, *Insert*-tragende Plasmide aufgereinigt, mittels Sanger-Methode sequenziert und die Sequenzen im Genom rückkartiert.

Für jede Zelllinie wurden 16 Sequenzen generiert und insgesamt 91 Sequenzen guter Qualität weiterverwendet. 84 dieser Sequenzen konnten eindeutig im Genom rückkartiert werden und gingen in die weitere Analyse ein. Es wurden drei Typ I- und vier Typ II Proviren als transkribiert identifiziert (vgl. **Tab. 18**). Die meisten der generierten cDNA-Sequenzen konnten dem *rec* mRNA-produzierenden *ERVK-6*-Locus auf Chromosom 7p22.1 zugeordnet werden. *ERVK-6*-Transkripte wurde hierbei sowohl in Melanom- als auch in Melanozyten-Zelllinien identifiziert, wohingegen mit dem zuvor beschriebenen *gag*-Amplikon *ERVK-6*-Transkripte ausschließlich in cDNAs der Melanom-Proben identifiziert wurden (vgl. hierzu auch **Tab. 15**). Auch die *np9* mRNA-produzierenden Loci *ERVK-7* (1q22), *ERVK-5* (3q12.3) und *ERVK-24* (22q11.21) waren sowohl in den Melanom- als auch in den Melanozyten-Zelllinien transkribiert.

**Tabelle 18:** In Melanom- und Melanozyten-Zelllinien als transkribiert identifizierte *rec* oder *np9* mRNA-produzierende HERV-K(HML-2)-Loci.\*

Typ	HML-2-Locus					Melanom-Zelllinien				Melanozyten	
	Name	Chr	Start	Ende	Bande	SK-Mel-25	SK-Mel-28	MEWO	WM3734a	Benno	Oskar
<i>np9</i>	<i>ERVK-7</i>	1	153863081	153872260	1q22				1	3	
	<i>ERVK-5</i>	3	102893427	102902549	3q12.3	1		1		2	
	<i>ERVK-24</i>	22	17306187	17315361	22q11.21			1		1	
<i>rec</i>	-	2	187093878	187095344	2q32.1					1	
	<i>ERVK-4</i>	3	127091992	127101129	3q21.2				2		
	<i>ERVK-6</i>	7	4588583	4606557	7p22.1	16	14	13	13	1	12
	<i>ERVK-17</i>	10	101570559	101577735	10q24.2					2	

\*Es wurde mit einem in der *rec/np9*-Region lokalisierten spezifischen Amplikon die Transkription von Typ I und II HML-2-Proviren untersucht, die retrovirale *np9*- oder *rec*-mRNA produzieren. Das Amplikon wurde aus cDNA der Melanom-Zelllinien SK-Mel-25, SK-Mel-28, MEWO und WM3734a sowie der Melanozyten-Zelllinien Benno und Oskar amplifiziert, cDNA-Sequenzen mittels Sanger-Methode generiert und im menschlichen hg18 Referenzgenom rückkartiert (vgl. Text). Angegeben sind die Chromosomenpositionen der transkribierten HML-2-Loci, deren offizielle HGNC-Namen (vgl. Mayer *et al.*, 2011) sowie für jede Zelllinie die absolute Anzahl an generierten Sequenzen, die eindeutig den verschiedenen Loci zugeordnet werden konnten.

Vergleicht man die Aminosäuresequenzen der von den identifizierten Transkripten kodierten Proteine mit bei *UniProt* hinterlegten Sequenzen für Rec- und Np9-Proteine (*acc. no.* Q69383, P61571-61576, P61578-61583), zeigen sich nur wenige Sequenzunterschiede. Eine aus SK-Mel-28 cDNA generierte *rec* cDNA-Sequenz, die auf Chromosom 7p22.1 (*ERVK-6*) rückkartiert wurde, zeigte eine 1 bp Deletion im 5'-Bereich des *rec*-ORFs, was in einer

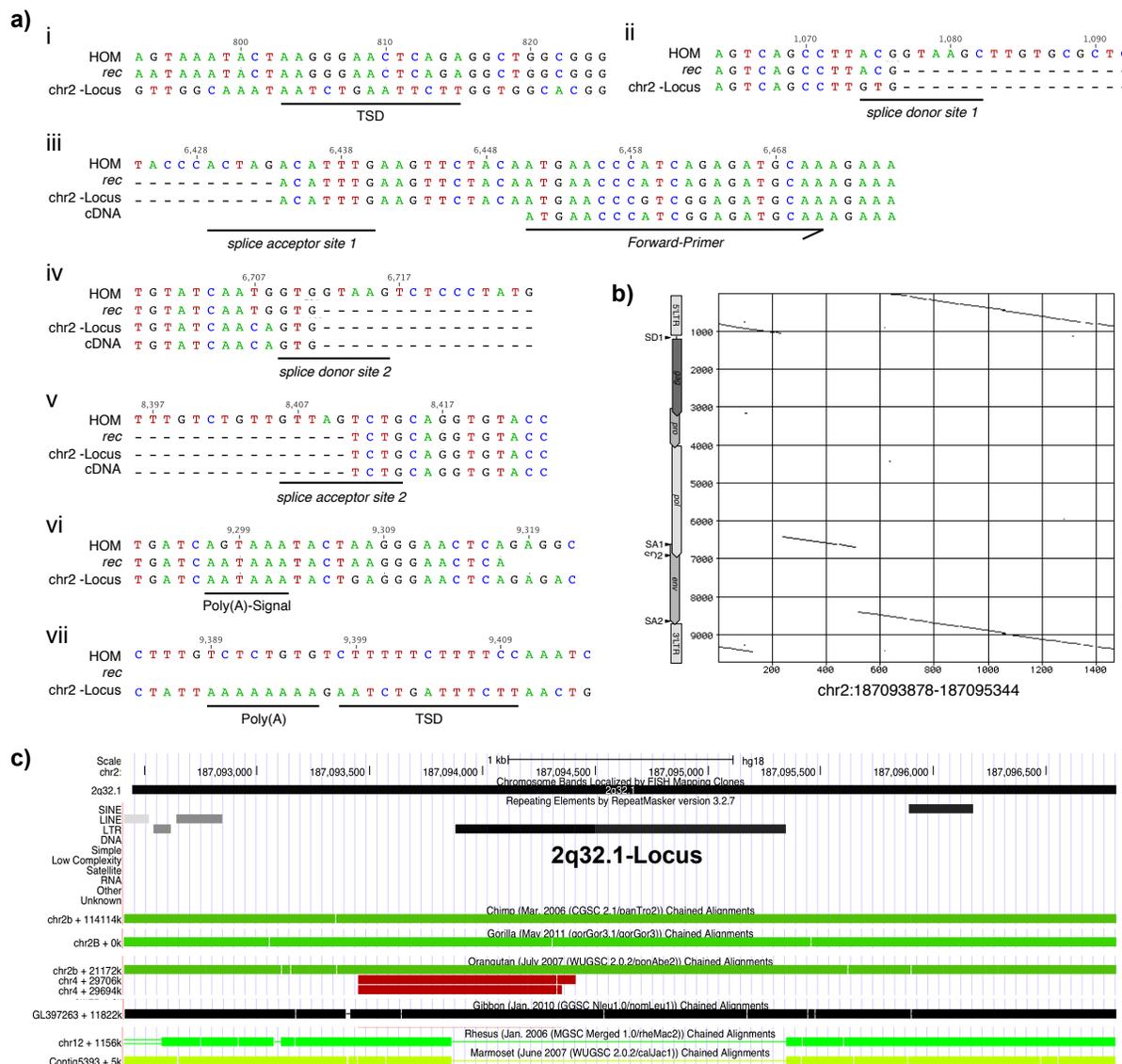
Verschiebung des Leserahmens resultiert. Zwei weitere Sequenzen, die von WM3734a cDNA amplifiziert und auf Chromosom 7p22.1 rückkartiert wurden, wiesen *Nonsense*-Mutationen auf, die in einem verfrühten Stopp-Codon resultieren. Bei den beschriebenen Sequenzen handelt es sich jedoch vermutlich um Mutanten, die *in vitro* während der cDNA-Synthese und/oder PCR entstanden.

Insgesamt wurden somit drei *rec* und vier *np9* mRNA-produzierende HML-2-Loci als transkribiert identifiziert, wobei sowohl in Melanom- als auch Melanozyten-Zelllinien kodierende Loci für beide Proteine transkribiert waren. Zwei der in der Melanozyten-Zelllinie Benno identifizierten Loci wurden aufgrund ihrer besonderen Struktur im Folgenden näher charakterisiert.

#### 4.1.11. Charakterisierung der Struktur zweier durch reverse Transkription und Reintegration gespleißter mRNA entstandener HML-2-Loci

Im Zuge der im vorangehenden Abschnitt beschriebenen Identifizierung transkribierter *rec* oder *np9* mRNA-produzierender HML-2-Loci wurden zwei verkürzte Typ II-Loci auf den Chromosomen 2q32.1 und 10q24.2 in der Melanozyten-Zelllinie Benno als transkribiert identifiziert.

Der stark verkürzte Locus auf Chromosom 2q32.1 besitzt eine interne Sequenz von weniger als 500 bp, die sich aus drei proviralen Bereichen zusammensetzt (vgl. **Abb. 24a, b**). Es handelt sich hierbei um einen etwa 100 bp langen Bereich stromauf der *env splice donor (SD) site* unmittelbar stromab des 5'-LTRs, einen ~280 bp großen Bereich zwischen der *env splice acceptor (SA) site* und der *rec SD site* sowie einen etwa 95 bp großen Bereich stromab der *rec/np9 SA site*. Alle *splice sites* liegen in unvollständiger Form vor, vergleichbar mit mRNA nach dem Spleißprozess. Auch die LTRs liegen in stark verkürzter Form vor. Vom 5'-LTR sind nur etwa 90 bp des 3'-Endes vorhanden und dem 3'-LTR fehlen ~150 bp an dessen 3'-Ende. Eine genauere Betrachtung der Sequenz deutet darauf hin, dass es sich bei dem Locus um mRNA handelt, die nach dem Muster von *rec*-mRNA gespleißt, revers transkribiert und im Genom reintegriert wurde. Zwei charakteristische Sequenzbereiche deuten hierbei auf eine reverse Transkription durch die L1-Maschinerie hin. Etwa 90 bp stromab des Poly(A)-Signals des 3'-LTRs wird der Locus von acht Adenosin-Nukleotiden flankiert, Zudem findet sich zu beiden Seiten des Locus die *target site duplication (TSD)* AATCTGAATTCTT (5') bzw. AATCTGAITTCTT (3'), die der *target site* Konsensussequenz von L1 ähnelt (Jurka, 1997; Ostertag und Kazazian, 2001).



**Abbildung 24:** Struktur eines durch L1-Retrotransposition entstandenen HERV-K(HML-2)-Locus auf Chromosom 2q32.1., der in der Melanozyten-Zelllinie Benno als transkribiert identifiziert wurde. **a)** Sequenzvergleich der Teilbereiche des genomischen 2q32.1-Locus mit einer diesem Locus zugeordneten cDNA, einer publizierten *rec* mRNA-Sequenz (*Genbank acc. no.* X72790) sowie der Sequenz relevanter Bereiche des HERV-K(HML-2.HOM)-Provirus (HOM, *Genbank acc. no.* AF074086.2). Die Nummern über dem *Alignment* bezeichnen die Position in HERV-K(HML-2.HOM). Der Locus repräsentiert eine gespleißte und anschließend durch die L1-Maschinerie retrotransponierte mRNA, wobei die *splice donor* (SD) und *splice acceptor* (SA) sites (ii-v) denen von *rec*-mRNA entsprechen. Der Locus wird von *target site duplications* (TSDs) flankiert (i, vii) und weist an seinem 3'-Ende ein Poly(A)-Signal (vi) auf. Die Position des zur Amplifikation verwendeten *Forward-Primers* ist eingezeichnet (iii). **b)** *Dot-Matrix*-Vergleich (*window size*: 30, *minimum score*: 50 %, *jump*: 1) des Locus auf Chromosom 2q32.1 mit der HERV-K(HML-2.HOM)-Sequenz. Die Positionen der LTRs, der *gag*-, *pro*-, *pol*- und *env*-ORFs sowie der SD und SA sites der Vergleichssequenz sind neben der y-Achse eingezeichnet. **c)** Der retrotransponierte 2q32.1-Locus im „RepeatMasker v.3.2.7 Annotation“-Track des UCSC Genome Browser (zusammengesetzt aus Annotationen in hg18 und hg19). Der 2q32.1-Locus ist im oberen Bereich durch einen schwarzen Balken in der endogene Retroviren enthaltenden Annotation von „LTR“-Elementen angezeigt. Im unteren Bereich finden sich in grün und schwarz homologe Sequenzbereiche in anderen Primatengenomen. Eine homologe provirale Sequenz ist in Schimpanse, Gorilla, Orang-Utan und Gibbon vorhanden, fehlt jedoch im Rhesusaffen und in Marmosetten. Zur weiteren Erklärung, s.Text.

Im „Retroposed Genes V5, Including Pseudogenes“-Track des UCSC Genome Browsers (hg19 Assembly; vgl. Pei et al., 2012) ist der Locus nicht als prozessiertes Pseudogen annotiert. Vergleiche zu anderen Primatengenomen deuten darauf hin, dass das Provirus vermutlich vor der evolutiven Aufspaltung der Hominoidea generiert wurde (vgl. Abb. 24c). Homologe, also eine sehr ähnliche HML-2-Sequenz enthaltende Bereiche sind auch in den Genomen von Schimpanse, Gorilla, Orang-Utan und Gibbon vorhanden, nicht jedoch im Rhesusaffen und Marmosetten, wobei letztere zur Gruppe der Neuweltaffen gehören, deren Genome generell keine HERV-K(HML-2)-homologen Sequenzen aufweisen (Mayer und Meese, 2005).

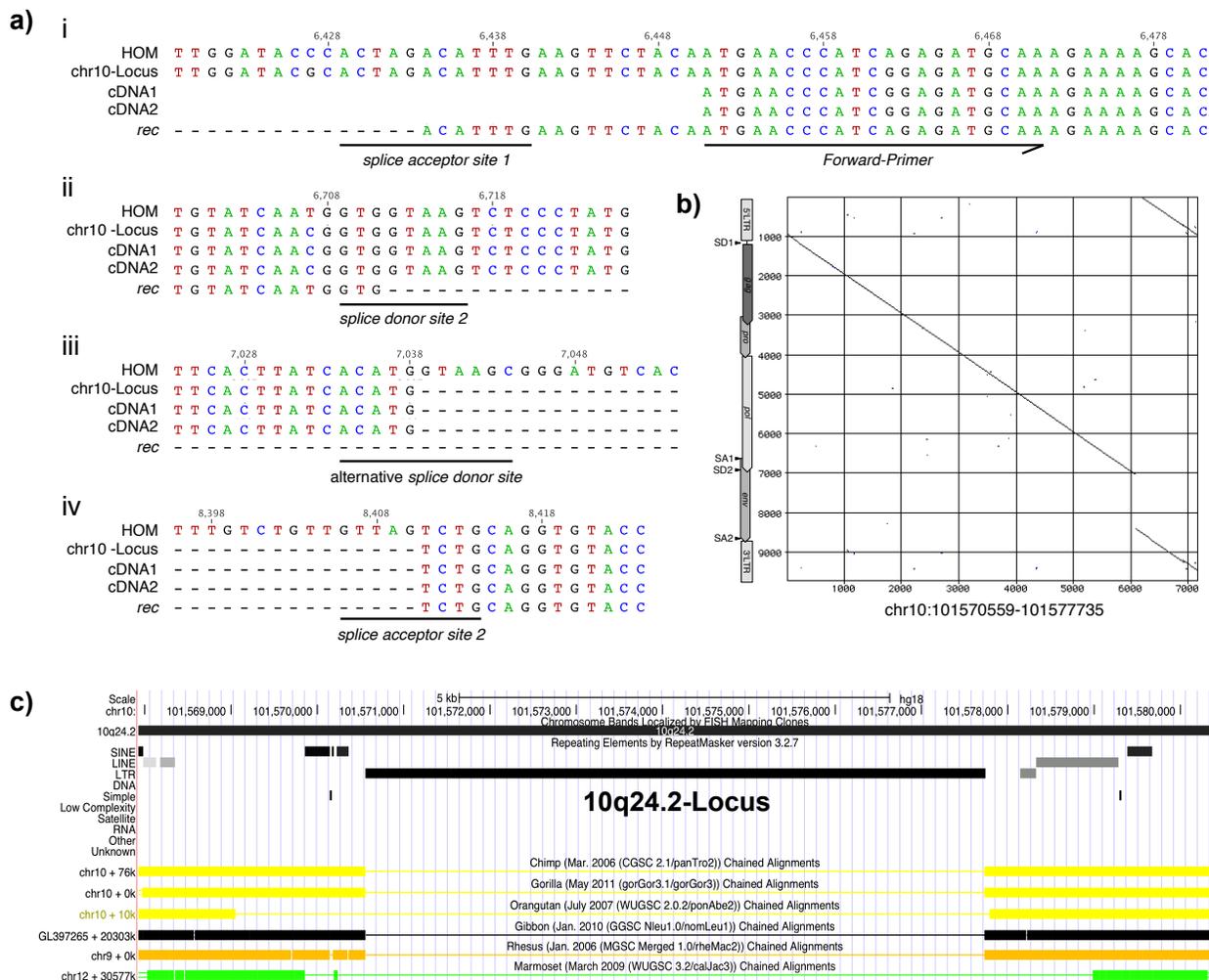
Für den 2q32.1-Locus wurde ein potentieller ORF von 408 bp vorhergesagt, der für ein Protein von 135 Aminosäuren kodiert. Die 60 N-terminalen Aminosäuren des hypothetischen Proteins zeigen eine 90 %-ige Sequenzidentität (93,33 % Ähnlichkeit) zu beschriebenen Rec- und Env-Proteinsequenzen (*UniProt acc. no.* Q69383 und Q69384), die wiederum in ihren 85 N-terminalen Aminosäuren identisch sind. In den zentralen und C-terminalen 75 Aminosäuren besteht keine signifikante Ähnlichkeit zu Rec, Env (vgl. Abb. 25) oder anderen bekannten nicht-HML-2-Proteinen .

Rec	MNPSEMQRKAPRRRRHRNRAPLTHKMNKMTVEEQMKLPSTKKAEPPTWAQLKLTQLA	60	Rec	-----	
10q24.2	MNPSEMQRKAPRRRRHHSRAPLTHKMNKMTVEEQMKLPSTKKAEPPTWAQLKLTQLA	60	10q24.2	-----	
Env	MNPSEMQRKAPRRRRHRNRAPLTHKMNKMTVEEQMKLPSTKKAEPPTWAQLKLTQLA	60	Env	TRDRKPFYTTDLNSSLTVPLQSCVKPPYMLVVGNIIVKPDSTITCENCRLLTCDSTFN	419
2q32.1	MNPSEMQRKAPRRRRHRSRAPSAAHNMNMVSEEQMKLPSTKKAEPPTWAQLKLTQLA	60	2q32.1	-----	
	*****.**:***:*****:*****				
Rec	TKYLENTK-VTQTPESMLLAALMIVS-----PMPAGAAAANYTYWAYVFPPLIRAVTWM	85	Rec	-----	
10q24.2	TKYLENTK-VTQTPESMLLAALMIVSTVVSLEPMPAGAAAANYTYWAYVFPPLIRAVTWM	119	10q24.2	-----	
Env	TKYLENTK-VTQTPESMLLAALMIVSHVVSLEPMPAGAAAANYTYWAYVFPPLIRAVTWM	119	Env	WQHRILLVRAREGVWIPVSMDRPWEASPSVHILTEVLKGVLNRSKRFIFTLIAIVMGLIA	479
2q32.1	KKKPREHKGDTNSREHAACSFECDIN-----	86	2q32.1	-----	
	. * . : * * * : * : .				
Rec	DNPTEVYVNDVSVVWVPGPTDDRCPAKPEEGBMINSIGYRYPICLGRAPGCLMPAVQNW	179	Rec	-----	
10q24.2	DNPTEVYVNDVSVVWVPGPTDDRCPAKPEEGBMINSIGYRYPICLGRAPGCLMPAVQNW	179	10q24.2	-----	
Env	DNPTEVYVNDVSVVWVPGPIDDRCPAKPEEGBMINSIGYHYPPICLGRAPGCLMPAVQNW	179	Env	VTATAAVAGVALHSSVQSVNFVNDWQKNSTRLWNSQSSIDQKLANQINDLRQTVIWMGDR	539
2q32.1	-----		2q32.1	-----	
Rec	-----MVSAGV-----PNSN---	95	Rec	-----	
10q24.2	LVEVPTVSPISRFTYHMSAGV-----PNSN---	204	10q24.2	-----	
Env	LVEVPTVSPICRFTYHMSGSLRPRVNYLQDFSYQSLKFRPKGKPCKEIPKESKNT	239	Env	LMSLEHRFQLOCDWNTSDFCITPQIYNESEHHWDMVRRHLQGREDNLTLDISKLEQIFE	599
2q32.1	-----SVCRCQTQQLRRER-----PSRTS---	104	2q32.1	-----	
	: * . :				
Rec	---EETAT-----IENGP-----	105	Rec	-----	
10q24.2	---EETAT-----IENGP-----	214	10q24.2	-----	
Env	VLVWEECVANSAVILQNEFGTIIIDWAPRGQFYHNCSGQTQSCPSAQVSPAVSDLTESL	299	Env	ASKAHLNLPVGTATAGVADGLANLNPVTWVKTIGSTTIINLILLVCLFCLLLVCRCTQ	659
2q32.1	---HDDNGG-----FVEKKGEMWGE-----RDIRLLLCL	131	2q32.1	-----	
	: : :				
Rec	-----		Rec	-----	
10q24.2	-----		10q24.2	-----	
Env	DKHKHKLQSFYPEWEGEKISTPRPKIVSPVSGPEHPELWRLTVASHHIRIWSGNQTL	359	Env	QLRRSDHRERAMTMAVLSKRKGGVNGKSKRDQIVTVSV	699
2q32.1	CRK-----	135	2q32.1	-----	

**Abbildung 25:** Multiples Alignment von Aminosäuresequenzen der hypothetischen Proteine, die von den beiden durch reverse Transkription und Reintegration gespleißter mRNA entstandenen HML-2-Loci auf den Chromosomen 2q32.1 und 10q24.2 kodiert werden. Die Proteinsequenzen wurden mit denen von Rec- und Env-Proteinen (*UniProt acc. no.* Q69383 und Q69384) aliniert. Im Falle des hypothetischen Proteins des durch L1-Retrotransposition entstandenen Locus auf Chromosom 2q32.1 zeigen die 60 N-terminalen Aminosäuren eine signifikante Ähnlichkeit zu Env und Rec. Das hypothetische Protein des durch einen retroviralen Mechanismus entstandenen Locus auf Chromosom 10q24.2 ist in seinen ersten 200 Aminosäuren identisch zum Env-Protein, das wiederum in den ersten 85 Aminosäuren mit Rec übereinstimmt. Die letzten 14 Aminosäuren des vom Chromosom 10q24.2-Locus kodierten Proteins stimmen mit Rec überein. Zur weiteren Erklärung, s. Text.

In der Melanozyten-Zelllinie Benno wurde ein zweiter Typ II Locus auf Chromosom 10q24.2 als transkribiert identifiziert, der durch eine 1.372 bp große Deletion im *env*-ORF gekennzeichnet ist (vgl. Abb. **26a, b**). Dem Locus fehlt ein Intron-Bereich im *env*-ORF, wobei der 3'-Bereich mit der Sequenz gespleißter *rec*-mRNA übereinstimmt, der Intron-Bereich der *rec splice acceptor site* somit fehlt. Das 5'-Ende des fehlenden *env*-Bereiches entspricht von seiner Struktur her einer *splice donor site*, der der Intron-Bereich fehlt. Die reguläre *rec SD site* etwa 330 bp stromab ist vollständig vorhanden. Der Locus repräsentiert somit eine weitere eventuell partiell gespleißte Variante eines HML-2-Transkriptes, wobei im Gegensatz zum regulären *rec*-Spleißprozess eine alternative *splice donor site* weiter stromab genutzt wurde und die mRNA nach einem retroviralen Mechanismus in cDNA revers transkribiert und reintegriert wurde. Vom 5'-LTR sind lediglich 27 bp des 3'-Endes vorhanden, der 3'-LTR ist hingegen mit 969 bp komplett. TSDs sind nicht vorhanden. Es handelt sich wahrscheinlich um ein evolutiv relativ junges Provirus, da homologe retrovirale Sequenzen in anderen Primatengenomen nicht vorhanden sind (vgl. **Abb. 26c**). Der Locus entstand somit vermutlich erst nach der Spaltung der evolutiven Linien von Mensch und Schimpanse vor etwa 6 Mio. Jahren.

Für den Locus lässt sich ein hypothetisches 214 Aminosäuren großes Protein vorhersagen. Die N-terminalen 196 Aminosäuren entsprechen hierbei bis auf vier Aminosäureunterschiede einem HML-2 Env-Protein (*UniProt acc. no. Q69384*), das in seinen N-terminalen 85 Aminosäuren wiederum zu HML-2 Rec-Protein (*UniProt acc. no. Q69383*) identisch ist. Die C-terminalen 18 Aminosäuren sind ebenfalls identisch zu Rec. Bei dem kodierten hypothetischen Protein handelt es sich somit um ein chimäres Protein, das in Teilen Rec und Env ähnelt (vgl. **Abb. 25**).

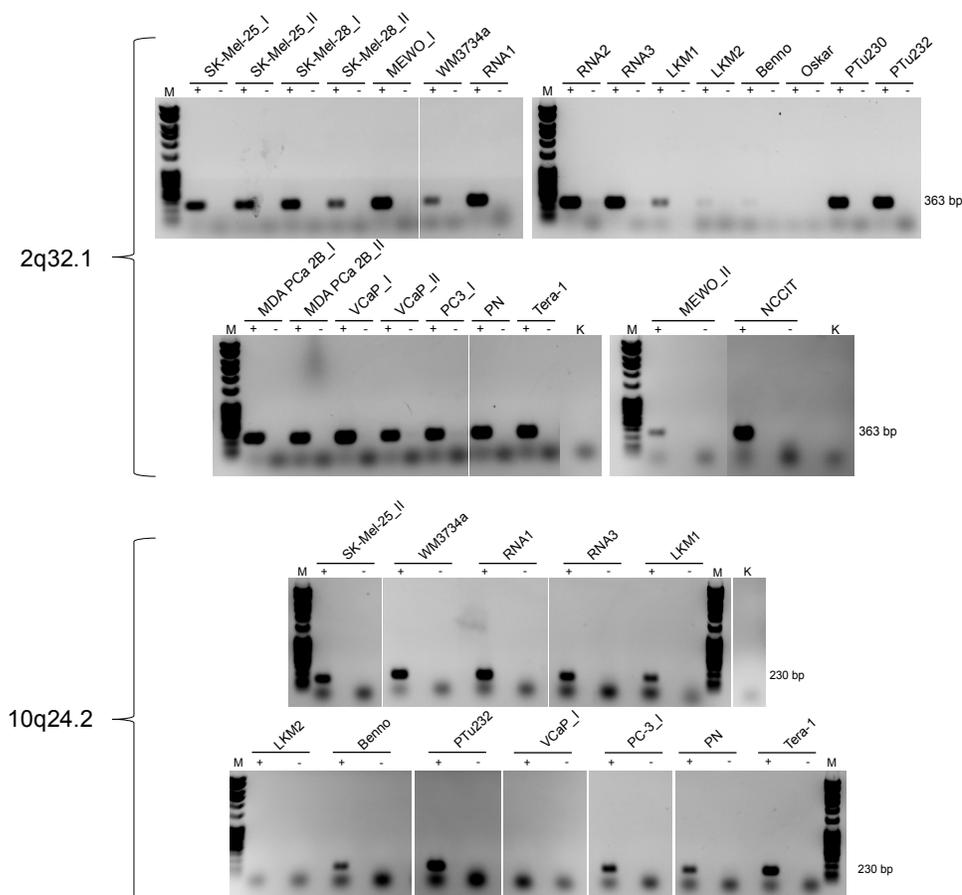


**Abbildung 26:** Struktur eines durch reverse Transkription und Reintegration gespleißter mRNA entstandenen HML-2-Locus auf Chromosom 10q24.2, der in der Melanozyten-Zelllinie Benno als transkribiert identifiziert wurde. **a)** Sequenzvergleich des genomischen 10q24.2-Locus mit den beiden dem Locus zugeordneten generierten cDNA-Sequenzen, einer *rec* mRNA-Sequenz (*Genbank acc. no.* X72790) sowie der Sequenz des HERV-K(HML-2.HOM)-Provirus (HOM, *Genbank acc. no.* AF074086.2). Die Nummern über dem *Alignment* beziehen sich auf die Position in der HERV-K(HML-2.HOM)-Sequenz. Der Locus repräsentiert vermutlich eine gespleißte mRNA, die durch einen retroviralen Mechanismus revers transkribiert und reintegriert wurde. Die Position des zur Amplifikation verwendeten *Forward-Primers* ist eingezeichnet (i). Dem Locus fehlt, ähnlich der *rec*-mRNA, ein Intron-Bereich im *env*-ORF. Die *splice acceptor site* stimmt hierbei mit der von *rec* überein (iv). Das 5'-Ende des fehlenden *env*-Introns unterscheidet sich jedoch von der *rec* mRNA-Struktur, da hier eine alternative *splice donor site* vorhanden ist, der der Intron-Bereich fehlt (iii). Die reguläre *rec SD site* etwa 330 bp stromauf ist vollständig vorhanden (ii). **b)** *Dot-Matrix*-Vergleich des 10q24.2-Locus mit der HERV-K(HML-2.HOM)-Sequenz (vgl. Text zu **Abb. 25b**). **c)** „RepeatMasker v.3.2.7 Annotation“-Track des 10q24.2-Locus im UCSC *Genome Browser* (zusammengesetzt aus Annotationen in hg18 und hg19). Der 10q24.2-Locus ist im oberen Bereich durch einen schwarzen Balken in der endogene Retroviren enthaltenden Annotation von „LTR“-Elementen angezeigt. Im unteren Bereich findet sich in gelb, schwarz, orange und grün die graphische Darstellung eines Sequenzvergleiches zu anderen Primatengenomen, in denen die flankierenden Sequenzen zwar vorhanden sind, die homologe HML-2-Sequenz selbst jedoch fehlt. Zur weiteren Erklärung, s. Text.

#### 4.1.12. Transkriptionsanalyse der durch reverse Transkription und Reintegration gespleißter mRNA entstandenen HML-2-Loci auf den Chromosomen 2q32.1 und 10q24.2

Alle zur Verfügung stehenden cDNA-Proben wurden in einem sensitiven PCR-Screening hinsichtlich der Transkription der beiden durch reverse Transkription und Reintegration gespleißter mRNA entstandenen HML-2-Loci auf den Chromosomen 2q32.1 und 10q24.2 analysiert. Hierfür wurden Locus-spezifische *Primer* verwendet, die ein 363 bp (2q32.1) bzw. 230 bp (10q24.2) großes Produkt generieren. Die Spezifität der *Primer* wurde zuvor durch Amplifikation der PCR-Produkte von genomischer DNA getestet, wobei alle 16 bzw. 15 von den klonierten PCR-Produkten generierte DNA-Sequenzen eindeutig den HML-2-Loci auf den Chromosomen 2q32.1 und 10q24.2 zugeordnet werden konnten.

Es konnten spezifische Transkripte beider HML-2-Loci in nahezu allen zur Verfügung stehenden cDNA-Proben generiert werden (vgl. **Abb. 27**). Beide Loci sind in zahlreichen Melanom-, Melanozyten- sowie Keimzelltumor- und Prostatakarzinom-Proben, zumindest schwach, transkribiert. Es wäre somit also denkbar, dass auch die potentiell von den Loci kodierten Proteine in verschiedenen Zellen exprimiert werden und dort u.U. eine biologische Funktion erfüllen.



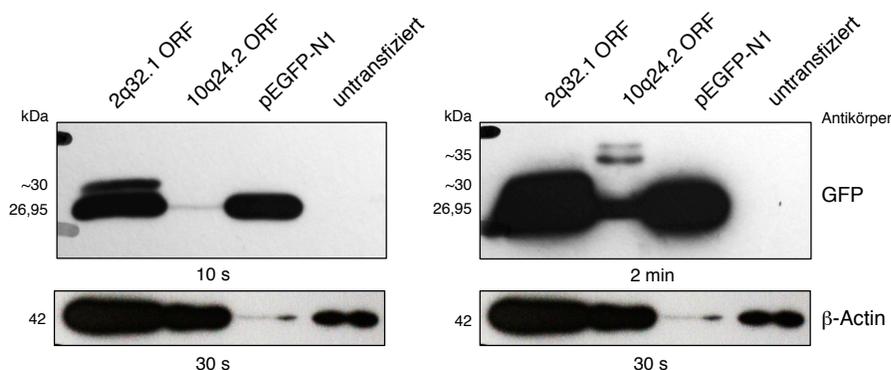
**Abbildung 27:** Ergebnis der gelelektrophoretischen Auftrennung spezifischer RT-PCR-Produkte zweier durch reverse Transkription und Reintegration gespleißter mRNA gebildeter HML-2-Loci auf den Chromosomen 2q32.1 und 10q24.2. Transkripte beider Loci wurden mit Locus-spezifischen *Primern* aus cDNA der Melanom-Zelllinien SK-Mel-25, SK-Mel-28, MEWO, WM3734a, drei Melanomen (RNA1-3), zwei Lymphknoten-Metastasen von Melanom-Patienten (LKM1, 2), den Melanozyten-Zelllinien Benno und Oskar, den Prostatakarzinom-Zelllinien PTu230, PTu232, MDA PCa 2B, VCaP, PC-3 und PN sowie den GCT-Zelllinien Tera-1 und NCCIT amplifiziert. Für einige Zelllinien stand RNA bzw. cDNA aus zwei unterschiedlichen Passagen der jeweiligen Zelllinien zur Verfügung, die mit I bzw. II gekennzeichnet sind. Im Falle des Vorhandenseins von cDNA-Sequenzen der Loci ergaben sich in den Proben spezifische PCR-Produkte von 363 bp (2q32.1) bzw. 230 bp (10q24.2). Zur weiteren Erklärung, s.Text.

#### 4.1.13. Nachweis von GFP-Fusionsproteinen der von den HML-2-Loci auf den Chromosomen 2q32.1 und 10q24.2 kodierten Proteine in transient transfizierten Zellen

Aufgrund der im vorangehenden Abschnitt beschriebenen Transkription der beiden durch reverse Transkription und Reintegration gespleißter mRNA entstanden HML-2-Loci auf den Chromosomen 2q32.1 und 10q24.2 in nahezu allen daraufhin untersuchten Proben wurde analysiert, ob die beiden von diesen Loci kodierten hypothetischen Proteinen tatsächlich translatiert werden können.

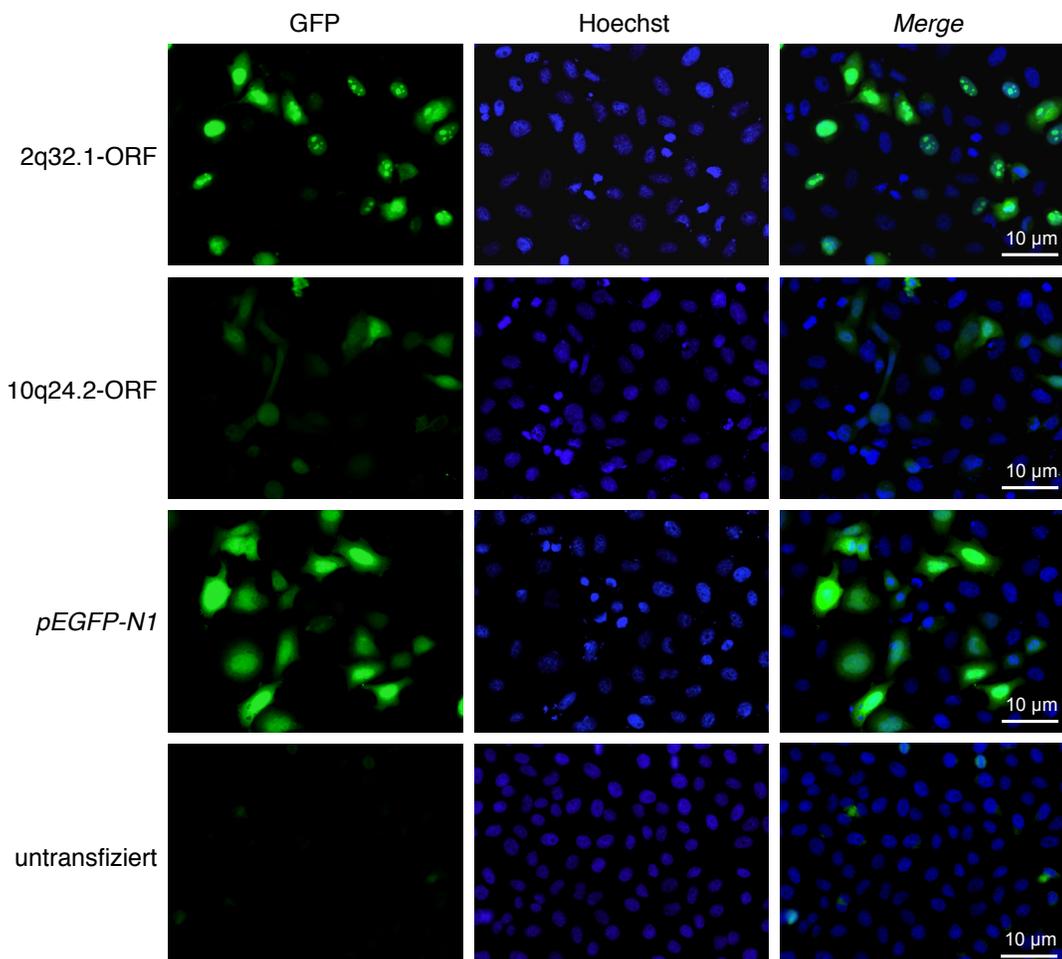
408 bp (2q32.1) bzw. 645 bp (10q24.2) große Produkte, die die ORFs beider Loci sowie Bereiche der 5'UTRs umfassten, wurden aus *pGEM<sup>®</sup>-T Easy*-Vektorkonstrukten amplifiziert, in denen der komplette 2q32.1-Locus bzw. der ORF- und 3'-LTR-Bereich des 10q24.2-Locus einkloniert war. Die verwendeten Vektorkonstrukte wurden im Rahmen von Biermann (2013) generiert. Für die Amplifikation wurden *Primer* mit *Bam*HI-Schnittstelle an ihrem 5'-Ende verwendet. Die ORF-Amplikons wurden in *pGEM<sup>®</sup>-T Easy* zwischenkloniert und über *Bam*HI-Verdau *in-frame* stromauf des GFP-ORFs in den zuvor mit *Bam*HI linearisierten Zielvektor *pEGFP-N1* einkloniert. Die korrekte Sequenz der einklonierten ORFs wurde durch Sanger-Sequenzierung beider Stränge bestätigt. Zellen der Melanozyten-Zelllinie MEWO wurden mit den ORF-Konstrukten transient transfiziert und nach 24 h die Expression von GFP-Fusionsproteinen mittels *Western Blot* und Fluoreszenzmikroskopie untersucht.

Aus einem Teil der Zellen wurde Gesamtprotein isoliert, mittels SDS-PAGE aufgetrennt, geblottet und potentielle GFP-Fusionsproteine mit einem spezifischen GFP-Antikörper nachgewiesen (vgl. **Abb. 28**). Als Positivkontrolle der Transfektion diente Lysat aus Zellen, die mit dem Insert-losen *pEGFP-N1*-Vektor transfiziert wurden. In allen transfizierten Zellen zeigte sich eine Proteinbande bei ca. 27 kDa, was der Größe von GFP (26,95 kDa) entspricht. In den Lysaten der mit den ORF-Konstrukten transfizierten Zellen traten weitere, größere Proteinbanden auf, die jedoch nicht den erwarteten Größen potentieller Fusionsproteine entsprachen. Das GFP-Fusionsprotein des 2q32.1-ORFs hat eine erwartete Größe von 43,33 kDa, es zeigte sich jedoch lediglich eine Bande für ein deutlich kleineres Protein bei etwa 30 kDa. Im Lysat der Zellen, die mit dem 10q24.2 ORF-Konstrukt transfiziert wurden, war eine Doppelbande bei etwa 34 und 36 kDa nachweisbar. Die theoretische Größe des erwarteten GFP-Fusionsproteins beträgt 51,57 kDa.



**Abbildung 28:** Nachweis von HERV-K(HML-2) GFP-Fusionsproteinen in Zellysaten transient transfizierter Zellen mittels *Western Blot*. Die ORFs der durch reverse Transkription und Reintegration gespleißter mRNA entstandenen HML-2-Loci auf den Chromosomen 2q32.1 und 10q24.2 wurden *in-frame* stromauf des GFP-ORFs in den *pEGFP-N1*-Vektor kloniert und Zellen der Melanom-Zelllinie MEWO mit den Vektorkonstrukten transient transfiziert. Zellysate wurde mittels SDS-PAGE aufgetrennt, geblottet und Fusionsproteine mit einem GFP-spezifischen Antikörper detektiert. Als Transfektionskontrolle dienten Zellen, die mit dem Insert-losen *pEGFP-N1*-Vektor transient transfiziert wurden, als Negativkontrolle dienten untransfizierte Zellen (vgl. Text). Proteingrößen und Belichtungszeiten sind angegeben. Als Ladekontrolle diente das Haushaltsgen  $\beta$ -Actin. Nach Entwicklung der Ladekontrolle stellte sich heraus, dass, obwohl in allen Spuren die vermeintlich gleiche Proteinmenge aufgetragen wurde, die vorangehende Protein-Konzentrationsbestimmung nicht zuverlässig war und die Mengen an aufgetragenem Protein stark variierten. Insbesondere in den Kontrollen war die tatsächlich aufgetrennte Proteinmenge deutlich geringer. Im Falle der mit den ORF-Konstrukten transfizierten Zellen waren die Unterschiede jedoch weniger stark ausgeprägt und erklären nicht die starken Unterschiede der Bandenintensitäten zwischen beiden ORF-Konstrukten bei Detektion der Fusionsproteine mit dem GFP-Antikörper bzw. in den nachfolgenden Fluoreszenzmikroskopischen Untersuchungen (s.u.).

Die Expression von GFP-Fusionsproteinen wurde zusätzlich fluoreszenzmikroskopisch nachgewiesen (vgl. **Abb. 29**). Als Positivkontrolle der Transfektion dienten auch hier Zellen, die mit dem GFP-exprimierenden, Insert-losen *pEGFP-N1*-Vektor transient transfiziert wurden. In den mit *pEGFP-N1* transfizierten Zellen zeigte sich eine gleichmäßige ubiquitäre Grünfluoreszenz der Zellen incl. Zellkern. Im Falle der transienten Transfektion mit dem 2q32.1 ORF-Konstrukt ergab sich kein einheitliches Bild. Einige Zellen zeigten eine ubiquitäre grüne Fluoreszenz, die in anderen Zellen hingegen ausschließlich auf den Zellkern beschränkt war, mit besonders starker Fluoreszenz der Nucleoli. Die mit dem 10q24.2 ORF-Konstrukt transfizierten Zellen zeigten eine ubiquitäre Grünfluoreszenz, die jedoch deutlich schwächer ausgebildet war und erst bei einer längeren Belichtungszeit von 500 ms sichtbar wurde, im Gegensatz zu 200 ms bei den mit dem 2q32.1-ORF-Konstrukt und den mit *pEGFP-N1* transfizierten Zellen.



**Abbildung 29:** Fluoreszenzmikroskopischer Nachweis der Expression von HERV-K(HML-2) GFP-Fusionsproteinen. Die ORFs der HML-2-Loci auf den Chromosomen 2q32.1 und 10q24.2 wurden *in-frame* stromauf des GFP-ORFs in den *pEGFP-N1*-Vektor kloniert und Zellen der Melanom-Zelllinie MEWO mit den Vektorkonstrukten transient transfiziert. Exprimierte GFP-Fusionsproteine wurden über ihre Immunfluoreszenz detektiert. Zellkerne wurden mit *Hoechst Stain Solution 33258* angefärbt. Die Belichtungszeit betrug für Hoechst jeweils 10 - 20 ms, für GFP 200 ms (für das 10q24.2-Konstrukt betrug die GFP-Belichtungszeit aufgrund der deutlich schwächeren Fluoreszenz 500 ms). Zur weiteren Erklärung, s. Text.

Mittels *Western Blot* und Immunfluoreszenz konnten in transient transfizierten Zellen somit GFP-Fusionsproteinen nachgewiesen werden. Es scheint, dass die Translation der hypothetischen Proteine, die von den beiden durch reverse Transkription und Reintegration gespleißter mRNA entstanden HML-2-Loci auf den Chromosomen 2q32.1 und 10q24.2 kodiert werden, somit zumindest *in vitro* möglich ist. Die von den erwarteten Größen abweichenden Banden im *Western Blot* sowie die zusätzlichen GFP-Banden deuten hierbei auf eine mögliche weitere Prozessierung der Fusionsproteine und eine Spaltung dieser hin. Die schwächeren Banden im *Western Blot* sowie die deutlich schwächere Fluoreszenz lassen auf eine deutlich geringere Expression des GFP-Fusionsproteins des 10q24.2-ORFs oder eine deutlich verringerte Proteinmengen nach Prozessierung schließen.

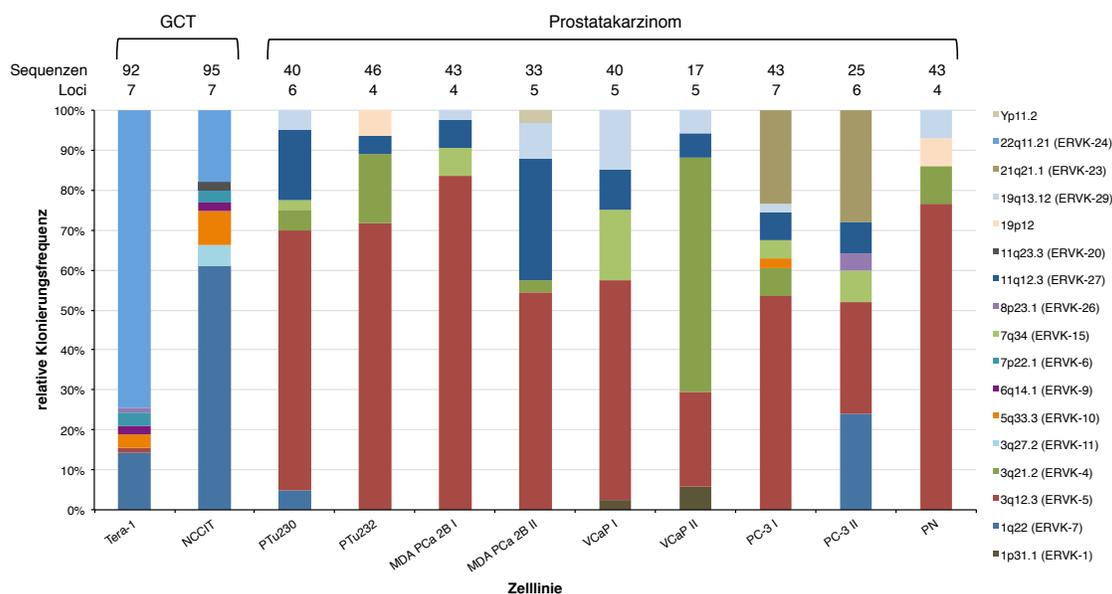
#### 4.1.14. Identifizierung transkribierter HERV-K(HML-2)-Loci in Keimzelltumor- und Prostatakarzinom-Proben

Neben Melanomen wurde eine potentielle Rolle humaner endogener Retroviren auch in weiteren Tumorerkrankungen, wie Keimzelltumoren (*germ cell tumours*, GCTs) (Sauter *et al.*, 1995, 1996; Herbst *et al.*, 1996; Boller *et al.*, 1997; Bieda *et al.*, 2001; Kleiman *et al.*, 2004) und Prostatakarzinomen (Ishida *et al.*, 2008; Goering *et al.*, 2011; Agoni *et al.*, 2013, Reis *et al.*, 2013) diskutiert, in denen retrovirale Transkripte und Proteine detektiert wurden. Analog den Melanom-Proben wurde die HML-2 *gag*-Transkription daher auch in Zelllinien dieser Krebsarten untersucht. Hierfür stand RNA aus den GCT- Zelllinien Tera-1 und NCCIT sowie den sechs Prostatakarzinom-Zelllinien PTu230, PTu232, MDA PCa 2B, VCaP, PC-3 und PN zur Verfügung. Im Falle der Zelllinien MDA PCa 2B, VCaP und PC-3 wurde RNA aus zwei unterschiedlichen Zellpassagen untersucht, die im Folgenden mit I und II gekennzeichnet sind. Aus cDNA aller Zelllinien wurde das HML-2 *gag*-Amplikon amplifiziert (vgl. **Anhang 2c-e**) und, wie zuvor beschrieben, Sanger-Sequenzen generiert und im menschlichen hg18 Referenzgenom rückkartiert.

Insgesamt wurden für alle Zelllinien 517 Sequenzen generiert, die eindeutig rückkartierbar waren. Es wurden 17 verschiedene Loci als transkribiert identifiziert, von denen 14 bereits in der Melanom-Studie detektiert wurden (vgl. **Tab. 15** und **16**). Zusätzlich wurde ein bisher nicht als transkribiert beschriebener Locus auf Chromosom Yp11.2 (chrY:6886441-6893384) mit einer cDNA-Sequenz in der Prostatakarzinom-Zelllinie MDA PCa 2B II detektiert. Dem etwa 7 kb großen Locus fehlt der 5'-LTR sowie ~500 bp im proviralen 5'-Bereich sowie ein ~1 kb großer interner Bereich, der etwa Position 2.800 bis 3.900 der bei *Repbase* hinterlegten HERVK-Konsensussequenz entspricht, was Teile der *pro*- und *pol*-ORFs umfasst. Des Weiteren wurde ein etwa 10 kb großer, mehr oder weniger vollständiger Locus auf chr19p12 mit je drei cDNA-Sequenzen in den Prostatakarzinom-Zelllinien PTu232 und PN detektiert sowie der ~9,5 kb große *ERVK-20* Locus (11q23.3) mit zwei Sequenzen in der GCT-Zelllinie NCCIT.

Die GCT- und Prostatakarzinom-Zelllinien unterschieden sich insgesamt in ihrem Transkriptionsmuster voneinander (vgl. **Abb. 30**; die relativen Klonierungsfrequenzen sowie ein Vergleich zu allen Proben der Melanom-Studie finden sich in **Anhang 6** und **7**). In den GCT-Zelllinien zeigten die Loci *ERVK-7* (1q22) und *ERVK-24* (22q11.21) die stärkste Transkription mit relativen Klonierungsfrequenzen von 14,13 % (*ERVK-7*) und 72,83 % (*ERVK-24*) in Tera-1 bzw. 61,05 % (*ERVK-7*) und 17,89 % (*ERVK-24*) in NCCIT. In den Prostatakarzinom-Zelllinien war der *ERVK-7*-Locus hingegen nicht dominant bzw. *ERVK-24* gar nicht transkribiert.

Stattdessen zeigte hier insbesondere der *ERVK-5*-Locus (3q12.3) sehr hohe relative Klonierungsfrequenzen von bis zu 83,72 % (MDA PCa 2B I). Weitere stark transkribierte Loci waren u.a. *ERVK-4* (3q21.2; max. 58,82 % in VCaP II) und *ERVK-23* (21q21.1; max. in PC-3 I bzw. II).



**Abbildung 30:** Transkriptionsprofile der in Keimzelltumor- und Prostatakarzinom-Zelllinien als transkribiert identifizierten HERV-K(HML-2)-Loci. Das HML-2 *gag*-Amplikon wurde aus cDNA der Keimzelltumor-Zelllinien Tera-1 und NCCIT sowie der Prostatakarzinom-Zelllinien PTu230, PTu232, MDA PCa 2B, VCaP, PC-3 und PN amplifiziert und anschließend generierte cDNA-Sequenzen wurden im Genom rückkartiert (vgl. Text). Für die Zelllinien MDA PCa 2B, VCaP und PC-3 wurde jeweils cDNA aus zwei unterschiedlichen Zellpassagen (I und II) untersucht. Angegeben ist die relative Klonierungsfrequenz jedes Locus, die sich für jede Zelllinie aus der Anzahl an Sequenzen, die eindeutig dem Locus zugeordnet werden konnten, dividiert durch die Gesamtanzahl an generierten Sequenzen pro Zelllinie berechnete. Über dem Diagramm ist für jede Zelllinie die Anzahl an eindeutig rückkartierbaren Sequenzen sowie die Anzahl an transkribierten Loci angegeben. Die Farbgebung der verschiedenen Loci entspricht der in **Abbildung 19**.

Der in Melanom- und Melanozyten-Zelllinien dominante *ERVK-14*-Locus (7q22.1; vgl. **Tab. 15** und **Abb. 18**) wurde in den GCT- und Prostatakarzinom-Zelllinien nicht detektiert. Ähnlich verhielt es sich mit dem *ERVK-6*-Locus auf Chromosom 7p22.1, der mit teilweise hohen relativen Klonierungsfrequenzen in den Melanom-Proben transkribiert war, hier jedoch in den Prostatakarzinom-Zelllinien gar nicht und in den GCT-Zelllinien nur mit niedrigeren relativen Klonierungsfrequenzen von 5,43 % (Tera-1) und 3,16 % (NCCIT) als transkribiert detektiert wurde.

Jede Zelllinie zeigte ein leicht unterschiedliches Transkriptionsmuster in den schwächer und stärker transkribierten Loci. Die beiden unterschiedlichen Passagen der Zelllinien MDA PCa 2B, VCaP und PC-3 lieferten jeweils sehr ähnliche Transkriptionsprofile. Unterschiede sind hier vermutlich größtenteils auf die geringere Anzahl eindeutig rückkartierbarer Sequenzen (17 bis

33; vgl. **Abb. 30**) der jeweils zweiten Zellpassage zurückzuführen und würden vermutlich durch Generierung weiterer Sequenzen minimiert.

Ähnlich den Melanom-Proben zeigen somit auch verschiedene GCT- und Prostatakarzinom-Zelllinien spezifische, leicht unterschiedliche Transkriptionsmuster. Insgesamt unterscheiden sich die Transkriptionsprofile jedoch in GCT- und Prostatakarzinom-Zelllinien voneinander sowie von den Melanom-Proben, insbesondere was die Transkription dominanter Loci betrifft. So zeigen in den GCT-Zelllinien die Loci *ERVK-7* und *ERVK-24* die stärkste Transkription, in den Prostatakarzinom-Zelllinien der *ERVK-5*-Locus, die in den jeweils anderen Zelllinien sowie den Melanom-Proben weniger dominant vertreten sind.

#### **4.1.15. HERV-K(HML-2) *rec/np9*-Transkription in Normalgeweben und Identifizierung transkribierter, potentiell bisher unbekannter HML-2-Loci**

In Abschnitt 4.1.9. wurde bereits die Transkription *rec* und *np9* mRNA-produzierender HML-2-Loci in Melanom-Proben sowie Melanozyten-Zelllinien gezeigt. Aufgrund der potentiellen Rolle der akzessorischen HML-2-Proteine Rec und Np9 in der Tumorgenese, die in früheren Studien diskutiert wurde (Armbruster *et al.*, 2004; Pece *et al.*, 2004; Denne *et al.*, 2007) sowie einer beschriebenen spezifischen Expression des Np9 Proteins in Tumorgewebe und transformierten Zellen (Armbruster *et al.*, 2002) stellt sich die Frage, ob sich auch in Normalgeweben *rec*- und *np9*-spezifische Transkripte finden oder ob die Transkription auf spezifische Tumorgewebe und Zelllinien beschränkt ist.

Zur Untersuchung der HML-2-Transkription in Normalgeweben stand cDNA aus 16 menschlichen Geweben zur Verfügung (Herz, Gehirn, Plazenta, Lunge, Leber, Skelettmuskel, Niere, Pankreas, Milz, Thymus, Prostata, Hoden, Eierstock, Dünndarm, Dickdarm, Leukozyten). In Kooperation mit apl.-Prof. Dr. rer. nat. Klaus Roemer (José Carreras Zentrum für Immun- und Gentherapie, Universitätsklinikum des Saarlandes, Homburg) wurden von Frau Dr. rer. nat. Kristina Heyne *rec/np9*-spezifische PCR-Produkte (vgl. Abschnitt 4.1.9.) aus allen cDNA-Proben amplifiziert. Die PCR-Produkte wurden in *pCR*<sup>®</sup>-2.1-*TOPO* ligiert und mit diesen Plasmiden transformierte Bakterienklone wurden für die vorliegende Arbeit zur Verfügung gestellt. Inserttragende Plasmide wurden mittels Kolonie-PCR identifiziert und die Plasmide aus diesen aufgereinigt, sequenziert und mittels *BLAT* rückkartiert. Aufgrund des von *BLAT* verwendeten Algorithmus wird hierbei nur der größte Bereich der Sequenz stromab der *rec splice acceptor site* berücksichtigt.

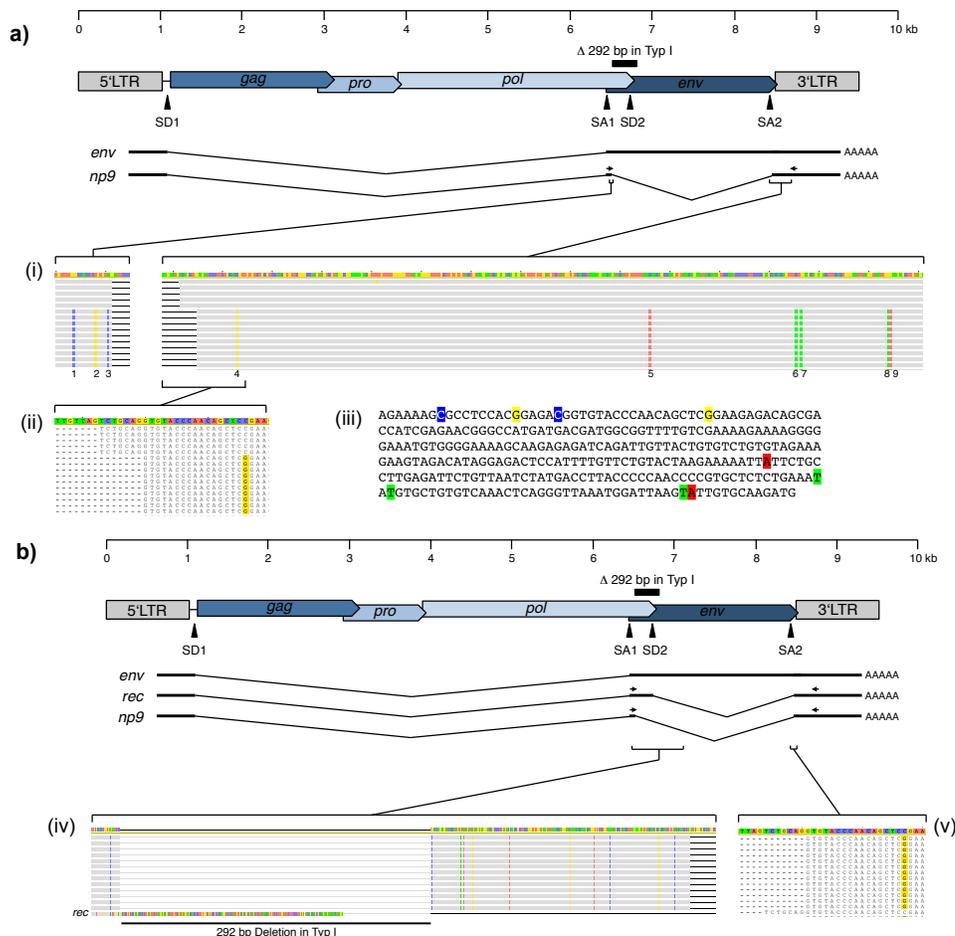
Zwar wurden die *np9* und *rec* PCR-Produkte im Gel getrennt ausgeschnitten und sequenziert, jedoch stellte sich bei der späteren Analyse der Sequenzen heraus, dass aufgrund nicht ausreichender gelelektrophoretischer Trennung der Banden auch *rec*-spezifische PCR-Produkte unter den *np9*-repräsentierenden PCR-Produkten zu finden waren, und umgekehrt, weswegen im Folgenden alle Sequenzen zusammen betrachtet werden.

Für alle 16 Gewebeproben wurden insgesamt 661 eindeutig rückkartierbare cDNA-Sequenzen mit nur einem besten *BLAT-Match* generiert. Hiervon konnten jedoch nur 459 Sequenzen (69,14 %) eindeutig mit weniger als sechs Nukleotidunterschieden einem HML-2-Locus zugeordnet werden. Eine, verglichen mit den vorangegangenen Analysen, sehr hohe Zahl an Sequenzen zeigten mehr als sechs Nukleotidunterschiede zu ihrem besten *BLAT-Match*.

Es trat hierbei in 15 der 16 Gewebe-Proben eine auffällig hohe Zahl von insgesamt 176 Sequenzen (26,63 %) auf, die als besten *BLAT-Match* auf den *ERVK-5*-Locus auf Chromosom 3q12.3 rückkartiert wurden, wobei alle Sequenzen jedoch zwischen sechs und zehn Nukleotidunterschiede zu *ERVK-5* aufwiesen. Die Sequenzunterschiede fanden sich hierbei größtenteils immer an den gleichen Positionen (vgl. **Abb. 31a**). Eine ähnliche Sequenz wurde zuvor im Zuge der *rec/np9*-Analyse auch aus cDNA der Melanozyten-Zelllinie Benno amplifiziert (vgl. Abschnitt 4.1.9.), ging dort jedoch aufgrund der hohen Anzahl von sechs Nukleotidunterschieden nicht in die weitere Analyse ein. Im Gegensatz zu Sequenzen, die in fast allen Geweben identifiziert wurden und mit weniger als sechs Nukleotidunterschieden eindeutig *ERVK-5* zugeordnet werden konnten, waren diese Sequenzen zudem um sieben Nukleotide verkürzt. Eine genauere Betrachtung zeigt, dass es sich hierbei um eine alternative *np9*-Spleißvariante handelt, bei der eine alternative *splice acceptor site* sieben Nukleotide stromab der regulären *np9/rec SA site* benutzt wird, was den Größenunterschied erklärt.

Des Weiteren traten in neun Proben insgesamt 23 Sequenzen (3,48 %) auf, die als besten *BLAT-Match* dem *ERVK-18*-Locus auf Chromosom 1q23.3 zugeordnet wurden, wobei alle Sequenzen zwischen 22 und 28 Nukleotidunterschiede zu diesem Locus aufwiesen. Wie schon bei dem zuvor beschriebenen Locus, der mit >6 Nukleotidunterschieden *ERVK-5* zugeordnet wurde, fanden sich auch hier die Sequenzunterschiede zu *ERVK-18* größtenteils jeweils an den gleichen Positionen (vgl. **Abb. 31b**). Ein genauerer Sequenzvergleich deutet darauf hin, dass es sich bei diesen cDNA-Sequenzen um eine weitere Spleißvariante eines Typ I Provirus handelt. Die eigentliche *np9 splice donor site* ist komplett vorhanden, sowie ein 244 bp langer Bereich, stromab der 292 bp Deletionsregion in Typ I Proviren, der in *np9*-mRNA normalerweise als

Intron-Teil gespleißt wird. Zudem wird eine alternative *splice acceptor site* sieben Nukleotide stromab der *np9/rec SA site* genutzt.



**Abbildung 31:** Aus cDNA gesunder Gewebe generierte, deutlich variante *rec/np9*-Sequenzen zweier potentieller neuer HML-2-Loci. Aus cDNA menschlicher Normalgewebe wurde ein *rec/np9*-spezifisches Amplikon generiert, kloniert und sequenziert, wobei eine hohe Anzahl an Sequenzen generiert wurde, die nicht eindeutig einem genomischen HML-2-Locus zugeordnet werden konnten. Der beste *BLAT-Match* von je einer dieser Sequenzpopulationen waren die Loci *ERVK-5* (3q12.3; a) und *ERVK-18* (1q23.3; b), wobei sich in allen Fällen die gleichen charakteristischen Sequenzunterschiede zu diesen Loci zeigten. a) Struktur des HERV-K(HML-2.HOM)-Provirus (Genbank acc. no. AF074086.2) und der Spleißmuster von *env*- und *np9*-mRNA (SD = splice donor site, SA = splice acceptor site). Darunter finden sich Ausschnitte eines multiplen Sequenz-Alignments (i) des *ERVK-5*-Locus (bunte oberste Sequenz) mit cDNA-Sequenzen, die dem Locus zugeordnet werden konnten. Die cDNA-Sequenzen sind hierbei als graue Balken verdeutlicht, wobei schwarze horizontale Linien jeweils dem in den cDNA-Sequenzen gespleißten Intron-Bereich entsprechen. Bunte Bereiche kennzeichnen Nukleotidunterschiede zur *ERVK-5*-Sequenz. Es sind exemplarisch fünf Sequenzen gezeigt, die eindeutig *ERVK-5* zugeordnet werden konnten, sowie zehn weitere Sequenzen, die durch insgesamt neun Sequenzunterschiede (1-9) zu *ERVK-5* gekennzeichnet sind. Diesen Sequenzen fehlt zudem, verglichen mit den eindeutig rückkartierbaren Sequenzen, ein sieben Nukleotide langer Abschnitt im Bereich der *splice acceptor site* 2 (SA2), die im unteren Alignment (ii) im Detail gezeigt ist. Exemplarisch ist in (iii) eine generierte cDNA-Sequenz mit hervorgehobenen charakteristischen Nukleotidunterschieden zu *ERVK-5* gezeigt. b) Eine weitere cDNA-Sequenzvariante im Vergleich zur Struktur des HERV-K(HML-2.HOM)-Provirus und der Spleißmuster von *env*-, *rec*- und *np9*-mRNA. Es sind Ausschnitte eines multiplen Sequenz-Alignments (iv) des *ERVK-18* Typ I-Locus (bunte obere Sequenz) mit *rec*-mRNA (bunte untere Sequenz) sowie 13 cDNA-Sequenzen gezeigt, die dem *ERVK-18*-Locus als besten *BLAT-Match* zugeordnet wurden und durch charakteristische Sequenzunterschiede zu diesem gekennzeichnet sind. Die gefundenen cDNA Sequenzen zeichnen sich durch ein Spleißmuster aus, das sich von *rec*- und *np9*-mRNA unterscheidet. Die cDNA-Sequenzen zeigen die für Typ I Proviren auf genomischer Ebene spezifische 292 bp Deletion, weisen stromab dieser jedoch einen zusätzlichen 244 bp langen Bereich auf, der in *np9*-mRNA üblicherweise als Intron-Teil gespleißt wird. Auch die *splice acceptor site* ist verglichen mit der in *rec/np9*-mRNA um sieben bp verschoben (v). Zur weiteren Erklärung, s. Text.

Es wurden noch weitere cDNA-Sequenzen generiert, die aufgrund einer jeweils hohen Anzahl an Sequenzunterschieden keinem HML-2-Locus eindeutig zugeordnet werden konnten. Diese traten jedoch nur in geringer Anzahl in einer oder wenigen Proben auf und wurden daher nicht weiter analysiert, da es sich hierbei vermutlich um Rekombinanten oder durch Sequenzierfehler bedingte Sequenzvarianten handelt.

Insgesamt wurden in den gesunden Proben, incl. der beiden beschriebenen potentiellen neuen Sequenzvarianten, 19 verschiedene HML-2-Loci als transkribiert identifiziert. Darunter fanden sich elf *np9* mRNA-produzierende Typ I-Loci sowie acht *rec* mRNA-produzierende Typ II-Loci (vgl. **Tab. 19**). Drei der Loci auf Chromosom 4q32.3, 5p13.3 und 5q15 wurden bisher nicht als transkribiert identifiziert, die restlichen Loci wurden bereits im Rahmen der Melanom-Studien als transkribiert beschrieben.

**Tabelle 19:** Absolute Anzahl an cDNA-Sequenzen der in 16 menschlichen Normalgeweben als transkribiert identifizierter *rec* oder *np9* mRNA-produzierender HML-2-Loci.\*

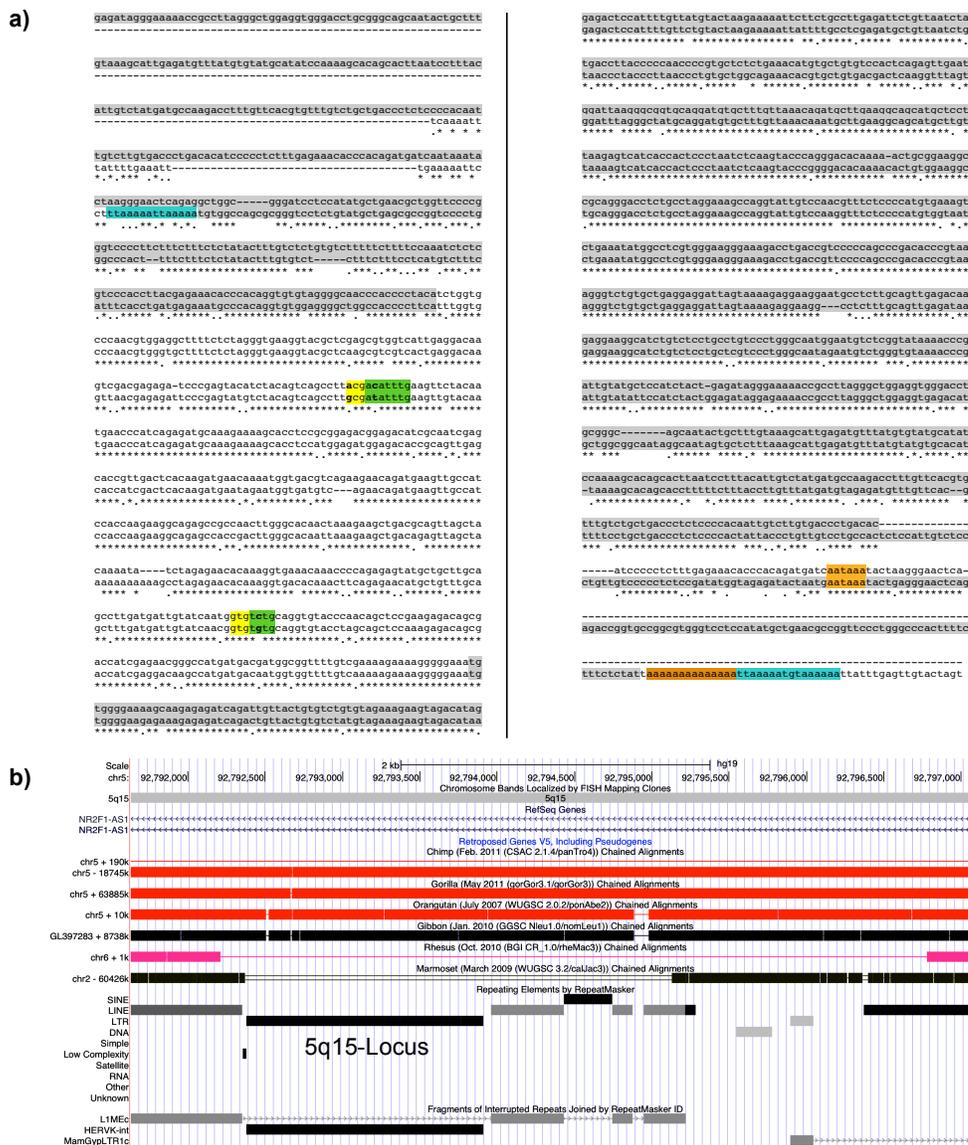
Locus	HGNC-Name	Bande	Gewebeprobe																Protein
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
neues PV, ähnl. <i>ERVK-5</i>	-	-	24	16	11	2	4	18	14	21	2	6	15	3	22	12	10		Np9
neues PV, ähnl. <i>ERVK-18</i>	-	-	9	2	1			3				4	1			1			
chr1:75615359-75621731	<i>ERVK-1</i>	1p31.1								1									
chr1:153863081-153872260	<i>ERVK-7</i>	1q22	1		11	15	21	2	1	1	22	11	4	7		10	8	24	
chr1:158927199-158936430	<i>ERVK-18</i>	1q23.3										3			1		1	2	
chr2:187093879-187095344	-	2q32.1	10	2	6	5	4	4		13	13	12	3	1	5	4	8	17	
chr3:102893427-102902549	<i>ERVK-5</i>	3q12.3	1	6		7	3	1	2	1	1	2	4	7		3	5		
chr3:114225814-114234972	<i>ERVK-3</i>	3q13.2											1						
chr4:166136289-166143518 <sup>#</sup>	-	4q32.3												1					
chr21:18855530-18863833	<i>ERVK-23</i>	21q21.1																	
chr22:17306187-17315361	<i>ERVK-24</i>	22q11.21		1															
chr3:127091992-127101129	<i>ERVK-4</i>	3q21.2														1			
chr5:30522516-30531962 <sup>#</sup>	-	5p13.3										1							
chr5:92818136-92819668 <sup>#</sup>	-	5q15	2	13	10	11	1			4	7	1	11	7	10	9	7		
chr6:78483381-78492802	<i>ERVK-9</i>	6q14.1				1		15											
chr7:4588583-4606557	<i>ERVK-6</i>	7p22.1				1						1					2		
chr10:101570559-101577735	<i>ERVK-17</i>	10q24.2				1		8		1				3			1		
chr12:57007509-57016965	<i>ERVK-21</i>	12q14.1				2	1			20			1		1	1	1		
chr19:32820338-32829201	<i>ERVK-27</i>	19q12		2															
gesamt			23	26	32	41	37	25	23	21	43	31	29	27	17	29	34	43	

\*Mit einem *rec/np9*-spezifischen Amplikon wurde die Transkription von *np9* oder *rec* mRNA-produzierenden Typ I- und II Proviren untersucht. Das Amplikon wurde von cDNA aus 16 menschlichen Normalgeweben amplifiziert und cDNA-Sequenzen im hg18 Referenzgenom rückkartiert. 1: Herz, 2: Gehirn, 3: Plazenta, 4: Lunge, 5: Leber, 6: Skelettmuskel, 7: Niere, 8: Pankreas, 9: Milz, 10: Thymus, 11: Prostata, 12: Hoden, 13: Eierstock, 14: Dünndarm, 15: Dickdarm, 16: Leukozyten. Angegeben sind die Chromosomenpositionen der transkribierten Loci, sowie deren offizielle HGNC-Namen (vgl. Mayer *et al.*, 2011) und Chromosomenbanden. Angegeben ist für jede Probe die absolute Anzahl an cDNA-Sequenzen, die den verschiedenen Loci eindeutig zugeordnet werden konnten, wobei die Gesamtanzahl an Sequenzen für jede Probe unter der Tabelle angegeben ist. Mit „#“ markierte Loci wurden bisher nicht als transkribiert identifiziert. Es wurden zudem Sequenzen zweier potentieller neuer, nicht im Referenzgenom annotierter Proviren (PV) detektiert, deren bester *BLAT-Match* die Loci *ERVK-5* bzw. *ERVK-18* darstellen, die sich jedoch jeweils durch charakteristische Nukleotidunterschiede zu diesen Loci auszeichnen. Zur weiteren Erklärung, s. Text.

Auch in menschlichen Normalgeweben konnten somit insgesamt 19 verschiedene *rec* und *np9* mRNA-produzierende HML-2-Loci als transkribiert identifiziert werden. Verschiedene Sequenzpopulationen, die nicht eindeutig rückkartierbar waren, deuteten hierbei auf die Transkription potentieller neuer HML-2-Loci hin, die nicht im Referenzgenom annotiert sind, oder repräsentieren allelische Varianten bekannter Proviren.

#### 4.1.16. Charakterisierung eines weiteren durch Retrotransposition gespleißter mRNA entstandenen HML-2-Locus auf Chromosom 5q15

Unter den *rec* mRNA-produzierenden Typ II HML-2-Loci fand sich u.a. der schon in Melanozyten-Zelllinien detektierte Locus auf Chromosom 10q24.2 (vgl. Abschnitt 4.1.10.) sowie ein weiterer stark verkürzter Locus auf Chromosom 5q15. Dieser Locus war in 13 der untersuchten Geweben transkribiert und ähnelt, analog dem HML-2-Locus auf Chromosom 2q32.1, einer retrotransponierten *rec*-mRNA, die innerhalb eines L1MEc-Elementes inseriert wurde (vgl. **Abb. 32a**). Der Locus besteht lediglich aus einem ~150 bp großen Abschnitt des 3'-Bereiches des 5'-LTRs, einem sich daran anschließenden ~110 bp großen internen proviralen Bereich, der bis zur unvollständigen *splice donor site* 1 reicht, ~280 bp zwischen der *splice acceptor site* 1 und der *rec SD site* 2 sowie dem Bereich stromab der *rec SA site* 2. Der sich daran anschließende 5'-LTR ist nahezu vollständig, es fehlen lediglich die letzten 90 bp der U5-Region. Gegenüber *rec*-mRNA zeigen sich zahlreiche kürzere Insertionen und Deletionen und eine Translation der Sequenz des Locus liefert lediglich verkürzte hypothetische Proteine ohne signifikante Ähnlichkeit zu Rec oder anderen HML-2-Proteinen. Wie bei dem retrotransponierten Locus auf Chromosom 2q32.1 findet sich auch hier knapp 90 bp stromab des Poly(A)-Signals des 3'-LTRs eine Wiederholung von 14 Adenosin-Nukleotiden und der Locus wird von TSDs (5'-TTAAAAAT(G)TAAAAAA-3') flankiert. Auch der 5q15-Locus ist im *UCSC Genome Browser* nicht als prozessiertes Pseudogen annotiert. Das Provirus wurde vermutlich vor der evolutiven Aufspaltung der Hominoidea generiert (vgl. **Abb. 32b**), da sich homologe Sequenzen hoher Ähnlichkeit auch in den Genomen von Schimpanse, Gorilla, Orang-Utan und Gibbon finden, jedoch nicht im Rhesusaffen und Marmosetten.



**Abbildung 32:** Struktur eines weiteren durch Retrotransposition entstandenen HML-2-Locus auf Chromosom 5q15, der in verschiedenen menschlichen Normalgeweben als transkribiert identifiziert wurde. Der Locus repräsentiert wahrscheinlich eine gespleißte *rec*-mRNA, die revers transkribiert und durch die L1-Maschinerie reintegriert wurde. **a)** Sequenzvergleich des 5q15-Locus mit einer *rec* mRNA-Sequenz (*Genbank acc. no. X72790*). Die LTR-Anteile beider Sequenzen sind grau unterlegt, die beiden *rec splice donor* bzw. *acceptor sites* sind in gelb bzw. grün markiert. Dem Locus fehlen, wie der gespleißten *rec*-mRNA, die Introns im Bereich *gag*, *pro* und *pol* sowie in *env*. Der Locus weist an seinem 3'-Ende im 3'-LTR ein Poly(A)-Signal (hellorange) sowie einen Poly(A)-Schwanz (dunkelorange) auf und wird von *target site duplications* (türkis) flankiert. **b)** Der retrotransponierte 5q15-Locus im „RepeatMasker v.3.2.7 Annotation“-Track des UCSC Genome Browser (hg18). Der 5q15-Locus ist im unteren linken Bereich durch einen schwarzen Balken als LTR-Repeat annotiert. Im oberen Teil der Abbildung ist die graphische Darstellung eines Sequenzvergleiches mit anderen Primatengenomen dargestellt. Der homologe HML-2-Locus ist in Schimpanse, Gorilla, Orang-Utan und Gibbon vorhanden. In Rhesus und Marmosetten fehlt der Locus sowie in den beiden letzteren weitere *Repeat*-reiche flankierende Bereiche. Der 5q15-Locus wurde von entsprechenden Algorithmen nicht als retrotransponierte Sequenz identifiziert bzw. annotiert (Annotation „Retroposed Genes V5 including Pseudogenes“, in blau).

#### 4.1.17. HERV-K(HML-2)-Sequenzvarianten konnten auch auf genomischer Ebene verifiziert werden

Im Zuge der *rec/np9*-Transkriptionsanalyse in menschlichen Normalgeweben wurden zwei stark vertretene Sequenzpopulationen generiert, die nicht eindeutig im Genom rückkartierbar waren und durch charakteristische Nukleotidunterschiede zu genomischen HML-2-Loci gekennzeichnet waren (vgl. Abschnitt 4.1.13.) Zur Klärung der Frage, ob es sich bei diesen cDNA-Sequenzen um *in vitro* generierte Varianten handelt, oder ob diese Sequenzvarianten, die sich von allen bekannten HML-2-Loci unterscheiden, auch auf genomischer Ebene nachweisbar sind, wurde versucht, vergleichbare Sequenzen mit den gleichen Unterschieden zu bekannten HML-2-Loci auch von genomischer DNA zu generieren.

Für beide Sequenzpopulationen wurde hierfür das reguläre *rec/np9*-Amplikon sowie ein leicht verkürztes spezifisches PCR-Produkt von genomischer DNA generiert, wobei *nested-PCR-Primer* eingesetzt wurden, die mit ihren 3'-Enden auf charakteristischen Nukleotidunterschieden der oben beschriebenen Sequenzpopulationen lagen. Die PCR-Produkte wurden in *pGEM<sup>®</sup>-T Easy* kloniert und aus beiden Richtungen sequenziert. Wie bereits von cDNA wurden auch hier Sanger-Sequenzen generiert, die nicht eindeutig im Genom rückkartiert werden konnten und die die gleichen charakteristischen Nukleotidunterschiede zu *ERVK-5* bzw. *ERVK-18* aufwiesen (Beispielsequenzen finden sich in **Anhang 8**).

Die Sequenzen waren somit auch auf genomischer Ebene verifizierbar, womit ausgeschlossen werden konnte, dass es sich um *in vitro* generierte Varianten handelte. Bei solchen Sequenzvarianten mit charakteristischen Unterschieden zu annotierten HML-2-Loci handelt es sich entweder um allelische Varianten eines bekannten Provirus oder um neue, bisher unbeschriebene HML-2-Loci, die nicht in der hg18 oder hg19 Referenzsequenz annotiert sind.

## 4.2. Charakterisierung der HERV-W-Transkription im Kontext von Multipler Sklerose

Zahlreiche frühere Studien diskutierten eine potentielle Rolle der HERV-W-Gruppe im Kontext von Multipler Sklerose (MS). Es wurden spezifisch HERV-W-Transkripte sowie retrovirale Proteine in MS-Patienten berichtet (Roebke *et al.*, 2010; Antony *et al.*, 2004) sowie potentiell inflammatorische Effekte retroviraler Proteine (Rolland *et al.*, 2006; Kremer *et al.*, 2013). Insbesondere für das Syncytin-1-Protein wurde eine erhöhte Expression in MS-Patienten sowie eine demyelinisierende Wirkung in Mausmodellen berichtet (Antony *et al.*, 2004, 2007a). Auch eine potentiell erhöhte HERV-W-Expression im Kontext von MS wurde berichtet, wobei verschiedene Studien hier teils unterschiedliche Ergebnisse lieferten (Antony *et al.*, 2006, 2007b; Mameli *et al.*, 2007) und potentielle Fehler in einigen Studien diskutiert wurden (Garson *et al.*, 2009) (vgl. Abschnitt 2.3.2.1. der Einleitung).

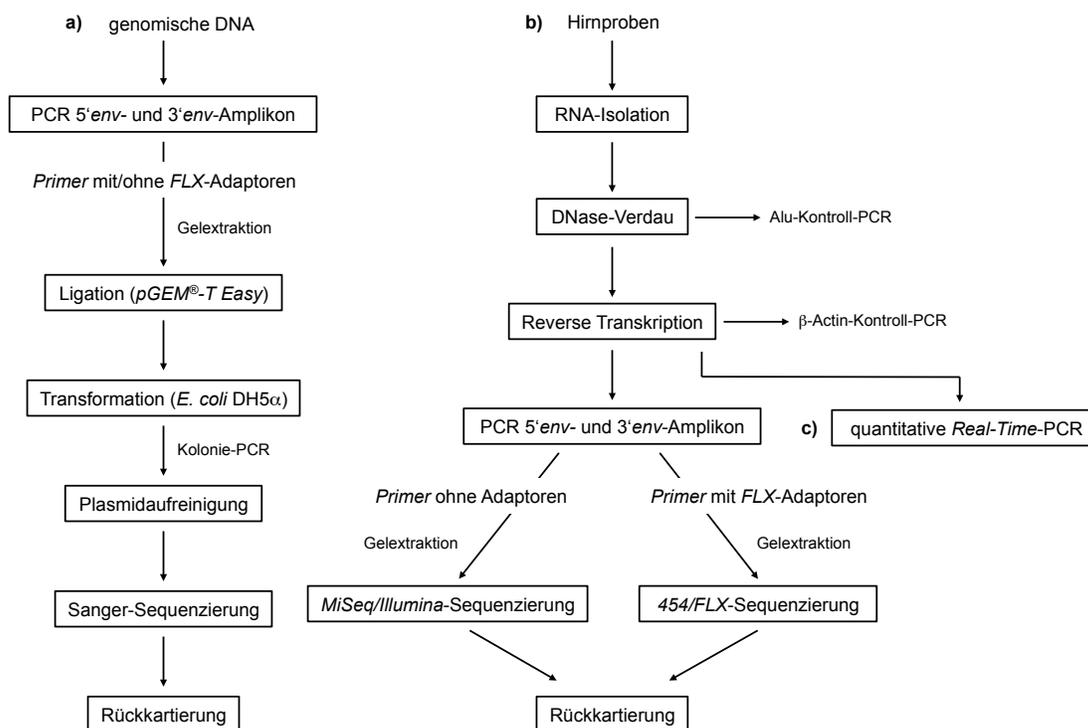
Als Beitrag zum besseren Verständnis der potentiellen Rolle von HERV-W in der Pathogenese von Multipler Sklerose wurden umfangreiche Transkriptionsanalysen durchgeführt und transkribierte HERV-W-Loci in Hirnproben von MS-Patienten und gesunden Kontrollen identifiziert. Es wurden detaillierte Transkriptionsmuster erstellt und transkribierte Loci wurden hinsichtlich ihrer Struktur und Genomlokalisierung weiter analysiert. Zur Detektion potentieller Unterschiede der HERV-W-Transkriptmengen zwischen MS-Patienten und gesunden Kontrollen wurden die HERV-W-Transkriptmengen quantifiziert. Des Weiteren wurden zur Klärung der Frage des Ursprungs retroviraler Transkripte ausgewählte LTR-Elemente transkribierter Loci hinsichtlich ihrer potentiellen Promotoraktivität untersucht.

### 4.2.1. Arbeitsablauf zur Identifizierung transkribierter HERV-W-Loci

Mittels *Next Generation Sequencing* (NGS) wurden transkribierte HERV-W-Loci in je sieben Hirnproben aus sechs MS-Patienten sowie aus sieben nicht-dementen Kontrollen identifiziert, um eine mögliche biologische Involvierung von HERV-W in der Pathogenese von Multipler Sklerose weiter zu untersuchen. Der Arbeitsablauf ist in **Abbildung 33** gezeigt.

Die zur Generierung HERV-W Gruppen-spezifischen RT-PCR-Produkte verwendeten *Primer* wurden zunächst hinsichtlich der Amplifikation möglichst vieler HERV-W-Loci in einer PCR optimiert und auf genomischer DNA getestet (vgl. **Abb. 33a**). Analog HERV-K(HML-2) (vgl. Absatz 4.1.1. und 4.1.2.) wurden auch hier die untersuchten, im HERV-W *env*-ORF lokalisierten Amplikons 5'*env* und 3'*env* (s.u.) von genomischer DNA amplifiziert, die PCR-Produkte

gelelektrophoretisch aufgetrennt, aufgereinigt, in *pGEM<sup>®</sup>-T Easy* kloniert und in *E. coli* DH5a transformiert. Insert-tragende Plasmide wurden in einer Kolonie-PCR identifiziert, die Plasmide aus positiven Klonen aufgereinigt und Sequenzen mittels Sanger-Sequenzierung generiert und im menschlichen hg18 Referenzgenom rückkartiert.

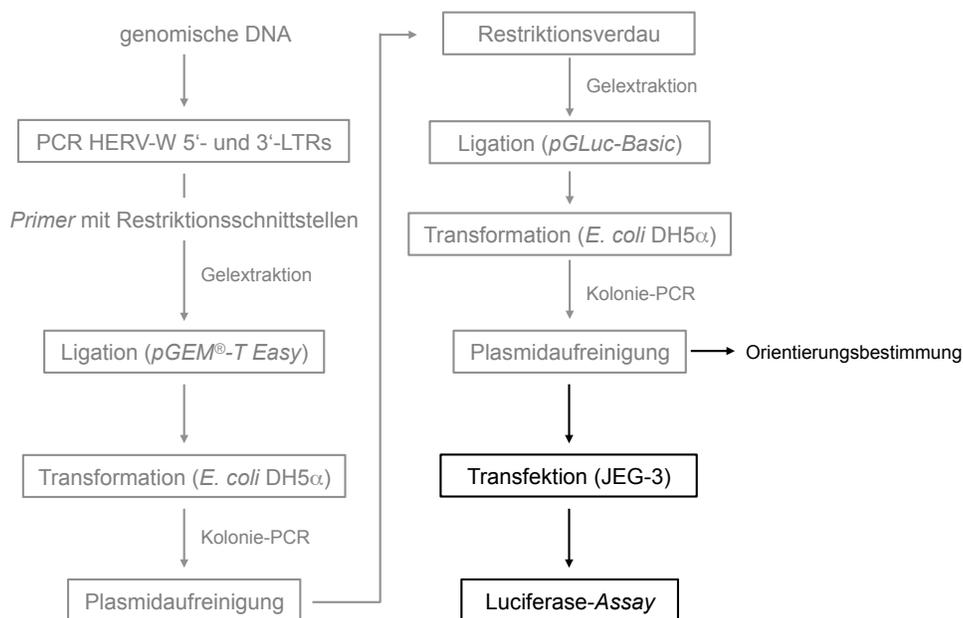


**Abbildung 33:** Arbeitsablauf zur Identifizierung transkribierter HERV-W-Loci in Hirnproben von MS-Patienten und gesunden Kontrollen sowie zur Quantifizierung der HERV-W-Transkriptmengen. Zur Erklärung, s. Text.

Aus den untersuchten Hirnproben aus *Plaques* von MS-Patienten sowie der weißen Substanz gesunder Kontrollen wurde die Gesamt-RNA isoliert und einem DNase-Verdau unterzogen, dessen Erfolg im Anschluss mittels Alu-Element spezifischer PCR überprüft wurde (vgl. **Abb. 33b**). Die DNase-behandelte RNA wurde in cDNA umgeschrieben, wobei für jedes RNA-Sample ein Ansatz ohne Enzym als RT(-) Negativkontrolle diente. Der Erfolg der reversen Transkription sowie die Abwesenheit von DNA in den Negativkontrollen wurde mittels  $\beta$ -Actin-spezifischer PCR überprüft. Es wurden nur Proben weiter verwendet, die sowohl nach der DNase-Behandlung als auch nach der reversen Transkription negativ auf Reste genomischer DNA getestet wurden. Von der generierten cDNA wurden im Anschluss die im 5'- und 3'-Bereich des *env*-ORFs lokalisierten Amplikons 5'*env* und 3'*env* amplifiziert und mittels *Next Generation Sequencing* Sequenzen generiert und im menschlichen hg18 Referenzgenom rückkartiert. Es wurden hierbei alle Proben mittels 454/FLX-Technologie sequenziert sowie für je vier MS- und Kontrollproben zusätzlich Sequenzen mittels *MiSeq/Illumina*-Technologie generiert.

Zur Detektion möglicher Unterschiede in den Transkriptmengen wurde die HERV-W 5'env-Transkription insgesamt sowie die spezifischen Transkriptmengen der Loci *ERVWE1/ERVW-1/syncytin-1* und *ERVWE2/ERVW-2* in alle Proben mittels *Real-Time-PCR* quantifiziert (vgl. **Abb. 33c**).

In weiteren Experimenten wurde die Promotoraktivität ausgesuchter LTR-Elemente transkribierter HERV-W-Loci in Reportergen-Assays untersucht (vgl. **Abb. 34**). Hierfür wurden *pGLuc-Basic*-Vektorkonstrukte verwendet, die im Rahmen von Richter (2012) generiert wurden. Die *pGLuc-Basic*-Konstrukte enthielten als putative Promotorsequenz HERV-W LTR-Elemente in *Sense* und/oder *Antisense*-Orientierung stromauf des Luciferase-Gens des Vektors. Die LTRs wurden im Rahmen von Richter (2012) mit *PCR-Primern*, die an ihrem 5'-Ende mit für die Klonierungen passenden Restriktionsschnittstellen versehen waren aus genomischer DNA amplifiziert, in *pGEM<sup>®</sup>-T Easy* zwischenkloniert und danach in den entsprechend linearisierten *pGLuc-Basic*-Vektor umkloniert. Die Orientierung der LTRs in den hier verwendeten Plasmid-Konstrukten wurde für die vorliegende Arbeit erneut durch Sanger-Sequenzierung bestätigt. Zellen der plazentalen Chorionkarzinom-Zelllinie JEG-3 wurden mit den LTR-Konstrukten transient transfiziert und nach 48 h die potentielle Promotoraktivität der LTRs mittels Luciferase-Assay bestimmt.



**Abbildung 34:** Arbeitsablauf zur Untersuchung der potentiellen Promotoraktivität spezifischer HERV-W 5'- und 3'-LTR-Elemente. Die grau unterlegten Arbeiten wurden im Rahmen von Richter (2012) durchgeführt. Zur Erklärung, s. Text.

#### 4.2.2. Amplifikationsstrategie zur Identifizierung transkribierter HERV-W-Loci sowie Optimierung und Test der zur Amplifikation gewählten PCR-Primer

Die zur PCR verwendeten *Primer* wurden hinsichtlich der Amplifikation möglichst vieler HERV-W-Loci in einer PCR optimiert. Wie bereits für HERV-K(HML-2) beschrieben (vgl. Abschnitt 4.1.2.) wurde auch hier zunächst eine Sequenzzusammenstellung genomischer HERV-W-Loci erstellt. Da vielen HERV-W-Loci, bedingt durch ihre Entstehung durch einen L1-vermittelten Retrotranspositionsmechanismus, die U5-Region des 3'-LTRs sowie unterschiedlich große Teile des proviralen 5'-Bereiches fehlen, wurde eine *BLAT*-Suche im hg18 Referenzgenom nicht mit der kompletten proviralen Sequenz durchgeführt, sondern lediglich mit der Sequenz des *env*-ORFs der bei *Repbases* hinterlegten HERV17-Konsensussequenz (entspricht HERV-W), da dieser Bereich in den meisten Loci vorhanden ist. Die DNA-Sequenzen der resultierenden Loci wurden mit dem *UCSC Table Browser* zusammengestellt und ein multiples Sequenz-*Alignment* mit *MAFFT* generiert (vgl. **Anhang 9**). In einer früheren Studie von Laufer *et al.* (2009) wurden bereits transkribierte HERV-W-Loci in PBMCs von MS-Patienten und gesunden Kontrollen identifiziert. Das in dieser Studie verwendete etwa 690 bp große Amplikon ist in vielen der hier zusammengestellten Loci nicht bzw. nicht vollständig vorhanden, weswegen in der vorliegenden Arbeit ein kürzeres, etwa 280 bp großes Amplikon im 5'-Bereich von *env* gewählt wurde, das im Folgenden als 5'*env* bezeichnet wird. Das 5'*env*-Amplikon überlappt mit dem 3'-Bereich des Amplikons der Studie von Laufer *et al.* (2009) und entspricht etwa dem Bereich von nt 7548/7567 bis 7821 der bei *Repbases* hinterlegten HERV17-Konsensussequenz. Da die Mehrheit der zusammengestellten HERV-W-Loci noch größere Deletionen im 5'-Bereich von *env* aufweisen, die auch die Position des 5'*env*-Amplikons betreffen, wurde ein zweites, etwa 330 bp großes Amplikon im 3'-Bereich von *env* gewählt, im Folgenden als 3'*env* bezeichnet, das nt 8.349 bis 8.587 der *Repbases* HERV17 Konsensussequenz entspricht. Die genauen Positionen beider Amplikons sind im *Alignment* in **Anhang 9** eingezeichnet. Um die Amplifikation möglichst vieler HERV-W-Loci zu unterstützen, wurden, wie schon für HERV-K(HML-2) beschrieben (vgl. Abschnitt 4.1.2.), Gemische verschiedener *Forward*- und *Reverse-Primer*-Varianten eingesetzt, die Nukleotidunterschiede in den *Primer*-bindenden Regionen zwischen den Loci berücksichtigen. Im Falle von 5'*env* wurden zwei teilweise überlappende *Forward-Primer* und zwei *Reverse-Primer*-Varianten eingesetzt, zur Amplifikation von 3'*env* vier *Forward*- und sieben *Reverse-Primer*-Varianten. Toleriert man bis zu zwei Nukleotidunterschiede innerhalb der *Primer*-bindenden Regionen, die sich nicht unmittelbar am 3'-Ende des *Primers* befinden, sollten theoretisch mindestens 23 (5'*env*) bzw. 147 (3'*env*) verschiedene HERV-W-Loci mit den *Primer*-Varianten amplifiziert werden können, wobei 20 der

Loci durch beide Amplikons detektiert werden können. Um die Amplifikation von Loci mit Nukleotidunterschieden zu den verwendeten *Primern* zu begünstigen, wurden die PCRs unter niedrig stringenten Bedingungen bei niedriger *Primer-Annealing*-Temperatur durchgeführt.

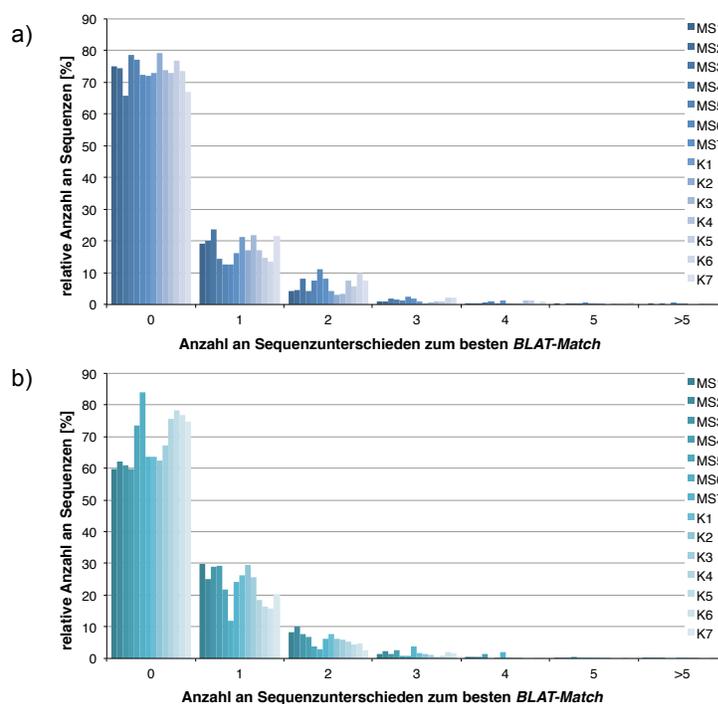
Die gewählten *Primer* wurden zunächst auf genomischer DNA getestet. Um einen möglichen Einfluss der später verwendeten 454/FLX-Adaptoren auf die PCR-Amplifikation auszuschließen, wurden die Amplikons 5'*env* und 3'*env* unter Verwendung von *Primern* mit und ohne Adaptoren an ihrem 5'-Ende von genomischer DNA amplifiziert. Die PCR-Produkte wurden gelelektrophoretisch aufgetrennt, aufgereinigt, in *pGEM<sup>®</sup>-T Easy* kloniert und in *E. coli* DH5a transformiert. *Insert*-tragende Plasmide wurden im Anschluss in Kolonie-PCRs identifiziert, aufgereinigt und sequenziert. Die generierten Sanger-Sequenzen wurden mittels *BLAT* (Kent, 2002) im menschlichen hg18 Referenzgenom rückkartiert. Wie bereits im Detail für HERV-K(HML-2) beschrieben (vgl. Abschnitt 4.1.3.), wurden auch im Falle von HERV-W die *Primer*-bindenden-Regionen vor der Rückkartierung entfernt und es gingen nur eindeutig rückkartierbare Sequenzen mit einem einzigen besten *BLAT-Match* in die weitere Analyse ein. Aufgrund der geringeren Größen der 5'*env*- und 3'*env*-Amplikons wurde die Grenze bei drei erlaubten Nukleotidunterschieden zum besten *BLAT-Match* gesetzt, wodurch Sequenzunterschiede von bis zu ~1 % toleriert wurden, die durch Fehler bei der PCR-Amplifikation und Sequenzierreaktion oder durch interindividuelle Sequenzunterschiede bzw. Einzelnukleotidpolymorphismen (SNPs) begründbar wären.

In den Experimenten zur Verifizierung der PCR-*Primer* bzw. der Amplifikationsstrategie konnten 136 (5'*env*) bzw. 149 (3'*env*) von genomischer DNA generierte Sequenzen eindeutig im Genom rückkartiert werden (vgl. **Abb. 35**). Hierbei wurden 22 bzw. 83 verschiedene HERV-W-Loci mit den Amplikons 5'*env* bzw. 3'*env* identifiziert. Die mehr oder weniger gleichmäßigen relativen Klonierungsfrequenzen sprechen hierbei gegen eine präferentielle Amplifikation spezifischer Loci durch die gewählten PCR-*Primer*.



einer lokalen *BLAT*-Suche im humanen hg18 Referenzgenom rückkartiert und die besten *BLAT-Matches* wurden gefiltert, sortiert und ausgezählt (vgl. Material und Methoden, Abschnitt 3.2.3.2.2.; eine detaillierte Übersicht über die Anzahl gefilterter *454/FLX*-Sequenzen findet sich in **Anhang 11**). Das hohe evolutionäre Alter von *HERV-W* und die damit verbundene höhere Anzahl an Sequenzunterschieden zwischen den Loci ermöglichte hierbei eine relativ eindeutige Zuordnung der cDNA Sequenzen zu einem spezifischen *HERV-W*-Locus, im Gegensatz zur evolutionär jüngeren *HERV-K*(*HML-2*)-Gruppe, deren Loci deutlich ähnlichere Sequenzen aufweisen.

Eine Übersicht über die Nukleotidunterschiede der generierten cDNA-Sequenzen zu ihrem jeweils besten *BLAT-Match* findet sich in **Abbildung 36**. Die Mehrheit von durchschnittlich 74 % (*5'env*) bzw. 69 % (*3'env*) der Sequenzen konnte mit Null Nukleotidunterschieden einem *HERV-W*-Locus zugeordnet werden. Weitere durchschnittlich 18 % (*5'env*) bzw. 23 % (*3'env*) zeigten einen Nukleotidunterschied, 6 % (*5'env* und *3'env*) zwei Nukleotidunterschiede, und 1 % (*5'env*) bzw. 2 % (*3'env*) drei Nukleotidunterschiede zum besten *BLAT-Match*. Die restlichen Sequenzen, die mehr als drei Nukleotidunterschiede aufwiesen, wurden verworfen, wodurch Sequenzunterschiede bis zu etwa 1 % toleriert wurden.



**Abbildung 36:** Sequenzunterschiede der aus Hirnproben von MS-Patienten und gesunden Kontrollen generierten *HERV-W* *5'env* und *3'env* cDNA-Sequenzen zu ihrem jeweils besten *BLAT-Match*. Gezeigt ist die absolute Anzahl an Nukleotidunterschieden von ~56.000 eindeutig rückkartierbaren Sequenzen des *5'env*- (a) und *3'env*-Amplifikons (b) zu ihrem besten *BLAT-Match*. Sequenzen mit mehr als drei Nukleotidunterschieden wurden verworfen.

Insgesamt konnten 92 % der generierten Sequenzen mit weniger als drei Nukleotidunterschieden eindeutig einem HERV-W-Locus zugeordnet werden. Bei den restlichen etwa 8 % handelt es sich vermutlich um Sequenzartefakte oder potentiell *in vitro* generierte rekombinante cDNA-Sequenzen von Transkripten verschiedener HERV-W-Loci (Flockerzi *et al.*, 2007).

#### **4.2.4. Identifizierung transkribierter HERV-W-Loci in Hirnproben von MS-Patienten und gesunden Kontrollen mittels 454/FLX-Sequenzierung**

Durch 454/FLX-Sequenzierung wurden insgesamt etwa 140 HERV-W-Loci in den untersuchten Hirnproben als transkribiert identifiziert. Mit dem 5'*env*-Amplikon wurden 25 verschiedene und mit dem 3'*env*-Amplikon 137 verschiedene Loci detektiert, wobei 14 der Loci durch beide Amplikons identifiziert wurden. Eine Übersicht der mit dem 5'*env*-Amplikon identifizierten Loci findet sich in **Tabelle 20**. Die entsprechende deutlich umfangreichere Tabelle für das 3'*env*-Amplikon findet sich in **Anhang 12** sowie im *Supplemental Material* zu Schmitt *et al.* (2013). Die relativen cDNA-Frequenzen der verschiedenen HERV-W-Loci, die sich für jede Probe aus der Anzahl der cDNA-Sequenzen berechnete, die eindeutig einem Locus zugeordnet werden konnten, dividiert durch die Gesamtanzahl an eindeutig rückkartierbaren Sequenzen der Probe, wurden als relatives Maß der Transkription der verschiedenen Loci herangezogen.

Die Gesamtanzahl an transkribierten Loci war in den von MS-Patienten stammenden Proben gegenüber den Kontrollen minimal erhöht. In den MS-Proben wurden im Durchschnitt 15 bzw. 68 Loci mit den Amplikons 5'*env* bzw. 3'*env* identifiziert, in den Kontrollen 12 bzw. 60 Loci.

Die relativen cDNA-Frequenzen der verschiedenen Loci unterschieden sich stark. Für die meisten Loci, beispielsweise auf Chromosom 2q22.1, 11p15.4 oder 20q13.2, wurden nur sehr wenige Sequenzen generiert, was auf eine sehr schwache Transkription hindeutet (vgl. hierzu auch Abschnitt 5.3.2. der Diskussion). Dagegen stammte die Mehrheit der generierten cDNA-Sequenzen von vergleichsweise wenigen Loci mit deutlich höherer relativer Transkriptionsfrequenz. Im Fall des 5'*env*-Amplikons stammten etwa 96 % der generierten Sequenzen von nur sieben verschiedenen Loci. Im Falle des 3'*env*-Amplikons wurden 83 % der Sequenzen 25 verschiedenen HERV-W-Loci zugeordnet. Die übrigen 4 % (5'*env*) bzw. 17 % (3'*env*) der generierten cDNA Sequenzen verteilten sich auf die übrigen 18 (5'*env*) bzw. 112 (3'*env*) Loci, wobei jeder Locus hierbei cDNA-Frequenzen von <1 % aufwies.

**Tabelle 20:** cDNA-Frequenzen und Struktur der in Hirnproben von MS-Patienten und gesunden Kontrollen mit dem 5'env Amplikon als transkribiert identifizierten HERV-W-Loci.\*

Chr	HERV-W-Locus		LTR-Struktur		P	MS-Proben										Kontrollen							ENCODE Caltech RNA-Seq Datensatz											
	Start	Ende	Bande	Strang		5'-LTR	3'-LTR	M51	M52	M53	M54	M55	M56	M57	K1	K2	K3	K4	K5	K6	K7	MW	1	2	3	4	5	6	7					
1	42182714	42188567	1p34.2	+	-	-	1,36	1,72	0,89	0,05		0,16	1,13	0,13	0,89						0,72	1	2	3	4	5	6	7						
1	55149269	55157786	1p32.3 <sup>b</sup>	+	-	+	0,19	0,42	0,14			1,54	0,02	0,05							0,35	5	10	1	3	2	4	7						
1	210095762	210098886	1q32.3	-	-	+	0,32		0,07	0,1			0,14	0,06	0,05						0,12	2						5						
2	136675926	136677399	2q22.1	-	-	+															0,05	6						2						
2	175898116	175899503	2q31.1 <sup>b</sup>	+	-	+															0,03							1						
3	27017519	27022905	3p24.1	+	-	+	0,46	0,23	0,22				0,03		0,05	0,14					0,24	2	1	1	1	1	1	4						
3	97868008	97876978	3q11.2	+	-	+	0,19	0,05	0,05												0,09	3	2	1	2	1	2	3						
3	143021321	143023045	3q23 <sup>b</sup>	-	-	+	0,02	0,19	0,2												0,12	1	1	1	1	1	1	2						
4	74010156	74017767	4q13.3	-	+	+	0,17	0,23	0,02	0,15											0,14	3	1	3	2	2	2	5						
4	139762390	139767803	4q31.1	-	-	+	0,61	0,65	1,01	0,35											0,97	1,21	0,74	0,14				3						
4	165796233	165798754	4q32.3	+	-	+	0,58	0,19	0,19				2,16	2,91	0,04	0,13					0,99	1	1	1	2	1	1	1						
5	44146946	44150936	5p12	+	+	+	0,02						0,03								0,03							2						
5	56851606	56854473	5q11.2 <sup>b</sup>	-	-	+	0,8	1,35	2,35				0,43	0,49	0,25	0,06	0,05	0,14			0,17	0,61	1	5	1	10	1	1						
5	107937705	107942182	5q21.3	-	-	+	0,1	0,09	0,34												0,18	2	1	1	3	2	2	2						
6	106782704	106790382	6q21 <sup>b</sup>	+	-	+	13,15	10,34	13,66	12,52	8,75	11,15	33,27	14,07	9,84	8,57	8,06	11,79	9,03	6,77	12,21	3	11	3	8	2	3	13						
7	91935248	91945442	7q21.2 <sup>b</sup>	-	+	+	38,44	44,2	35,98	31,72	40,9	38,74	24,72	38,55	46,95	52,31	58,42	56,6	59,66	54,82	44,43	11	3	11	10	5	8	7						
9	113138513	113140280	9q31.3	-	-	+	0,42	0,6	0,95	1,21											0,67													
11	9325928	9327740	11p15.4	-	-	+			0,14												0,08	1	4	1	2	2	2							
12	49582525	49589417	12q13.13	-	+	+	0,22	1,39	1,03	0,6	0,91	2,16									2,12	1,14	0,95	1,29	1,18	2,73	0,85	1,27	2	5	2	4	4	5
14	445626438	44562664	14q21.3 <sup>b</sup>	-	-	+	20,57	24,91	21,25	29,23	26,11	18,07	16,85	24,89	26,21	22,19	26,19	20,38	17,94	22,54	6	1	2	2	1	2	2	3						
15	53384371	53391866	15q21.3 <sup>b</sup>	-	-	+	9,14	4,64	9,08	15,06	9,66	12,99	15,62	7,5	4,63	2,68	0,58	3,3	0,84	2,2	6,99	16	17	7	5	10	14	15						
17	32764002	32767735	17q12 <sup>b</sup>	-	-	+	1,19	2,74	3,31	0,55	0,1	0,43	1,15	1,88	1,65	2,1	1,58	0,59			1,44	1	1	1	1	1	2	1						
20	53398912	53402997	20q13.2	+	-	+															0,05	3	9	2	2	3	1	1						
X	7627209	7632109	Xp22.31	+	-	+	0,29			0,05	0,05										0,13	3	2	2	2	2	1	2						
X	106182016	106184757	Xq22.3 <sup>b</sup>	-	-	+	12,18	6,26	9,73	8,58	12,07	13,53	3,19	7,84	6,85	8,94	3,02	5,07	4,2	16,92	8,46	2	8	2	43	8	9	2						
							Gesamtanzahl an Sequenzen										18							18										
							Anzahl verschiedene Loci										12							12										
							durchschnittliche Anzahl verschiedener Loci										9							9										

\*Das 5'env-Amplikon wurde aus cDNA von je sieben MS- und Kontroll-Proben (K) amplifiziert und mittels 454/FLX-Technologie Sequenzen generiert und im menschlichen Referenzgenom rückkartiert (vgl. Text). Angegeben sind in den ersten sechs Spalten für jeden Locus die hg18 Chromosomenposition und -bande (hochgestellte Kleinbuchstaben kennzeichnen Banden, in denen in den weiteren Analysen mehr als ein Locus als transkribiert identifiziert wurde), der Strang sowie der offizielle HGNC-Name (vgl. Mayer et al., 2011). Fett geschriebene Loci wurden sowohl mit dem 5'env- als auch mit dem 3'env-Amplikon als transkribiert identifiziert. Die Struktur der LTRs der Loci ist in den folgenden beiden Spalten beschrieben. „+“ und „-“ kennzeichnen hierbei das Vorhandensein bzw. Fehlen des kompletten LTRs. Das Fehlen bestimmter Bereiche der LTRs ist mit „Δ“ gekennzeichnet, wobei sich die Zahlen auf die Position in der bei Repbase hinterlegten LTR17 Referenzsequenz beziehen. In der Spalte „P“ sind Loci mit einem „+“ gekennzeichnet, die durch L1-vermittelte Retrotransposition generierte Pseudogene darstellen. Für alle als transkribiert identifizierten Loci sind die relativen cDNA-Frequenzen in [%] in den verschiedenen Proben angegeben. Diese berechnet sich für jeden Locus in jeder Probe aus der Anzahl an eindeutig dem Locus zugeordneten cDNA-Sequenzen dividiert durch die Gesamtanzahl an eindeutig rückkartierbaren Sequenzen, die für die Probe generiert wurden. Unter „MW“ ist ein Mittelwert der relativen cDNA-Frequenzen für jeden Locus angegeben. Unter der Tabelle findet sich für jede Probe die Gesamtzahl an eindeutig rückkartierbaren Sequenzen sowie die Anzahl an transkribierten Loci. In den rechten Spalten ist die Anzahl an ENCODE RNA-Seq-Sequenzen aus sieben verschiedenen Zelllinien angegeben, die mit den hier identifizierten Loci überlappen. 1: GM12878, 2: 1 H1 hESC, 3: HUVEC, 4: HeLa-S3, 5: Hep G2, 6: K562, 7: NHEK (zur weiteren Erklärung, s. Text)

#Loci, die zuvor in PBMCs von MS-Patienten und gesunden Kontrollen als transkribiert beschrieben wurden (vgl. Lauffer et al., 2009).

§Im 5'-LTR findet sich eine MER11A-Insertion.

Im Zuge der offiziellen Benennung transkribierter HERV-Loci wurden, neben sechs bereits beschriebenen HERV-W-Loci (vgl. Laufer *et al.*, 2009; Mayer *et al.*, 2011), 23 weiteren der hier als transkribiert identifizierten Loci mit cDNA-Frequenzen von über 1 % offizielle HGNC-Namen zugewiesen (*ERVW-7* bis *ERVW-29*; vgl. **Tab. 20** und **Anhang 12**). Diese Namen werden im Folgenden zur Benennung der Loci verwendet. Zusätzlich ist in Klammern die jeweilige Chromosomenbande angegeben, wobei durch angehängte hochgestellte Buchstaben verdeutlicht wurde, wenn in der betreffenden Bande mehrere Loci als transkribiert identifiziert wurden.

Im Falle des 5'*env*-Amplikons zeigte der *ERVWE1/ERVW-1/syncytin-1*-Locus auf Chromosom 7q21.1 in allen Proben die höchsten cDNA-Frequenzen von durchschnittlich 44,43 %, gefolgt von den Loci *ERVW-26* (14q21.3,  $\emptyset = 22,54$  %), *ERVW-17* (6q21<sup>a</sup>,  $\emptyset = 12,21$  %), *ERVWE2/ERVW-2* (Xq22.3<sup>b</sup>,  $\emptyset = 8,46$  %) und *ERVW-4* (15q21.3,  $\emptyset = 6,99$  %). Für das 3'*env*-Amplikon zeigte sich ein ähnliches Bild mit wenigen, relativ stark transkribierten Loci, wobei sich die genauen cDNA-Frequenzen jedoch unterschieden, vermutlich aufgrund der deutlich höheren Anzahl an mit diesem Amplikon als transkribiert identifizierten Loci. Die höchsten cDNA-Frequenzen von durchschnittlich 9,8 % zeigte der *ERVW-13*-Locus auf Chromosom 2q13, gefolgt von den Loci *ERVW-20* (8q21.11,  $\emptyset = 9$  %), *ERVW-9* (1q25.2,  $\emptyset = 8,2$  %), *ERVW-12* (2p16.2,  $\emptyset = 7,6$  %), *ERVWE1/ERVW-1/syncytin-1* (7q21.2,  $\emptyset = 7$  %), *ERVW-10* (1q32.1,  $\emptyset = 6,1$  %) und *ERVW-4* (15q21.3,  $\emptyset = 5$  %). Weitere 18 Loci zeigten cDNA-Frequenzen von 1 - 3,6 %, darunter der *ERVWE2/ERVW-2*-Locus (Xq22.3<sup>b</sup>,  $\emptyset = 1,8$  %). Die relativen cDNA-Frequenzen der restlichen 112 Loci lagen jeweils unter 1 %. Einige der stärker transkribierten Loci wurden bereits in einer früheren Studie von Laufer *et al.* (2009) in PBMCs von MS-Patienten und gesunden Kontrollen als transkribiert beschrieben. Dies waren die Loci *ERVW-17* (6q21<sup>a</sup>), *ERVW-26* (14q21.3), *ERVW-4* (15q21.3) sowie *ERVWE1/ERVW-1/syncytin-1* (7q21.2) und *ERVWE2/ERVW-2* (Xq22.3<sup>b</sup>).

In den MS-Proben und Kontrollen konnten somit durch *454/FLX*-Sequenzierung etwa 140 verschiedene HERV-W-Loci als transkribiert identifiziert werden. Hierbei stammte die Mehrheit der generierten cDNA-Sequenzen von vergleichsweise wenigen verschiedenen Loci, was auf eine relativ starke Transkription dieser hindeutet.

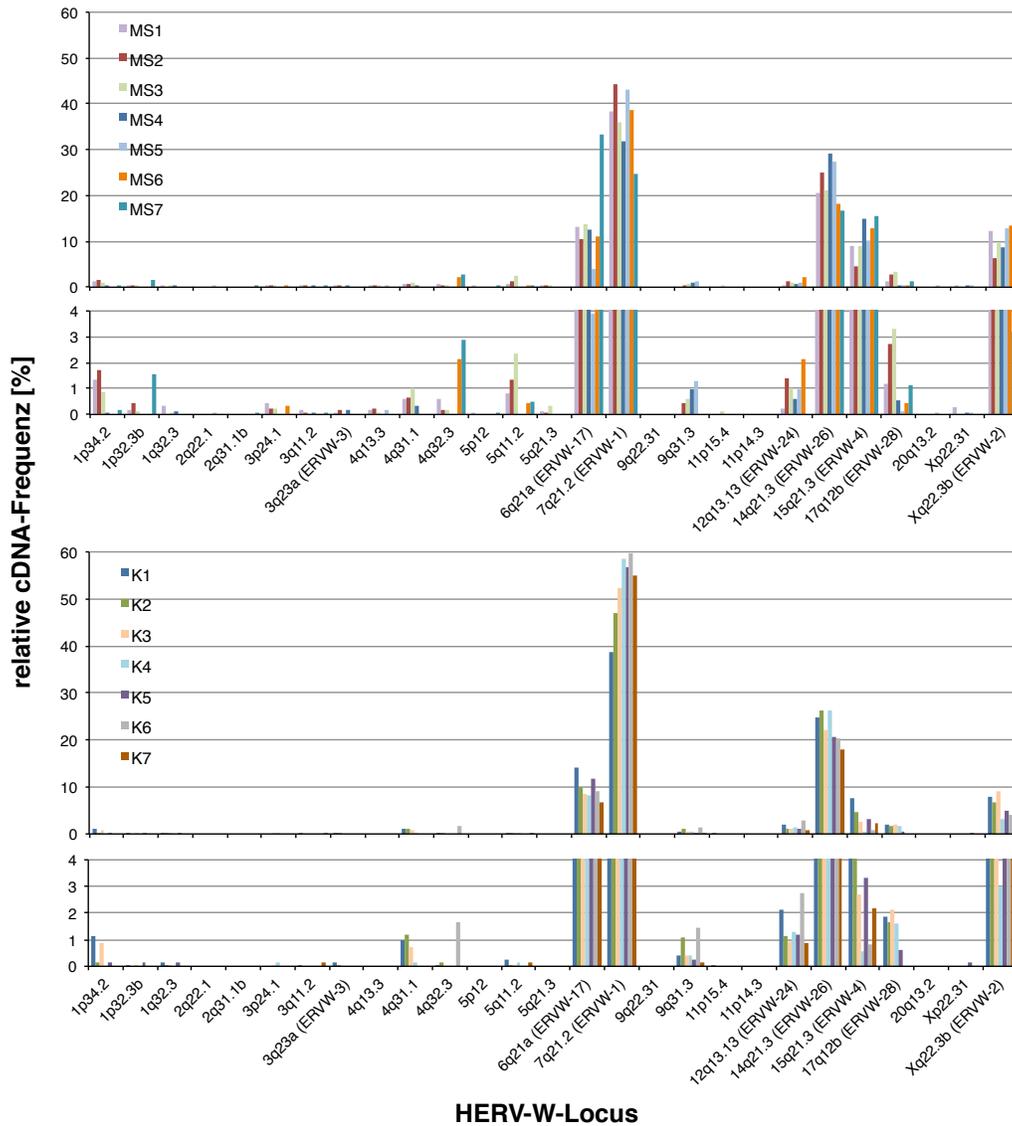
#### 4.2.5. Die HERV-W-Transkriptionsmuster in Hirnproben von MS-Patienten und gesunden Kontrollen unterscheiden sich nicht signifikant

Vergleicht man die Transkriptionsmuster der mit den Amplikons 5'*env* und 3'*env* durch 454/FLX-Sequenzierung identifizierten HERV-W-Loci zwischen den MS- und Kontrollproben, zeigten sich keine signifikanten Unterschiede. Einige wenige Loci waren ausschließlich in einigen MS- oder Kontrollproben transkribiert, darunter Loci auf Chromosom 2q22.1 (0,05 % in MS5; 5'*env*), 2q31.1<sup>b</sup> (0,03 % in MS7; 5'*env*) und 5p12 (0,02 bzw. 0,03 % in MS1 und MS7; 5'*env*) sowie 9q22.31 (1,22 % in K1; 3'*env*), wobei die niedrigen relativen cDNA-Frequenzen auf eine sehr niedrige Transkriptionsrate der Loci hindeuten.

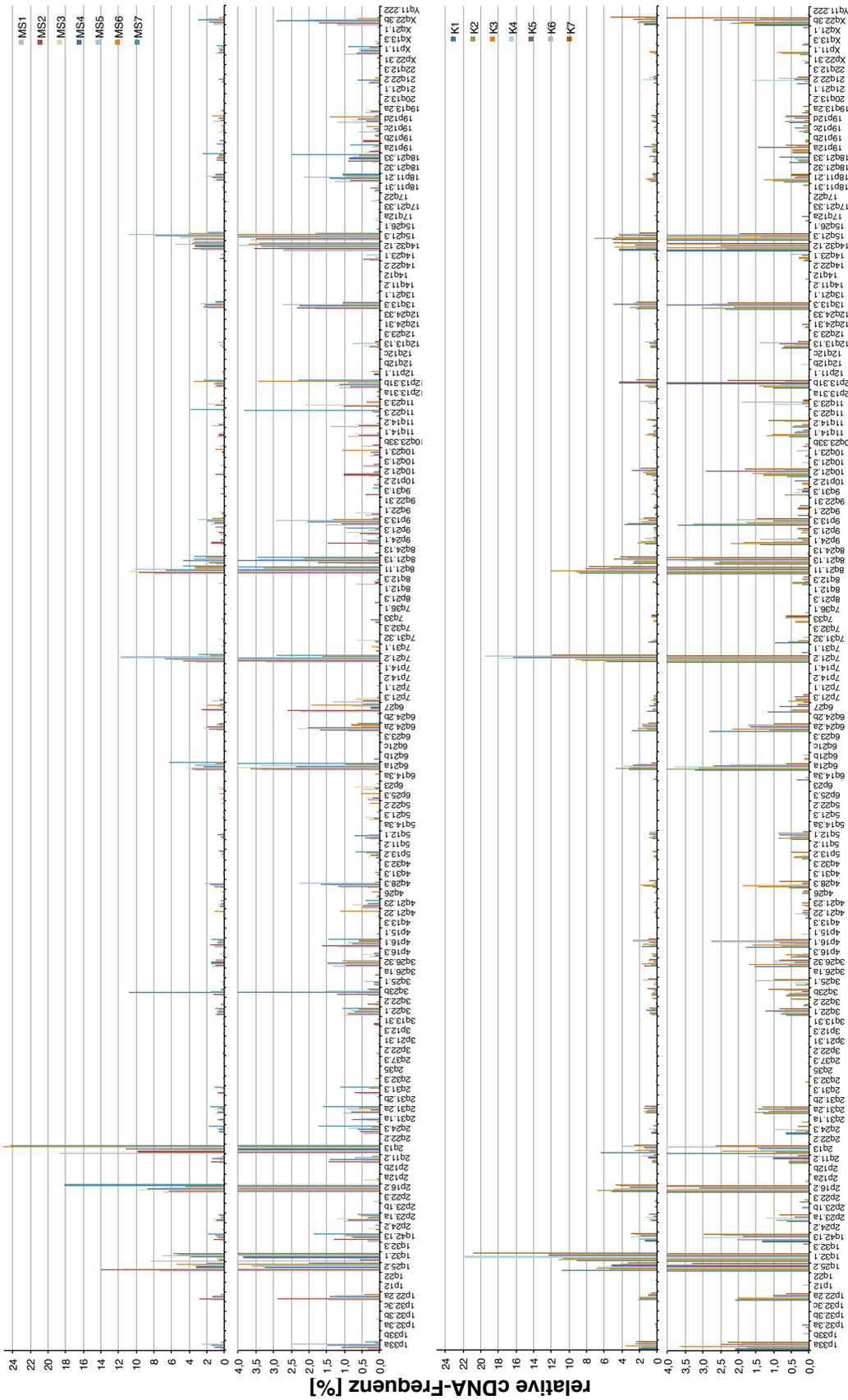
Das Transkriptionsmuster der niedrig transkribierten Loci war in allen Proben leicht unterschiedlich, wobei jedoch kein charakteristischer Unterschied zwischen MS- und Kontrollproben ersichtlich war. In einigen Proben zeigten einige Loci, die sonst nur schwach transkribiert waren, leicht erhöhte cDNA-Frequenzen. Beispielsweise zeigte ein Locus auf Chromosom 5q11.2 in drei MS-Proben 5'*env* cDNA-Frequenzen von 0,8 % bis 2,4 %, wohingegen die Transkription in fünf der Kontrollproben niedrigere Werte von maximal 0,25 % erreichte. Solche minimalen Unterschiede einiger Loci zwischen MS-Proben und Kontrollen sind vermutlich auf die generell niedrigen cDNA-Frequenzen der betreffenden Loci zurückzuführen.

Stark transkribierte Loci wurden sowohl in MS- als auch Kontrollproben detektiert. So zeigten beispielsweise im 5'*env*-Datensatz der *ERVW-26* (14q21.3)-Locus durchschnittliche cDNA-Frequenzen von 21,87 % in MS und 23,86 % in Kontrollen und die Loci *ERVWE1/ERVW-1/syncytin-1* (7q21.2) und *ERVWE2/ERVW-2* (Xq22.3<sup>b</sup>) zeigten mittlere Frequenzen von 35,49 % (*ERVW-1*) und 8,92 % (*ERVW-2*) in MS bzw. 46,11 % (*ERVW-1*) und 7,71 % (*ERVW-2*) in Kontrollen. Im 3'*env*-Datensatz zeigte beispielsweise der *ERVW-4* (15q21.3)-Locus durchschnittliche cDNA-Frequenzen von 5,08 % in MS und von 4,96 % in Kontrollen und der *ERVW-20* (8q21.11)-Locus Frequenzen von 8,81 % in MS und 9,25 % in Kontrollen.

Insgesamt zeigten sich in MS und Kontrollen sehr ähnliche HERV-W-Transkriptionsmuster (vgl. **Abb. 37** und **38**).



**Abbildung 37:** Relative cDNA-Frequenzen der mit dem 5'*env*-Amplikon in Hirnproben von MS-Patienten und gesunden Kontrollen als transkribiert identifizierten HERV-W-Loci. Das Amplikon wurde aus cDNA der Proben amplifiziert und mittels 454/FLX-Technologie wurden cDNA-Sequenzen generiert und rückkartiert (vgl. Text). Angegeben sind die relativen cDNA-Frequenzen in [%], die sich für jeden Locus in jeder Multiple Sklerose- (MS) (oberes Diagramm) und Kontrollprobe (K) (unteres Diagramm) aus der Anzahl an Sequenzen, die eindeutig dem bestimmten Locus zugeordnet werden konnten, dividiert durch die Gesamtzahl an rückkartierbaren Sequenzen der Probe berechneten (vgl. **Tab. 20**). Im oberen Teil jedes Diagramms findet sich jeweils eine Gesamtübersicht und darunter die Detailansicht der niedriger transkribierten Loci mit cDNA-Frequenzen <math>< 4\%</math>. Die Loci sind nach ihrer Chromosomen-Bande und ihrem offiziellen HGNC-Namen (vgl. Mayer *et al.*, 2011) benannt bzw. geordnet.



HERV-W-Locus

Abbildung 38: Relative cDNA-Frequenzen der mit dem 3'env-Amplikon in Hirnproben von MS-Patienten und gesunden Kontrollen als transkribiert identifizierten HERV-W Loci. Zur weiteren Erklärung vgl. Text zu Abbildung 37.

Zur Untersuchung potentiell signifikanter Transkriptionsunterschiede zwischen MS- und Kontrollproben wurde die mittlere relative Transkription aller Loci in MS- und Kontrollproben statistisch mittels Wilcoxon-Mann-Whitney-Test (U-Test) untersucht. Hier zeigten sich einige wenige Ausnahmen hinsichtlich der Transkriptionsunterschiede zwischen MS- und Kontrollproben. Der *ERVW-4*-Locus (15q21.3) zeigte für das 5'*env*-Amplikon statistisch signifikant höhere relative cDNA-Frequenzen von  $\bar{\emptyset}$  10,88 % in MS verglichen mit  $\bar{\emptyset}$  3,1 % in den Kontrollproben ( $p = 0,0082$ ). Für das 3'*env*-Amplikon zeigte sich für diesen Locus jedoch kein signifikanter Unterschied ( $p = 1$ ). Ebenfalls war die niedrigere Transkription des *ERVWE1/ERVW-1/syncytin-1*-Locus in MS ( $\bar{\emptyset}$  36,39 %) gegenüber den Kontrollen ( $\bar{\emptyset}$  52,47 %) für die 5'*env*-Daten signifikant ( $p = 0,014$ ), nicht jedoch für die 3'*env*-Daten (MS  $\bar{\emptyset}$  5,25 %, Kontrollen  $\bar{\emptyset}$  12,72 %;  $p = 0,069$ ). Im Falle der nur mit dem 3'*env*-Amplikon identifizierten Loci zeigten sich einige signifikante Unterschiede. Der *ERVW-10*-Locus (1q32.1) zeigte eine signifikant höhere Transkription in den Kontrollen ( $\bar{\emptyset}$  14,02 %) gegenüber den MS-Proben ( $\bar{\emptyset}$  4,55 %;  $p = 0,015$ ) und der *ERVW-13*-Locus (2q13) zeigte eine signifikant höhere Transkription in den MS-Proben ( $\bar{\emptyset}$  16,97 %) gegenüber den Kontrollen ( $\bar{\emptyset}$  2,47 %;  $p = 0,015$ ). Der *ERVW-17*-Locus (6q21<sup>a</sup>) zeigte für das 5'*env*-Amplikon in der Probe MS7 mit 33 % eine deutlich höhere cDNA-Frequenz als in den anderen Proben ( $\bar{\emptyset}$  11,6 %) und den Kontrollen ( $\bar{\emptyset}$  9,7 %). Ähnliche Unterschiede zeigten sich für diesen Locus auch in den mittels *MiSeq/Illumina* generierten Daten (s.u.), wo die Transkription in MS7 bei 24 % lag, gegenüber  $\bar{\emptyset}$  11,5 % in den restlichen MS-Proben und  $\bar{\emptyset}$  12,75 % in den Kontrollen.

Jede untersuchte Probe zeigte somit ein leicht unterschiedliches Muster transkribierter HERV-W-Loci, wobei die Unterschiede hauptsächlich in den schwächer transkribierten Loci zu finden sind. Die Loci mit hohen relativen cDNA-Frequenzen finden sich sowohl in MS-Proben als auch in den gesunden Kontrollen. Insgesamt zeigten sich keine charakteristischen Unterschiede im Transkriptionsmuster zwischen den MS-Proben und Kontrollen.

#### **4.2.6. Durch *MiSeq/Illumina*-Sequenzierung konnten HERV-W-Transkriptionsmuster in MS-Proben und gesunden Kontrollen bestätigt und weitere transkribierte Loci identifiziert werden**

Zur Überprüfung der Zuverlässigkeit der Sequenzierung bzw. der Reproduzierbarkeit der Transkriptionsmuster wurden transkribierte HERV-W-Loci in je vier der MS- und Kontrollproben (MS4-7 und K4-7) zusätzlich mittels *MiSeq/Illumina*-Technik sequenziert. Hierzu wurden die Amplikons 5'*env* und 3'*env* aus der schon zur Amplifikation der 454/*FLX*-Amplikons

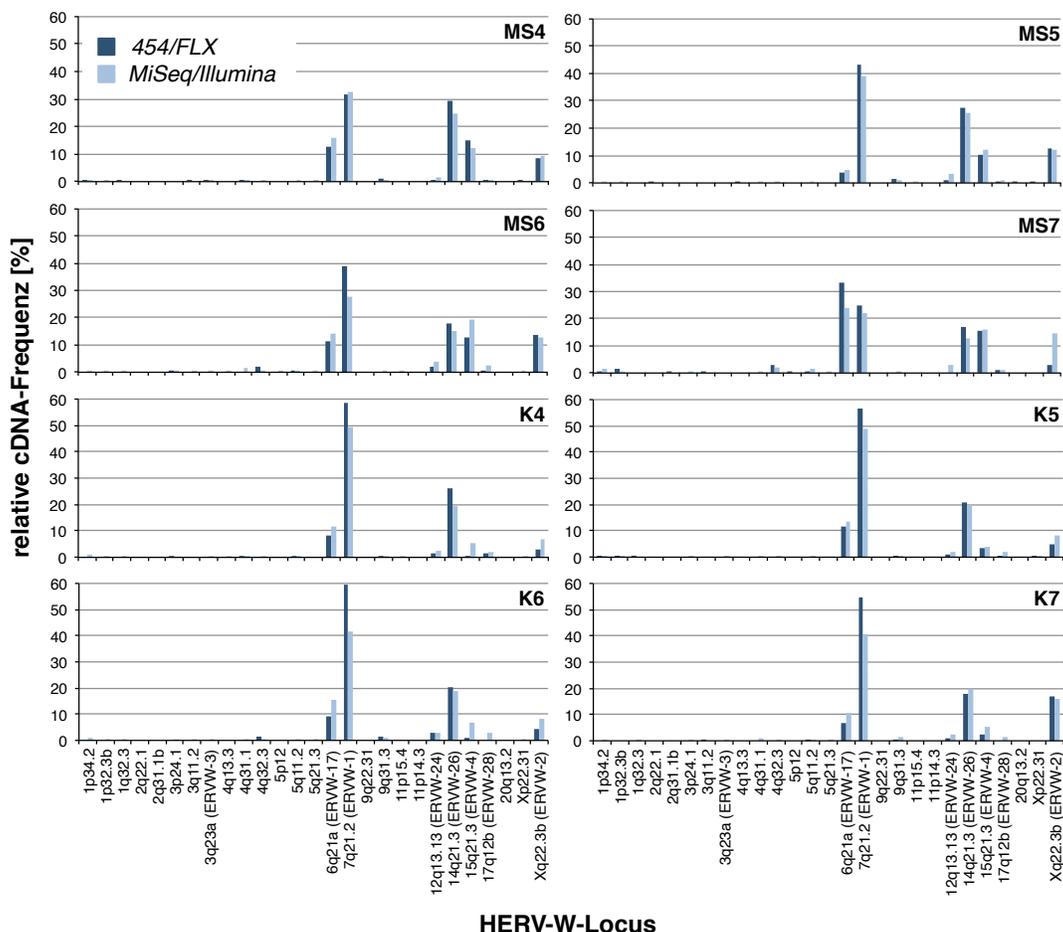
verwendeten cDNA der Proben amplifiziert und aufgereinigt (vgl. hierzu Abschnitt 4.2.1. und **Anhang 13**). Zur Amplifikation wurden in diesem Fall PCR-Primer ohne *Adaptor*-Sequenzen verwendet, diese wurden erst nachträglich im Zuge der *Library Preparation* ligiert.

Insgesamt wurden über 8,5 Millionen Sequenzen generiert. Wie im Detail im Material und Methoden Teil beschrieben (vgl. Abschnitt 3.2.3.2.2.), wurden Sequenzen <185 bp aus dem Datensatz gelöscht, *Primer*-Sequenzen wurden entfernt und die cDNA-Sequenzen in einer lokalen *BLAT*-Suche im humanen hg18 Referenzgenom rückkartiert. Sequenzen mit mehr als einem besten *BLAT-Match* oder mehr als drei Nukleotidunterschieden zu ihrem besten *BLAT-Match* wurden verworfen. Es blieben ~6,2 Millionen Sequenzen, die in die weitere Analyse eingingen (eine detaillierte Übersicht über die Anzahl gefilterter Sequenzen jeder Probe findet sich in **Anhang 14**).

Die *MiSeq/Illumina*-Sequenzierung lieferte insgesamt ein ähnliches Bild transkribierter HERV-W-Loci wie die *454/FLX*-Sequenzierung (vgl. **Abb. 39**). Insgesamt wurden 183 verschiedene HERV-W-Loci als transkribiert identifiziert, 27 Loci mit dem 5'*env*-Amplikon und 176 mit dem 3'*env*-Amplikon, wobei 20 der Loci mit beiden Amplikons identifiziert wurden (die cDNA-Frequenzen der mit dem 5'*env*-Amplikon identifizierten Loci finden sich in **Anhang 15**, die Frequenzen der mit dem 3'*env*-Amplikon identifizierten Loci in **Anhang 12**). In den MS-Proben waren im Schnitt 18 (5'*env*) bzw. 124 (3'*env*) verschiedene Loci transkribiert, in den Kontrollen 20 (5'*env*) bzw. 117 (3'*env*) Loci. Die verglichen mit den *454/FLX*-Ergebnissen deutlich höhere Anzahl an identifizierten Loci ist vermutlich auf die deutlich höhere Zahl an generierten und letztendlich rückkartierten cDNA-Sequenzen zurückzuführen.

Insgesamt waren die relativen cDNA-Frequenzen der detektierten Loci sehr ähnlich zu den Daten der *454/FLX*-Sequenzierung (vgl. Abschnitt 4.2.4. und 4.2.5. und **Anhang 12** und **15**). Die Mehrheit der Loci zeigte niedrige cDNA-Frequenzen, was auf eine sehr niedrige Transkriptionsrate hindeutet, wohingegen der Großteil an Sequenzen von relativ wenigen Loci mit deutlich stärkerer Transkription stammte, darunter u.a. der *ERVWE1/ERVW-1/syncytin-1*-Locus. Für die MS- und Kontrollproben zeigten sich sehr ähnliche Transkriptionsmuster für die stark transkribierten Loci.

Die mittels *454/FLX*-Sequenzierung ermittelten Transkriptionsmuster in MS-Proben und Kontrollen konnten somit insgesamt durch zusätzliche *MiSeq-Illumina*-Sequenzierung von je vier MS- und Kontrollproben bestätigt werden. Aufgrund der deutlich höheren Zahl generierter Sequenzen konnten zudem insgesamt noch etwa 40 weitere Loci identifiziert werden.



**Abbildung 39:** Relative cDNA-Frequenzen des aus Hirnproben von MS-Patienten und gesunden Kontrollen amplifizierten HERV-W 5'env-Amplikons nach 454/FLX- sowie MiSeq/Illumina-Sequenzierung. Gezeigt sind für jede Probe die relative Frequenzen in [%] aller in dieser als transkribiert identifizierten Loci, die der Anzahl an cDNA-Sequenzen, die dem jeweiligen Locus zugeordnet werden konnten, dividiert durch die Gesamtzahl an eindeutig rückkartierbaren Sequenzen der Probe entsprechen. In dunkelblau sind die Ergebnisse der 454/FLX-Sequenzierung gezeigt (vgl. Tab. 20 und Abb. 34), in hellblau die der MiSeq/Illumina-Sequenzierung derselben vier Multiple Sklerose (MS) und Kontrollproben (K).

#### 4.2.7. Struktur und Genomlokalisierung der transkribierten HERV-W-Loci

Zur weiteren Charakterisierung wurden die als transkribiert identifizierten HERV-W-Loci hinsichtlich ihrer Struktur und Genomlokalisierung weiter analysiert. Insbesondere ist hierbei das Vorhandensein von LTRs von Interesse, die aufgrund der potentiellen Promotorsequenzen die Transkription eines Locus erklären können. Im Falle des Fehlens von flankierenden LTR-Sequenzen können HERV-W-Transkripte auch durch Promotorsequenzen benachbarter Gene generiert worden sein.

Die meisten der als transkribiert identifizierten HERV-W-Loci sind unvollständig, da sie durch einen L1-vermittelten Retrotranspositionsmechanismus entstanden (vgl. Abb. 40). Solche

prozessierten Pseudogene zeichnen sich durch Deletion der U3-Region des 5'-LTRs oder größerer Bereiche im 5'-Bereich des Provirus aus, sowie durch Deletion der U5 Region im 3'-LTR. Lediglich 35 der mehr als 170 als transkribiert identifizierten HERV-W-Loci stellen Proviren dar, die noch beide LTRs besitzen. Aufgrund des hohen evolutiven Alters besitzen alle HERV-W-Loci Deletionen unterschiedlicher Länge innerhalb der proviralen Region oder Insertionen zahlreicher repetitiver Elemente. Eine Ausnahme zu solchen stark geschädigten retroviralen Strukturen stellt der *ERVWE1/ERVW-1/syncytin-1*-Locus (7q21.2) dar, der noch einen physiologisch bedeutsamen intakten Leserahmen für das Syncytin-1 Protein besitzt.

Betrachtet man die chromosomale Lokalisation der transkribierten Loci, so ist die Mehrheit nicht näher mit zellulären Genen assoziiert, d.h. sie finden sich nicht intronisch in Genen, sondern sind oft mehrere bis viele kb von diesen entfernt (eine detaillierte Übersicht über die Genomlokalisierungen und benachbarte Gene der Loci findet sich in **Anhang 15** und **16**). Einige Loci liegen hingegen intronisch in zellulären Genen, darunter der *ERVW-17*-Locus (6q21<sup>a</sup>), der in einem ~46 kb großen Intron des *ATG5*-Gens lokalisiert ist und im Falle der *454/FLX*-Sequenzierung des 5'*env*-Amplikons eine vergleichsweise hohe cDNA-Frequenz von Ø 12 % zeigte. Der *ERVW-26*-Locus (14q21.3), der bei *454/FLX*-Sequenzierung des 5'*env*-Amplikons eine durchschnittliche cDNA-Frequenz von Ø 23 % zeigte, liegt in einem 16 kb großen Intron des *FAM179B*-Gens. Die Loci *ERVW-9* (1q25.2), *ERVW-12* (2p16.2) und *ERVW-10* (1q32.1), die bei *454/FLX*-Sequenzierung des 3'*env*-Amplikons durchschnittliche cDNA-Frequenzen von 8 %, 7,6 % und 6 % zeigten, liegen intronisch in den Genen *RASAL2*, *ASB3* und *LOC284581*. Auch unter den niedrig transkribierten Loci finden sich Beispiele, die in Introns von Genen lokalisiert sind, darunter ein Locus auf Chromosom 1p34.2 (Ø 1,6 % für 5'*env*; *454/FLX*) im *ACACA*-Gen oder ein Locus auf Chromosom 8q12.3 (Ø 0,2 % für 3'*env*; *454/FLX*) in *CYP7B1*.

In einigen Fällen sind intronisch gelegene HERV-W-Loci Teile von (alternativen) Transkripten der entsprechenden Gene. Im Falle des *FAM179B*-Gens ist eine mRNA-Sequenz (*Genbank acc. no.* CR749557) als alternative Spleißvariante mit HERV-W-Anteilen annotiert, die einen 65 bp langen HERV-W-Bereich aufweist, wobei der *ERVW-26*-Locus *splice donor* und *acceptor sites* besteuert. Ein Locus auf Chromosom 22q22.2, der bei *454/FLX*-Sequenzierung des 3'*env*-Amplikons als schwach transkribiert identifiziert wurde, bildet den Transkriptionstart, und *splice donor* und *acceptor sites* für 5'UTR bzw. die Exons 1 und 2 des *IGSF5*-Gens.



**Abbildung 40:** Struktur der in Hirnproben von MS-Patienten und gesunden Kontrollen als transkribiert identifizierten HERV-W-Loci. Gezeigt ist die graphische Darstellung eines multiplen Sequenz-Alignments aller als transkribiert identifizierten HERV-W-Loci mit den bei *Repbase* hinterlegten Konsensussequenzen für LTR17/HERV17/LTR17, die als oberste Sequenz in der Abbildung dargestellt ist. Retrovirale Bereiche sind durch schwarze Balken verdeutlicht, Lücken durch schwarze horizontale Linien. Die Positionen der LTRs, der retroviralen ORFs für *gag*, *pro*, *pol* und *env* sowie der hier untersuchten Amplikons 5'*env* und 3'*env* sind über dem *Alignment* eingezeichnet. Zur weiteren Erklärung, s. Text.

Weitere HERV-W-Loci sind, bezüglich der Transkriptionsrichtung der Gene, stromab in relativ enger Nachbarschaft zu Genen lokalisiert. Hierzu zählt der Locus auf Chromosom 19p12<sup>c</sup> ( $\emptyset = 0,2\%$  für 3'env; 454/FLX), der etwa 100 bp stromab des *ZNF99*-Gens lokalisiert ist, der *ERVW-27*-Locus (14q32.12;  $\emptyset = 3,4\%$  für 3'env; 454/FLX) etwa 300 bp stromab des *C14orf159*-Gens und der *ERVW-3*-Locus (3q23a;  $\emptyset = 0,07\%$  für 3'env; *MiSeq/Illumina*) 2,7 kb stromab des *GRK7*-Gens. Ein weiteres Beispiel ist der stärker transkribierte *ERVWE2/ERVW-2*-Locus (Xq22.3b;  $\emptyset = 8,5\%$  für 5'env; 454/FLX), der etwa 7 kb stromab des *RBM41*-Gens liegt.

Zudem finden sich einige transkribierte Loci zwischen zwei benachbarten Genen, die beide jeweils in Richtung des HERV-W-Locus transkribiert werden. Dazu zählt der *ERVW-29*-Locus (18p11.21;  $\emptyset = 1,1\%$  für 3'env; 454/FLX), der 3,9 kb stromab des *LDLRAD4*- und 5,5 kb stromauf des *FAM210A*-Gens liegt, sowie ein Locus auf Chromosom 2q31.2<sup>b</sup> ( $\emptyset = 0,004\%$  für 3'env; 454/FLX), der 1 kb stromab des *DFNB59*-Gens und und 1,8 kb stromauf des *FKBP7*-Gens lokalisiert ist.

Aufgrund der Bildung durch L1-vermittelte Retrotransposition weisen die meisten der als transkribiert identifizierten HERV-W-Loci charakteristische Deletionen in den LTR-Bereichen sowie teilweise noch größere Deletionen im 5'-Bereich auf. Dennoch könnten auch potentiell vorhandene Promotorsequenzen in verkürzten LTRs eine Transkription des betreffenden Locus erklären. Weitere Loci sind innerhalb oder in direkter Nachbarschaft von Genen lokalisiert, wobei die Detektion von HERV-Transkripten hier durch ungespleißte RNA oder (alternative) Gen-Transkripte erklärbar wäre.

#### 4.2.8. Identifizierung transkribierter HERV-W-Loci in *ENCODE RNA-Seq* Datensätzen

Aufgrund der im Vergleich zu früheren Studien (Laufer *et al.*, 2009) sehr hohen Anzahl an detektierten HERV-W-Loci wurde nach weiteren Belegen für deren tatsächliche Transkription gesucht. Die als transkribiert identifizierten Loci wurden hierzu mit Transkriptom-Daten des *Encyclopedia of DNA Elements* (ENCODE)-Projektes verglichen (ENCODE Project Consortium, 2004). Daten von im Rahmen dieses Projektes generierten RNA-Sequenzen sind im *UCSC Genome Browser* (Kent *et al.*, 2002) hinterlegt. Mittels *UCSC Table Browser* (Karolchik *et al.*, 2004) wurden aus Poly(A)-RNA generierte und im Genom eindeutig rückkartierte 75 bp lange *RNA-Seq*-Sequenzen gefiltert, die mit den Chromosomenkoordinaten der als transkribiert identifizierten HERV-W-Loci überlappen. Hierbei wurden Daten aus den Zelllinien GM12878, 1 H1 hESC, HUVEC, HeLa-S3, Hep G2, K562 und NHEK analysiert.

Mit Ausnahme von sechs Loci auf Chromosom 2q37.3, 5q14.3b, 9q31.3, 10q21.3, 14q2 und 22q12.3 waren alle Loci auch durch *RNA-Seq*-Sequenzen in mindestens einer der untersuchten Zelllinien vertreten (vgl. **Tab. 20** für das 5'*env*-Amplikon; die Daten für die mit dem 3'*env*-Amplikon identifizierten Loci finden sich in **Anhang 16** sowie im *Supplemental Material* zu Schmitt *et al.*, 2013). Die Anzahl an in jeder der Zelllinien transkribierten HERV-W-Loci war hierbei vergleichbar mit der Anzahl transkribierter Loci in den untersuchten Hirnproben. Im Falle der 5'*env*-Daten waren im Durchschnitt 19 verschiedene HERV-W-Loci in den Hirnproben transkribiert und auch in den *RNA-Seq*-Datensätzen der Zelllinien waren durchschnittlich 19 dieser Loci vertreten. Im Falle des 3'*env*-Amplikons, wo im Schnitt 121 verschiedene Loci in den Hirnproben als transkribiert identifiziert wurden, waren durchschnittlich 111 der Loci in den *RNA-Seq*-Datensätzen vertreten. Es überlappten bis zu 135 annotierte *RNA-Seq*-Sequenzen (13q21.1 in 1 H1 hESC) mit in den Hirnproben als transkribiert identifizierten HERV-W-Loci. Mit den Chromosomenkoordinaten der zuvor beschriebenen relativ stark transkribierten Loci *ERVWE1/ERVW-1/syncytin-1* und *ERVWE2/ERVW-2* überlappten zwischen drei (1 H1 hESC) und elf (GM12878 und HUVEC) bzw. zwischen zwei (GM12878, HUVEC und NHEK) und 43 (HeLa-S3) *RNA-Seq*-Sequenzen. In einigen Fällen waren weitere Loci, die zuvor als relativ stark transkribiert beschrieben wurden, auch durch eine vergleichsweise hohe Anzahl an *RNA-Seq*-Sequenzen repräsentiert, beispielsweise *ERVW-12* ( $\emptyset = 6,71\%$  für 3'*env*; 454/FLX) mit Sequenz-Anzahlen zwischen 14 (HeLa-S3) und 76 (Hep G2), *ERVW-1* ( $\emptyset = 44,43\%$  für 5'*env*; 454/FLX) mit Sequenz-Anzahlen zwischen drei (1 H1 hESC) und elf (HUVEC) oder *ERVW-4* ( $\emptyset = 9,41\%$  bzw.  $4,71\%$  für 5'*env* bzw. 3'*env*; 454/FLX) mit Sequenz-Anzahlen zwischen fünf (HeLa-S1) und 17 (1 H1 hESC).

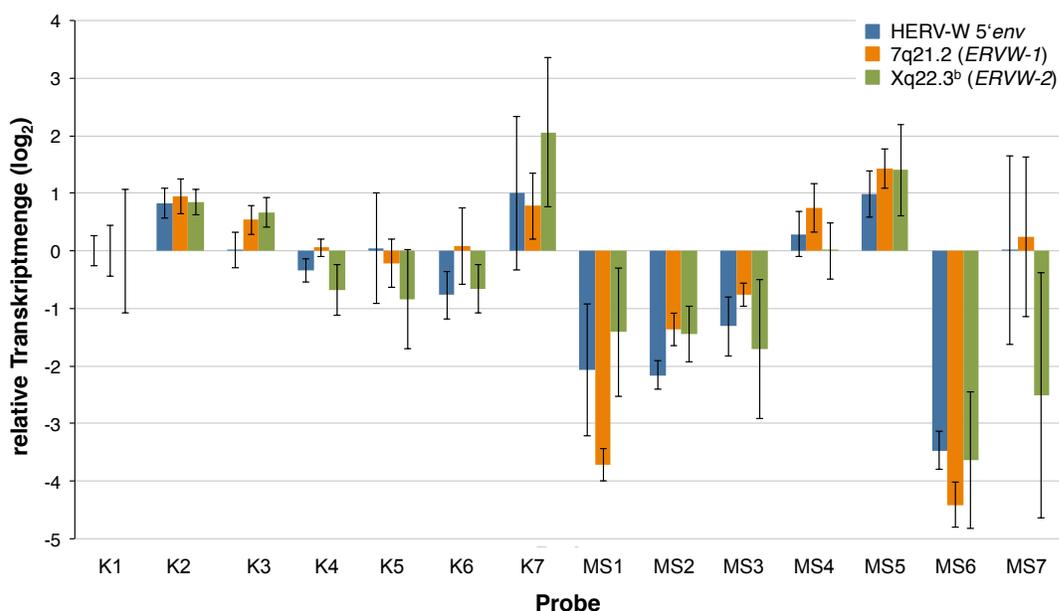
Die im Zuge des *ENCODE*-Projektes generierten *RNA-Seq*-Daten liefern somit zusätzliche Hinweise für die Transkription einer vergleichsweise hohen Anzahl verschiedener HERV-W-Loci, auch in verschiedenen Zelllinien ohne direkten Bezug zu Multipler Sklerose.

#### **4.2.9. Die Quantifizierung der HERV-W-Transkriptmengen lieferte keine spezifischen Transkriptionsunterschiede zwischen MS-Proben und gesunden Kontrollen**

Da in früheren Studien eine erhöhte HERV-W Transkription im Kontext von MS berichtet wurde (Antony *et al.*, 2006, 2007; Mameli *et al.*, 2007), die Ergebnisse der Studien jedoch teils widersprüchliche Ergebnisse lieferten, wurden in der vorliegenden Arbeit die relativen HERV-W-Transkriptmengen in allen MS- und Kontrollproben mittels *Real-Time-PCR* quantifiziert. Es wurde hierbei sowohl die Gesamt-Transkriptmengen von HERV-W, basierend auf dem 5'*env*-

Amplikon, sowie die spezifischen Transkriptmengen der Loci *ERVWE1/ERVW-1/syncytin-1* und *ERVWE2/ERVW-2* durch vorher etablierte Locus-spezifische *Primer* (de Parseval *et al.*, 2003; Radonic *et al.*, 2004; Laufer *et al.*, 2009) quantifiziert. Zur Normalisierung dienten die Haushaltsgene *G6PDH* und *RPII* und die Transkription in allen Proben wurde mit der gesunden Kontrolle K1 verglichen und relativ zu dieser berechnet.

Die drei untersuchten Amplikons zeigten für die meisten Proben eine gleichmäßig erhöhte oder erniedrigte Transkription verglichen mit K1 (vgl. **Abb. 41**). Zwischen den Proben zeigten sich jedoch interindividuelle Unterschiede und die Transkription unterschied sich um das bis zu ~52-fache. Die relative Transkriptmenge des *ERVWE1/ERVW-1/syncytin-1*-Locus in MS6 betrug beispielsweise nur 5 % der Menge in K1, wohingegen die Transkriptmenge in MS5 gegenüber K1 um das ~2,5-fache erhöht war.



**Abbildung 41:** Quantifizierung der HERV-W Transkriptmengen mittels *Real-Time-PCR*. Die relative Transkriptmenge von HERV-W 5'env sowie der Loci *ERVWE1/ERVW-1/syncytin-1* und *ERVWE2/ERVW-2* wurde in je sieben Multiple Sklerose (MS) und Kontrollproben (K) quantifiziert (vgl. Text). Zur Normalisierung dienten die Haushaltsgene *G6PDH* und *RPII* und als Referenz die Transkriptmenge in der Kontrolle K1. Alle Proben wurden in Triplikaten gemessen. Die schwarzen Balken verdeutlichen die minimal und maximal gemessenen Werte der relativen Transkriptmengen.

Insgesamt zeigte sich kein einheitliches Bild, das die MS- von den Kontrollproben unterschied. Fünf der MS-Proben zeigten gegenüber der Kontrolle K1 niedrigere relative Transkriptmengen aller drei Amplikons, bzw. im Falle von MS7 eine niedrigere relative Transkriptmenge des *ERVWE2/ERVW-2* Locus. Die restlichen drei MS-Proben zeigten erhöhte relative Transkriptmengen aller Amplikons gegenüber K1, die jedoch mit denen der Kontrollen K2, K3 oder K7 vergleichbar waren. Der statistische Vergleich der relativen Transkriptmengen aller

Amplikons zwischen MS- und Kontrollproben lieferte keinen signifikanten Unterschied (Wilcoxon-Mann-Whitney-Test, bereinigt,  $p > 0,22$ ).

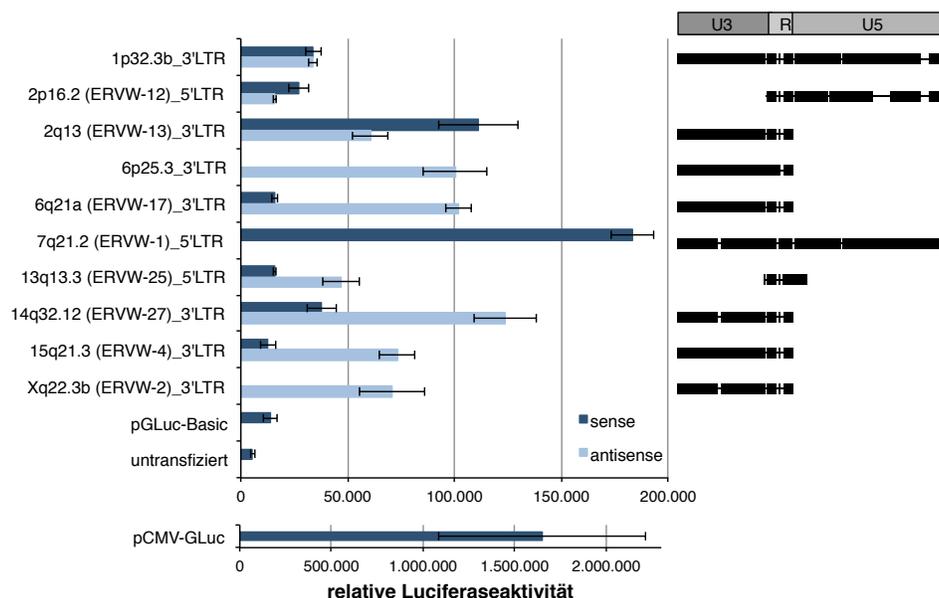
Die Quantifizierung der HERV-W-Transkriptmengen in den untersuchten Hirnproben deutet somit auf große interindividuelle Unterschiede bzgl. der relativen HERV-W- Gesamt-Transkription sowie der locuspezifischen HERV-W-Transkriptionsstärken hin. Es zeigten sich keine signifikanten Unterschiede in der Transkription zwischen MS- und Kontrollproben. Dies steht im Gegensatz zu früheren Studien, die eine erhöhte HERV-W-Transkription in MS beschrieben (Antony *et al.*, 2006, 2007; Mameli *et al.*, 2007). Die hier beobachteten Ergebnisse bestätigen jedoch die im Zuge der Arbeit identifizierten ähnlichen cDNA-Transkriptionsmuster in MS-Proben und gesunden Kontrollen.

#### 4.2.10. Untersuchung der Promotoraktivität von LTR-Elementen diverser HERV-W-Loci

Zur weiteren Klärung der Frage, wo die Transkription zuvor identifizierter HERV-W-Loci initiiert werden kann, wurde die potentielle Promotoraktivität verschiedener 5'- und 3'-LTRs getestet. Eine Transkription sollte hierbei entweder von HERV-W-Locus-flankierenden LTRs oder von unbekannt benachbarten Promotorsequenzen initiiert werden. Einige der als transkribiert identifizierten Loci besitzen beide Vollängen-LTRs, wohingegen anderen die U3 Region des 5'-LTRs fehlt, in der eine Promotorsequenz lokalisiert ist, oder der komplette 5'-LTR deletiert ist. Viele Loci besitzen ausschließlich einen 3'-LTR, dem die U5-Region fehlt. Ob solche vollständig oder teilweise deletierten LTRs eine Transkription initiieren können, wurde mittels Luciferase-Assay getestet.

Es wurden 5'- und 3'-LTRs von stark und schwach transkribierten HERV-W-Loci ausgewählt, von genomischer DNA amplifiziert und in Promotorposition in den *pGLuc-Basic*-Vektor kloniert. Die Vektorkonstrukte wurden im Rahmen von Richter (2012) generiert (vgl. **Anhang 17**). Die Orientierung der LTRs in den *pGLuc-Basic*-Vektoren wurde im Rahmen der vorliegenden Arbeit erneut durch Sanger-Sequenzierung bestätigt. Insgesamt wurden 17 Konstrukte von zehn verschiedenen LTRs in *Sense*- und/oder *Antisense*-Orientierung getestet. Zellen der menschlichen Chorionkarzinomzelllinie JEG-3, die bereits in früheren Experimenten zum Nachweis der Promotoraktivität von HERV-W-LTRs verwendet wurden (Prudhomme *et al.*, 2004), wurden mit den Vektorkonstrukten transfiziert und die Promotoraktivität wurde mittels Luciferase-Assay bestimmt.

Die relativen Luciferase-Werte eines Experimentes, in dem alle Konstrukte in Triplikaten gemessen wurden, sind exemplarisch in **Abbildung 42** gezeigt. Zwei Wiederholungen des Experimentes lieferten ein vergleichbares Ergebnis (die Diagramme sind in **Anhang 18** gezeigt). Sowohl 5'- als auch 3'-LTRs der meisten HERV-W-Loci zeigten eine Promotoraktivität in *Sense*- und/oder *Antisense*-Orientierung. Lediglich vier LTR-Konstrukte zeigten keine Promotoraktivität, die relative Luciferaseaktivität war hier gleich zu der des zur Kontrolle transfizierten promotorlosen *pGLuc-Basic*-Vektors oder zu untransfizierten Zellen. Die Promotoraktivität aller anderen Vektorkonstrukte war leicht bis stark erhöht gegenüber den Negativkontrollen. Die stärkste Promotoraktivität zeigte der 5'-LTR des *ERVWE1/ERVW-1/syncytin-1*-Locus in *Sense*-Orientierung. Die relative Luciferaseaktivität erreichte Werte von etwa 10 % des als Positivkontrolle transfizierten *pCMV-GLuc*-Vektors, der Luciferase unter Kontrolle des starken CMV-Promotors exprimiert. Die 3'-LTRs der Loci *ERVW-17* (6q21<sup>a</sup>), *ERVW-27* (14q32.12) und *ERVW-4* (15q21.3) zeigten eine Promotoraktivität auch in *Antisense*-Orientierung. Im Fall des *ERVW-27* 3'-LTRs war diese höher als in *Sense*-Orientierung, im Fall von *ERVW-17* und *ERVW-4* wurde in *Sense*-Orientierung keine nennenswerte Promotoraktivität nachgewiesen.



**Abbildung 42:** Promotoraktivität von 5'- und 3'-LTRs von als transkribiert identifizierten HERV-W-Loci. LTRs von zehn verschiedenen HERV-W-Loci wurden von genomischer DNA amplifiziert, in *Sense*- (dunkelblau) und/oder *Antisense*-Orientierung (hellblau) in Promotorposition des *pGLuc-Basic* Vektors kloniert und die Promotoraktivität nach Transfektion der humanen Chronionkarzinomzelllinie JEK-3 mittels Luciferase-Assay bestimmt (vgl. Text). Gezeigt ist die relative Luciferaseaktivität jedes Vektorkonstruktes. Als Negativkontrolle wurde der promotorlose *pGLuc-Basic* Vektor transfiziert sowie untransfizierte Zellen getestet. Als Positivkontrolle diente der *pCMV-GLuc* Vektor, der die Luciferase unter Kontrolle des starken CMV-Promotors exprimiert (man beachte die unterschiedliche Skalierung des unteren Diagramms). Die Fehlerbalken zeigen die Standardabweichungen der Messung jedes Konstruktes in Triplikaten an. Gezeigt ist ein ausgewähltes Experiment, wobei zwei Wiederholungen ein vergleichbares Bild lieferten (vgl. **Anhang 18**). Auf der rechten Seite ist die Struktur der LTRs, verglichen mit der bei *Repbase* hinterlegten 780 bp langen LTR17-Konsensussequenz gezeigt. Die U3-, R- und U5-Regionen sind schematisch dargestellt. Die U3-Region besitzt hierbei in ihrem 3'-Bereich eine TATA-Box und die R-Regionen der 5'- und 3'-LTRs stellen die Start- und Endpunkte der HERV-W-Transkription dar.

Mit Ausnahme des 3'-LTRs des HERV-W-Locus auf Chromosom 1p32.3 und des 5'-LTRs des *ERVWE1/ERVW-1/syncytin-1*-Locus waren alle getesteten LTRs unvollständig. Hierbei zeigen die im Experiment getesteten LTRs der Loci *ERVW-13* (2q13), *ERVW-17* (6q21), 6p25.3, *ERVW-27* (14q32.12), *ERVW-4* (15q21.3) und *ERVWE2/ERVW-2* (Xq22.3<sup>b</sup>) eine Deletion der U5-Region des 3'-LTRs, wie sie für Loci charakteristisch ist, die durch L1-vermittelte Retrotransposition entstanden. Dem im Experiment getesteten 5'-LTR des *ERVW-12* (2p16.2) Locus fehlt die U3-Region und der getestete 5'-LTR des *ERVW-25* (13q13.3)-Locus besteht nur aus einem Rest der U3-Region und der kompletten R-Region. Beide LTRs zeigen jedoch, zumindest in *Antisense*-Orientierung, Promotoraktivität.

Die Ergebnisse deuten somit darauf hin, dass die meisten HERV-W-LTRs, auch wenn diese durch Deletionen verkürzt sind, Promotoraktivität in *Sense*- und/oder *Antisense*-Orientierung besitzen. Die Transkription eines Großteils der in der vorliegenden Arbeit identifizierten HERV-W-Loci könnte somit von den Resten der flankierenden 5'- und/oder 3'-LTRs in *Sense*- und/oder *Antisense*-Orientierung initiiert worden sein.

## 5. Diskussion

Humane endogene Retroviren (HERVs) stellen einen wesentlichen Bestandteil des menschlichen Transkriptoms dar und Transkripte unterschiedlicher HERV-Gruppen wurden in allen daraufhin untersuchten menschlichen Geweben detektiert (s. u.a. Frank *et al.*, 2005, 2008; Seifart *et al.*, 2005). Hierbei ist sicherlich nicht nur die Transkription der Proviren selbst von Bedeutung sowie potentiell von diesen kodierte retrovirale Proteine, sondern auch der Einfluss, den HERVs durch ihre Transkriptionsinitiations- und -terminationssignale sowie weitere regulatorische Elemente auf benachbarte Genomregionen ausüben (vgl. Jern und Coffin, 2008; Cohen *et al.*, 2009). Eine detailliertere Beschreibung spezifischer HERV-Transkriptionsmuster sowie potentieller Unterschiede in der spezifischen Transkription in verschiedenen Geweben ist somit ein wichtiger Schritt zum besseren Verständnis der Transkriptionsregulation der HERV-Loci selbst sowie deren Einfluss auf die Regulation und Transkription benachbarter Gene.

Auch im Zuge des besseren Verständnisses der potentiellen Rolle von HERVs in verschiedenen Krankheiten, für deren Pathogenese ein möglicher Einfluss humaner endogener Retroviren diskutiert wird, ist die Identifizierung transkribierter HERV-Loci sowie die Analyse ihrer Struktur von besonderem Interesse. Insbesondere auch, was ihre Kodierungskapazität für retrovirale Proteine betrifft, die eine potentielle biologische Funktion erfüllen könnten. Aufgrund früherer Studien, in denen HERV-spezifische Transkripte sowie retrovirale Proteine in krankem Gewebe und Zelllinien detektiert wurden, wurde u.a. eine mögliche Involvierung der HERV-K(HML-2)-Gruppe im Zusammenhang mit diversen Krebsarten sowie der HERV-W-Gruppe in der Pathogenese von Multipler Sklerose (MS) diskutiert (vgl. *Reviews* von Ruprecht *et al.*, 2008; Antony *et al.*, 2011; und Abschnitt 2.3.1.4/5 und 2.3.2.1. der Einleitung).

In den meisten Studien, die sich mit der Transkription humaner endogener Retroviren befassten, wurden lediglich allgemeine Unterschiede in der Transkription verschiedener HERV-Gruppen in gesunden und kranken Geweben untersucht (vgl. z.B. Yin *et al.*, 1997; Frank *et al.*, 2005, 2008; Seifarth *et al.*, 2005; Hu *et al.*, 2006; Muradrasoli *et al.*, 2006; Haupt *et al.*, 2011). Sofern HERVs tatsächlich für die Krankheitspathogenese von Bedeutung sind, wären jedoch ausschließlich die spezifischen genomischen Proviren der jeweiligen Gruppen von Interesse, die in den verschiedenen Geweben aktiv sind und von denen die zuvor detektierten Transkripte und translatierten Proteine stammen. Die Identifizierung transkribierter Loci wäre somit ein wichtiger Schritt zur weiteren Klärung der potentiellen Rolle von HERVs in der Krankheitspathogenese. Insbesondere sind hierbei auch spezifische Unterschiede bezüglich der Transkription in krankem und gesundem Gewebe von Interesse. In diesem Zusammenhang trägt die Identifizierung

transkribierter Loci zur Klärung der wichtigen Frage bei, ob es spezifische Transkriptionsmuster oder einzelne transkribierte Loci gibt, die kranke Gewebe von gesunden Vergleichsgeweben unterscheiden. Da frühere Studien erhöhte HERV-Transkriptmengen in verschiedenen Geweben beschrieben (Antony *et al.*, 2006, 07b; Mameli *et al.*, 2007; Stengel *et al.*, 2010; Schanab *et al.*, 2011), könnte durch Identifizierung transkribierter Loci geklärt werden, ob solche veränderten Transkriptmengen aus der Aktivierung einzelner spezifischer Loci oder aus einer mehr oder weniger willkürlichen Aktivierung verschiedener Loci resultieren. Des Weiteren könnte die Transkription spezifischer HERV-Loci als Biomarker dienen oder Hinweise auf genetische und/oder epigenetische Veränderungen in spezifischen Genomregionen liefern, die auch benachbarte Gene in diesen Regionen betreffen können. Auch die Analyse transkribierter Loci in krankem und gesundem Vergleichsgewebe bezüglich ihrer Kodierungskapazität für retrovirale Proteine ist, im Hinblick auf deren potentielle Funktion in der Krankheitspathogenese, von Interesse.

Bisher befassten sich vergleichsweise wenige Studien mit der spezifischen Identifizierung transkribierter HERV-Loci. Einige Studien versuchten, mittels bioinformatischer Methoden die Aktivität verschiedener HERVs anhand der vorhandenen Informationen in EST-Datenbanken abzuschätzen (Stauffer *et al.*, 2004; Oja *et al.*, 2007). Hierdurch konnte zwar die Transkription zahlreicher Proviren nachgewiesen werden, jedoch kam es aufgrund der teilweise hohen Sequenzähnlichkeit innerhalb verschiedener HERV-Gruppen sowie der teilweise schlechten Sequenzqualität von ESTs oft zu Problemen in der eindeutigen Zuordnung der Transkripte zu genomischen HERV-Loci. Als zusätzliches Problem stellte sich heraus, dass die in den EST-Datenbanken hinterlegten Informationen für eine umfangreiche Beschreibung der HERV-Aktivität nicht ausreichend sind, da beispielsweise niedrig transkribierte Loci in solchen cDNA-Bibliotheken generell unterrepräsentiert sind (Flockerzi *et al.*, 2008). In weiteren Studien wurde die Locus-spezifische HERV-Transkription mittels *Microarray* untersucht (Gimenez *et al.*, 2010; Pérot *et al.*, 2012). Hier zeigte sich jedoch, dass insbesondere bei starken Hybridisierungs-Signalen und/oder sehr ähnlichen Loci, die sich in direkt benachbarten Basen unterscheiden, das Risiko der Kreuzhybridisierung nicht auszuschließen ist, weswegen es sich in diesen Fällen empfiehlt, die Ergebnisse mittels Locus-spezifischer *Real-Time-PCR* zu validieren, was jedoch mit einem entsprechend hohen zusätzlichen Arbeitsaufwand verbunden ist (Gimenez *et al.*, 2010).

Ein weiterer molekulargenetischer Ansatz zur Identifizierung transkribierter Loci besteht in der Generierung HERV-Gruppen-spezifischer cDNA-Produkte und deren anschließender Sequenzierung und Zuordnung der erhaltenen Sequenzen zu genomischen Loci auf Grundlage

spezifischer Nukleotidunterschiede der einzelnen Loci im Vergleich zueinander (Flockerzi *et al.*, 2008; Frank *et al.*, 2008; Ruprecht *et al.*, 2008a; Laufer *et al.*, 2009). In der Studie von Flockerzi *et al.* (2008) wurden auf diese Weise insgesamt 23 transkribierte HERV-K(HML-2)-Loci in Hirnproben aus Patienten mit bipolaren Störungen und Schizophrenie, gesunden Hirnproben sowie in verschiedenen Hirn- und Keimzelltumoren identifiziert. Frank *et al.* (2008) identifizierten sechs transkribierte Loci der HERV-K(HML-6)-Gruppe in Mammatumoren und gesunden Vergleichsgeweben. In der Studie von Ruprecht *et al.* (2008a) wurden in retroviralen Partikeln der GCT-Zelllinie Tera-1 Transkripte von insgesamt acht verschiedenen HML-2-Loci detektiert. Laufer *et al.* (2009) identifizierten in PBMCs aus vier MS-Patienten und vier gesunden Kontrollen insgesamt sieben transkribierte Loci der HERV-W-Gruppe.

Analog zu diesen Studien von Flockerzi *et al.* (2008), Frank *et al.* (2008), Ruprecht *et al.* (2008a) und Laufer *et al.* (2009) wurden in der vorliegenden Arbeit transkribierte Loci der HERV-Gruppen HERV-W und HERV-K(HML-2) durch Generierung Gruppen-spezifischer RT-PCR-Produkte identifiziert, die sequenziert und anhand charakteristischer Nukleotidunterschiede im Genom rückkartiert wurden. Die Amplifikationsstrategie wurde hierfür im Vergleich zu den vorangehenden Arbeiten optimiert, um die Anzahl auf diese Weise theoretisch detektierbarer Loci weiter zu erhöhen.

### **5.1. Etablierung einer optimierten Amplifikationsstrategie zur Identifizierung transkribierter HERV-Loci**

Sowohl für HERV-K(HML-2) als auch für HERV-W wurden verbesserte Amplifikationsstrategien etabliert. Gegenüber den Studien von Flockerzi *et al.* (2008), Frank *et al.* (2008), Ruprecht *et al.* (2008a) und Laufer *et al.* (2009) wurden neue Amplikonbereiche zur RT-PCR-Amplifikation gewählt, die zum einen die theoretisch eindeutige Rückkartierbarkeit der generierten RT-PCR-Sequenzen gewährleisten (s.u.) und zum anderen die Detektion möglichst vieler verschiedener Loci in einer PCR erlaubten. Die Anzahl theoretisch amplifizierbarer Loci wurde zudem erhöht, indem zur Amplifikation Gemische verschiedene *Primer*-Varianten eingesetzt wurden. Diese berücksichtigen Nukleotidunterschiede in den *Primer*-bindenden Regionen, die zwischen den verschiedenen Loci einer HERV-Gruppe vorkommen können. Zusätzlich wurden die Amplifikationen unter niedrig stringenten PCR-Bedingungen bei niedriger *Primer-Annealing*-Temperatur durchgeführt, um auch die Amplifikation solcher Loci zu begünstigen, für die keine 100%-ige Übereinstimmung zu einer der *Primer*-Varianten besteht.

In der Studie von Flockerzi *et al.* (2008) wurde die HERV-K(HML-2)-Transkription durch Amplifikation und Sequenzierung zweier Amplikons in der proviralen *gag*- und *env*-Region untersucht. Im Gegensatz hierzu wurde in der vorliegenden Arbeit lediglich ein Amplikon im zentralen Bereich des *gag*-ORFs gewählt. Ein Sequenzvergleich ergab, dass (fast) alle Loci, die in der Studie von Flockerzi *et al.* (2008) durch die *gag*- und *env*-spezifischen Primer erkannt wurden, auch durch *Primer*-Varianten des in der vorliegenden Arbeit gewählten *gag*-Amplikons allein detektiert werden können. Die einzige Ausnahme bildet ein Locus auf Chromosom 4q32.2, der eine Deletion im Bereich des hier gewählten *gag Reverse-Primers* aufweist, wobei der Locus jedoch bisher nicht als transkribiert beschrieben wurde. Durch Wahl eines zweiten Amplikons im *env*-ORF hätte sich die Zahl amplifizierbarer Loci somit nicht signifikant erhöht und hätte lediglich doppelten Arbeitsaufwand sowie doppelte Kosten bedeutet, weswegen die Analyse auf das optimierte *gag*-Amplikon beschränkt wurde.

Im Fall von HERV-W wurde ein ~280 bp großer Amplikonbereich innerhalb des *env*-ORFs gewählt. Ein großer Teil der genomischen HERV-W-Loci weist, aufgrund ihrer Entstehung durch L1-vermittelte Retrotransposition, unterschiedlich große Deletionen im proviralen 5'-Bereich auf (Costas, 2002; Pavlíček *et al.*, 2002), der *env*-Bereich ist hingegen in vergleichsweise vielen Loci noch vorhanden. Bereits in der Studie von Laufer *et al.* (2009) wurde ein etwa 690 bp großes Amplikon im *env*-ORF untersucht. Die *Primer* dieser früheren Studie wurden so gewählt, dass eine zuvor publizierte MSR-Seqenz (*GenBank acc. no.* AF331500) sowie der *env*-Bereich des *syncytin-1/ERVWE1/ERVW-1*-Locus auf Chromosom 7q21.2 amplifiziert werden können. Ein Sequenzvergleich zeigt jedoch, dass die amplifizierte Region in vielen HERV-W-Loci nicht oder nur zum Teil vorhanden ist (vgl. **Anhang 9**). Um die Anzahl theoretisch detektierbarer Loci zu erhöhen, wurde ein verkürzter Amplikonbereich von etwa 280 bp gewählt (s.u.), der mit dem 3'-Bereich des Amplikons der Studie von Laufer *et al.* (2009) überlappt und der in einer größeren Zahl von Loci vorhanden ist. Da ein großer Teil der genomischen HERV-W-Loci noch größere Deletionen aufweist und die betreffenden Loci lediglich den 3'-Bereich des *env*-ORFs besitzen, wurde ein zweites, etwa 330 bp großes Amplikon in diesem Bereich gewählt, um auch Transkripte dieser Loci zu detektieren.

Durch Test der für beide HERV-Gruppen gewählten *Primer* auf genomischer DNA konnte die grundsätzlich mögliche Amplifikation einer großen Anzahl verschiedener HERV-Loci bestätigt werden. Für das HML-2 *gag*-Amplikon wurden 18 verschiedene Loci der HML-2-Gruppe amplifiziert (vgl. **Abb. 16**). Da im Falle von HML-2 theoretisch 39 verschiedene Loci von den verwendeten *Primer*-Varianten erkannt werden, deutet die Anzahl von 18 verschiedenen amplifizierten Loci zunächst auf eine möglicherweise präferentielle Amplifikation bestimmter Loci

hin. Die Verteilung korrelierte jedoch nicht mit in späteren Experimenten generierten relativen cDNA-Klonierungsfrequenzen. So zeigten beispielsweise HML-2-Loci auf Chromosom 3q27.2 (*ERVK-11*), 19q12 (*ERVK-28*) und 22q11.21 (*ERVK-24*) relativ hohe Klonierungsfrequenzen bei Amplifikation von genomischer DNA, die Loci wurden jedoch bei späterer Amplifikation von cDNA mit deutlich niedrigeren Frequenzen detektiert (vgl. **Tab. 15**). Des Weiteren konnten bei Amplifikation von cDNA neun weitere HML-2-Loci amplifiziert werden, die von genomischer DNA nicht amplifiziert wurden, was zeigt, dass mehr als die dort detektierten 18 Loci amplifiziert werden können. Im Falle von HERV-W konnten für das 5'*env*- bzw. 3'*env*-Amplikon 22 bzw. 83 verschiedene Loci von genomischer DNA amplifiziert werden (vgl. **Abb. 35**). Eine relativ gleichmäßige Verteilung der amplifizierten Loci spricht hierbei gegen eine spezifische bzw. bevorzugte Amplifikation einzelner Loci

### 5.1.1. Rückkartierung generierter cDNA-Sequenzen anhand charakteristischer Nukleotidunterschiede zwischen genomischen HERV-Loci

In den HERV-Loci der verschiedenen Gruppen haben sich im Laufe der Zeit zahlreiche Mutationen angesammelt, anhand derer sich die Loci untereinander unterscheiden. Die untersuchten HERV-W- und HML-2-spezifischen Amplikonbereiche wurden so gewählt, dass der amplifizierte Bereich genügend charakteristische Nukleotidunterschiede umfasst, um generierte cDNA-Sequenzen eindeutig genomischen Loci zuzuordnen.

Im Fall der evolutiv älteren HERV-W-Gruppe genügten hierfür die 280 und 330 bp großen Amplikonbereiche 5'*env* und 3'*env* im *env*-ORF, die im Anschluss mittels *Next Generation Sequencing* analysiert wurden. Im Fall der HML-2-Gruppe, die zahlreiche evolutiv junge Proviren umfasst, ist die Anzahl an Mutationen, die sich im Laufe der Zeit angesammelt haben, deutlich geringer als in evolutiv älteren HERV-Gruppen. Aufgrund dieser höheren Sequenzähnlichkeiten vieler HML-2-Loci war es notwendig, einen vergleichsweise großen Amplikonbereich von ~620 bp zu wählen, damit dieser genügend charakteristische Sequenzunterschiede umfasst, um eine eindeutige Zuordnung der theoretisch amplifizierbaren Sequenzen zu genomischen HML-2-Loci zu gewährleisten (vgl. **Abb. 15**). Selbst hier finden sich jedoch Loci, die sich in der amplifizierten Region lediglich in einer Nukleotidposition unterscheiden. Aufgrund der Größe des Amplikons wurden im Fall von HML-2 cDNA-Sequenzen durch Klonierung und anschließende Sanger-Sequenzierung generiert. Ein *Next Generation Sequencing*-Ansatz hätte zwar eine deutlich umfangreichere Transkriptionsanalyse durch die Generierung einer größeren Sequenzanzahl erlaubt, jedoch wäre hier eine eindeutige Zuordnung der generierten Sequenzen

aufgrund kürzerer Leseweiten deutlich erschwert bzw. in vielen Fällen gar nicht möglich gewesen. Durch ein noch größeres Amplikon hätten u.U. weitere Nukleotidunterschiede erfasst werden können, jedoch wären in diesem Falle auch die Grenzen der zuverlässigen Sanger-Sequenzierung erreicht. Ein noch längeres Amplikon hätte, um eine komplette Sequenz in guter Qualität zu generieren, aus beiden Richtungen sequenziert werden müssen, was mit doppelten Sequenzierkosten verbunden gewesen wäre.

Die eindeutige Rückkartierung generierter cDNA-Sequenzen wurde in vielen Fällen zusätzlich durch Nukleotidunterschiede zu den genomischen proviralen Sequenzen erschwert. Die cDNA-Sequenzen, die von RNA-Transkripten eines spezifischen Locus generiert wurden, sind in vielen Fällen nicht identisch zur genomischen Referenzsequenz dieses Locus. Solche Nukleotidunterschiede können auf Fehlern im Zuge der reversen Transkription, Amplifikation und/oder Sanger-Sequenzierung beruhen, oder auf Rekombination zwischen Transkripten verschiedener HERV-Loci während der PCR oder reversen Transkription zurückzuführen sein (vgl. Meyerhans *et al.*, 1990; Judo *et al.*, 1998; Mader *et al.*, 2001; Cocquet *et al.*, 2006; Flockerzi *et al.*, 2007). In früheren Studien (s.u.) wurden generierte cDNA-Sequenzen auf solche Rekombinationsereignisse hin untersucht, worauf jedoch in der vorliegenden Arbeit verzichtet wurde. Insbesondere im Falle von HML-2 ist eine sichere Identifizierung der an der Rekombination beteiligten Transkripte aufgrund der hohen Sequenzähnlichkeiten vieler Loci schwierig und es kommt in vielen Fällen mehr als ein möglicher Rekombinationspartner in Frage. Im Falle von HERV-W war eine solche Analyse aufgrund der enormen Datenmenge nicht möglich. Des Weiteren können Nukleotidunterschiede interindividuelle Sequenzunterschiede bzw. Einzelnukleotidpolymorphismen repräsentieren, aufgrund derer sich generierte cDNA-Sequenzen von den Proviren im Referenzgenom unterscheiden (vgl. Mayer *et al.*, 2005; de Parseval *et al.*, 2005). Es ist hier bei der Rückkartierung von HERV cDNA-Sequenzen grundsätzlich anzunehmen, dass im Zuge der Rückkartierung der beste *BLAT-Match* korrekt ist, sofern sich der nächstbeste *BLAT-Match* durch mindestens einen weiteren Nukleotidunterschied von diesem unterscheidet. Nicht eindeutig rückkartierbare Sequenzen, mit mehreren besten *BLAT-Matches* mit gleicher Anzahl an Nukleotidunterschieden, wurden verworfen. In solchen Fällen wäre die generierte cDNA-Sequenz zwar ein Indiz für die Transkription einer der möglichen proviralen Sequenzen, jedoch müsste mittels zusätzlicher Locus-spezifischer PCRs geklärt werden, welcher der möglichen HERV-Loci tatsächlich der transkribierte ist. In vielen Fällen wurde in der betreffenden Probe einer der möglichen Loci durch weitere cDNA-Sequenzen eindeutig als transkribiert identifiziert. Hier könnte man somit davon ausgehen, dass wahrscheinlich auch die leicht variante Sequenz von diesem Locus stammt, nun aber aufgrund

*ex vivo* generierter Sequenzunterschiede noch zu einem weiteren Locus mit gleicher Anzahl an Nukleotidunterschieden passt. In weiteren Fällen wurden jedoch beide möglichen Loci aufgrund weiterer cDNA-Sequenzen eindeutig als transkribiert identifiziert, weswegen auf solche Spekulationen verzichtet und die betreffenden Sequenzen verworfen wurden.

Die Festlegung eines eindeutigen Schwellenwertes bzgl. der Anzahl an Sequenzunterschieden, die eine zuverlässige Zuordnung zu einem genomischen Locus noch erlauben, ist schwierig und stellte in früheren Studien im Falle vieler Loci ein Problem dar (Flockerzi *et al.*, 2008; Laufer *et al.*, 2009). In der HML-2-Studie von Flockerzi *et al.* (2008) wurden bis zu 18 Nukleotidunterschiede zur genomischen Sequenz toleriert und Sequenzen mit mehr Unterschieden als potentielle Rekombinanten verworfen. In der HERV-W-Studie von Laufer *et al.* (2009) wurden alle generierten Sequenzen mit mindestens einem Unterschied zur Vergleichssequenz auf mögliche Rekombinationsereignisse hin untersucht. In der vorliegenden Arbeit wurde ein niedrigerer Wert von sechs Nukleotidunterschieden im Fall des ~620 bp langen HML-2-Amplikons bzw. drei Nukleotidunterschieden im Fall der 280 bzw. 330 bp großen HERV-W-Amplikons festgelegt und Sequenzen mit mehr Unterschieden, die Rekombinanten, unbekannte allelische Varianten oder Sequenzartefakte repräsentieren könnten, verworfen. Hierdurch wurden im Falle beider untersuchter HERV-Gruppen jeweils Sequenzunterschiede bis zu etwa 1 % toleriert. Es zeigte sich in den späteren Experimenten, dass nur relativ wenige cDNA-Sequenzen mehr Nukleotidunterschiede aufwiesen. Im Falle von HML-2 wurden weniger als 3 %, bzw. nach UVB-Bestrahlung 9 %, der generierten Sanger-Sequenzen verworfen, für HERV-W waren es 8 % der 454/FLX-Sequenzen. Für die überwiegende Mehrheit war eine eindeutige Rückkartierung und somit die sichere Identifizierung des jeweiligen Locus als transkribiert möglich (vgl. **Abb. 17, 20 und 36**).

Lediglich im Falle der Amplifikation des HML-2 *gag*-Amplikons von genomischer DNA wurde knapp die Hälfte der generierten Sequenzen aufgrund einer hohen Anzahl von bis zu 26 Nukleotidunterschieden zu ihrem besten *BLAT-Match* verworfen. Eine solche hohe Anzahl an nicht eindeutig rückkartierbaren Sequenzen wurde bei der späteren Amplifikation von cDNA nicht beobachtet. Dies könnte auf eine erhöhte Wahrscheinlichkeit für Rekombinationsereignisse während der PCR zurückzuführen sein, da im Falle von genomischer DNA als *Template* mehr potentielle Rekombinationspartner im Reaktionsansatz vorhanden sind als im Falle von cDNA. Durch eine geringere Ausgangsmenge an genomischer DNA während der PCR könnte hier die Anzahl an Rekombinationsereignissen u.U. verringert werden.

### 5.1.2. Die mögliche Bedeutung von *Antisense*-Transkripten

Ziel der vorliegenden Arbeit war eine möglichst ausführliche Beschreibung transkribierter Loci der HERV-Gruppen HERV-K(HML-2) und HERV-W in verschiedenen Geweben und Zelllinien, unabhängig von der Transkriptionsrichtung der Loci. Die gewählte Methode der cDNA-Generierung durch Verwendung von *Random-Primern* erlaubt keine Unterscheidung zwischen *Sense*- und *Antisense*-Transkripten spezifischer Loci, da beide gleichermaßen in cDNA umgeschrieben werden. Zur Klärung der Transkriptionsrichtung wäre eine Wiederholung der Experimente nötig, wobei in je zwei verschiedenen Ansätzen Strang-spezifisch cDNA jeder RNA-Probe generiert werden müsste (vgl. hierzu auch Roebke *et al.*, 2010). Hierbei würde jeder Ansatz unterschiedliche spezifische *Primer* enthalten, die jeweils nur an Transkripte einer Orientierung, *sense* oder *antisense*, binden und deren reverse Transkription in cDNA ermöglichen. Durch Vergleich der als transkribiert identifizierten Loci in beiden Ansätzen, bzw. das Fehlen von Loci im einen oder anderen Fall, wäre eine Aussage über deren ursprüngliche Transkriptionsrichtung möglich. Eine solch umfassende zusätzliche Analyse konnte in der vorliegenden Arbeit aus Zeit- und Kostengründen nicht mehr durchgeführt werden. Zudem waren in vielen Fällen die zur Verfügung stehenden Ausgangsmengen an RNA stark begrenzt und hätten nicht für zwei weitere RT-Ansätze incl. der nötigen Kontrollen ausgereicht.

Im biologischen Kontext könnten auch *Antisense* mRNA-Transkripte von Bedeutung sein, beispielsweise durch Regulierung der Expression spezifischer Zielgene mittels RNA-Interferenz (Gogvadze *et al.*, 2009; Gosenca *et al.*, 2012). Da die meisten HML-2-Loci einen 5'-LTR besitzen und dieser normalerweise Promotoraktivität besitzt, stellen Loci, die ausschließlich in *Antisense*-Richtung transkribiert werden, für diese Gruppe vermutlich eine Ausnahme dar (Vinogradova *et al.*, 2001; Buzdin *et al.*, 2006). Für HML-2 wurde eine bidirektionale Promotoraktivität solitärer LTRs gezeigt (Domansky *et al.*, 2000), weswegen für HML-2-Loci jedoch auch eine *Antisense*-Transkription ausgehend vom 3'-LTR denkbar wäre. In der vorliegenden Arbeit konnte auch für die HERV-W-Gruppe eine bidirektionale Promotoraktivität verschiedener 5'- und 3'-LTRs gezeigt werden (s.u.). Da vielen HERV-W-Loci aufgrund der Generierung durch L1-vermittelte Retrotransposition der 5'-LTR fehlt, wäre hier die Initiation von *Antisense*-Transkripten ausgehend vom 3'-LTR denkbar. Im Falle der Deletion beider flankierender LTRs ist eine Initiation der HERV-Transkription ausgehend von Promotorsequenzen stromauf oder stromab des Provirus möglich, was im letzteren Fall ebenfalls in der Generierung von *Antisense*-Transkripten resultieren kann.

## 5.2. Transkriptionsprofile der HERV-K(HML-2)-Gruppe und weitere Charakterisierung transkribierter HML-2-Loci in Krebs und Normalgeweben

In früheren Studien wurden in zahlreichen untersuchten Tumorproben verschiedener Krebsarten und verschiedenen Krebs-Zelllinien, darunter Melanome und Keimzelltumoren, HML-2 mRNA-Transkripte sowie retrovirale Proteine nachgewiesen (vgl. z.B. Sauter *et al.*, 1995, 1996; Boller *et al.*, 1997; Bieda *et al.*, 2001, Muster *et al.*, 2003, Büscher *et al.*, 2005, 2006; Reiche *et al.*, 2010; und Abschnitt 2.3.1.4./5 der Einleitung). Die potentielle direkte oder indirekte Rolle, die HML-2 in der Krebspathogenese spielt, ist hierbei noch unklar. Die Identifizierung transkribierter HML-2-Loci in Normal- und Tumorgewebe und eine Analyse ihrer Lokalisation in möglicherweise krankheitsrelevanten Genomregionen sowie ihrer Struktur, insbesondere hinsichtlich LTRs und ihrer Kodierungskapazität für retrovirale Proteine, sind somit ein wichtiger Schritt zum besseren Verständnis der potentiellen Rolle von HML-2 in der Krankheitspathogenese. Ebenso ist die Erstellung detaillierter Transkriptionsprofile sowie die Untersuchung potentieller Unterschiede in der Transkription spezifischer Loci zwischen verschiedenen Geweben von großem Interesse.

### 5.2.1. Melanom-Proben und Melanozyten-Zelllinien zeigen interindividuell unterschiedliche Muster transkribierter HML-2-Loci

In den verschiedenen Melanom-Proben, Melanom-Zelllinien sowie Melanozyten-Zelllinien konnten insgesamt 23 HML-2-Loci als transkribiert identifiziert werden, darunter auch Proviren, über deren Aktivität zuvor nichts bekannt war (vgl. **Tab. 15**). Diese Zahl ist vergleichbar mit der Anzahl verschiedener transkribierter Loci in Keimzelltumoren und Hirnproben aus der Studie von Flockerzi *et al.* (2008). Neunzehn der Loci wurden hierbei in beiden Studien detektiert. Die vier zusätzlichen hier detektierten Loci wurden zuvor noch nicht als transkribiert beschrieben. Ihre als EST zu definierenden Sequenzen wurden an die *Genbank*-Datenbank übermittelt und die Loci in Übereinstimmung mit dem *HUGO Gene Nomenclature Committee* mit den offiziellen Locus-Bezeichnungen *ERVK-25* bis *ERVK-28* versehen (vgl. Mayer *et al.*, 2011).

Siebzehn der als transkribiert identifizierten HML-2-Loci wurden kürzlich auch in einer *Microarray*-Studie detektiert (Perot *et al.*, 2012). Die in der vorliegenden Arbeit detektierten Loci *ERVK-1*, *ERVK-7*, *ERVK-18*, *ERVK-6*, *ERVK-14* und *ERVK-26* wurden in der *Microarray*-Studie nicht identifiziert. Dies könnte mit dem generellen *Design* des *Microarrays* sowie der Hybridisierungs-Sonden zusammenhängen, die u.U. eine spezifische Detektion dieser Loci nicht erlauben. Es ist zudem anzumerken, dass in dieser *Microarray*-Studie zwar zahlreiche Tumoren

sowie gesundes Vergleichsgewebe, jedoch keine Melanom-spezifischen Proben untersucht wurden, was ebenfalls Unterschiede erklären könnte (vgl. hierzu auch Abschnitt 5.2.3. zur HML-2-Transkription in GCT- und Prostatakarzinom-Proben).

Insgesamt unterscheiden sich die Transkriptionsmuster in allen untersuchten Melanom-Proben und Melanozyten-Zelllinien. In jeder Probe ist eine spezifische Auswahl an HML-2-Loci transkribiert, die nur zum Teil mit den transkribierten Loci in anderen Proben überlappen. Die exemplarische Wiederholung des Experimentes mit RNA einer weiteren Passage von zwei der untersuchten Zelllinien lieferte ein sehr ähnliches Bild wie im Fall der ersten Passage (vgl. **Abb. 19**). Die Loci mit höchster relativer Klonierungsfrequenz waren hierbei gleich und Unterschiede fanden sich lediglich bei den schwächer transkribierten Loci, wobei diese vermutlich durch Generierung noch größerer Sequenzanzahlen weiter minimiert werden würden. Das vergleichbare Transkriptionsprofil in unterschiedlichen Zellpassagen deutet darauf hin, dass sich die Zelllinien durch ein unterschiedliches, jedoch für die betreffende Zelllinie jeweils stabiles Transkriptionsmuster charakteristischer HERV-Loci auszeichnen.

Der *ERVK-6*-Locus auf Chromosom 7p22.1 war hierbei von besonderem Interesse, da dieser ausschließlich in sieben Melanom-Proben mit teilweise hoher relativer Klonierungsfrequenz detektiert wurde, jedoch nicht in den Melanozyten-Zelllinien. In den untersuchten Prostatakarzinom-Zelllinien (vgl. Abschnitt 5.2.3.) sowie in im Rahmen einer Kooperation mit Frau Prof. Dr. Christine Leib-Mösch (Institut für Virologie, Helmholtz Zentrum, München) untersuchten Mammatumoren und gesunden Vergleichsgeweben (Daten nicht gezeigt) wurde der *ERVK-6* Locus ebenfalls nicht als transkribiert detektiert. In den untersuchten GCT-Zelllinien (vgl. Abschnitt 5.2.3.) zeigte *ERVK-6* lediglich eine schwächere relative Transkription, die schon in den GCT-Geweben der Studie von Flockerzi *et al.* (2008) für das dort untersuchte *gag*-Amplikon beschrieben wurde. Der Locus stellt somit, auch aufgrund seiner Struktur und potentiellen Kodierungskapazität für Gag- und Env-Proteine (vgl. Abschnitt 5.2.2.), einen interessanten Kandidaten für weitere Analysen einer größeren Probenanzahl bzgl. seiner potentiellen Rolle als Melanom-Biomarker dar.

Da in früheren Studien mehrere HERV-K(HML-2)-Loci als polymorph beschrieben wurden (Barbulescu *et al.*, 1999; Hughes und Coffin, 2004; Macfarlane und Simmonds 2004; Mayer *et al.*, 2005) könnten Transkriptionsunterschiede zwischen untersuchten Proben bzw. das Fehlen von Transkripten eines solchen polymorphen Provirus in der vorangehenden Transkriptionsanalyse schlicht auf ein Fehlen des Provirus in der genomischen DNA der entsprechenden Probe zurückzuführen sein. Aus diesem Grund wurden die zur Verfügung

stehende genomische DNA verschiedener Zelllinien auf das Vorhandensein polymorpher Vollängen-Proviren untersucht. Von sieben HML-2-Loci, die zuvor als polymorph in der menschlichen Population beschrieben wurden, konnte in den zur Verfügung stehenden genomischen DNAs in den meisten Fällen mindestens eine allelische Kopie des jeweiligen Vollängen-Provirus nachgewiesen werden. Die beobachteten Transkriptionsunterschiede sind somit nicht auf genomischer Ebene durch das Fehlen der Proviren in den betreffenden Proben begründbar.

Generell wird die unterschiedliche Transkription verschiedener Loci ausgehend von deren LTRs oder anderen benachbarten Promotorsequenzen vermutlich durch das Zusammenspiel epigenetischer Faktoren sowie von Transkriptionsfaktoren reguliert. Unterschiedliche epigenetische Modifikationen verschiedener Loci in den untersuchten Geweben können hierbei in unterschiedlichen Transkriptionsprofilen resultieren. Um die Frage zu klären, warum spezifische Loci in den verschiedenen Geweben transkribiert werden oder nicht, wären zusätzliche Experimente zur Untersuchung des LTR-Methylierungsstatus (vgl. Lavie *et al.*, 2005) sowie spezifischer aktivierender oder reprimierender Histonmodifikationen nötig. Die LTR-Elemente spezifischer Loci könnten nach Bisulfit-Behandlung der genomischen DNA einer Probe (Frommer *et al.*, 1992) amplifiziert und sequenziert werden, was Aufschluss über deren Methylierung in der Probe liefert. Histonmodifikationen sowie mögliche Interaktionen mit Transkriptionsfaktoren könnten in weiterführenden Studien mittels Chromatin-Immunopräzipitation und anschließender Sequenzierung (*ChIP-seq*) näher charakterisiert werden (Barski und Zhao, 2009)

### **5.2.2. Unter den in Melanom-Proben und Melanozyten-Zelllinien transkribierten HML-2-Loci finden sich strukturell intakte Proviren mit vollständigen LTRs sowie ORFs für retrovirale Gag- und Env-Proteine**

Die meisten transkribierten HML-2-Loci zeichnen sich durch eine mehr oder weniger vollständige Struktur mit internen Sequenzen und 5'- und 3'-LTRs aus (vgl. **Tab. 16**). Die Transkription könnte in diesen Fällen vom jeweiligen 5'-LTR initiiert worden sein, der die proviralen Promotorsequenzen enthält. Es wäre jedoch auch, insbesondere im Falle der Loci, denen der 5'-LTR fehlt, eine Transkriptionsinitiation durch eine bidirektionale Promotoraktivität des 3'-LTRs denkbar (vgl. Domansky *et al.*, 2000). Die Loci *ERVK-14* (7q22.1) und *ERVK-17* (10q24.2), denen der 5'-LTR fehlt, liegen intronisch und in *Antisense*-Orientierung innerhalb von Genen. Die detektierten Transkripte könnten in diesen beiden Fällen auch auf alternative

Spleißprozesse der jeweiligen Gene zurückzuführen sein. Der *ERVK-14*-Locus liegt intronisch im *LHFPL3*-Gen, wobei der 5'-Bereich des Locus in 14 annotierten EST-Sequenzen vorhanden ist und dort ein zusätzliches Exon bildet (Kim und Hahn, 2010). *ERVK-17* ist intronisch im *MRP2*-Gen lokalisiert, einem Mitglied der *MRP*-Unterfamilie, die den Transport von Anti-Krebs-Medikamenten vermittelt und zur Arzneimittelresistenz beitragen (Kruh und Belinsky, 2003). Für andere HML-2-Loci ist eine Transkriptionsinitiation durch noch unbekannte Promotorsequenzen in der näheren Umgebung der Loci denkbar, oder eine *read-through* Transkription benachbarter Gene. So liegt beispielsweise der *ERVK-15*-Locus auf Chromosom 7q32, dessen 5'-LTR nahezu vollständig deletiert ist, etwa 1,6 kb stromauf des Haushaltsgens *SSBP1*, für das eine Beteiligung an der mitochondrialen Biogenese gezeigt wurde und das eine Untereinheit eines DNA-Einzelstrang bindenden Komplexes darstellt, der an der Erhaltung der Genomstabilität beteiligt ist (Tiranti *et al.*, 1995; Huang *et al.*, 2009). Zur Klärung der Transkriptionsinitiation der detektierten HML-2-Loci müsste die potentielle Promotoraktivität der proviralen LTRs der betreffenden Loci sowie ggf. flankierender Sequenzen in spezifischen Reporter-Gen-Assays detailliert untersucht werden, wie sie in der vorliegenden Arbeit für LTRs der HERV-W-Gruppe durchgeführt wurden (vgl. Abschnitt 5.3.5.).

Einige der als transkribiert identifizierten HML-2-Loci besitzen vollständige ORFs für die retroviralen Proteine Gag und Env, denen eine potentielle Funktion in der Tumorphagenese zukommen könnte. Hierbei wurden die drei Loci *ERVK-21* (12q14.1), *ERVK-28* (19q12) und *ERVK-6* (7p22.1), die einen ORF für ein Volllänge Env-Protein besitzen, ausschließlich in Melanom-Proben identifiziert. *ERVK-21* und *ERVK-28* waren jeweils lediglich in einer der Proben transkribiert, wohingegen der *ERVK-6*-Locus in sieben der Melanom-Proben, jedoch nicht in beiden Melanozyten-Zelllinien als transkribiert identifiziert wurde und daher von besonderem Interesse ist. In einer früheren Studie wurde gespleißte *env*-mRNA sowie Env-Protein in den Melanom-Zelllinien SK-Mel-25 und MEWO detektiert (Büscher *et al.*, 2005). Da in der vorliegenden Arbeit in beiden Zelllinien der Volllänge Env-kodierende Locus *ERVK-6* als transkribiert identifiziert wurde, könnten in der früheren Studie identifizierte mRNA und Proteine auf diesen Locus zurückzuführen sein. Eine potentielle *in vivo* Translation des von *ERVK-6* kodierte Env-Proteins könnte zudem Env-spezifische Antikörper erklären, die in Seren von einigen Melanom-Patienten detektiert wurden (Hahn *et al.*, 2008), jedoch im Vergleich zu Seren von GCT-Patienten in allgemein niedrigerer Häufigkeit und Titern (Sauter *et al.*, 1996). Eine Korrelation zwischen der Transkription Env-kodierender HML-2-Loci und dem Vorhandensein spezifischer Antikörper wurde im Falle von GCTs in der Studie von Flockerzi *et al.* (2008) nicht gefunden. Da der *ERVK-6*-Locus zudem einen ORF für ein Volllänge

Gag-Protein besitzt, könnte auch in früheren Studien detektiertes Gag-Protein in Melanom-Metastasen von diesem Locus stammen sowie Gag-spezifische Antikörper in Seren von einigen Melanom-Patienten auf von diesem Locus exprimiertes Gag-Protein zurückzuführen sein (Büscher *et al.*, 2005, Hahn *et al.*, 2008). Zur weiteren Klärung der potentiellen Rolle von *ERVK-6* in Melanomen wären jedoch weiterführende umfangreiche Experimente nötig, in denen die Transkription des Locus in einer größeren Anzahl Melanom-spezifischer Proben sowie gesunder Vergleichsgewebe untersucht werden müsste, um die möglicherweise spezifische oder erhöhte Transkription in Melanomen weiter zu bestätigen. Des Weiteren müsste mittels *Western Blot* oder spezifischer Fluoreszenzfärbungen die *in vivo* Translation der kodierten Proteine in den Proben weiter untersucht werden, um mögliche Unterschiede zwischen Tumor- und Vergleichsgewebe zu detektieren. Die weiteren Gag-kodierenden Loci *ERVK-9* (6q14.1) und *ERVK-28* (19q12) wurden jeweils nur in einer der Melanom-Proben detektiert und waren dort nur schwach transkribiert. Drei weitere Loci mit vollständigem *gag*-ORF, *ERVK-5* (3q12.3), *ERVK-10* (5q33.3) und *ERVK-24* (22q11.21), waren sowohl in Melanom-Proben als auch Melanozyten-Zelllinien transkribiert und zeigten keine auffälligen Unterschiede bzgl. ihrer relativen Klonierungsfrequenz. Auch im Falle von GCTs wurden Gag-kodierende Loci in Tumorproben aus Patienten mit und ohne Antikörper-Immunantwort gegen das Protein detektiert (Flockerzi *et al.*, 2008). Die Bildung spezifischer Antikörper gegen HML-2 Gag- und Env-Proteine scheint somit einem komplexeren Regulationsmechanismus zu unterliegen, der nicht allein durch Aktivierung Protein-kodierender HML-2-Loci bestimmt wird (Flockerzi *et al.*, 2008).

### **5.2.3. Die HML-2 *gag*-Transkription unterscheidet sich zwischen Keimzelltumor- und Prostatakarzinom-Zelllinien sowie von der Transkription in Melanom-Proben und Melanozyten-Zelllinien**

In den beiden untersuchten GCT-Zelllinien wurden insgesamt neun verschiedene HML-2-Loci als transkribiert identifiziert, in den sechs untersuchten Prostatakarzinom-Zelllinien waren 12 verschiedene Loci transkribiert. Vier Loci, *ERVK-7*, *ERVK-5*, *ERVK-10* und *ERVK-26*, wurden hierbei in beiden Tumorarten detektiert. Drei der in GCTs und Prostatakarzinomen gefundenen Loci wurden im Rahmen der zuvor beschriebenen Analyse der Melanom-Proben und Melanozyten-Zelllinien nicht identifiziert. Es handelte sich hierbei um den *ERVK-20*-Locus auf Chromosom 11q23.3, der in der GCT-Zelllinie NCCIT detektiert wurde, sowie um die zuvor noch nicht als transkribiert identifizierten Loci auf Chromosom 19p12 und Yp11.2, die in zwei bzw. einer der Prostatakarzinom-Zelllinien detektiert wurden.

Insgesamt unterschieden sich die Transkriptionsprofile in GCT- und Prostatakarzinom-Zelllinien untereinander sowie von denen in Melanom-Proben und Melanozyten-Zelllinien (vgl. **Anhang 7**). In GCTs zeigten die Loci *ERVK-7* und *ERVK-24* die höchsten Klonierungsfrequenzen, wohingegen *ERVK-7* in den Prostatakarzinom-Zelllinien schwächer bzw. *ERVK-24* gar nicht detektiert wurden. Auch in den Melanom-Proben und Melanozyten-Zelllinien zeigte der *ERVK-24* Locus nur schwache relative Klonierungsfrequenzen, wohingegen der *ERVK-7*-Locus in einigen Melanom-Proben stärker transkribiert war. Die Loci *ERVK-7* und *ERVK-24* zeigten bereits in der Studie von Flockerzi *et al.* (2008) (dort als c1\_B und c22\_A benannt) in den untersuchten GCT-Proben eine vergleichsweise starke Transkription. Auch in der Studie von Ruprecht *et al.* (2008a) konnten die meisten Transkripte aus retroviralen Partikeln der GCT-Zelllinie Tera-1 dem *ERVK-24*-Locus auf Chromosom 22q11.21 zugeordnet werden. Die Loci *ERVK-4* (c3\_C), *ERVK-20* (c11\_B) und *ERVK-23* (c21\_A), die in den von Flockerzi *et al.* (2008) untersuchten GCT-Proben ebenfalls stärker transkribiert waren, wurden in den beiden in der vorliegenden Arbeit analysierten GCT-Zelllinien jedoch nur mit niedrigen relativen Klonierungsfrequenzen bzw. *ERVK-4* und *ERVK-23* gar nicht detektiert. In der *Microarray*-Studie von Pérot *et al.* (2012) wurde der *ERVK-7*-Locus in Hodentumoren, gesundem Vergleichsgewebe sowie den restlichen untersuchten Proben, nicht detektiert. Dies könnte, wie oben bereits erwähnt, auf das *Design* der Hybridisierungs-Sonden des *Microarrays* zurückzuführen sein, die u.U. eine spezifische Detektion dieses Locus nicht erlauben. Der *ERVK-24*-Locus war hingegen auch in der Studie von Pérot *et al.* (2012) nur in Tumorgewebe differentiell exprimiert und stellt somit einen interessanten Locus im Zusammenhang mit Keimzelltumoren dar. In den Prostatakarzinom-Zelllinien war insbesondere der *ERVK-5*-Locus dominant vertreten, der in allen Zelllinien hohe relative Klonierungsfrequenzen zeigte. Dieser wurde auch in einigen Melanom-Proben und Melanozyten-Zelllinien mit mittlerer relativer Transkriptionsfrequenz detektiert.

Wie bereits der im Zusammenhang mit der Melanom-Studie erwähnte *ERVK-6*-Locus, konnten somit auch in GCT- und Prostatakarzinom-Zelllinien Loci identifiziert werden, die dominant vertreten waren. Auch hier könnte durch Untersuchung einer größeren Probenanzahl eine möglicherweise spezifische oder stark erhöhte Transkription der Loci weiter untersucht werden, was diese als mögliche Biomarker identifizieren könnte.

#### **5.2.4. Einige Melanom-Proben und beide Keimzelltumor-Zelllinien zeigen, verglichen mit Melanozyten, eine erhöhte HML-2 gag-Transkription**

Mittels *Real-Time*-PCR konnten in etwa der Hälfte der Melanom-Proben, verglichen mit den Melanozyten-Zelllinien, erhöhte HML-2 *gag*-Transkriptmengen nachgewiesen werden. Die Transkription war in der Melanom-Zelllinie WM3734a, drei Melanom RNA-Proben sowie einer Melanom-Lymphknoten-Metastase erhöht, verglichen mit der Transkription in beiden Melanozyten-Zelllinien (für die zweite Lymphknoten-Metastase konnten aus nicht bekannten Gründen auch nach mehrfacher Wiederholung keine brauchbaren Ergebnisse generiert werden, weswegen diese Probe nicht in die Analyse einging). Insgesamt waren die HML-2 *gag*-Transkriptmengen jedoch überall deutlich schwächer als in den GCT-Zelllinien NCCIT und Tera-1, in denen bereits in früheren Studien eine stark erhöhte HML-2-Transkription beschrieben wurde (siehe u.a. Florl *et al.*, 1999, Ruprecht *et al.*, 2008a). Solche erhöhten HML-2-Transkriptmengen in Melanomen und anderen Tumoren könnten, sofern sich unter den transkribierten Loci solche mit ORF für retrovirale Proteine finden, zur Expression von Proteinen in nachweisbaren Mengen führen und in früheren Studien detektierte HML-2-Proteine erklären (s. z.B. Büscher *et al.*, 2005; Reiche *et al.* 2010).

#### **5.2.5. Die Bestrahlung von Melanom- und Melanozyten-Zelllinien mit UVB resultierte in veränderten Transkriptionsprofilen sowie verringerten HML-2 gag-Transkriptmengen**

Aufgrund der mutagenen Wirkung von UV-Strahlung und der daraus resultierenden Rolle als Risikofaktor bei der Entstehung von Melanomen (Hill *et al.*, 2013) sowie einer kürzlich beschriebenen spezifisch erhöhten HML-2-Transkription und erhöhten Env-Expression in Melanom-Zellen nach UVB-Bestrahlung (Schanab *et al.*, 2011) wurde der Einfluss von UVB-Bestrahlung auf die HML-2-Transkription in Melanom- und Melanozyten-Zelllinien untersucht.

Eine Bestrahlung mit einer UVB-Dosis von 200 mJ/cm<sup>2</sup>, was einer 2 - 3 stündigen Sommer-Sonnen-Exposition entspricht (Schanab *et al.*, 2011), resultierte in der vorliegenden Arbeit in allen untersuchten Melanom- und Melanozyten-Zelllinien in einem leicht veränderten Transkriptionsprofil. Die Transkription des zuvor dominant vertretenen *ERVK-14*-Locus auf Chromosom 7q22.1 wurde in fünf der Melanom-Zelllinien und beiden Melanozyten-Zelllinien vermindert. Abgesehen davon kam es jedoch zu keiner auffälligen Aktivierung oder Deaktivierung einzelner spezifischer Loci, auch im Vergleich zwischen Melanom- und Melanozyten-Zelllinien.

In der Studie von Schanab *et al.* (2011) wurde nach UVB-Bestrahlung eine erhöhte HML-2 *pol*-Transkription in Melanom-Zelllinien beschrieben, wohingegen die Transkription in Melanozyten-Zelllinien verringert wurde. In der vorliegenden Arbeit konnte eine solche Erhöhung der Transkription in Zellen, die mit der UVB-Dosis bestrahlt wurden, die in der Studie von Schanab *et al.* (2011) in der stärksten Veränderung resultierte, nicht beobachtet werden. Im Gegenteil resultierte die UVB-Bestrahlung sowohl in Melanom- als auch in Melanozyten-Zelllinien deutlich verringerten HML-2 *gag*-Transkriptmengen (vgl. **Abb. 22**). Diese Unterschiede könnten darin begründet sein, dass in der Studie von Schanab *et al.* (2011) ein unterschiedliches Amplikon im HML-2 *pol*-ORF analysiert wurde. Ein Vergleich der theoretisch mit den *Primern* der Studie von Schanab *et al.* (2011) amplifizierbaren Loci mit denen, die vom hier verwendeten *gag*-Amplikon erkannt werden, ergab jedoch keine auffälligen Unterschiede. Es werden theoretisch etwa die gleichen Loci erkannt, wodurch es zu keiner selektiven Amplifikation spezifischer deregulierter Loci mit dem *pol*- oder *gag*-Amplikon kommen sollte. Der Sequenzvergleich ergab ebenfalls, dass die Unterschiede in den relativen Transkriptmengen nicht in der Amplifikation spezifischer Loci begründbar ist, deren Transkription nach UVB-Bestrahlung stark erhöht sein könnte, die jedoch aufgrund von Deletionen mit den *gag*-spezifischen *Primern* nicht detektiert werden können. Die Ursache für die beobachteten Unterschiede bleibt somit unklar. Zur Klärung könnte die Identifizierung der spezifischen Loci beitragen, die tatsächlich mit den *Primern* aus Schanab *et al.* (2011) amplifiziert werden, sowie deren Vergleich zu den mit dem hier verwendeten *gag*-Amplikon amplifizierten Loci.

### **5.2.6. Transkripte *rec* und *np9* mRNA-produzierender HML-2-Loci finden sich in Melanom- und Melanozyten-Zelllinien sowie menschlichen Normalgeweben**

Frühere Studien diskutierten eine mögliche Rolle der von verschiedenen HML-2-Loci kodierten akzessorischen Proteine Rec und Np9 in der Tumorgenese (vgl. z.B. Boese *et al.*, 2000; Armbruster *et al.*, 2004; Galli *et al.*, 2005; Denne *et al.*, 2007). Die Transkription von *rec* und *np9* mRNA-produzierenden HML-2-Loci konnte in der vorliegenden Arbeit sowohl in Melanom- und Melanozyten-Zelllinien als auch in zahlreichen gesunden menschlichen Geweben nachgewiesen werden (vgl. **Tab. 18** und **19**). Insgesamt wurden in den untersuchten Geweben und Zelllinien neun verschiedene Proviren des Typ I sowie acht verschiedene Typ II Proviren als transkribiert identifiziert.

In einer früheren Studie von Büscher *et al.* (2006) wurden *rec* und *np9* mRNA-Transkripte in Melanomen sowie *rec*-mRNA in Melanozyten identifiziert. Rec-Protein wurde jedoch nur in

wenigen Melanomen detektiert. Np9-Protein sowie Antikörper gegen die beiden akzessorischen Proteine konnten nicht nachgewiesen werden (Büscher *et al.*, 2006). Das Vorhandensein von *rec* und *np9* mRNA-Transkripten bedeutet somit nicht zwangsläufig, dass auch die entsprechenden Proteine in nachweisbarer Menge exprimiert werden. Die vorliegende Arbeit beschränkte sich auf die Identifizierung transkribierter potentiell Rec/Np9-kodierender Loci, wobei eine Quantifizierung der *rec/np9*-Transkriptmengen sowie die Analyse einer möglichen Proteinexpression nicht durchgeführt wurden. Es wäre somit durchaus möglich, dass Melanozyten und menschliche Normalgewebe, verglichen mit Melanomen, eine niedrigere Rec- und/oder Np9-Proteinexpression zeigen oder, dass Transkripte oder Proteine in Normalgewebe weniger stabil sind und schneller degradiert werden. Dies müsste in detaillierteren proteinbiochemischen und molekularbiologischen Experimenten getestet werden, wobei die vorliegende Arbeit Hinweise darauf liefert, welche Loci bzw. welche von diesen potentiell kodierte Proteine hierbei von Interesse sind. Sofern sich Rec- und/oder Np9-Proteine in allen Proben mit entsprechender mRNA-Transkription nachweisen ließen, zu denen auch zahlreiche menschliche Normalgewebe zählten, würde dies hingegen eher gegen eine direkte Rolle während der Melanom-Pathogenese sprechen. Es wäre jedoch auch denkbar, dass die Proteine erst im Zusammenspiel mit weiteren, möglicherweise spezifisch im Zuge der Tumorphagenese exprimierten Proteinen, eine biologische Funktion ausüben.

### 5.2.7. Identifizierung von Transkripten potentieller bisher unbekannter HML-2-Loci

Im Zuge der Identifizierung transkribierter *rec/np9* mRNA-produzierender HML-2-Loci in menschlichen Normalgeweben wurde eine auffällig hohe Anzahl an Sequenzen generiert, deren eindeutige Zuordnung zu einem genomischen HML-2-Locus aufgrund einer hohen Anzahl an Nukleotidunterschieden nicht möglich war.

Eine Sequenzpopulation wurde im Zuge der Rückkartierung als besten *BLAT-Match* dem *ERVK-5* Locus auf Chromosom 3q12.3 zugeordnet, die Sequenzen zeichneten sich jedoch durch jeweils mehr als sechs Nukleotidunterschiede gegenüber diesem Locus aus. Die Sequenzunterschiede fanden sich hierbei im Vergleich zu diesem immer an den gleichen Positionen und die generierten Sequenzen waren gegenüber solchen, die eindeutig dem *ERVK-5* Locus zugeordnet werden konnten, um sieben Nukleotide verkürzt. Ein genauerer Sequenzvergleich zeigte, dass es sich bei diesen Sequenzen um ein alternativ gespleißtes *np9*-Transkript handelt, bei dem eine alternative *splice acceptor site* sieben Nukleotide stromab der regulären *np9/rec SA site* genutzt wird (vgl. **Abb. 31**).

Eine weitere Sequenzpopulation lieferte im Zuge der Rückkartierung als besten *BLAT-Match* den *ERVK-18*-Locus auf Chromosom 1q23.2, wobei hier jeweils über 22 Sequenzunterschiede zu diesem Locus vorhanden waren, die sich größtenteils an den gleichen Positionen fanden. Auch bei diesen cDNA-Sequenzen handelt es sich um eine Spleißvariante eines Typ I Provirus, die einen zusätzlichen 244 bp langen Bereich besitzt, der sonst in Typ I Proviren stromab der charakteristischen 292 bp Deletionsregion folgt und in *np9*-mRNA normalerweise als Intronbereich gespleißt wird. Eine Suche mit einer cDNA-Sequenz der beschriebenen Sequenzpopulation in der *NCBI* Nukleotid-Datenbank mit dem *BLAST*-Algorithmus identifizierte eine im Zuge eines *whole genome shotgun sequencing* Experimentes generierte, etwa 4,2 kb lange Sequenz (*Genbank acc. no. ABBA01159463.1*), die identisch zur hier beschriebenen cDNA-Sequenz ist und die gleichen Nukleotidunterschiede zu *ERVK-18* aufweist.

Um auszuschließen, dass es sich bei den beschriebenen Sequenzpopulationen um ein *in vitro* generiertes Phänomen handelt, wurde das Vorhandensein beider Sequenzpopulationen auf genomischer Ebene verifiziert. Bei solchen Sequenzpopulationen mit charakteristischen Nukleotidunterschieden zu genomischen HML-2-Loci könnte es sich um allelische Varianten bekannter Proviren handeln, wobei über solche Polymorphismen innerhalb verschiedener HML-2-Loci jedoch noch relativ wenig bekannt ist (Mayer *et al.*, 2005; de Parseval *et al.*, 2005). Des Weiteren könnte es sich um neue, bisher nicht beschriebene, polymorphe Loci handeln, die nicht in der menschlichen hg18/hg19 Referenzsequenz annotiert sind. Es wurden kürzlich noch weitere polymorphe HML-2-Loci beschrieben, deren genaue Natur jedoch bisher noch nicht endgültig geklärt ist (Contreras-Galindo *et al.*, 2013). Um dies weiter zu untersuchen, müsste zunächst versucht werden, die genomische Lokalisation solcher Loci zu bestimmen. Hierfür könnten große *whole genome sequencing* Datensätze, wie sie beispielsweise im Rahmen des 1000-Genome-Projektes (1000 Genomes Project Consortium, 2010) generiert wurden, gezielt nach diesen Sequenzen durchsucht werden, um die den Locus flankierenden Sequenzen zu identifizieren und so seine Genomlokalisierung zu bestimmen. Dies würde zeigen, ob es sich hierbei um einen bereits bekannten HERV-Locus handelt, die Sequenzen also vermutlich allelische Varianten repräsentieren, oder ob für die spezifische Position noch kein Provirus, oder ggf. nur ein solitärer LTR, annotiert ist und die Sequenzen somit ein neues polymorphes Provirus repräsentieren. Im Anschluss könnte die DNA verschiedener Individuen in PCRs mit spezifischen, den Locus flankierenden Primern, wie sie auch in dieser Arbeit zur Bestimmung des Allelstatus polymorpher HML-2-Loci durchgeführt wurden, gezielt auf das Vorhandensein des Provirus untersucht werden, auch um eine Aussage über dessen Häufigkeit und Allelfrequenz zu treffen.

### 5.2.8. Unter den transkribierten HML-2-Loci finden sich stark verkürzte retrovirale Loci, die durch reverse Transkription und Reintegration gespleißter mRNA entstanden sind und potentiell für Rec- und Env-ähnliche Proteine kodieren

Im Zuge der *rec/np9*-Analysen wurde in der Melanozyten-Zelllinie Benno sowie in menschlichen Normalgeweben ein stark verkürzter HML-2-Locus auf Chromosom 2q32.1 als transkribiert identifiziert (vgl. **Abb. 24**). Der Locus besitzt die Struktur eines prozessierten Pseudogens, das durch L1-vermittelte Retrotransposition eines *rec* mRNA-Transkriptes generiert wurde. Solche retrotransponierten Loci wurden bereits für die Gruppen HERV-H und HERV-W beschrieben (Goodchild *et al.*, 1995, Pavlíček *et al.*, 2002). Im Falle der HML-2-Gruppe wurde zuvor jedoch noch kein Locus beschrieben, der durch einen solchen Mechanismus generiert wurde, wobei der Locus trotz seiner stark verkürzten Struktur transkribiert wird. Die tatsächliche Transkription des Locus wird durch fünf EST-Sequenzen (*Genbank acc. no.* AW968573, AA167225, BF921025, GD144208 und AA491129) unterstützt, die stromauf des Locus annotiert sind und mit diesem teilweise überlappen. Ein Sequenzvergleich mit anderen Primatengenomen deutet darauf hin, dass der 2q32.1-Locus vor der evolutiven Aufspaltung der Hominoidea gebildet wurde, da in den Genomen anderer Spezies jeweils der homologe Locus fehlt.

Der verkürzte 2q32.1-Locus wurde im Zuge der initialen Sequenz-Zusammenstellung genomischer HML-2-Loci durch *BLAT*-Suche mit der *Repbase* HERVK-Konsensussequenz (vgl. Absatz 4.1.2. der Ergebnisse und **Anhang 1**) nicht identifiziert. Auch in einer kürzlich veröffentlichten Studie, in der HML-2 Vollängen-Proviren und verkürzte Loci im menschlichen Genom beschrieben wurden, war der 2q32.1-Locus nicht enthalten (Subramanian *et al.*, 2011). Durch spezifische *BLAT*-Suchen im menschlichen Referenzgenom mit Sequenzen gespleißter *rec*- oder *np9*-Transkripte könnten unter Umständen weitere solche retrotransponierte HML-2-Loci im menschlichen Genom identifiziert werden.

In der Melanozyten-Zelllinie Benno sowie verschiedenen menschlichen Normalgeweben wurde ein weiterer verkürzter HML-2-Locus des Typs II auf Chromosom 10q24.2 als transkribiert identifiziert (vgl. **Abb. 26**). Dieser wurde vermutlich durch reverse Transkription und Reintegration eines HML-2 mRNA-Transkriptes durch einen retroviralen Mechanismus generiert. Der Locus weist eine knapp 1,4 kb große Deletion im *env*-ORF auf, die ungefähr dem intronischen Bereich innerhalb von *env* entspricht, der in *rec*-mRNA gespleißt wird, wobei jedoch eine unterschiedliche *splice donor site* stromab der regulären *rec SD site* genutzt wurde. Der Locus wurde vermutlich erst nach der evolutiven Spaltung von Mensch und Schimpanse gebildet, da der homologe Locus in anderen Primatengenomen fehlt.

Die Loci auf Chromosom 2q32.1 und 10q24.2 besitzen ORFs für Rec- und Env-ähnliche Proteine (vgl. **Abb. 25**). Das hypothetische Protein des 2q32.1-Locus besitzt lediglich in seinem N-Terminus Ähnlichkeit zu zuvor beschriebenen Rec- und Env-Proteinsequenzen, das hypothetische Protein des 10q24.2-Locus stellt ein chimäres Protein dar, das in Teilen Rec und Env ähnelt. Da es sich auch bei Np9 um eine Proteinvariante handelt, die aus der Deletion eines 292 bp langen Bereiches an der Grenze von *pol* zu *env* in Proviren des Typs I resultiert (Armbruster *et al.*, 2002), wäre es auch im Falle der von den Loci auf Chromosom 2q32.1 und 10q24.2 hypothetisch kodierten Proteine denkbar, dass diese translatiert werden und von biologischer Bedeutung sind.

Die Translation der theoretisch von den Loci auf Chromosom 2q32.1 und 10q24.2 kodierten Proteine wurde daher weiter untersucht. GFP-Fusionsproteine wurden in MEWO-Zellen exprimiert und mittels *Western Blot* und Immunfluoreszenz detektiert (vgl. **Abb. 28** und **29**). Die im *Western Blot* mit einem Anti-GFP-Antikörper detektierten Proteine entsprachen hierbei jedoch nicht den erwarteten Größen. Die nachweisbaren Fusionsproteine waren deutlich kleiner und im Falle des 10q24.2-Fusionsproteins trat eine Doppelbande auf. Dies deutet auf eine mögliche weitere Prozessierung der GFP-Fusionsproteine bzw. auf die Abspaltung spezifischer N-terminaler Polypeptidsequenzen hin, was in zusätzlichen Experimenten durch massenspektrometrische Bestimmung der im *Western Blot* nachgewiesenen Proteine und Identifizierung möglicher Spaltungsstellen genauer untersucht werden müsste. In beiden Fällen war zudem eine Proteinbande nachweisbar, deren Größe der von GFP allein entspricht. Dies könnte ein weiterer Hinweis darauf sein, dass es posttranslational zu weiteren Prozessierungen der Fusionsproteine kommt, die letztendlich in der Abspaltung der retroviralen Polypeptide vom GFP resultieren. Die abgespaltenen, von HML-2 stammenden Polypeptidanteile wären hierbei mittels GFP-Antikörpern nicht nachweisbar.

Mittels Immunfluoreszenz konnten im Falle des vom Locus auf Chromosom 2q32.1 kodierten Proteins GFP-Signale in einigen Fällen ubiquitär in den Zellen oder spezifisch nur im Zellkern bzw. den Nukleoli nachgewiesen werden. Um was es sich bei diesen detektierten Proteinen genau handelt, bzw. wie groß der HML-2-Anteil dieser tatsächlich ist, lässt sich hierbei jedoch aufgrund der von der erwarteten Größe abweichenden Proteinbande im *Western Blot* nicht sagen. Die unterschiedliche zelluläre Lokalisation könnte möglicherweise auf eine unterschiedliche Prozessierung und das potentielle Vorhandensein spezifischer Zellkern-Lokalisationssignale zurückzuführen sein. GFP-Fusionsproteine des vom Locus auf Chromosom 10q24.2 kodierten Proteins, über deren tatsächlichen HML-2-Anteil ebenfalls keine genaue Aussage getroffen werden kann, konnten in der gesamten Zelle nachgewiesen werden, wobei

die Grünfluoreszenz hier deutlich schwächer ausgeprägt war, was durch eine schwächere Expression des Fusionsproteins oder eine verringerte Menge nach Prozessierung begründbar wäre. Solche Prozessierungen bzw. die im Experimente entstanden Proteine müssten in spezifisch zu wählenden proteinbiochemischen Experimenten eingehender untersucht werden.

In menschlichen Normalgeweben fanden sich Transkripte eines weiteren stark verkürzten Typ II Locus auf Chromosom 5q15, der, ähnlich dem HML-2-Locus auf Chromosom 2q32.1, einem retrotransponierten *rec* mRNA-Transkript ähnelt, das innerhalb eines L1-Elementes inseriert wurde. Aufgrund zahlreicher Insertionen und Deletionen gegenüber *rec*-mRNA liefert eine *in silico* Translation der Sequenz lediglich stark verkürzte hypothetische Proteine, die keine signifikante Ähnlichkeit zu HML-2-Proteinen aufweisen. Die tatsächliche Translation solcher hypothetischer Proteine müsste mittels *Western Blot* oder Immunfluoreszenz näher untersucht werden. Zur weiteren Klärung einer potentiellen biologischen Funktion könnten potentielle Interaktionspartner beispielsweise durch Co-Immunopräzipitation oder *Two-Hybrid Screening* (vgl. z.B. Kaboord und Perr, 2008; Reece-Hoyes *et al.*, 2012) und anschließende massenspektrometrische Bestimmung identifiziert werden.

### **5.3. Transkriptionsprofile der HERV-W-Gruppe und Charakterisierung transkribierter Loci in Hirnproben von Multiple Sklerose Patienten und gesunden Kontrollen**

Für die HERV-W-Gruppe wird aufgrund der Ergebnisse früherer Studien eine potentielle Rolle in der Pathogenese von Multipler Sklerose diskutiert. Sowohl eine erhöhte HERV-W-Transkription als auch das Vorhandensein oder die Überexpression retroviraler Proteine sowie z.T. eine demyelinisierende oder inflammatorische Wirkung dieser wurden in verschiedenen Studien im Zusammenhang mit MS beschrieben (vgl. Roland *et al.*, 2006; Antony *et al.*, 2004, 2006, 2007; Mameli *et al.*, 2007; Roebke *et al.*, 2010; und Abschnitt 2.3.2.1. der Einleitung). Als weiterer Schritt zur Klärung der potentiellen Rolle von HERV-Transkripten oder Proteinen in MS ist eine detaillierte Analyse transkribierter Loci von großem Interesse, insbesondere was potentielle Unterschiede in der Transkription verschiedener Loci zwischen MS-Geweben und gesunden Vergleichsgeweben betrifft. Da sich unter den potentiell transkribierten HERV-W-Loci auch die Protein-codierenden Loci *ERVWE1/ERVW-1/syncytin-1* auf Chromosom 7q21.2 und *ERVWE2/ERVW-2* auf Chromosom Xq22.3 finden, sind insbesondere auch Unterschiede in der Transkription dieser beiden Loci von Interesse, welche sich auf die mögliche Expression der kodierten Proteine auswirken könnten.

### 5.3.1. Die cDNA Sequenzierung mittels *Next Generation Sequencing* erlaubt die Generierung umfassender Datensätze und die Detektion einer großen Anzahl auch schwach transkribierter Loci

Durch Generierung HERV-W Gruppen-spezifischer RT-PCR-Produkte sowie deren anschließende Sequenzierung mittels *Next Generation Sequencing* (NGS) und Rückkartierung im menschlichen Referenzgenom konnte die bisher umfangreichste Übersicht über transkribierte HERV-W-Loci in *Plaques* von MS-Patienten sowie in der weißen Substanz von Kontrollpersonen ohne neurologische Krankheiten erstellt werden.

Im Schnitt wurden für jede untersuchte Probe 2.200 *454/FLX*-Sequenzen bzw. 348.000 *MiSeq/Illumina*-Sequenzen analysiert. Durch Generierung einer solch hohen Anzahl an Sequenzen wird das Risiko potentiell verfälschter Transkriptionsmuster gegenüber weniger umfangreichen Studien, in denen nur eine geringe Anzahl an cDNA-Sequenzen generiert wird, deren Auswahl aus der Gesamt-cDNA immer mehr oder weniger zufällig erfolgt, minimiert. Zudem erlaubt der NGS-Ansatz und die damit verbundene hohe Anzahl generierter cDNA-Sequenzen auch die umfangreiche Detektion schwach transkribierter Loci, die in der Gesamt-cDNA generell unterrepräsentiert sind. Dies zeigt sich auch in der Tatsache, dass durch *MiSeq/Illumina*-Sequenzierung, mit der gegenüber *454/FLX* erneut eine 160-fach höhere Anzahl an Sequenzen pro untersuchter Probe generiert wurde, auch die Anzahl als transkribiert identifizierter Loci um weitere 40 Loci erweitert wurde.

Aufgrund des hohen evolutiven Alters der HERV-W-Gruppe und der damit verbundenen höheren Anzahl an Sequenzunterschieden zwischen den Loci war, trotz der vergleichsweise kurzen NGS-Leseweiten, eine sichere Zuordnung der generierten Sequenzen zu genomischen Loci möglich. Im Zuge der Rückkartierung wiesen, verglichen mit dem besten *BLAT-Match*, die nächstbesten Treffer immer mehrere zusätzliche Nukleotidunterschiede auf, im Gegensatz zu HML-2, wo der Unterschied oft nur ein Nukleotid betraf. Die sichere Rückkartierung wurde weiter durch die Tatsache unterstützt, dass die zwei unabhängig verwendeten NGS-Methoden, *454/FLX* sowie *MiSeq/Illumina*, insgesamt sehr ähnliche Transkriptionsprofile lieferten (vgl. **Abb. 39**). Die hier verwendete Strategie eignet sich somit zur umfangreichen Identifizierung transkribierter HERV-W-Loci in einem biologischen Gewebe von Interesse. Eine gute Unterscheidbarkeit der genomischen Loci vorausgesetzt, wäre mit der Methode auch die umfangreiche Transkriptionsanalyse weiterer HERV-Gruppen mit hoher Kopienzahl im menschlichen Genom denkbar, für die eine potentielle Rolle in der Pathogenese verschiedener Krankheiten diskutiert wird, beispielsweise HERV-H (vgl. Antony *et al.*, 2011) oder HERV-E (vgl.

z.B. Wang-Johanning *et al.*, 2003). Auf diese Weise können im großen Maßstab potentielle Transkriptionsunterschiede zwischen kranken und Normalgeweben detektiert und ggf. spezifisch transkribierte oder auffällige Loci identifiziert werden, die für weiterführende detailliertere Studien von Interesse sind.

Des Weiteren ermöglicht ein NGS-Ansatz zur Identifizierung transkribierter Loci und die damit verbundene hohe Anzahl generierter Sequenzen die Identifikation biologisch relevanter Polymorphismen oder somatischer Mutationen innerhalb der Sequenz transkribierter HERV-Loci. Die detaillierte bioinformatische Analyse von NGS-Datensätzen in Hinblick auf häufiger auftretende Sequenzvarianten könnte somit Hinweise auf allelische Varianten spezifischer Loci liefern. Insbesondere sind hierbei sicherlich Polymorphismen innerhalb proteinkodierender Bereiche von Interesse. Spezifische Mutationen innerhalb von ORFs könnten beispielsweise in der Generierung eines verfrühten Stopp-Codons resultieren, wodurch die Expression eines Proteins mit potentiell biologischer Funktion in diesen Fällen unterbunden wird.

### **5.3.2. In Hirnproben von MS-Patienten und gesunden Kontrollen findet sich eine hohen Anzahl transkribierter HERV-W-Loci**

In einer früheren Studie von Laufer *et al.* (2009) wurden bereits sieben HERV-W-Loci in PBMCs von MS-Patienten und gesunden Kontrollen als transkribiert identifiziert. In der vorliegenden Arbeit konnte durch Analyse zweier optimierter Amplikons, die Deletionen im 5'-Bereich des *env*-ORF vieler HERV-W-Loci berücksichtigen (vgl. Abschnitt 5.1. der Diskussion), und die Generierung umfangreicher NGS-Datensätze insgesamt etwa 180 transkribierte HERV-W-Loci in Hirnproben von MS-Patienten und gesunden Kontrollen detektiert werden. Die Detektion einer solch hohen Anzahl transkribierter Loci wurde insbesondere durch Verwendung des zusätzlichen Amplikons im 3'-Bereich von *env* ermöglicht, da ein großer Teil der HERV-W-Loci relativ große Deletionen im 5'-Bereiche aufweist (vgl. **Anhang 9**).

Ein großer Teil der generierten cDNA-Sequenzen stammte von relativ wenigen Loci, was auf eine relativ starke Transkription dieser hindeutet. Im Falle des 5'*env* 454/*FLX*-Datensatzes waren dies sieben dominante Loci, die 96 % der generierten cDNA-Sequenzen ausmachten, im 3'*env* 454/*FLX*-Datensatz stammten 83 % der Transkripte von 25 verschiedenen Loci. Unter diesen stärker transkribierten Loci, die in allen MS- sowie auch Kontrollproben transkribiert waren, fanden sich u.a. der *ERVWE1/ERVW-1/syncytin-1*-Locus auf Chromosom 7q21.1, der *ERVWE2/ERVW-2*-Locus auf Chromosom Xq22.3 sowie die Loci *ERVW-4* (15q21.3) und

*ERVW-17* (6q21). Diese vier Loci wurden bereits in der Studie von Laufer *et al.* (2009) anhand ihrer cDNA-Klonierungsfrequenzen als dominant in PBMCs von MS-Patienten und gesunden Kontrollen beschrieben.

Die übrigen NGS-cDNA Sequenzen verteilten sich auf eine große Anzahl an Loci, die relative Transkriptionsraten von weniger als 1 % aufwiesen, was auf eine schwächere Transkription dieser Loci hindeutet. Die Transkriptionsmuster dieser schwächer transkribierten Loci unterschieden sich in allen untersuchten Proben und auch zwischen den Proben MS4 und MS5, die vom selben Patienten stammten. Im Falle einiger Loci, wie beispielsweise des Locus auf Chromosom 1q22, der im Zuge der *454/FLX*-Sequenzierung lediglich in einer Kontrollprobe mit einer Sequenz detektiert wurde, ist eine DNA-Kontamination nicht komplett auszuschließen. Auch wenn alle RNA-Proben einem strengen DNase-Verdau unterzogen wurden, dessen Erfolg durch mehrere spezifische Kontrollen in den nachfolgenden Experimenten überprüft wurde, ist es nicht komplett auszuschließen, dass doch noch minimalste Reste genomischer DNA in den RNA-Proben vorhanden waren. HERV-W-Loci, die in solchen NGS-Datensätzen von mehreren Zehntausend bzw. mehreren Millionen Sequenzen durch wenige oder teilweise eine einzige Sequenz detektiert werden, könnten auf solche minimalen Überreste genomischer DNA zurückzuführen sein. Es ist jedoch schwer, in diesem Fall eine genaue Grenze für die falsch-positive Detektion transkribierter Loci zu ziehen, zumal viele Bereiche des menschlichen Genoms anscheinend schwach transkribiert sind und durch wenige Sequenzen in *RNA-Seq*-Datensätzen repräsentiert sind (van Bakel *et al.*, 2010; Hangauer *et al.*, 2011). Eine ähnlich schwache Transkription wäre durchaus auch für HERV-W-Loci denkbar, die evtl. in solchen schwach transkribierten Genomregionen lokalisiert sind. Des Weiteren sind zahlreiche der hier detektierten Loci mit niedrigen relativen Transkriptionsfrequenzen intronisch innerhalb von Genen lokalisiert (vgl. hierzu die Tabellen im **Anhang 15** und **16**). cDNA solcher Loci könnte demnach auch von ungespleißter hnRNA generiert worden sein.

Dennoch zeigt eine hohe Anzahl HERV-W-Loci höhere relative cDNA-Frequenzen, was auf eine tatsächliche Transkription dieser Loci in Hirngewebe hindeutet. Den 23 Loci, die mit höheren *454/FLX* cDNA-Frequenzen von über 1 % identifiziert wurden und die noch nicht in der Studie von Laufer *et al.* (2009) beschrieben waren, wurden im Zuge der offiziellen Benennung transkribierter HERV-Loci die offiziellen HGNC-Namen *ERVW-7* bis *ERVW-29* zugewiesen (vgl. **Tab. 20** und **Anhang 12**). Setzt man diesen Wert von 1 % als Grenze, sind insgesamt 28 Loci mit höherer relativer Frequenz in Hirngewebe von MS-Patienten und gesunden Kontrollen transkribiert, und weitere ~120 Loci mit niedrigerer Frequenz.

Die Transkription einer hohen Anzahl verschiedener HERV-W-Loci, auch in Proben ohne Bezug zu Multipler Sklerose, wird durch Daten aus *RNA-Seq*-Experimenten unterstützt, die im Rahmen des *ENCODE* Projektes (ENCODE Project Consortium, 2004) aus diversen menschlichen Zelllinien generiert wurden. Auch in diesen Datensätzen finden sich in jeder der hier analysierten sieben Zelllinien ohne Bezug zu MS über 90 verschiedene der in der vorliegenden Arbeit identifizierten Loci als transkribiert und nur wenige Loci sind in den *RNA-Seq*-Datensätzen nicht repräsentiert (vgl. **Tab. 20** und **Anhang 16**).

### **5.3.3. Die HERV-W *env*-Transkriptionsmuster in MS und gesunden Vergleichsgeweben unterscheiden sich nicht signifikant**

Insgesamt zeigten sich keine signifikanten Transkriptionsprofile, die die MS-Proben von den gesunden Vergleichsgeweben unterscheiden. Dies deckt sich mit den Ergebnissen der Studie von Laufer *et al.* (2009), in der ebenfalls ähnliche Transkriptionsmuster der sieben dort identifizierten HERV-W-Loci in PBMCs von MS-Patienten und gesunden Kontrollen beschrieben wurden. Die Loci *ERVWE1/ERVW-1/syncytin-1* und *ERVWE2/ERVW-2*, die insbesondere aufgrund ihrer Genprodukte im Kontext von MS von Bedeutung sein könnten und die in der vorliegenden Arbeit zu den relativ stark transkribierten Loci zählten, zeigten in MS- und Kontrollproben vergleichbare relative Transkriptionsfrequenzen. Dies wurde auch durch Quantifizierung der Locus-spezifischen Transkriptmengen mittels *Real-Time-PCR* weiter bestätigt, wo sich kein signifikanter Unterschied zeigte (vgl. Abschnitt 5.3.4.).

Der *ERVWE1/ERVW-1/syncytin-1*-Locus zeigte vergleichbare hohe relative Transkriptionsfrequenzen in MS und gesundem Gewebe. Im Falle der 5'*env*-Daten zeigte sich zwar eine statistisch signifikant niedrigere Transkription des *ERVWE1/ERVW-1/syncytin-1*-Locus, die Unterschiede waren jedoch für den 3'*env*-Datensatz nicht signifikant. Solche Unterschiede, die für einige HERV-W-Loci beobachtet wurden und die in den unterschiedlichen Datensätzen gegensätzliche Ergebnisse lieferten, sind vermutlich eher auf stochastische Schwankungen im Experiment zurückzuführen und nicht auf tatsächlich signifikante Unterschiede in der Transkription der Loci. Für den *ERVWE2/ERVW-2*-Locus sowie weitere Loci mit höherer relativer Transkription zeigten sich keine signifikanten Unterschiede. Einige Loci wurden ausschließlich in MS-Proben identifiziert, darunter beispielsweise Loci auf Chromosom 2q22.1 oder 12q23.2, jedoch waren diese ausnahmslos sehr schwach und in nur wenigen Proben transkribiert. Die biologische Signifikanz solcher minimalen Unterschiede ist jedoch noch unklar und müsste spezifisch weiter untersucht werden, indem beispielsweise durch sensitive Locus-

spezifische RT-PCRs potentielle Unterschiede der Transkription dieser Loci in MS- und Vergleichsgewebe analysiert werden.

Generell kann die unterschiedliche Transkription verschiedener HERV-W-Loci auf einen unterschiedlichen epigenetischen Status in den entsprechenden Genomregionen hindeuten. Der reprimierende Einfluss von DNA-Methylierung auf die Transkription des *ERVWE1/ERVW-1/syncytin-1*-Locus wurde bereits beschrieben (Matouskova *et al.*, 2006; Gimenez *et al.*, 2009; Trejbalová *et al.*, 2011) und entsprechende Analysen des epigenetischen Status weiterer Loci könnte potentielle Unterschiede bzgl. Methylierung und aktivierenden/reprimierenden Histonmodifikationen spezifischer Genomregionen in MS- und gesundem Gewebe zeigen.

Da die in der vorliegenden Arbeit untersuchten Hirnproben aus entzündlichen *Plaques* von MS-Patienten stammen, könnten, neben Transkripten neuronaler Zellen, auch Transkripte anderer Zellen, die spezifisch im Entzündungsprozess aktiviert wurden, in die Analyse mit eingegangen sein. Eine detaillierte Untersuchung solcher potentiellen Zelltyp-spezifischen Transkriptionsunterschiede wäre durch *single-cell RNA-Seq* Analyse einzelner Zellen möglich (Tang *et al.*, 2010; Hashimshony *et al.*, 2012).

#### **5.3.4. Die Quantifizierung der HERV-W-Transkriptmengen liefert kein einheitliches Bild für MS- und Kontrollproben und deutet auf interindividuelle Unterschiede hin**

Die Quantifizierung der HERV-W-Transkriptmengen insgesamt mittels *Real-Time*-PCR, sowie die Quantifizierung der Transkriptmengen der Loci *ERVWE1/ERVW-1/syncytin-1* und *ERVWE2/ERVW-2* mit Locus-spezifischen *Primern* (vgl. de Parseval *et al.*, 2003; Laufer *et al.*, 2009; Roebke *et al.*, 2010) lieferte keine signifikanten Unterschiede zwischen MS- und gesundem Gewebe. Es ergab sich kein einheitliches Bild für MS- und gesunde Proben, da einige der MS-Proben relative HERV-W-Transkriptmengen zeigten, die gegenüber den Kontrollen erniedrigt waren, wohingegen in anderen MS-Proben die relativen Transkriptmengen erhöht war. Dies steht im Gegensatz zu früheren Studien, die eine erhöhte HERV-W *env*-Transkription in MS beschrieben, jedoch teilweise gegensätzliche Ergebnisse lieferten. So wurde eine erhöhte Transkription von *env*-mRNA in Gehirnproben (Antony *et al.*, 2006, 2007b) und PBMCs (Mameli *et al.*, 2007) von MS-Patienten gegenüber gesunden Kontrollen beschrieben. Eine erhöhte *env* mRNA-Transkription in PBMCs wurde in der Studie von Antony

*et al.* (2006) nicht gefunden. Bezüglich der Studien von Antony *et al.*, (2006, 2007b) wurden jedoch kürzlich potentielle technische Fehler diskutiert (Garson *et al.*, 2009).

Statt genereller Unterschiede zwischen MS und gesundem Gewebe deuten die Ergebnisse der Quantifizierungen eher auf interindividuelle Unterschiede in der HERV-W-Transkription hin. Solche interindividuellen Transkriptionsunterschiede wurden für HERV-W auch in *Microarray*-Experimenten (Frank *et al.*, 2005, 2008) beschrieben, sowie in einer Studie, in der die HERV-W *gag*-Transkription durch Analyse der Schmelztemperaturen in semi-quantitativen PCRs untersucht wurde (Nellåker *et al.*, 2009). In früheren Studien wurde gezeigt, dass im Fall zellulärer Gene interindividuelle Unterschiede auf genetischen Variationen beruhen können (Cheung *et al.*, 2003, 2010). Solche Variationen könnten auch die Transkription von HERV-W-Loci beeinflussen, was in weiterführenden *GOG*E (*genetics of gene expression*)-Studien detailliert untersucht werden könnte, die darauf abzielen, mittels spezifischer bioinformatischer Methoden regulatorische Regionen und DNA-Sequenzvarianten in umfangreichen Datensätzen zu identifizieren, die in einer individuell unterschiedlichen Genexpression resultieren (vgl. z.B. Cheung und Spielman, 2009) Es wäre zudem denkbar, dass sich die Transkription von HERV-W in verschiedenen Hirnregionen unterscheidet. Dies wird im Falle der Proben MS4 und MS5 deutlich, die vom selben Patienten stammen, sich jedoch in ihren relativen Transkriptmengen unterscheiden und auch ein leicht unterschiedliches Transkriptionsmuster spezifischer HERV-W-Loci zeigen. Auf solche Transkriptionsunterschiede in verschiedenen Hirnregionen deuten auch die Ergebnisse der zuvor erwähnten Schmelztemperatur Studie von Nellåker *et al.* (2009) hin. Es wäre denkbar, dass zuvor beschriebene MS-spezifische Unterschiede in der HERV-Transkription auf solchen inter- und intraindividuellen Unterschieden in der HERV-Transkription beruhen, was in spezifischen weiterführenden Studien, in denen detaillierte Transkriptionsprofile verschiedener Hirnregionen erstellt und verglichen werden, getestet werden müsste.

Die Ergebnisse der vorliegenden Arbeit deuten insgesamt darauf hin, dass, sofern HERV-W in der Krankheitspathogenese von MS involviert sein sollte, dies nicht auf eine deregulierte Transkription spezifischer Loci, insbesondere der proteincodierenden Loci *ERVWE1/ERVW-1/syncytin-1* und *ERVWE2/ERVW-2*, zurückzuführen ist. Beide Loci sind sowohl in MS als auch in Kontrollgewebe transkribiert und die entsprechenden Proteine, Syncytin-1 und N-Trenv, könnten auch dort, zumindest potentiell, translatiert werden. Weiterführende Studien zur Expression der Proteine in verschiedenen Geweben sind hierbei sicherlich von großem biologischem Interesse. Auch wenn Transkripte beider Loci in MS und Kontrollgeweben detektiert werden, wäre es dennoch denkbar, dass die tatsächliche Translation nur im Fall von MS stattfindet oder die Proteine nur unter bestimmten zusätzlichen Bedingungen

stabil und funktionsfähig sind. Wäre eine Expression auch in Normalgeweben nachweisbar, würde dies hingegen gegen eine direkt pathogene Wirkung der Proteine sprechen.

### **5.3.5. HERV-W-LTRs zeigen trotz zahlreicher Deletionen oft noch Promotoraktivität und könnten die Transkription der entsprechenden Loci induzieren**

In Reportergeren-Assays konnte eine Promotoraktivität zahlreicher HERV-W-LTRs gezeigt werden, auch wenn diese LTRs meist große Deletionen aufweisen. Es ist anzumerken, dass die Reportergeren-Assays im Rahmen der vorliegenden Arbeit in der plazentalen Chorionkarzinom-Zelllinie JEG-3 durchgeführt wurden, die schon in einer früheren Studie die HERV-W LTR-Transkription unterstützte (Prudhomme *et al.*, 2004). Über eine vergleichbare Promotoraktivität in für MS relevanten neuronalen Zellen, die in die vorangehenden Transkriptionsanalysen eingingen, kann keine Aussage getroffen werden. Frühere Studien zur Untersuchung der HERV-W LTR-Aktivität befassten sich hauptsächlich mit plazentalen und Tumor-Zelllinien (Prudhomme *et al.*, 2004; Gimenez *et al.*, 2009).

Die Ergebnisse der Reportergeren-Assays deuten darauf hin, dass detektierte HERV-W-Transkripte vieler Loci potentiell von ihren eigenen Promotoren initiiert werden könnten, die in den Überresten der flankierenden LTRs vorhanden sind. Hierbei zeigten, mit Ausnahme des *ERVW-13* 3'-LTRs, die getesteten 3'-LTRs, denen die U5 Region fehlt, eine höhere Aktivität in *antisense*-Orientierung. Frühere Studien beschrieben ebenfalls eine Promotoraktivität der U3- und R-Regionen von HERV-W-LTRs (Gimenez *et al.*, 2010) sowie der R/U5-Region (Lee *et al.*, 2003). Hierbei ist anzumerken, dass sich auch im Falle der HERV-W-Loci, die durch L1-vermittelte Retrotransposition generiert wurden, normalerweise noch eine Kopie der R-Region sowie die U3-Region im 3'-LTR findet, die in diesem Fall, zumindest potentiell, eine *Antisense*-Transkription des betreffenden Locus initiieren könnten. Der tatsächliche Beitrag der U3-, R- und U5-Regionen unvollständiger HERV-W-LTRs zur *Sense*- und/oder *Antisense*-Transkription spezifischer Loci müsste jedoch in weiterführenden Studien untersucht werden. Hierbei könnten Reportergeren-Konstrukte einer größeren Zahl von als transkribiert identifizierten HERV-W-LTRs sowie spezifischer Bereiche aus diesen generiert werden, deren potentielle Promotoraktivität in Reportergeren-Assays bestimmt werden könnte. Des Weiteren müsste auch eine Promotoraktivität der LTRs in neuronalen Zellen bestätigt werden.

Die Ergebnisse der Reportergeren-Studien deuten somit, neben der Promotoraktivität strukturell intakter 5'- und 3'-LTRs, auch auf eine potentielle Promotoraktivität von unvollständigen

HERV-W-LTRs transkribierter Loci hin. Dies ist auch im Hinblick auf den potentiellen Einfluss, den HERV-W-Loci auf benachbarte zelluläre Gene ausüben, von großer Bedeutung. Viele LTRs retroviraler Loci, auch wenn diesen als prozessierte Pseudogene spezifische LTR-Bereiche fehlen, sowie auch solitäre LTRs könnten durch ihre potentielle Promotoraktivität alternative Transkripte benachbarter zellulärer Gene initiieren, wobei eine Deregulation der Transkription dieser Gene in bestimmten Zelltypen oder Krankheiten durch eine veränderte LTR-Aktivität nicht auszuschließen ist.

In einigen Fällen könnte die Transkription spezifischer HERV-W-Loci, trotz potentieller Promotoraktivität ihrer LTRs, auf ihre Nähe zu zellulären Genen zurückzuführen sein. So könnten beispielsweise *read-through*-Transkripte der Gene *ZNF99*, *C14orf159*, *GRK7* oder *RBM41*, welches etwa 9,5 kb stromauf des *ERVWE2/ERVW-2*-Locus liegt und in dessen Richtung transkribiert wird, zur Transkription von HERV-W-Sequenzen in direkter Nachbarschaft stromab der Gene beitragen. Das Gleiche gilt für Loci, die zwischen zwei zellulären Genen lokalisiert sind. Es finden sich insgesamt jedoch nur wenige Loci, für die aufgrund ihrer Nähe zu zellulären Genen eine solche indirekte Transkription wahrscheinlich wäre.

### **5.3.6. Zuvor publizierte MSRV-Sequenzen können auf Rekombinationen transkribierter HERV-W-Loci zurückzuführen sein**

Aufgrund früherer Studien, in denen retrovirale Sequenzen in MS-Patienten detektiert wurden, wurde die mögliche Rolle eines potentiellen exogenen Retrovirus, MSRV, in der Krankheitspathogenese von MS diskutiert (Blond *et al.*, 1999; Mameli *et al.*, 2007, 2009; Dolei und Perron 2009; Garcia-Montojo *et al.*, 2013; vgl. auch Abschnitt 2.3.2.1. der Einleitung). Eine genaue Untersuchung der beschriebenen MSRV-Sequenzen spricht jedoch dafür, dass es sich bei diesen Sequenzen schlicht um Transkripte endogener HERV-W-Loci oder *in vitro* generierte Rekombinanten aus solchen Transkripten handelt (Laufer *et al.*, 2009). Die beschriebenen MSRV-Sequenzen könnten auf Transkripte der Loci auf Chromosom 3q23, Xq22.3, 15q21.3 (*ERVW-4*) und 3q26.32 (*ERVW-5*) oder auf Rekombinanten zwischen Transkripten der Loci 3p12.3 und 18q21.32 sowie zwischen *Xq22.3/ERVWE2/ERVW-2* und 5p12 zurückzuführen sein. In der Studie von Laufer *et al.* (2009) konnten jedoch nur die Loci auf Chromosom 15q21.3 (*ERVW-4*) und *Xq22.3B/ERVWE2/ERVW-2* als in PBMCs von MS-Patienten und gesunden Kontrollen transkribiert identifiziert werden. In der vorliegenden Arbeit konnte erstmals die Transkription aller dieser Loci in Hirnproben von MS-Patienten und gesunden Kontrollen gezeigt werden.

Auch wenn einige der Loci nur mit relativ niedriger Transkriptionsfrequenz in Hirnproben detektiert wurden, was nicht zwangsläufig eine gleichermaßen niedrige Transkription in den im Zuge der MSR-V-Studien untersuchten Proben bedeuten muss, unterstützt ihre Transkription dennoch die Annahme, dass alle beschriebenen MSR-V-Sequenzen, die sich nicht direkt einem genomischen HERV-W-Locus zuordnen lassen, nun durch *in vitro* Rekombination verschiedener transkribierter HERV-W-Loci erklärt werden können.

## 6. Literaturverzeichnis

- 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA (2010) **A map of human genome variation from population-scale sequencing.** *Nature* 467(7319):1061-73
- Adami HO, Bergström R, Möhner M, Zatoński W, Storm H, Ekblom A, Tretli S, Teppo L, Ziegler H, Rahu M (1994) **Testicular cancer in nine northern European countries.** *Int J Cancer* 59(1):33-8
- Adelman MK, Marchalonis JJ (2002) **Endogenous retroviruses in systemic lupus erythematosus: candidate lupus viruses.** *Clin Immunol* 102(2):107-16
- Agoni L, Guha C, Lenz J (2013) **Detection of Human Endogenous Retrovirus K (HERV-K) Transcripts in Human Prostate Cancer Cell Lines.** *Front Oncol* 3:180
- Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (eds) (2002) **Molecular Biology of the Cell. 5th Edition.** Garland Science, Taylor & Francis Group, New York
- Andersson AC, Svensson AC, Rolny C, Andersson G, Larsson E (1998) **Expression of human endogenous retrovirus ERV3 (HERV-R) mRNA in normal and neoplastic tissues.** *Int J Oncol* 12(2):309-13
- Andersson ML, Lindeskog M, Medstrand P, Westley B, May F, Blomberg J (1999) **Diversity of human endogenous retrovirus class II-like sequences.** *J Gen Virol* 80 (Pt 1):255-60
- Antony JM, van Marle G, Opii W, Butterfield DA, Mallet F, Yong VW, Wallace JL, Deacon RM, Warren K, Power C (2004) **Human endogenous retrovirus glycoprotein-mediated induction of redox reactants causes oligodendrocyte death and demyelination.** *Nat Neurosci* 7(10):1088-95
- Antony JM, Izad M, Bar-Or A, Warren KG, Vodjgani M, Mallet F, Power C (2006) **Quantitative analysis of human endogenous retrovirus-W env in neuroinflammatory diseases.** *AIDS Res Hum Retroviruses* 22(12):1253-9
- Antony JM, Ellestad KK, Hammond R, Imaizumi K, Mallet F, Warren KG, Power C (2007a) **The human endogenous retrovirus envelope glycoprotein, syncytin-1, regulates neuroinflammation and its receptor expression in multiple sclerosis: a role for endoplasmic reticulum chaperones in astrocytes.** *J Immunol* 179(2):1210-24
- Antony JM, Zhu Y, Izad M, Warren KG, Vodjgani M, Mallet F, Power C (2007b) **Comparative expression of human endogenous retrovirus-W genes in multiple sclerosis.** *AIDS Res Hum Retroviruses* 23(10):1251-6.
- Antony JM, Deslauriers AM, Bhat RK, Ellestad KK, Power C (2011). **Human endogenous retroviruses and multiple sclerosis: innocent bystanders or disease determinants?** *Biochim Biophys Acta* 1812(2):162-76
- Armbruester V, Sauter M, Krautkraemer E, Meese E, Kleiman A, Best B, Roemer K, Mueller-Lantzsch N (2002) **A novel gene from the human endogenous retrovirus K expressed in transformed cells.** *Clin Cancer Res* 8(6):1800-7
- Armbruester V, Sauter M, Roemer K, Best B, Hahn S, Nty A, Schmid A, Philipp S, Mueller A, Mueller-Lantzsch N (2004) **Np9 protein of human endogenous retrovirus K interacts with ligand of numb protein X.** *J Virol* 78(19):10310-9
- Bannert N, Kurth R (2004) **Retrotelements and the human genome: new perspectives on an old relation.** *Proc Natl Acad Sci U S A* 101 Suppl 2:14572-9
- Bannert N, Kurth R (2006) **The evolutionary dynamics of human endogenous retroviral families.** *Annu Rev Genomics Hum Genet* 7:149-73
- Barbulescu M, Turner G, Seaman MI, Deinard AS, Kidd KK, Lenz J (1999) **Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans.** *Curr Biol* 9(16):861-8
- Barski A, Zhao K (2009) **Genomic location analysis by ChIP-Seq.** *J Cell Biochem* 107(1):11-8
- Beimforde N, Hanke K, Ammar I, Kurth R, Bannert N (2008) **Molecular cloning and functional characterization of the human endogenous retrovirus K113.** *Virology* 371(1):216-25
- Belshaw R, Dawson AL, Woolven-Allen J, Redding J, Burt A, Tristem M (2005) **Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity.** *J Virol* 79(19):12507-14
- Berkhout B, Jebbink M, Zsiros J (1999) **Identification of an active reverse transcriptase enzyme encoded by a human endogenous HERV-K retrovirus.** *J Virol* 73(3):2365-75

- Bessis D, Molès JP, Basset-Séguin N, Tesniere A, Arpin C, Guilhou JJ (2004) **Differential expression of a human endogenous retrovirus E transmembrane envelope glycoprotein in normal, psoriatic and atopic dermatitis human skin.** *Br J Dermatol* 151(4):737-45
- Bi S, Gavrilova O, Gong DW, Mason MM, Reitman M (1997) **Identification of a placental enhancer for the human leptin gene.** *J Biol Chem* 272(48):30583-8.
- Bieda K, Hoffmann A, Boller K (2001) **Phenotypic heterogeneity of human endogenous retrovirus particles produced by teratocarcinoma cell lines.** *J Gen Virol* 82(Pt 3):591-6
- Biermann, Jana (2013) **Analyse putativer Promotorbereiche und Translation Retrovirus-ähnlicher Leserahmen von zwei transkribierten Loci der HERV-K(HML-2) Gruppe.** Bachelorarbeit am Institut für Humangenetik, Universitätsklinikum Homburg
- Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J (2010) **Galaxy: a web-based genome analysis tool for experimentalists.** *Curr Protoc Mol Biol* Chapter 19:Unit 19.10.1-21
- Blomberg J, Ushameckis D, Jern P (2005) **Evolutionary aspects of human endogenous retroviral sequences (HERVs) and disease.** In: Sverdlov E (ed) *Retroviruses and Primate Genome Evolution.* Landes Bioscience, Austin, Texas, pp. 204–238
- Blomberg J, Benachenhou F, Blikstad V, Sperber G, Mayer J (2009) **Classification and nomenclature of endogenous retroviral sequences (ERVs): problems and recommendations.** *Gene* 448(2):115-23
- Blond JL, Besème F, Duret L, Bouton O, Bedin F, Perron H, Mandrand B, Mallet F (1999) **Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family.** *J Virol* 73(2):1175-85
- Blond JL, Lavillette D, Cheynet V, Bouton O, Oriol G, Chapel-Fernandes S, Mandrand B, Mallet F, Cosset FL (2000) **An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor.** *J Virol* 74(7):3321-9
- Boeke JD, Stoye JP (1997) **Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements.** In: Coffin JM, Hughes SH, Varmus HE (eds) *Retroviruses.* Cold Spring Harbor Laboratory Press, Plainview, New York, pp343-435
- Boese A, Sauter M, Mueller-Lantzsch N (2000) **A rev-like NES mediates cytoplasmic localization of HERV-K cORF.** *FEBS Lett* 468(1):65-7
- Boissinot S, Chevret P, Furano AV (2000) **L1 (LINE-1) retrotransposon evolution and amplification in recent human history.** *Mol Biol Evol* 17(6):915-28
- Boller K, Frank H, Löwer J, Löwer R, Kurth R (1983) **Structural organization of unique retrovirus-like particles budding from human teratocarcinoma cell lines.** *J Gen Virol* 64 ( Pt 12):2549-59
- Boller K, König H, Sauter M, Mueller-Lantzsch N, Löwer R, Löwer J, Kurth R (1993) **Evidence that HERV-K is the endogenous retrovirus sequence that codes for the human teratocarcinoma-derived retrovirus HTDV.** *Virology* 196(1):349-53
- Boller K, Janssen O, Schuldes H, Tönjes RR, Kurth R (1997) **Characterization of the antibody response specific for the human endogenous retrovirus HTDV/HERV-K.** *J Virol* 71(6):4581-8
- Boller K, Schönfeld K, Lischer S, Fischer N, Hoffmann A, Kurth R, Tönjes RR (2008) **Human endogenous retrovirus HERV-K113 is capable of producing intact viral particles.** *J Gen Virol* 89(Pt 2):567-72
- Bronson DL, Fraley EE, Fogh J, Kalter SS (1979) **Induction of retrovirus particles in human testicular tumor (Tera-1) cell cultures: an electron microscopic study.** *J Natl Cancer Inst* 63(2):337-9
- Brooks-Wilson AR, Goodfellow PN, Povey S, Nevanlinna HA, de Jong PJ, Goodfellow PJ (1990) **Rapid cloning and characterization of new chromosome 10 DNA markers by Alu element-mediated PCR.** *Genomics* 7(4):614-20
- Büscher K, Trefzer U, Hofmann M, Sterry W, Kurth R, Denner J (2005) **Expression of human endogenous retrovirus K in melanomas and melanoma cell lines.** *Cancer Res* 65(10):4172-80
- Büscher K, Hahn S, Hofmann M, Trefzer U, Ozel M, Sterry W, Löwer J, Löwer R, Kurth R, Denner J (2006) **Expression of the human endogenous retrovirus-K transmembrane envelope, Rec and Np9 proteins in melanomas and melanoma cell lines.** *Melanoma Res* 16(3):223-34
- Burmeister T, Ebert AD, Pritze W, Loddenkemper C, Schwartz S, Thiel E (2004) **Insertional polymorphisms of endogenous HERV-K113 and HERV-K115 retroviruses in breast cancer patients and age-matched controls.** *AIDS Res Hum Retroviruses* 20(11):1223-9
- Buzdin A, Kovalskaya-Alexandrova E, Gogvadze E, Sverdlov E (2006) **GREM, a technique for genome-wide isolation and quantitative analysis of promoter active repeats.** *Nucleic Acids Res* 34(9):e67
- Callahan R, Drohan W, Tronick S, Schlom J (1982) **Detection and cloning of human DNA sequences related to the mouse mammary tumor virus genome.** *Proc Natl Acad Sci U S A* 79(18):5503-7

- Callahan R, Smith GH (2000) **MMTV-induced mammary tumorigenesis: gene discovery, progression to malignancy and cellular pathways.** *Oncogene* 19(8):992-1001
- Cheung VG, Conlin LK, Weber TM, Arcaro M, Jen KY, Morley M, Spielman RS (2003) **Natural variation in human gene expression assessed in lymphoblastoid cells.** *Nat Genet* 33(3):422-5
- Cheung VG, Spielman RS (2009) **Genetics of human gene expression: mapping DNA variants that influence gene expression.** *Nat Rev Genet* 10(9):595-604
- Cheung VG, Nayak RR, Wang IX, Elwyn S, Cousins SM, Morley M, Spielman RS (2010) **Polymorphic cis- and trans-regulation of human gene expression.** *PLoS Biol* 8(9) pii: e1000480
- Christensen T (2005) **Association of human endogenous retroviruses with multiple sclerosis and possible interactions with herpes viruses.** *Rev Med Virol* 15(3):179-211
- Cianciolo GJ, Copeland TD, Oroszlan S, Snyderman R (1985) **Inhibition of lymphocyte proliferation by a synthetic peptide homologous to retroviral envelope proteins.** *Science* 230(4724):453-5
- Clark MB, Amaral PP, Schlesinger FJ, Dinger ME, Taft RJ, Rinn JL, Ponting CP, Stadler PF, Morris KV, Morillon A, Rozowsky JS, Gerstein MB, Wahlestedt C, Hayashizaki Y, Carninci P, Gingeras TR, Mattick JS (2011) **The reality of pervasive transcription.** *PLoS Biol* 9(7): e1000625
- Cocquet J, Chong A, Zhang G, Veitia RA (2006) **Reverse transcriptase template switching and false alternative transcripts.** *Genomics* 88(1):127-31
- Cohen CJ, Lock WM, Mager DL (2009) **Endogenous retroviral LTRs as promoters for human genes: a critical assessment.** *Gene* 448(2):105-14
- Contreras-Galindo R, Kaplan MH, He S, Contreras-Galindo AC, Gonzalez-Hernandez MJ, Kappes F, Dube D, Chan SM, Robinson D, Meng F, Dai M, Gitlin SD, Chinnaiyan AM, Omenn GS, Markovitz DM (2013) **HIV infection reveals widespread expansion of novel centromeric human endogenous retroviruses.** *Genome Res* 23(9):1505-13
- Costas J, Naveira H (2000) **Evolutionary history of the human endogenous retrovirus family ERV9.** *Mol Biol Evol* 17(2):320-30
- Costas J (2002) **Characterization of the intragenomic spread of the human endogenous retrovirus family HERV-W.** *Mol Biol Evol* 19(4):526-33
- Currer R, Van Duyne R, Jaworski E, Guendel I, Sampey G, Das R, Narayanan A, Kashanchi F (2012) **HTLV tax: a fascinating multifunctional co-regulator of viral and cellular pathways.** *Front Microbiol* 3:406
- Dangel AW, Baker BJ, Mendoza AR, Yu CY (1995) **Complement component C4 gene intron 9 as a phylogenetic marker for primates: long terminal repeats of the endogenous retrovirus ERV-K(C4) are a molecular clock of evolution.** *Immunogenetics* 42(1):41-52
- De Jager PL, Jia X, Wang J, de Bakker PI, Ottoboni L, Aggarwal NT, Piccio L, Raychaudhuri S, Tran D, Aubin C, Briskin R, Romano S; International MS Genetics Consortium, Baranzini SE, McCauley JL, Pericak-Vance MA, Haines JL, Gibson RA, Naeglin Y, Uitdehaag B, Matthews PM, Kappos L, Polman C, McArdle WL, Strachan DP, Evans D, Cross AH, Daly MJ, Compston A, Sawcer SJ, Weiner HL, Hauser SL, Hafler DA, Oksenberg JR (2009) **Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci.** *Nat Genet* 41(7):776-82
- de Parseval N, Lazar V, Casella JF, Benit L, Heidmann T (2003) **Survey of human genes of retroviral origin: identification and transcriptome of the genes with coding capacity for complete envelope proteins.** *J Virol* 77(19):10414-22
- de Parseval N, Heidmann T (2005) **Human endogenous retroviruses: from infectious elements to human genes.** *Cytogenet Genome Res* 110(1-4):318-32
- Denne M, Sauter M, Armbruster V, Licht JD, Roemer K, Mueller-Lantzsch N (2007) **Physical and functional interactions of human endogenous retrovirus proteins Np9 and rec with the promyelocytic leukemia zinc finger protein.** *J Virol* 81(11):5607-16
- Dennis LK (1999) **Analysis of the melanoma epidemic, both apparent and real: data from the 1973 through 1994 surveillance, epidemiology, and end results program registry.** *Arch Dermatol* 135(3):275-80
- Dewannieux M, Blaise S, Heidmann T (2005) **Identification of a functional envelope protein from the HERV-K family of human endogenous retroviruses.** *J Virol* 79(24):15573-7
- Dolei A, Perron H (2009) **The multiple sclerosis-associated retrovirus and its HERV-W endogenous family: a biological interface between virology, genetics, and immunology in human physiology and disease.** *J Neurovirol* 15(1):4-13
- Domansky AN, Kopantzev EP, Snezhkov EV, Lebedev YB, Leib-Mosch C, Sverdlov ED (2000) **Solitary HERV-K LTRs possess bi-directional promoter activity and contain a negative regulatory element in the U5 region.** *FEBS Lett* 472(2-3):191-5

- Droege M, Hill B (2008) **The Genome Sequencer FLX System--longer reads, more applications, straight forward bioinformatics and more complete data sets.** *J Biotechnol* 136(1-2):3-10
- Dunn CA, Medstrand P, Mager DL (2003) **An endogenous retroviral long terminal repeat is the dominant promoter for human beta1,3-galactosyltransferase 5 in the colon.** *Proc Natl Acad Sci U S A* 100(22):12841-6
- Dunn CA, Romanish MT, Gutierrez LE, van de Lagemaat LN, Mager DL (2006) **Transcription of two human genes from a bidirectional endogenous retrovirus promoter.** *Gene* 366(2):335-42
- ENCODE Project Consortium (2004) **The ENCODE (ENCyclopedia Of DNA Elements) Project.** *Science* 306(5696):636-40
- ENCODE Project Consortium (2007) **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.** *Nature* 447(7146):799-816
- Ehlhardt S, Seifert M, Schneider J, Ojak A, Zang KD, Mehraein Y (2006) **Human endogenous retrovirus HERV-K(HML-2) Rec expression and transcriptional activities in normal and rheumatoid arthritis synovia.** *J Rheumatol* 33(1):16-23
- Esnault C, Maestre J, Heidmann T (2000) **Human LINE retrotransposons generate processed pseudogenes.** *Nat Genet* 24(4):363-7
- Esnault C, Heidmann O, Delebecque F, Dewannieux M, Ribet D, Hance AJ, Heidmann T, Schwartz O (2005) **APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses.** *Nature* 433(7024):430-3
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) **Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*.** *Nature* 391(6669):806-11
- Flockerzi A, Maydt J, Frank O, Ruggieri A, Maldener E, Seifarth W, Medstrand P, Lengauer T, Meyerhans A, Leib-Mösch C, Meese E, Mayer J (2007) **Expression pattern analysis of transcribed HERV sequences is complicated by ex vivo recombination.** *Retrovirology* 4:39
- Flockerzi A, Ruggieri A, Frank O, Sauter M, Maldener E, Kopper B, Wullich B, Seifarth W, Müller-Lantzsch N, Leib-Mösch C, Meese E, Mayer J (2008) **Expression patterns of transcribed human endogenous retrovirus HERV-K(HML-2) loci in human tissues and the need for a HERV Transcriptome Project.** *BMC Genomics* 9:354
- Flori AR, Löwer R, Schmitz-Dräger BJ, Schulz WA (1999) **DNA methylation and expression of LINE-1 and HERV-K provirus sequences in urothelial and renal cell carcinomas.** *Br J Cancer* 80(9):1312-21
- Frank O, Giehl M, Zheng C, Hehlmann R, Leib-Mösch C, Seifarth W (2005) **Human endogenous retrovirus expression profiles in samples from brains of patients with schizophrenia and bipolar disorders.** *J Virol* 79(17):10890-901
- Frank O, Verbeke C, Schwarz N, Mayer J, Fabarius A, Hehlmann R, Leib-Mösch C, Seifarth W (2008) **Variable transcriptional activity of endogenous retroviruses in human breast cancer.** *J Virol* 82(4):1808-18
- Frendo JL, Olivier D, Cheynet V, Blond JL, Bouton O, Vidaud M, Rabreau M, Evain-Brion D, Mallet F (2003) **Direct involvement of HERV-W Env glycoprotein in human trophoblast cell fusion and differentiation.** *Mol Cell Biol* 23(10):3566-74
- Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL (1992) **A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands.** *Proc Natl Acad Sci U S A* 89(5):1827-31
- Fuchs NV, Kraft M, Tondera C, Hanschmann KM, Löwer J, Löwer R (2011) **Expression of the human endogenous retrovirus (HERV) group HML-2/HERV-K does not depend on canonical promoter elements but is regulated by transcription factors Sp1 and Sp3.** *J Virol* 85(7):3436-48
- Galli UM, Sauter M, Lecher B, Maurer S, Herbst H, Roemer K, Mueller-Lantzsch N (2005) **Human endogenous retrovirus rec interferes with germ cell development in mice and may cause carcinoma in situ, the predecessor lesion of germ cell tumors.** *Oncogene* 24(19):3223-8
- Garcia-Montojo M, Dominguez-Mozo M, Arias-Leal A, Garcia-Martinez Á, De las Heras V, Casanova I, Faucard R, Gehin N, Madeira A, Arroyo R, Curtin F, Alvarez-Lafuente R, Perron H (2013) **The DNA copy number of human endogenous retrovirus-W (MSRV-type) is increased in multiple sclerosis patients and is influenced by gender and disease severity.** *PLoS One* 8(1):e53623
- Garson JA, Huggett JF, Bustin SA, Pfaffl MW, Benes V, Vandesompele J, Shipley GL (2009) **Unreliable real-time PCR analysis of human endogenous retrovirus-W (HERV-W) RNA expression and DNA copy number in multiple sclerosis.** *AIDS Res Hum Retroviruses* 25(3):377-8; author reply 379-81
- Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, Emanuelsson O, Zhang ZD, Weissman S, Snyder M (2007) **What is a gene, post-ENCODE? History and updated definition.** *Genome Res* 17(6):669-81
- Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A (2005) **Galaxy: a platform for interactive large-scale genome analysis.** *Genome Res* 15(10):1451-5
- Gifford R, Tristem M (2003) **The evolution, distribution and diversity of endogenous retroviruses.** *Virus Genes* 26(3):291-315

- Jimenez J, Montgiraud C, Oriol G, Pichon JP, Ruel K, Tsatsaris V, Gerbaud P, Frendo JL, Evain-Brion D, Mallet F (2009) **Comparative methylation of ERVWE1/syncytin-1 and other human endogenous retrovirus LTRs in placenta tissues.** DNA Res 16(4):195-211
- Jimenez J, Montgiraud C, Pichon JP, Bonnaud B, Arsac M, Ruel K, Bouton O, Mallet F (2010) **Custom human endogenous retroviruses dedicated microarray identifies self-induced HERV-W family elements reactivated in testicular cancer upon methylation control.** Nucleic Acids Res 38(7):2229-46
- Goecks J, Nekrutenko A, Taylor J; Galaxy Team (2010) **Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.** Genome Biol 11(8):R86
- Goering W, Ribarska T, Schulz WA (2011) **Selective changes of retroelement expression in human prostate cancer.** Carcinogenesis 32(10):1484-92
- Göttinger N, Sauter M, Roemer K, Mueller-Lantzsch N (1996) **Regulation of human endogenous retrovirus-K Gag expression in teratocarcinoma cell lines and human tumours.** J Gen Virol 77 ( Pt 12):2983-90
- Goff SP (2004) **Retrovirus restriction factors.** Mol Cell 16(6):849-59
- Gogvadze E, Stukacheva E, Buzdin A, Sverdlov E (2009) **Human-specific modulation of transcriptional activity provided by endogenous retroviral insertions.** J Virol 83(12):6098-105
- Goodchild NL, Wilkinson DA, Mager DL (1993) **Recent evolutionary expansion of a subfamily of RTVL-H human endogenous retrovirus-like elements.** Virology 196(2):778-88
- Goodchild NL, Freeman JD, Mager DL (1995) **Spliced HERV-H endogenous retroviral sequences in human genomic DNA: evidence for amplification via retrotransposition.** Virology 206(1):164-73
- Gordon R (2013) **Skin cancer: an overview of epidemiology and risk factors.** Semin Oncol Nurs 29(3):160-9
- Gosenca D, Gabriel U, Steidler A, Mayer J, Diem O, Erben P, Fabarius A, Leib-Mösch C, Hofmann WK, Seifarth W (2012) **HERV-E-mediated modulation of PLA2G4A transcription in urothelial carcinoma.** PLoS One 7(11):e49341
- Hahn S, Ugurel S, Hanschmann KM, Strobel H, Tondera C, Schadendorf D, Löwer J, Löwer R (2008) **Serological response to human endogenous retrovirus K in melanoma patients correlates with survival probability.** AIDS Res Hum Retroviruses 24(5):717-23
- Han JS, Boeke JD (2005) **LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression?** Bioessays 27(8):775-84
- Hangauer MJ, Vaughn IW, McManus MT (2013) **Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs.** PLoS Genet 9(6):e1003569
- Haraguchi S, Good RA, Cianciolo GJ, Engelman RW, Day NK (1997) **Immunosuppressive retroviral peptides: immunopathological implications for immunosuppressive influences of retroviral infections.** J Leukoc Biol 61(6):654-66
- Hartl DL (1996) **The most unkindest cut of all.** Nat Genet 12(3):227-9
- Hashimshony T, Wagner F, Sher N, Yanai I (2012) **CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification.** Cell Rep 2(3):666-73
- Haupt S, Tisdale M, Vincendeau M, Clements MA, Gauthier DT, Lance R, Semmes OJ, Turqueti-Neves A, Noessner E, Leib-Mösch C, Greenwood AD (2011) **Human endogenous retrovirus transcription profiles of the kidney and kidney-derived cell lines.** J Gen Virol 92(Pt 10):2356-66
- Herbst H, Sauter M, Mueller-Lantzsch N (1996) **Expression of human endogenous retrovirus K elements in germ cell and trophoblastic tumors.** Am J Pathol 149(5):1727-35
- Hill VK, Gartner JJ, Samuels Y, Goldstein AM (2013) **The genetics of melanoma: recent advances.** Annu Rev Genomics Hum Genet 14:257-79
- Horwich A, Shipley J, Huddart R (2006) **Testicular germ-cell cancer.** Lancet 367(9512):754-65
- Hu L, Hornung D, Kurek R, Ostman H, Blomberg J, Bergqvist A (2006) **Expression of human endogenous gammaretroviral sequences in endometriosis and ovarian cancer.** AIDS Res Hum Retroviruses 22(6):551-7
- Huang J, Gong Z, Ghosal G, Chen J (2009) **SOSS complexes participate in the maintenance of genomic stability.** Mol Cell 35(3):384-93
- Hughes DC (2001) **Alternative splicing of the human VEGFR-3/FLT4 gene as a consequence of an integrated human endogenous retrovirus.** J Mol Evol 53(2):77-9

- Hughes JF, Coffin JM (2004) **Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: implications for human and viral evolution.** Proc Natl Acad Sci U S A 101(6):1668-72
- International Human Genome Sequencing Consortium (2001) **Initial sequencing and analysis of the human genome.** Nature 409(6822):860-921
- Ishida T, Obata Y, Ohara N, Matsushita H, Sato S, Uenaka A, Saika T, Miyamura T, Chayama K, Nakamura Y, Wada H, Yamashita T, Morishima T, Old LJ, Nakayama E (2008) **Identification of the HERV-K gag antigen in prostate cancer by SEREX using autologous patient serum and its immunogenicity.** Cancer Immun 8:15
- Jensen S, Gassama MP, Heidmann T (1994). **Retrotransposition of the Drosophila LINE I element can induce deletion in the target DNA: a simple model also accounting for the variability of the normally observed target site duplications.** Biochem Biophys Res Commun 202(1):111-9
- Jern P, Sperber GO, Blomberg J (2005) **Use of endogenous retroviral sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy.** Retrovirology 2:50
- Jern P, Coffin JM (2008) **Effects of retroviruses on host genome function.** Annu Rev Genet 42:709-32
- Judo MS, Wedel AB, Wilson C (1998) **Stimulation and suppression of PCR-mediated recombination.** Nucleic Acids Res 26(7):1819-25
- Jurka J, Klonowski P (1996) **Integration of retroposable elements in mammals: selection of target sites.** J Mol Evol 43(6):685-9
- Jurka J (1997) **Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons.** Proc Natl Acad Sci U S A 94(5):1872-7
- Jurka J (2000) **Rebase update: a database and an electronic journal of repetitive elements.** Trends Genet 16(9):418-20
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) **Rebase Update, a database of eukaryotic repetitive elements.** Cytogenet Genome Res 110(1-4):462-7
- Jurka J, Kapitonov VV, Kohany O, Jurka MV (2007) **Repetitive sequences in complex genomes: structure and evolution.** Annu Rev Genomics Hum Genet 8:241-59
- Kakalacheva K, Münz C, Lünemann JD (2011) **Viral triggers of multiple sclerosis.** Biochim Biophys Acta 1812(2):132-40
- Kapitonov VV, Jurka J (1999) **The long terminal repeat of an endogenous retrovirus induces alternative splicing and encodes an additional carboxy-terminal sequence in the human leptin receptor.** J Mol Evol 48(2):248-51
- Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ (2004) **The UCSC Table Browser data retrieval tool.** Nucleic Acids Res 32(Database issue):D493-6
- Katoh K, Misawa K, Kuma K, Miyata T (2002) **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.** Nucleic Acids Res 30(14):3059-66
- Katoh K, Toh H (2008) **Recent developments in the MAFFT multiple sequence alignment program.** Brief Bioinform 9(4):286-98
- Katoh I, Mírová A, Kurata S, Murakami Y, Horikawa K, Nakakuki N, Sakai T, Hashimoto K, Maruyama A, Yonaga T, Fukunishi N, Moriishi K, Hirai H (2011) **Activation of the long terminal repeat of human endogenous retrovirus K by melanoma-specific transcription factor MITF-M.** Neoplasia 13(11):1081-92
- Katoh K, Standley DM (2013) **MAFFT multiple sequence alignment software version 7: improvements in performance and usability.** Mol Biol Evol 30(4):772-80
- Kazazian HH Jr (2004) **Mobile elements: drivers of genome evolution.** Science 303(5664):1626-32
- Kent WJ (2002) **BLAT--the BLAST-like alignment tool.** Genome Res 12(4):656-64
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D (2002) **The human genome browser at UCSC.** Genome Res 12(6):996-1006
- Keydar I, Ohno T, Nayak R, Sweet R, Simoni F, Weiss F, Karby S, Mesa-Tejada R, Spiegelman S (1984) **Properties of retrovirus-like particles produced by a human breast carcinoma cell line: immunological relationship with mouse mammary tumor virus proteins.** Proc Natl Acad Sci U S A 81(13):4188-92
- Kim DS, Hahn Y (2010) **Human-specific antisense transcripts induced by the insertion of transposable element.** Int J Mol Med 26(1):151-7
- Kitamura Y, Ayukawa T, Ishikawa T, Kanda T, Yoshiike K (1996) **Human endogenous retrovirus K10 encodes a functional integrase.** J Virol 70(5):3302-6

- Kleiman A, Senyuta N, Tryakin A, Sauter M, Karseladze A, Tjulandin S, Gurtsevitch V, Mueller-Lantzsch N (2004) **HERV-K(HML-2) GAG/ENV antibodies as indicator for therapy effect in patients with germ cell tumors.** *Int J Cancer* 110(3):459-61
- Kovalskaya E, Buzdin A, Gogvadze E, Vinogradova T, Sverdlov E (2006) **Functional human endogenous retroviral LTR transcription start sites are located between the R and U5 regions.** *Virology* 346(2):373-8
- Kremer D, Schichel T, Förster M, Tzekova N, Bernard C, van der Valk P, van Horssen J, Hartung HP, Perron H, Küry P (2013) **Human endogenous retrovirus type W envelope protein inhibits oligodendroglial precursor cell differentiation.** *Ann Neurol* 74(5):721-32
- Kruh GD, Belinsky MG (2003) **The MRP family of drug efflux pumps.** *Oncogene* 22(47):7537-52
- Landry JR, Rouhi A, Medstrand P, Mager DL (2002) **The Opitz syndrome gene Mid1 is transcribed from a human endogenous retroviral promoter.** *Mol Biol Evol* 19(11):1934-42
- Läufer G, Mayer J, Mueller BF, Mueller-Lantzsch N, Ruprecht K (2009) **Analysis of transcribed human endogenous retrovirus W env loci clarifies the origin of multiple sclerosis-associated retrovirus env sequences.** *Retrovirology* 6:37
- Lavie L, Medstrand P, Schempp W, Meese E, Mayer J (2004) **Human endogenous retrovirus family HERV-K(HML-5): status, evolution, and reconstruction of an ancient betaretrovirus in the human genome.** *J Virol* 78(16):8788-98
- Lavie L, Maldener E, Brouha B, Meese EU, Mayer J (2004) **The human L1 promoter: variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity.** *Genome Res* 14(11):2253-60
- Lavie L, Kitova M, Maldener E, Meese E, Mayer J (2005) **CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K(HML-2).** *J Virol* 79(2):876-83
- Lee WJ, Kwun HJ, Jang KL (2003) **Analysis of transcriptional regulatory sequences in the human endogenous retrovirus W long terminal repeat.** *J Gen Virol* 84(Pt 8):2229-35
- Lee YN, Bieniasz PD (2007) **Reconstitution of an infectious human endogenous retrovirus.** *PLoS Pathog* 3(1):e10
- Li S, Liu ZC, Yin SJ, Chen YT, Yu HL, Zeng J, Zhang Q, Zhu F (2013) **Human endogenous retrovirus W family envelope gene activates the small conductance Ca(2+)-activated K+ channel in human neuroblastoma cells through CREB.** *Neuroscience* 247:164-74
- Löwer J, Löwer R, Stegmann J, Frank H, Kurth R (1981) **Retrovirus particle production in three of four human teratocarcinoma cell lines.** *Haematol Blood Transfus* 26:541-4
- Löwer R, Löwer J, Frank H, Harzmann R, Kurth R (1984) **Human teratocarcinomas cultured in vitro produce unique retrovirus-like viruses.** *J Gen Virol* 65 ( Pt 5):887-98
- Löwer R, Boller K, Hasenmaier B, Korbmacher C, Müller-Lantzsch N, Löwer J, Kurth R (1993) **Identification of human endogenous retroviruses with complex mRNA expression and particle formation.** *Proc Natl Acad Sci U S A* 90(10):4480-4
- Löwer R, Tönjes RR, Korbmacher C, Kurth R, Löwer J (1995) **Identification of a Rev-related protein by analysis of spliced transcripts of the human endogenous retroviruses HTDV/HERV-K.** *J Virol* 69(1):141-9
- Luan DD, Korman MH, Jakubczak JL, Eickbush TH (1993) **Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition.** *Cell* 72(4):595-605
- Macfarlane C, Simmonds P (2004) **Allelic variation of HERV-K(HML-2) endogenous retroviral elements in human populations.** *J Mol Evol* 59(5):642-56
- Mader RM, Schmidt WM, Sedivy R, Rizovski B, Braun J, Kalipciyan M, Exner M, Steger GG, Mueller MW (2001) **Reverse transcriptase template switching during reverse transcriptase-polymerase chain reaction: artificial generation of deletions in ribonucleotide reductase mRNA.** *J Lab Clin Med* 137(6):422-8
- Mager DL, Hunter DG, Schertzer M, Freeman JD (1999) **Endogenous retroviruses provide the primary polyadenylation signal for two new human genes (HHLA2 and HHLA3).** *Genomics* 59(3):255-63
- Mager DL, Medstrand P (2003) **Retroviral Repeat Sequences.** In: Cooper D (ed) *Nature Encyclopedia of the Human Genome. Volume 5.* Macmillan Publishers Ltd, Hampshire, England, pp 57-63
- Magin C, Löwer R, Löwer J (1999) **cORF and RcRE, the Rev/Rex and RRE/RxRE homologues of the human endogenous retrovirus family HTDV/HERV-K.** *J Virol* 73(11):9496-507
- Makalowski W, Toda Y (2007) **Modulation of host genes by mammalian transposable elements.** *Genome Dyn* 3:163-74
- Mallet F, Bouton O, Prudhomme S, Cheynet V, Oriol G, Bonnaud B, Lucotte G, Duret L, Mandrand B (2004) **The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology.** *Proc Natl Acad Sci U S A* 101(6):1731-6

- Mameli G, Astone V, Arru G, Marconi S, Lovato L, Serra C, Sotgiu S, Bonetti B, Dolei A (2007) **Brains and peripheral blood mononuclear cells of multiple sclerosis (MS) patients hyperexpress MS-associated retrovirus/HERV-W endogenous retrovirus, but not Human herpesvirus 6.** J Gen Virol 88(Pt 1):264-74
- Mameli G, Poddighe L, Astone V, Delogu G, Arru G, Sotgiu S, Serra C, Dolei A (2009) **Novel reliable real-time PCR for differential detection of MSRVenv and syncytin-1 in RNA and DNA from patients with multiple sclerosis.** J Virol Methods 161(1):98-106
- Manghera M, Douville RN (2013) **Endogenous retrovirus-K promoter: a landing strip for inflammatory transcription factors?** Retrovirology 10:16
- Martin MA, Bryan T, Rasheed S, Khan AS (1981) **Identification and cloning of endogenous retroviral sequences present in human DNA.** Proc Natl Acad Sci U S A 78(8):4892-6
- Marrie RA (2004) **Environmental risk factors in multiple sclerosis aetiology.** Lancet Neurol 3(12):709-18
- Martin SL, Li WL, Furano AV, Boissinot S (2005) **The structures of mouse and human L1 elements reflect their insertion mechanism.** Cytogenet Genome Res 110(1-4):223-8
- Matoušková M, Blazková J, Pajer P, Pavlíček A, Hejnar J (2006) **CpG methylation suppresses transcriptional activity of human syncytin-1 in non-placental tissues.** Exp Cell Res 312(7):1011-20
- Mayer J, Meese E, Mueller-Lantzsch N (1997a) **Multiple human endogenous retrovirus (HERV-K) loci with gag open reading frames in the human genome.** Cytogenet Cell Genet 78(1):1-5
- Mayer J, Meese E, Mueller-Lantzsch N (1997b) **Chromosomal assignment of human endogenous retrovirus K (HERV-K) env open reading frames.** Cytogenet Cell Genet 79(1-2):157-61
- Mayer J, Meese E, Mueller-Lantzsch N (1998) **Human endogenous retrovirus K homologous sequences and their coding capacity in Old World primates.** J Virol 72(3):1870-5
- Mayer J, Sauter M, Rácz A, Scherer D, Mueller-Lantzsch N, Meese E (1999) **An almost-intact human endogenous retrovirus K on human chromosome 7.** Nat Genet 21(3):257-8
- Mayer J, Meese EU (2002) **The human endogenous retrovirus family HERV-K(HML-3).** Genomics 80(3):331-43
- Mayer J, Meese EU (2003) **Presence of dUTPase in the various human endogenous retrovirus K (HERV-K) families.** J Mol Evol 57(6):642-9
- Mayer J, Ehlhardt S, Seifert M, Sauter M, Müller-Lantzsch N, Mehraein Y, Zang KD, Meese E (2004) **Human endogenous retrovirus HERV-K(HML-2) proviruses with Rec protein coding capacity and transcriptional activity.** Virology 322(1):190-8
- Mayer J, Meese E (2005) **Human endogenous retroviruses in the primate lineage and their influence on host genomes.** Cytogenet Genome Res 110(1-4):448-56
- Mayer J, Stuhr T, Reus K, Maldener E, Kitova M, Asmus F, Meese E (2005) **Haplotype analysis of the human endogenous retrovirus locus HERV-K(HML-2.HOM) and its evolutionary implications.** J Mol Evol 61(5):706-15
- Mayer J, Blomberg J, Seal RL (2011) **A revised nomenclature for transcribed human endogenous retroviral loci.** Mob DNA 2(1):7
- Medstrand P, Blomberg J (1993) **Characterization of novel reverse transcriptase encoding human endogenous retroviral sequences similar to type A and type B retroviruses: differential transcription in normal human tissues.** J Virol 67(11):6778-87
- Medstrand P, Mager DL (1998) **Human-specific integrations of the HERV-K endogenous retrovirus family.** J Virol 72(12):9782-7
- Medstrand P, Landry JR, Mager DL (2001) **Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans.** J Biol Chem 276(3):1896-903
- Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddleloh JA, Mattick JS, Rinn JL (2011) **Targeted RNA sequencing reveals the deep complexity of the human transcriptome.** Nat Biotechnol 30(1):99-104
- Meyerhans A, Vartanian JP, Wain-Hobson S (1990) **DNA recombination during PCR.** Nucleic Acids Res 18(7):1687-91
- Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S, Keith JC Jr, McCoy JM (2000) **Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis.** Nature 403(6771):785-9
- Molès JP, Tesnière A, Guilhou JJ (2005) **A new endogenous retroviral sequence is expressed in skin of patients with psoriasis.** Br J Dermatol 153(1):83-9

- Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH Jr (1996) **High frequency retrotransposition in cultured mammalian cells.** *Cell* 87(5):917-27
- Moran, J.V. & Gilbert, N (2002). **Mammalian LINE-1 retrotransposons and related elements.** In: Craig N, Craggie R, Gellert M, Lambowitz (eds) *Mobile DNA II.* ASM, Washington DC, pp 836–869
- Moyes DL, Martin A, Sawcer S, Temperton N, Worthington J, Griffiths DJ, Venables PJ (2005) **The distribution of the endogenous retroviruses HERV-K113 and HERV-K115 in health and disease.** *Genomics* 86(3):337-41
- Moyes D, Griffiths DJ, Venables PJ (2007) **Insertional polymorphisms: a new lease of life for endogenous retroviruses in human disease.** *Trends Genet* 23(7):326-33
- Mueller-Lantzsch N, Sauter M, Weiskircher A, Kramer K, Best B, Buck M, Grässer F (1993) **Human endogenous retroviral element K10 (HERV-K10) encodes a full-length gag homologous 73-kDa protein and a functional protease.** *AIDS Res Hum Retroviruses* 9(4):343-50
- Mullis KB (1990) **Target amplification for DNA analysis by the polymerase chain reaction.** *Ann Biol Clin (Paris)* 48(8):579-82
- Muradrasoli S, Forsman A, Hu L, Blikstad V, Blomberg J (2006) **Development of real-time PCRs for detection and quantitation of human MMTV-like (HML) sequences HML expression in human tissues.** *J Virol Methods* 136(1-2):83-92
- Muster T, Waltenberger A, Grassauer A, Hirschl S, Caucig P, Romirer I, Födinger D, Seppel H, Schanab O, Magin-Lachmann C, Löwer R, Jansen B, Pehamberger H, Wolff K (2003) **An endogenous retrovirus derived from human melanoma cells.** *Cancer Res* 63(24):8735-41
- Nakagawa K, Brusica V, McColl G, Harrison LC (1997) **Direct evidence for the expression of multiple endogenous retroviruses in the synovial compartment in rheumatoid arthritis.** *Arthritis Rheum* 40(4):627-38
- Nellåker C, Li F, Uhrzander F, Tyrcha J, Karlsson H (2009) **Expression profiling of repetitive elements by melting temperature analysis: variation in HERV-W gag expression across human individuals and tissues.** *BMC Genomics* 10:532
- Ogasawara H, Naito T, Kaneko H, Hishikawa T, Sekigawa I, Hashimoto H, Kaneko Y, Yamamoto N, Maruyama N, Yamamoto N (2001) **Quantitative analyses of messenger RNA of human endogenous retrovirus in patients with systemic lupus erythematosus.** *J Rheumatol* 28(3):533-8
- Oja M, Peltonen J, Blomberg J, Kaski S (2007) **Methods for estimating human endogenous retrovirus activities from EST databases.** *BMC Bioinformatics* 8 Suppl 2:S11
- Okada N, Hamada M, Ogiwara I, Ohshima K (1997) **SINes and LINEs share common 3' sequences: a review.** *Gene* 205(1-2):229-43
- Ono M, Yasunaga T, Miyata T, Ushikubo H (1986) **Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome.** *J Virol* 60(2):589-98
- Ono M, Kawakami M, Takezawa T (1987) **A novel human nonviral retroposon derived from an endogenous retrovirus.** *Nucleic Acids Res* 15(21):8725-37
- Ostertag EM, Kazazian HH Jr (2001) **Biology of mammalian L1 retrotransposons.** *Annu Rev Genet* 35:501-38
- Ostertag EM, Goodier JL, Zhang Y, Kazazian HH Jr (2003) **SVA elements are nonautonomous retrotransposons that cause disease in humans.** *Am J Hum Genet* 73(6):1444-51
- Paty DW (1997) **MRI as a method to reveal in-vivo pathology in MS.** *J Neural Transm Suppl* 49:211-7
- Pavlicek A, Paces J, Elleder D, Hejnar J (2002) **Processed pseudogenes of human endogenous retroviruses generated by LINEs: their integration, stability, and distribution.** *Genome Res* 12(3):391-9
- Pece S, Serresi M, Santolini E, Capra M, Hulleman E, Galimberti V, Zurrida S, Maisonneuve P, Viale G, Di Fiore PP (2004) **Loss of negative regulation by Numb over Notch is relevant to human breast carcinogenesis.** *J Cell Biol* 167(2):215-21
- Pei B, Sisu C, Frankish A, Howald C, Habegger L, Mu XJ, Harte R, Balasubramanian S, Tanzer A, Diekhans M, Raymond A, Hubbard TJ, Harrow J, Gerstein MB (2012) **The GENCODE pseudogene resource.** *Genome Biol* 13(9):R51
- Pérot P, Mugnier N, Montgiraud C, Gimenez J, Jaillard M, Bonnaud B, Mallet F (2012) **Microarray-based sketches of the HERV transcriptome landscape.** *PLoS One* 7(6):e40194
- Perron H, Geny C, Genoulaz O, Pellat J, Perret J, Seigneurin JM (1991) **Antibody to reverse transcriptase of human retroviruses in multiple sclerosis.** *Acta Neurol Scand* 84(6):507-13
- Piotrowski PC, Duriagin S, Jagodzinski PP (2005) **Expression of human endogenous retrovirus clone 4-1 may correlate with blood plasma concentration of anti-U1 RNP and anti-Sm nuclear antibodies.** *Clin Rheumatol* 24(6):620-4

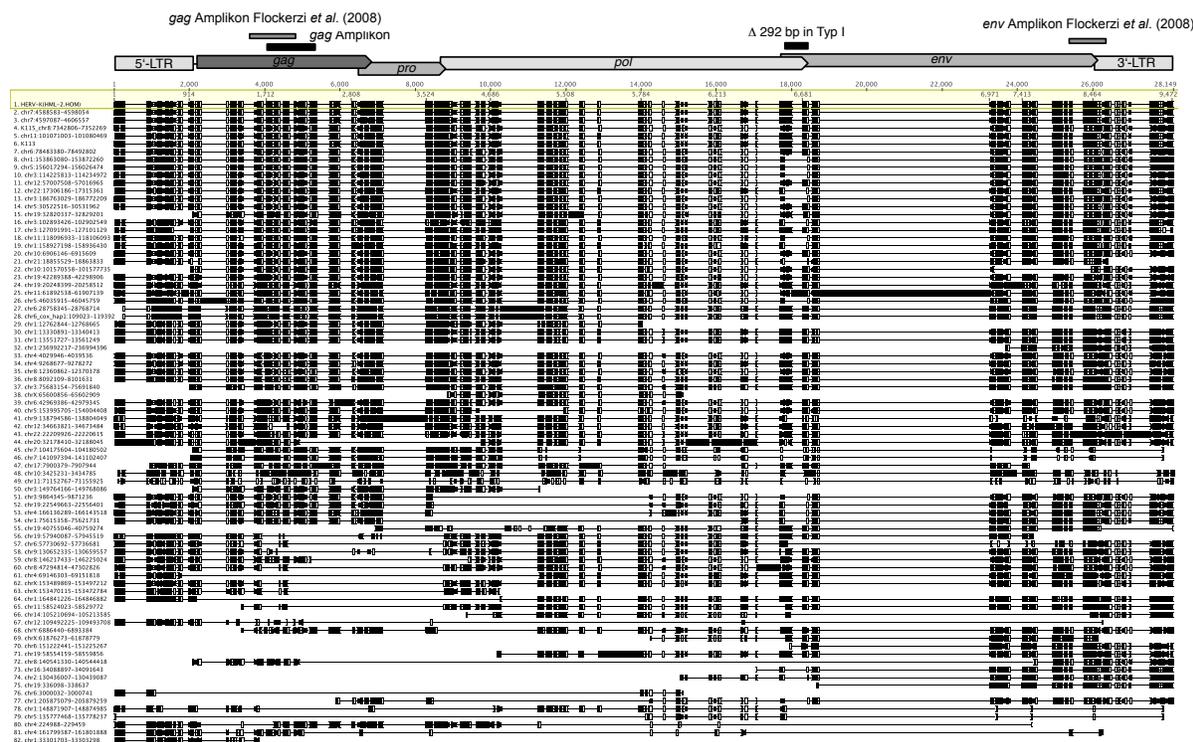
- Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC (1980) **Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma.** Proc Natl Acad Sci U S A 77(12):7415-9
- Prudhomme S, Oriol G, Mallet F (2004) **A retroviral promoter and a cellular enhancer define a bipartite element which controls env ERVWE1 placental expression.** J Virol 78(22):12157-68
- Radonić A, Thulke S, Mackay IM, Landt O, Siegert W, Nitsche A (2004) **Guideline to reference gene selection for quantitative real-time PCR.** Biochem Biophys Res Commun 313(4):856-62
- Reece-Hoyes JS, Marian Walhout AJ (2012) **Yeast one-hybrid assays: a historical and technical perspective.** Methods 57(4):441-7
- Reiche J, Pauli G, Ellerbrok H (2010) **Differential expression of human endogenous retrovirus K transcripts in primary human melanocytes and melanoma cell lines after UV irradiation.** Melanoma Res 20(5):435-40
- Reis BS, Jungbluth AA, Frosina D, Holz M, Ritter E, Nakayama E, Ishida T, Obata Y, Carver B, Scher H, Scardino PT, Slovin S, Subudhi SK, Reuter VE, Savage C, Allison JP, Melamed J, Jäger E, Ritter G, Old LJ, Gnjatic S (2013) **Prostate cancer progression correlates with increased humoral immune response to a human endogenous retrovirus GAG protein.** Clin Cancer Res 19(22):6112-25
- Reiss D, Mager DL (2007) **Stochastic epigenetic silencing of retrotransposons: does stability come with age?** Gene 390(1-2):130-5
- Reus K, Mayer J, Sauter M, Scherer D, Müller-Lantzsch N, Meese E (2001a) **Genomic organization of the human endogenous retrovirus HERV-K(HML-2.HOM) (ERV6) on chromosome 7.** Genomics 72(3):314-20
- Reus K, Mayer J, Sauter M, Zischler H, Müller-Lantzsch N, Meese E (2001b) **HERV-K(OLD): ancestor sequences of the human endogenous retrovirus family HERV-K(HML-2).** J Virol 75(19):8917-26
- Repaske R, O'Neill RR, Steele PE, Martin MA (1983) **Characterization and partial nucleotide sequence of endogenous type C retrovirus segments in human chromosomal DNA.** Proc Natl Acad Sci U S A 80(3):678-82
- Richter, Christin (2012). **Charakterisierung der Long Terminal Repeats transkribierter Elemente der humanen endogenen Retrovirus Gruppe HERV-W.** Bachelorarbeit am Institut für Humangenetik, Universitätsklinikum Homburg
- Roebke C, Wahl S, Laufer G, Stadelmann C, Sauter M, Mueller-Lantzsch N, Mayer J, Ruprecht K (2010) **An N-terminally truncated envelope protein encoded by a human endogenous retrovirus W locus on chromosome Xq22.3.** Retrovirology 7:69
- Rolland A, Jouvin-Marche E, Viret C, Faure M, Perron H, Marche PN (2006) **The envelope protein of a human endogenous retrovirus-W family activates innate immunity through CD14/TLR4 and promotes Th1-like responses.** J Immunol 176(12):7636-44
- Rozen S, Skaletsky H (2000) **Primer3 on the WWW for general users and for biologist programmers.** Methods Mol Biol. 2000;132:365-86.
- Ruprecht K, Ferreira H, Flockerzi A, Wahl S, Sauter M, Mayer J, Mueller-Lantzsch N (2008a) **Human endogenous retrovirus family HERV-K(HML-2) RNA transcripts are selectively packaged into retroviral particles produced by the human germ cell tumor line Tera-1 and originate mainly from a provirus on chromosome 22q11.21.** J Virol 82(20):10008-16
- Ruprecht K, Mayer J, Sauter M, Roemer K, Mueller-Lantzsch N (2008b) **Endogenous retroviruses and cancer.** Cell Mol Life Sci 65(21):3366-82
- Reynier F, Verjat T, Turrel F, Imbert PE, Marotte H, Mouglin B, Miossec P (2009) **Increase in human endogenous retrovirus HERV-K (HML-2) viral load in active rheumatoid arthritis.** Scand J Immunol 70(3):295-9
- Sanger F, Nicklen S, Coulson AR (1977) **DNA sequencing with chain-terminating inhibitors.** Proc Natl Acad Sci U S A 74(12):5463-7
- Sauter M, Schommer S, Kremmer E, Remberger K, Dölken G, Lemm I, Buck M, Best B, Neumann-Haefelin D, Mueller-Lantzsch N (1995) **Human endogenous retrovirus K10: expression of Gag protein and detection of antibodies in patients with seminomas.** J Virol 69(1):414-21
- Sauter M, Roemer K, Best B, Afting M, Schommer S, Seitz G, Hartmann M, Mueller-Lantzsch N (1996) **Specificity of antibodies directed against Env protein of human endogenous retroviruses in patients with germ cell tumors.** Cancer Res 56(19):4362-5
- Schanab O, Humer J, Gleiss A, Mikula M, Sturlan S, Grunt S, Okamoto I, Muster T, Pehamberger H, Waltzenberger A (2011) **Expression of human endogenous retrovirus K is stimulated by ultraviolet radiation in melanoma.** Pigment Cell Melanoma Res 24(4):656-65

- Seifarth W, Frank O, Zeilfelder U, Spiess B, Greenwood AD, Hehlmann R, Leib-Mösch C (2005) **Comprehensive analysis of human endogenous retrovirus transcriptional activity in human tissues with a retrovirus-specific microarray.** *J Virol* 79(1):341-52
- Serafino A, Balestrieri E, Pierimarchi P, Matteucci C, Moroni G, Oricchio E, Rasi G, Mastino A, Spadafora C, Garaci E, Vallebona PS (2009) **The activation of human endogenous retrovirus K (HERV-K) is implicated in melanoma cell malignant transformation.** *Exp Cell Res* 315(5):849-62
- Shen L, Wu LC, Sanlioglu S, Chen R, Mendoza AR, Dangel AW, Carroll MC, Zipf WB, Yu CY (1994) **Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and the C4B genes in the HLA class III region. Molecular cloning, exon-intron structure, composite retroposon, and breakpoint of gene duplication.** *J Biol Chem* 269(11):8466-76
- Shiramizu B, Herndier BG, McGrath MS (1994) **Identification of a common clonal human immunodeficiency virus integration site in human immunodeficiency virus-associated lymphomas.** *Cancer Res* 54(8):2069-72
- Sicat J, Sutkowski N, Huber BT (2005) **Expression of human endogenous retrovirus HERV-K18 superantigen is elevated in juvenile rheumatoid arthritis.** *J Rheumatol* 32(9):1821-31
- Sinkovics JG (1988) **Oncogenes and growth factors.** *Crit Rev Immunol* 8(4):217-98
- Schmitt K, Richter C, Backes C, Meese E, Ruprecht K, Mayer J (2013) **Comprehensive Analysis of Human Endogenous Retrovirus Group HERV-W Locus Transcription in Multiple Sclerosis Brain Lesions by High-Throughput Amplicon Sequencing.** *J Virol* 87(24):13837-52
- Sospedra M, Martin R (2005) **Immunology of multiple sclerosis.** *Annu Rev Immunol* 23:683-747
- Spencer TE, Mura M, Gray CA, Griebel PJ, Palmarini M (2003) **Receptor usage and fetal expression of ovine endogenous betaretroviruses: implications for coevolution of endogenous and exogenous retroviruses.** *J Virol* 77(1):749-53
- Spencer TE, Palmarini M (2012) **Endogenous retroviruses of sheep: a model system for understanding physiological adaptation to an evolving ruminant genome.** *J Reprod Dev* 58(1):33-7
- Sperber GO, Airola T, Jern P, Blomberg J (2007) **Automated recognition of retroviral sequences in genomic data--RetroTector.** *Nucleic Acids Res* 35(15):4964-76
- Stauffer Y, Theiler G, Sperisen P, Lebedev Y, Jongeneel CV (2004) **Digital expression profiles of human endogenous retroviral families in normal and cancerous tissues.** *Cancer Immunol* 4:2
- Stengel S, Fiebig U, Kurth R, Denner J (2010) **Regulation of human endogenous retrovirus-K expression in melanomas by CpG methylation.** *Genes Chromosomes Cancer* 49(5):401-11
- Stocking C, Kozak CA (2008) **Murine endogenous retroviruses.** *Cell Mol Life Sci* 65(21):3383-98
- Stoye JP (2012) **Studies of endogenous retroviruses reveal a continuing evolutionary saga.** *Nat Rev Microbiol* 10(6):395-406.
- Strick R, Ackermann S, Langbein M, Swiatek J, Schubert SW, Hashemolhosseini S, Koscheck T, Fasching PA, Schild RL, Beckmann MW, Strissel PL (2007) **Proliferation and cell-cell fusion of endometrial carcinoma are induced by the human endogenous retroviral Syncytin-1 and regulated by TGF-beta.** *J Mol Med (Berl)* 85(1):23-38
- Subramanian RP, Wildschutte JH, Russo C, Coffin JM (2011) **Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses.** *Retrovirology* 8:90.
- Sugimoto J, Matsuura N, Kinjo Y, Takasu N, Oda T, Jinno Y (2001) **Transcriptionally active HERV-K genes: identification, isolation, and chromosomal mapping.** *Genomics* 72(2):137-44
- Swergold GD (1990) **Identification, characterization, and cell specificity of a human LINE-1 promoter.** *Mol Cell Biol* 10(12):6718-29
- Takahara T, Ohsumi T, Kuromitsu J, Shibata K, Sasaki N, Okazaki Y, Shibata H, Sato S, Yoshiki A, Kusakabe M, Muramatsu M, Ueki M, Okuda K, Hayashizaki Y (1996) **Dysfunction of the Orleans reeler gene arising from exon skipping due to transposition of a full-length copy of an active L1 sequence into the skipped exon.** *Hum Mol Genet* 5(7):989-93
- Tamura K, Dudley J, Nei M, Kumar S (2007) **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 24(8):1596-9
- Tang F, Barbacioru C, Nordman E, Li B, Xu N, Bashkirov VI, Lao K, Surani MA (2010) **RNA-Seq analysis to capture the transcriptome landscape of a single cell.** *Nat Protoc* 5(3):516-35
- Taylor GM, Gao Y, Sanders DA (2001) **Fv-4: identification of the defect in Env and the mechanism of resistance to ecotropic murine leukemia virus.** *J Virol* 75(22):11244-8

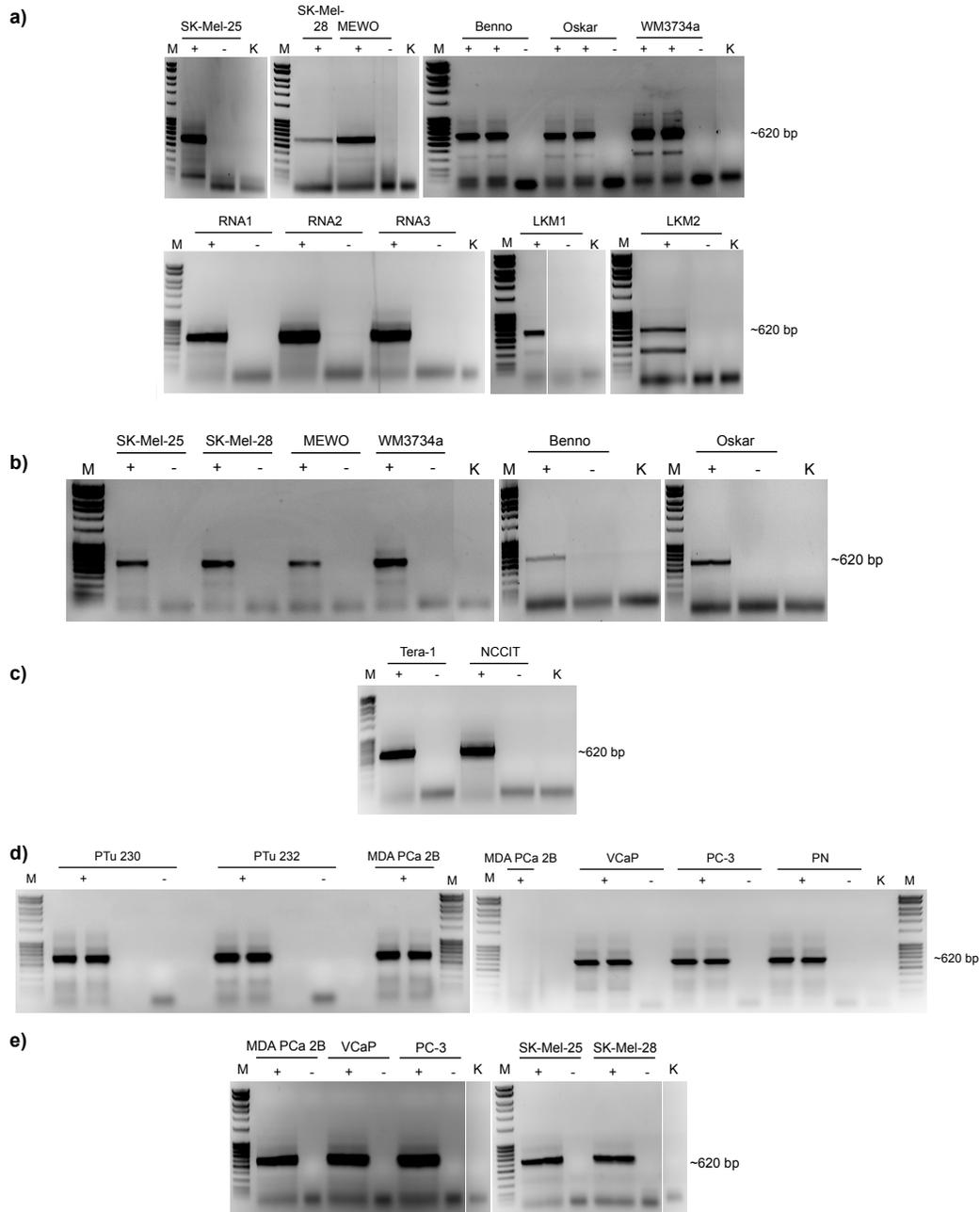
- Tiranti V, Rossi E, Ruiz-Carrillo A, Rossi G, Rocchi M, DiDonato S, Zuffardi O, Zeviani M (1995) **Chromosomal localization of mitochondrial transcription factor A (TCF6), single-stranded DNA-binding protein (SSBP), and endonuclease G (ENDOG), three human housekeeping genes involved in mitochondrial biogenesis.** *Genomics* 25(2):559-64
- Tönjes RR, Limbach C, Löwer R, Kurth R (1997) **Expression of human endogenous retrovirus type K envelope glycoprotein in insect and mammalian cells.** *J Virol* 71(4):2747-56
- Tönjes RR, Czauderna F, Kurth R (1999) **Genome-wide screening, cloning, chromosomal assignment, and expression of full-length human endogenous retrovirus type K.** *J Virol* 73(11):9187-95
- Trejbalová K, Blazková J, Matoušková M, Kucerová D, Pecnová L, Vernerová Z, Heráček J, Hirsch I, Hejnar J (2011) **Epigenetic regulation of transcription and splicing of syncytins, fusogenic glycoproteins of retroviral origin.** *Nucleic Acids Res* 39(20):8728-39
- Trinklein ND, Karaöz U, Wu J, Halees A, Force Aldred S, Collins PJ, Zheng D, Zhang ZD, Gerstein MB, Snyder M, Myers RM, Weng Z (2007) **Integrated analysis of experimental data sets reveals many novel promoters in 1% of the human genome.** *Genome Res* 17(6):720-31
- Turner G, Barbulescu M, Su M, Jensen-Seaman MI, Kidd KK, Lenz J (2001) **Insertional polymorphisms of full-length endogenous retroviruses in humans.** *Curr Biol* 11(19):1531-5
- van Bakel H, Nislow C, Blencowe BJ, Hughes TR (2010) **Most "dark matter" transcripts are associated with known genes.** *PLoS Biol* 8(5):e1000371
- Vasdev N, Moon A, Thorpe AC (2013) **Classification, epidemiology and therapies for testicular germ cell tumours.** *Int J Dev Biol* 57(2-4):133-9
- Vinogradova TV, Leppik LP, Nikolaev LG, Akopov SB, Kleiman AM, Senyuta NB, Sverdlov ED (2001) **Solitary human endogenous retroviruses-K LTRs retain transcriptional activity in vivo, the mode of which is different in different cell types.** *Virology* 290(1):83-90
- Vogt PK (1997) **Historical introduction to the general properties of retroviruses.** In: Coffin JM, Hughes SH, Varmus HE (eds) *Retroviruses*, Cold Spring Harbour Laboratory Press, New York, pp 1-5
- Voisset C, Blancher A, Perron H, Mandrand B, Mallet F, Paranhos-Baccalà G (1999) **Phylogeny of a novel family of human endogenous retrovirus sequences, HERV-W, in humans and other primates.** *AIDS Res Hum Retroviruses* 15(17):1529-33
- Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, Burgess SM, Brachmann RK, Haussler D (2007) **Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53.** *Proc Natl Acad Sci U S A* 104(47):18613-8
- Wang-Johanning F, Frost AR, Jian B, Azerou R, Lu DW, Chen DT, Johanning GL (2003) **Detecting the expression of human endogenous retrovirus E envelope transcripts in human prostate adenocarcinoma.** *Cancer* 98(1):187-97
- Weis S, Llenos IC, Sabuncian S, Dulay JR, Isler L, Yolken R, Perron H (2007) **Reduced expression of human endogenous retrovirus (HERV)-W GAG protein in the cingulate gyrus and hippocampus in schizophrenia, bipolar disorder, and depression.** *J Neural Transm* 114(5):645-55
- Westley B, May FE (1984) **The human genome contains multiple sequences of varying homology to mouse mammary tumour virus DNA.** *Gene* 28(2):221-7
- Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shiptsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, Ballinger DG, Sparks AB, Hartigan J, Smith DR, Suh E, Papadopoulos N, Buckhaults P, Markowitz SD, Parmigiani G, Kinzler KW, Velculescu VE, Vogelstein B (2007) **The genomic landscapes of human breast and colorectal cancers.** *Science* 318(5853):1108-13
- Yin H, Medstrand P, Andersson ML, Borg A, Olsson H, Blomberg J (1997) **Transcription of human endogenous retroviral sequences related to mouse mammary tumor virus in human breast and placenta: similar pattern in most malignant and nonmalignant breast tissues.** *AIDS Res Hum Retroviruses* 13(6):507-16
- Yoder JA, Walsh CP, Bestor TH (1997) **Cytosine methylation and the ecology of intragenomic parasites.** *Trends Genet* 13(8):335-40
- Yolken RH, Karlsson H, Yee F, Johnston-Wilson NL, Torrey EF (2000) **Endogenous retroviruses and schizophrenia.** *Brain Res Brain Res Rev* 31(2-3):193-9

## 7. Anhang

**Anhang 1:** Graphische Darstellung eines multiplen Sequenz-Alignments genomischer Loci der HERV-K(HML-2)-Gruppe. Die Sequenzen von HML-2-Loci wurden, wie im Material und Methoden- und Ergebnisteil beschrieben (vgl. Abschnitt 3.2.3.1. und 4.2.1.), aus dem menschlichen hg18 Referenzgenom zusammengestellt und aliniert. Alinierbare Bereiche der Proviren sind als schwarze Balken dargestellt, horizontale Linien repräsentieren Lücken im Vergleich zur Sequenz anderer Loci. Größere Insertionen in einigen Loci repräsentieren Insertionen von nicht-HML-2 Repeat-Sequenzen. Über dem Alignment ist die Nukleotidposition relativ zum gesamten Alignment sowie darunter die Nukleotidposition relativ zur HERV-K(HML-2.HOM)-Sequenz (*Genbank acc. no.* AF074086.2; gelb unterlegt) angegeben. Die Positionen der LTRs und ORFs der retroviralen Gene *gag*, *pro*, *pol* und *env* sowie die Position einer 292 bp großen Deletion in Typ I Proviren, verglichen mit Proviren des Typs II, sind eingezeichnet. Ebenso die Position des hier untersuchten Amplikons im *gag*-ORF sowie die der in der Studie von Flockerzi *et al.* (2008) untersuchten Amplikons in *gag* und *env*.



**Anhang 2:** Ergebnis der gelelektrophoretischen Auftrennung des aus verschiedenen Tumorproben amplifizierten HERV-K(HML-2) *gag*-Amplikons. Das Amplikon wurde aus cDNA der folgenden Proben amplifiziert (vgl. Abschnitte 4.1.4., 4.1.5., 4.1.7. und 4.1.14. der Ergebnisse): **a)** Den Melanom-Zelllinien SK-Mel-25, SK-Mel-28, MEWO und WM3734a, den Melanozyten-Zelllinien Benno und Oskar, drei Melanom-Proben (RNA1-3) sowie zwei Lymphknoten-Metastasen (LKM) von Melanom-Patienten. **b)** Den Melanom-Zelllinien SK-Mel-25, SK-Mel-28, MEWO und WM3734a sowie den Melanozyten-Zelllinien Benno und Oskar 24 h nach Bestrahlung aller Zelllinien mit 200 mJ/cm<sup>2</sup> UVB. **c)** Den Keimzelltumor-Zelllinien Tera-1 und NCCIT. **d)** Den Prostatakarzinom-Zelllinien PTu 230, PTu 232, MDA PCa 2B, VCaP, PC-3 und PN. **e)** Einer unterschiedlichen Passage der Prostatakarzinom-Zelllinien MDA PCa 2B, VCaP und PC-3 sowie der Melanom-Zelllinien SK-Mel-25 und SK-Mel-28. Das etwa 620 bp große Amplikon sowie ggf. weitere Nebenprodukte wurden ausgeschnitten und aufgereinigt. Als Negativkontrolle diente für jede cDNA-Probe (+) der jeweilige RT(-) Ansatz ohne Enzym (-). K = PCR-Wasserkontrolle, M = 100 bp DNA-Leiter (Carl Roth, Karlsruhe, Deutschland).



**Anhang 3:** Relative HERV-K(HML-2)-Klonierungsfrequenzen in einer zweiten Passage zweier Melanom-Zelllinien sowie in den Melanom- und Melanozyten-Zelllinien nach UVB-Bestrahlung. Das Amplikon im HERV-K(HML-2) gag-ORF wurde aus cDNA der Zelllinien amplifiziert und cDNA-Sequenzen generiert (vgl. Abschnitt 4.1.5. und 4.1.7. der Ergebnisse). Angegeben sind in beiden Tabellen, neben der Chromosomenposition und -bande des kompletten HML-2-Locus incl. LTRs, dessen offizieller HGNC-Name (vgl. Mayer *et al.*, 2011) sowie die absolute Anzahl generierter cDNA-Sequenzen des Locus (#) und seine relative Klonierungsfrequenz in den verschiedenen Proben in [%]. Die relative Klonierungsfrequenz berechnete sich in jeder Probe für jeden Locus aus der Anzahl der eindeutig rückkartierbaren cDNA-Sequenzen, die dem entsprechenden Locus zugeordnet werden konnten, dividiert durch die Gesamtanzahl an eindeutig rückkartierbaren Sequenzen, die für die jeweilige Probe generiert wurden. Unter der Tabelle findet sich jeweils die Gesamtanzahl eindeutig rückkartierbarer Sequenzen sowie die Anzahl transkribierter Loci. **a)** Relative Klonierungsfrequenzen von HML-2-Loci in zwei unterschiedlichen Passagen (I und II) der Melanom-Zelllinien SK-Mel-25 und SK-Mel-28. Passage I entspricht den Zellen, deren Ergebnisse bereits in Tabelle 15 aufgeführt sind. **b)** Relative Klonierungsfrequenzen von HML-2-Loci in den Melanom-Zelllinien SK-Mel-25, SK-Mel-28, MEWO und WM3734a und in den Melanozyten-Zelllinien Benno und Oskar im unbehandelten Zustand, sowie 24 h nach Bestrahlung mit 200 mJ/cm<sup>2</sup> UVB. Die Werte der unbehandelten Zellen entsprechen denen aus Tabelle 15.

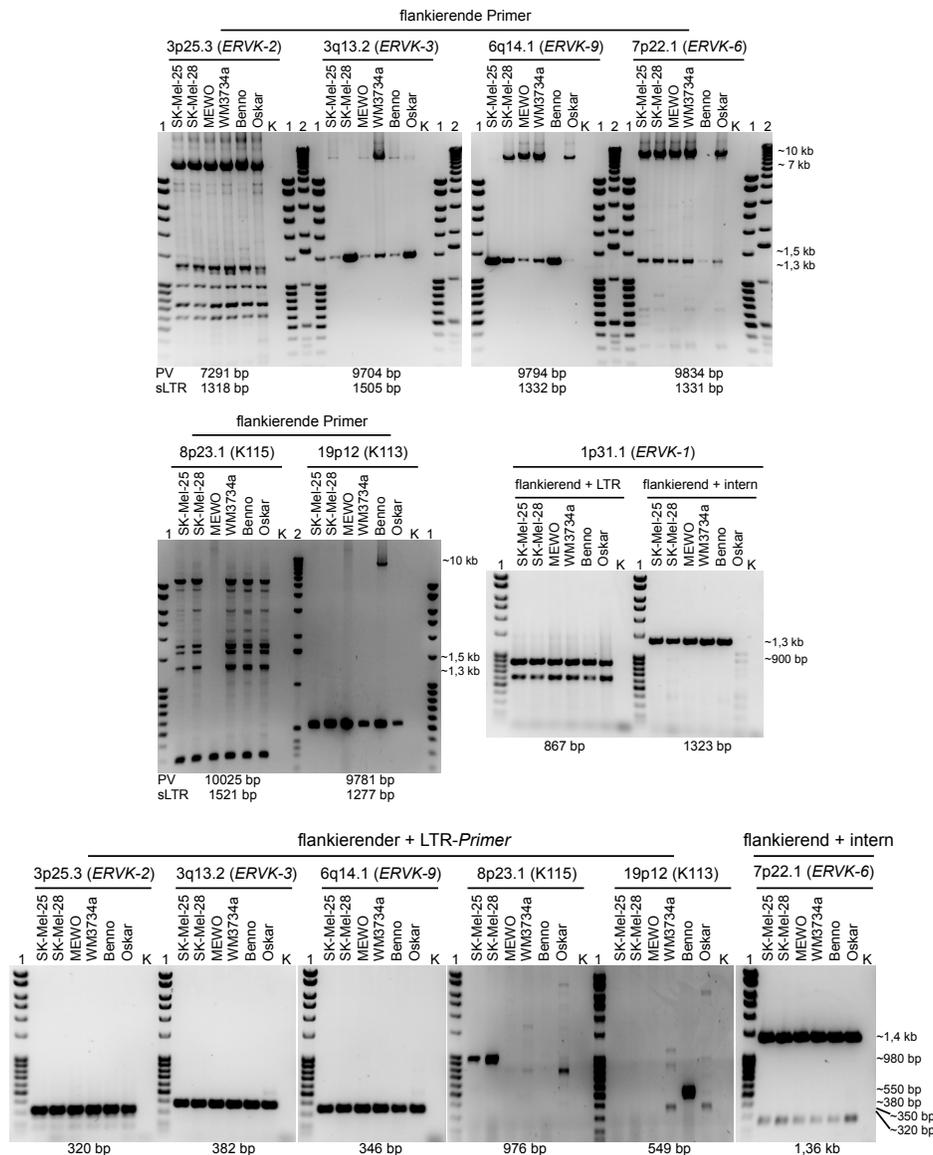
a)

Name	Chromosomenposition	Bande	SK-Mel-25				SK-Mel-28			
			Passage I		Passage II		Passage I		Passage II	
			#	%	#	%	#	%	#	%
ERVK-7	chr1:153869885-153870462	1q22	50	48,08	33	48,53	5	6,94	5	7,35
ERVK-5	chr3:102895226-102895803	3q12.3	10	9,62	14	20,59			1	1,47
ERVK-4	chr3:127093625-127094202	3q21.2	11	10,58	4	5,88			2	2,94
ERVK-13	chr3:149766678-149767256	3q24	1	0,96						
ERVK-11	chr3:186769838-186770436	3q27.2					1	1,39	1	1,47
ERVK-10	chr5:156024099-156024697	5q33.3					4	5,56	1	1,47
ERVK-6	chr7:4595679-4596276	7p22.1	22	21,15	9	13,24	54	75,00	51	75,00
ERVK-14	chr7:104179098-104179696	7q22.1					3	4,17	1	1,47
ERVK-15	chr7:141100874-141101452	7q34	1	0,96	2	2,94				
ERVK-26	chr8:12367993-12368503	8p23.1			1	1,47				
ERVK-16	chr10:6913237-6913817	10p14	2	1,92						
ERVK-17	chr10:101576298-101576896	10q24.2					2	2,78	5	7,35
ERVK-29	chr19:42296438-42297026	19q13.12	2	1,92	3	4,41	3	4,17		
ERVK-24	chr22:17307985-17308562	22q11.21	5	4,81	2	2,94			1	1,47
		Summe	104		68		72		68	
		# Loci	9		8		7		9	

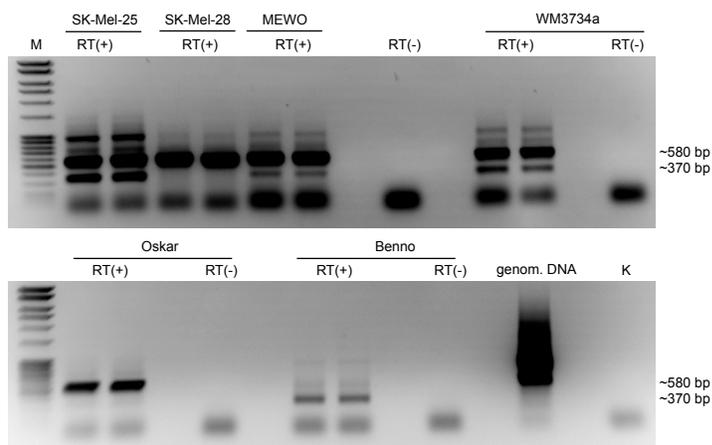
b)

Name	Chromosomenposition	Bande	Melanom-Zelllinien												Melanozyten-Zelllinien												
			SK-Mel-25				SK-Mel-28				MEWO				WM3734a				Benno		Oskar						
			unbeh.	UVB	unbeh.	UVB	unbeh.	UVB	unbeh.	UVB	unbeh.	UVB	unbeh.	UVB	unbeh.	UVB	unbeh.	UVB	unbeh.	UVB							
ERVK-7	chr1:153869885-153870462	1q22	50	48,08	18	22,22	5	6,94	12	14,29	8	11,27	3	5,00	2	2,56	5	6,67	1	1,49	11	16,67					
ERVK-2	chr3:9868910-9869488	3p25.3						2	2,38			5	7,04														
ERVK-5	chr3:102895226-102895803	3q12.3	10	9,62	14	17,28					6	6,00	2	2,82	2	2,25	1	1,67	11	14,10	2	2,67	4	5,97	4	6,06	
ERVK-3	chr3:114232604-114233182	3q13.2																					1	1,52			
ERVK-4	chr3:127093625-127094202	3q21.2	11	10,58	21	25,93			14	16,67	4	4,00	13	18,31	3	3,37	2	3,33	6	7,69	13	17,33			1	1,52	
ERVK-13	chr3:149766678-149767256	3q24	1	0,96																							
ERVK-11	chr3:186769838-186770436	3q27.2			3	3,70	1	1,39	4	5,56	5	5,95	1	1,00	8	11,27											
ERVK-10	chr5:156024099-156024697	5q33.3																									
ERVK-6	chr7:4595679-4596276	7p22.1	22	21,15	2	2,47	54	75,00	33	39,29	6	6,00	9	12,68			5	8,33									
ERVK-14	chr7:104179098-104179696	7q22.1							2	2,38	79	79,00	4	5,63	82	92,13	21	35,00	54	69,23	29	38,67	41	61,19	20	30,30	
ERVK-15	chr7:141100874-141101452	7q34	1	0,96	8	9,88			9	10,71	1	1,00	15	21,13	2	2,25	19	31,67	4	5,13	1	1,33			1	1,52	
ERVK-16	chr10:6913237-6913817	10p14	2	1,92																							
ERVK-17	chr10:101576298-101576896	10q24.2							2	2,38	2	2,38	1	1,41			5	8,33			11	14,67	8	11,94	10	15,15	
ERVK-27	chr11:61904491-61905059	11q12.3			1	1,23			2	2,38	1	1,00															
ERVK-29	chr19:42296438-42297026	19q13.12	2	1,92	14	17,28	3	4,17											1	1,28	2	2,67	1	1,49	5	7,58	
ERVK-24	chr22:17307985-17308562	22q11.21	5	4,81					3	3,57	1	1,00	3	4,23			2	3,33			7	9,33	1	1,49			
		Summe	104		81		72		84		100		71		89		60		78		75		67		66		
		# Loci	9		8		7		10		9		12		4		10		6		10		8		10		

**Anhang 4:** Überprüfung des Allelstatus von als polymorph beschriebenen HERV-K(HML-2)-Loci. Der Allelstatus wurde durch Locus-spezifische PCRs mit zur Verfügung stehender genomischer DNA der Melanom-Zelllinien SK-Mel-25, SK-Mel-28, MEWO und WM3734a sowie der Melanozyten-Zelllinien Benno und Oskar überprüft, wobei die Größe der PCR-Produkte Aufschluss über den Allelstatus des jeweiligen polymorphen Locus lieferte (vgl. Abschnitt 4.1.8. der Ergebnisse). Zur Amplifikation wurden den Locus flankierende *Primer* verwendet, wobei sich im Falle eines Volllängen-Provirus (PV) ein Produkt von ~7 - 10 kb ergab, im Falle eines solitären LTRs (sLTR) ein Produkt von ~1,3 - 1,5 kb, oder ein ~500 bp großes Produkt im Falle einer leeren Präintegrationsstelle, sofern das Provirus-Allel noch nicht in der Population fixiert ist. Das Vorhandensein eines Provirus oder solitären LTRs wurde durch eine zweite PCR mit Locus-flankierendem *Primer* und einem weiteren *Primer* im LTR, bzw. im Falle von *ERVK-6* (7p22.1) in der proviralen *gag*-Region, verifiziert. Im Falle des *ERVK-1* (1p31.1)-Locus, der von L1-Sequenzen flankiert wird, konnten keine spezifischen Locus-flankierenden *Primer* verwendet werden, weswegen hier ein flankierender sowie ein zweiter *Primer* innerhalb des *gag*-ORFs verwendet wurde. Gezeigt ist das Ergebnis der gelelektrophoretischen Auftrennung der Locus-spezifischen PCR-Produkte aus genomischer DNA der Zelllinien, wobei die jeweils zur Amplifikation verwendeten *Primer*-Kombinationen angegeben sind. Zudem sind die Größen der zu erwartenden Produkte unter den Gelbildern angegeben. Mit Ausnahme der polymorphen Loci *HERV-K113* und *HERV-K115* konnte für nahezu alle Loci in fast allen Proben ein ~10 kb großes PCR-Produkt generiert werden, wodurch das Vorhandensein mindestens einer allelischen Kopie des jeweiligen Volllängen-Provirus bestätigt wurde. K = PCR-Wasserkontrolle, 1 = 100 bp DNA-Leiter (Carl Roth, Karlsruhe, Deutschland), 2 = 1 kb DNA-Leiter (Invitrogen/Life Technologies, Carlsbad, USA).



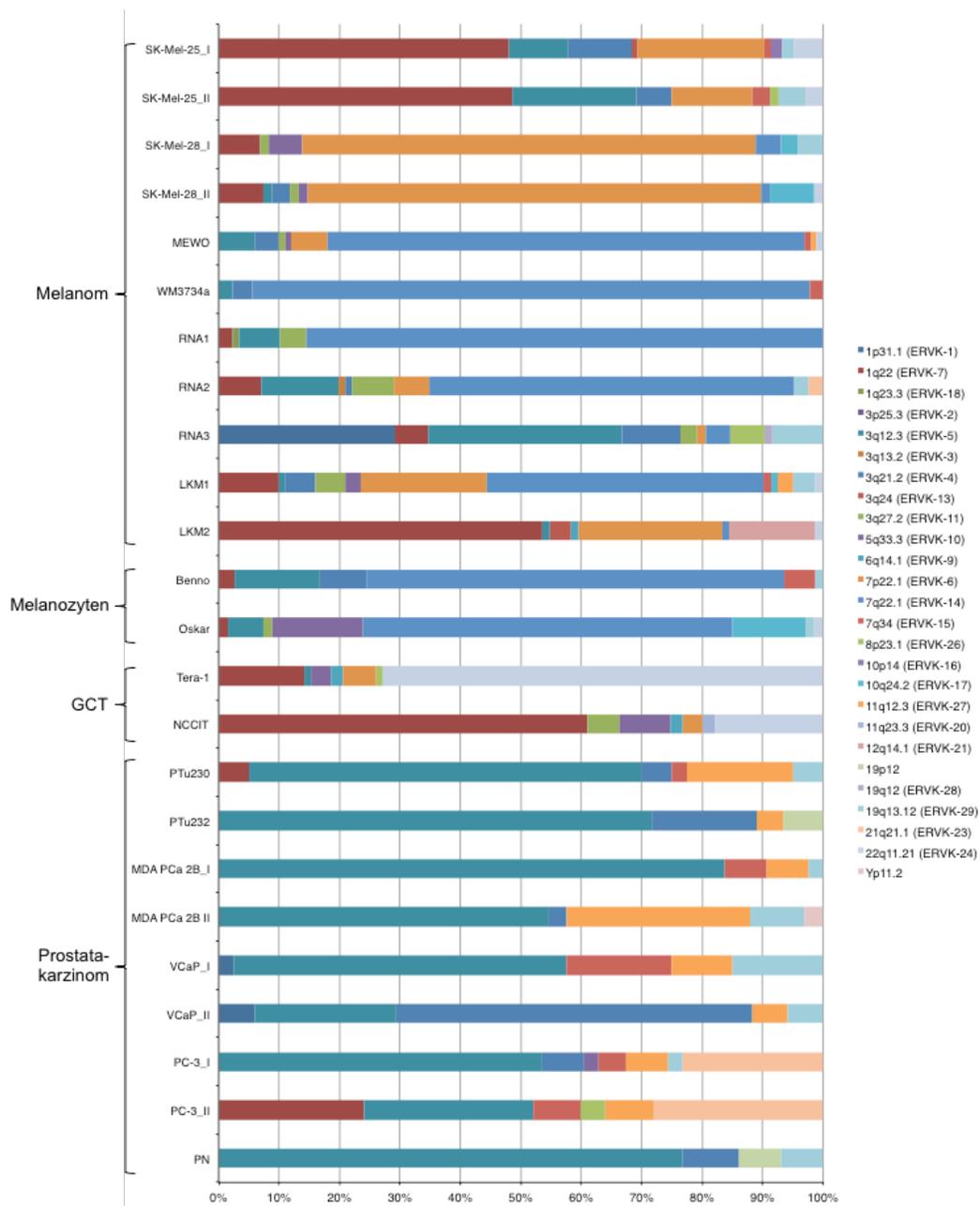
**Anhang 5:** Ergebnis der gelelektrophoretischen Auftrennung des aus verschiedenen Tumorproben amplifizierten HERV-K(HML-2) *rec/np9*-Amplikons. Die Amplikons wurde aus cDNA der Melanom-Zelllinien SK-Mel-25, SK-Mel-28, MEWO und WM3734a sowie der Melanozyten-Zelllinien Benno und Oskar amplifiziert (vgl. Abschnitt 4.1.10. der Ergebnisse). Die ~370 bp (*np9*) bzw. ~580 bp (*rec*) großen PCR-Produkte, sowie ggf. weitere auftretende Nebenprodukte wurden ausgeschnitten und aufgereinigt. Als Negativkontrolle diente für jede cDNA-Probe RT(+) der jeweilige RT(-) Ansatz ohne Enzym, als Positivkontrolle diente humane genomische DNA (es ist zu beachten, dass die PCR-Produkte von genomischer DNA, die die in *rec*- und *np9*-mRNA gespleißte Region beinhalten, entsprechend größer sind). K = PCR-Wasserkontrolle, M = 100 bp DNA-Leiter (Carl Roth, Karlsruhe, Deutschland).



**Anhang 6:** Relative Klonierungsfrequenzen der in Keimzelltumor- (GCT) und Prostatakarzinom-Zelllinien als transkribiert identifizierten HERV-K(HML-2)-Loci. Das Amplikon im HERV-K(HML-2) *gag*-ORF wurde aus cDNA der GCT-Zelllinien Tera-1 und NCCIT sowie der Prostatakarzinom-Zelllinien PTu230, PTu232, MDA PCa 2B (in der Tabelle kurz MDA), VCaP, PC-3 und PN amplifiziert, mittels Sanger-Methode Sequenzen generiert und im menschlichen Referenzgenom rückkartiert (vgl. Abschnitt 4.1.14. der Ergebnisse). Im Falle von MDA PCa 2B, VCaP und PC-3 stand cDNA aus zwei unterschiedlichen Passagen der Zelllinien (in der Tabelle jeweils bezeichnet als I und II) zur Verfügung. Angegeben sind, neben der Chromosomenposition und -bande des kompletten HML-2-Locus incl. LTRs dessen offizieller HGNC-Name (vgl. Mayer *et al.*, 2011), die absolute Anzahl (#) generierter Sequenzen jedes Locus sowie seine relative Klonierungsfrequenz in den verschiedenen Proben in [%]. Die relative Klonierungsfrequenz berechnete sich in jeder Probe für jeden Locus aus der Anzahl an eindeutig rückkartierbaren Sequenzen, die dem entsprechenden Locus zugeordnet werden konnten, dividiert durch die Gesamtanzahl an rückkartierbaren Sequenzen, die für die Probe generiert wurden.

Name	Chromosomenposition	Bande	GCT-Zelllinien				Prostatakarzinom-Zelllinien																		
			Tera-1		NCCIT		PTu230		PTu232		MDA I		MDA II		VCaP I		VCaP II		PC-3 I		PC-3 II		PN		
			#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%	
ERVK-1	chr1:75615359-75621731	1p31.1																							
ERVK-7	chr1:153863081-153872260	1q22	13	14,13	58	61,05	2	5,00																	
ERVK-5	chr3:102893427-102902549	3q12.3	1	1,09			26	65,00	33	71,74	36	83,72	18	54,55	22	55,00	4	23,53	23	53,49	7	28,00	33	76,74	
ERVK-4	chr3:127091992-127101129	3q21.2					2	5,00	8	17,39			1	3,03			10	58,82	3	6,98			4	9,30	
ERVK-11	chr3:186763030-186772209	3q27.2			5	5,26																			
ERVK-10	chr5:156017295-156026474	5q33.3	3	3,26	8	8,42													1	2,33					
ERVK-9	chr6:78483381-78492802	6q14.1	2	2,17	2	2,11																			
ERVK-6	chr7:4595679-4604779	7p22.1	3	3,26	3	3,16																			
ERVK-15	chr7:141097395-141102407	7q34					1	2,50			3	6,98			7	17,50			2	4,65	2	8,00			
ERVK-26	chr8:12360863-12370378	8p23.1	1	1,09																					
ERVK-27	chr11:61892538-61907139	11q12.3					7	17,50	2	4,35	3	6,98	10	30,30	4	10,00	1	5,88	3	6,98	2	8,00			
ERVK-20	chr11:118096934-118106093	11q23.3			2	2,11																			
-	chr19:20248399-20258512	19p12							3	6,52														3	6,98
ERVK-29	chr19:42289389-42298906	19q13.12					2	5,00			1	2,33	3	9,09	6	15,00	1	5,88	1	2,33				3	6,98
ERVK-23	chr21:18855530-18863833	21q21.1																					10	23,26	
ERVK-24	chr22:17306187-17315361	22q11.21	67	72,83	17	17,89																7	28,00		
-	chrY:6886441-6893384	Yp11.2											1	3,03											
		Summe	92		95		40		46		43		33		40		17		43		25		43		
		# Loci	7		7		6		4		4		5		5		5		7		6		4		

**Anhang 7:** Graphische Darstellung der relative Klonierungsfrequenzen von in Melanom-, Keimzelltumor- und Prostatakarzinom-Proben als transkribiert identifizierten HERV-K(HML-2)-Loci. Das HML-2 gag-Amplikon wurde aus cDNA der Melanom-Zelllinien SK-Mel-25, SK-Mel-28, MEWO und WM3734a, drei Melanomproben (RNA1-3), zwei Lymphknoten-Metastasen von Melanom-Patienten (LKM1, 2), den Melanozyten-Zelllinien Benno und Oskar, den Keimzelltumor (GCT)-Zelllinien Tera-1 und NCCIT sowie den Prostatakarzinom-Zelllinien PTu 230, PTu 232, MDA PCa 2B, VCaP, PC-3 und PN amplifiziert, mittels Sanger-Methode Sequenzen generiert und im menschlichen Referenzgenom rückkartiert (vgl. Abschnitt 4.1.4., 4.1.5. und 4.1.14. der Ergebnisse). Im Falle der Melanom-Zelllinien SK-Mel-25 und SK-Mel-28 sowie der Prostatakarzinom-Zelllinien MDA PCa 2B, VCaP und PC-3 stand jeweils cDNA aus zwei unterschiedlichen Passagen (in der Abbildung bezeichnet als I und II) der Zelllinien zur Verfügung. Gezeigt ist eine graphische Darstellung der relativen Klonierungsfrequenzen in [%], die sich in jeder Probe für jeden Locus aus der Anzahl der eindeutig diesem Locus zugeordneten cDNA-Sequenzen dividiert durch die Gesamtzahl an rückkartierbaren Sequenzen der Probe berechneten. Die Einzelergebnisse, aus denen sich das Diagramm zusammensetzt, sind bereits in den Abbildungen 18, 19 und 30 gezeigt, die relativen Klonierungsfrequenzen finden sich in Tabelle 15 und Anhang 3a und 6.



**Anhang 8:** Von genomischer DNA generierte Sequenzen potentieller neuer HERV-K(HML-2)-Loci oder allelischer Varianten bekannter Loci. Im Zuge der Identifizierung *rec* bzw. *np9* mRNA-produzierender Loci in menschlichen Normalgeweben wurde eine große Anzahl cDNA-Sequenzen zweier Sequenzpopulationen generiert, die im Zuge der Rückkartierung im menschlichen hg18 Referenzgenom keinem HML-2-Locus zugeordnet werden konnten (vgl. Abschnitt 4.1.15. und 4.1.17. der Ergebnisse). Die Sequenzpopulationen zeichneten sich immer durch eine hohe Anzahl an Nukleotidunterschieden zu ihrem besten *BLAT-Match* aus, welcher im einen Fall der *ERVK-5*-Locus (3q12.3) war, im zweiten Fall der *ERVK-18*-Locus (1q23.3). Die Sequenzunterschiede zu diesen Loci fanden sich hierbei größtenteils immer an den gleichen Positionen. Um auszuschließen, dass es sich bei diesen Sequenzen um ein *in vitro* generiertes Phänomen handelt, wurde das *rec/np9*-Amplikon auch von genomischer DNA amplifiziert und Sequenzen generiert. Zusätzlich wurde ein zweites Amplikon von genomischer DNA generiert und sequenziert, wobei hier in der PCR spezifische *nested-Primer* verwendet wurden, die mit ihrem 3'-Ende auf charakteristischen Nukleotidunterschieden der cDNA-Sequenzpopulationen lagen. Auch von genomischer DNA konnten Sequenzen generiert werden, die die gleichen charakteristischen Nukleotidunterschiede wie die cDNA-Sequenzen aufwiesen. **a)** Im Falle der *ERVK-5*-ähnlichen Sequenzen wurde unter Verwendung der regulären *rec/np9-Primer* eine Sequenz generiert, die sich durch die gleichen Nukleotidunterschiede wie die cDNA-Sequenzpopulation mit Ähnlichkeit zu *ERVK-5* auszeichnete. Gezeigt ist das Ergebnis der Sequenzierung des Amplikons aus genomischer DNA aus beiden Richtungen (T7 und SP6), wobei die Bereiche der genomischen Sequenz, die denen der zuvor generierten gespleißten cDNA-Sequenzen entsprechen, grau unterlegt sind. Die Nukleotidunterschiede sind bunt markiert, wobei die Farbgebung der aus Abbildung 31a entspricht. **b)** Im Falle der *ERVK-18*-ähnlichen Sequenzen konnten unter Verwendung der *nested-Primer* zahlreiche Sequenzen generiert werden, die sich durch die gleichen Nukleotidunterschiede wie die cDNA-Sequenzpopulation mit Ähnlichkeit zu *ERVK-18* auszeichneten und von denen exemplarisch eine gezeigt ist. Das sequenzierte *nested-PCR*-Produkt war hierbei innerhalb des 244 bp langen Bereiches lokalisiert, der auf die charakteristische Typ I 292 bp Deletion folgt und in *np9*-mRNA üblicherweise als Intron-Teil gespleißt wird. Die charakteristischen Nukleotidunterschiede sind bunt markiert, wobei die Farbgebung der aus Abbildung 31b entspricht.

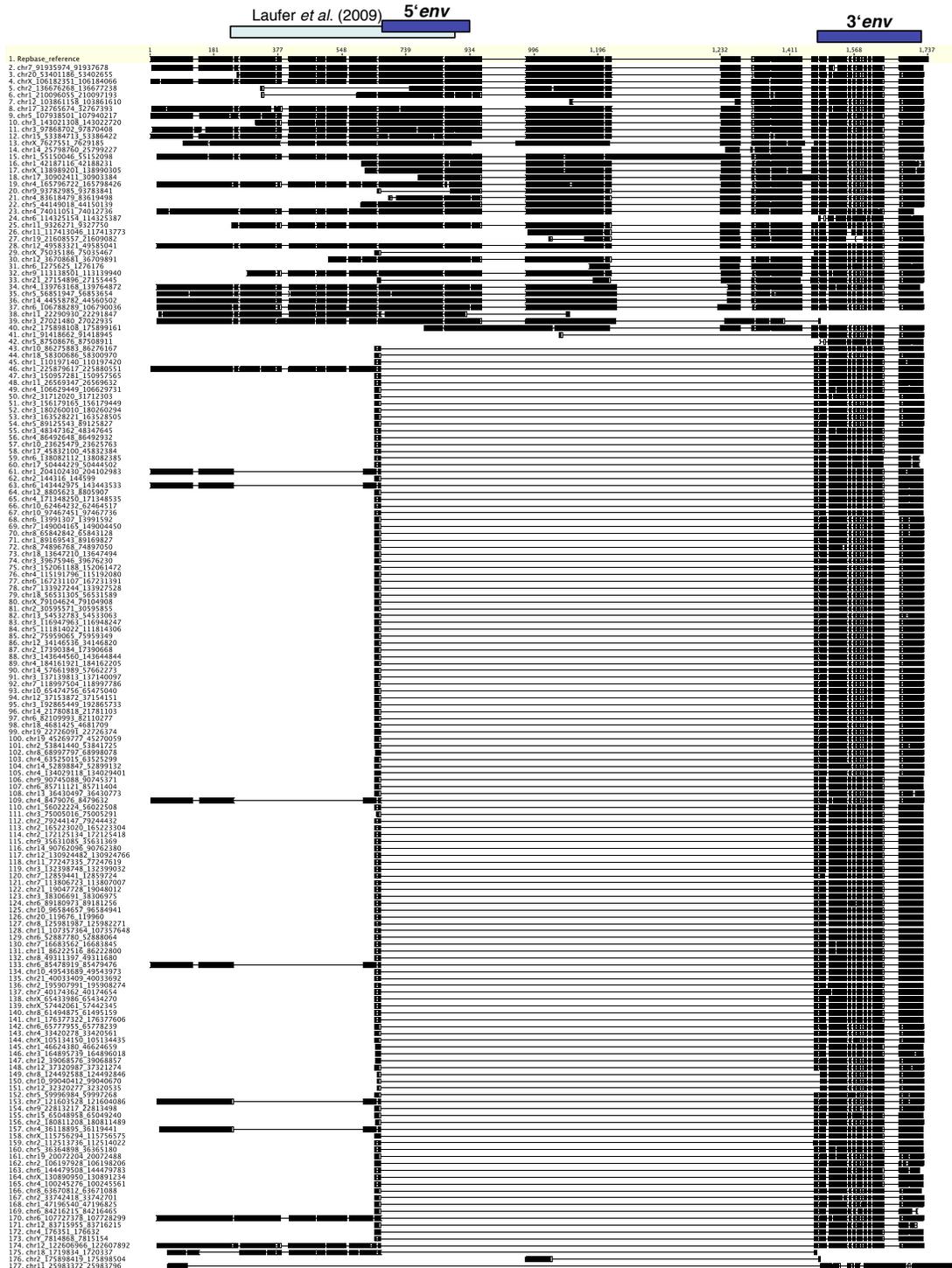
a)

```
>HS_KSch.flank12_T7
AGAAAAGCGCCTCCACGGAGACGGTAACACCACTCACATGGATGGATAATCCTATAGAATTAATGTTAAGGATAGTGTA
TGGGTACCTGGCCCCACAGATGATCACTGCCCTGCCAAACCTGAGGAAGAAGGGATGATGATAAATATTTCCATTGGGTA
TCGTTATCCTCCTATTTGCGTAGGAAGAGCACCAGGATGCTTAATGCCTGCATTCCAAAATTGGTTGGTAGAAGTACCTA
CTGCCGGTCTTAACAGTAGACTCACTTATCACATGGTAAGCGGGATGTCACTCAGGCCACGGGTAATTTATTTACAAGAC
TTTTCTTATCAAAGATCATTAATAATTTAGACCTAAGGGAAACCTTGCCCAAGGAAATTCCTCAAGGATCAAAAAATAC
AGAAGTTTTAGTTTGGGAAGAATGTGTGGCAATAGTGGCGGTGATATTACAAAACAATGAATCCGAACCTGTTATAGATT
GGGCATCTCGAGTCAATTTACCACAATTTGCTCAGGACAAACTCAGTCACTGTCCAAAGTGCACAAGTGGTCCAGCTGTT
GATAGCGACTTAACAGAAAGTCTAGACAAAACATAAGCATAAAAAATTACAGTCTTTCTACCCTTGGGTATGGGGAGAAAA
AGGAATCTCTAACCAGACCAAAAATAAAGTCTGTTTCTGGTCTGAAACATCCAGAATTTGTGGAGGCTTACTGCAG
CCTCATACCACATTAGGATTTGGTCTGGAATCAAACCTTAGAAACAAGAGATCGTAAGCCATTTTATACTATCTACC
>HS_KSCH.flank12_SP6
TATTAAGGCGTTTTAAATAGATCCGAAAGATTCATTTTACTTTAATTCAGTGATTATGGGATTAATTCAGTCACAG
CTATGGCTGCTGTGGCAGGAGTTGCATTTGCACTCTCTGTTCAGTCAGTAAAAATTTGTTAATGATTGACAAAAAATTC
ACATGATTGTGGAATTCACAATCTGGTATTGATCAAAAATTTGGCAAATCAAATTAATGATCTTAGACAAACTGTCAATTTG
GATGGGAGACAGACTCATGAGCTTAGAACATCGTTTCCAGTTACAGTGTGACTGGAATACATCAGATTTTTGTATTACAC
CCAAAATTTATAATGAGTCTGAGCATCACTGGGACATGGTTAGACGTCATCTACAGGGAAGAGAAGATAATCTCACTTTA
GACATTTCCAAGTTAAAAGAACAAATTTTCAAACATCAAAGCCCATTTAAATTTGGTGCCAGGAACTGAGGCAATTGC
AGGAGTTGCTGATGGCCTCGCAAATCTTAACCTGTCACTTGGGTTAAGACCATCAGAAGTACTACGATTGTAAGTCTCA
TATTAATCCTTGTGTGCTGTTTGTCTGTGTTGTTGCTCAGGTTGATCCCAACAGCTCGAAGAGACAGAGACCATCGA
GAACGGCCATGATGACGATGGCGGTTTTGTGCGAAAAGAAAAGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGT
TACTGTGCTGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTGTTCTGTACTAAGAAAAATTTCCTGCTTGAGAT
TCTGTTAATCTATGACCTTACCCCAACCCCGTCTCTCTGAAA
```

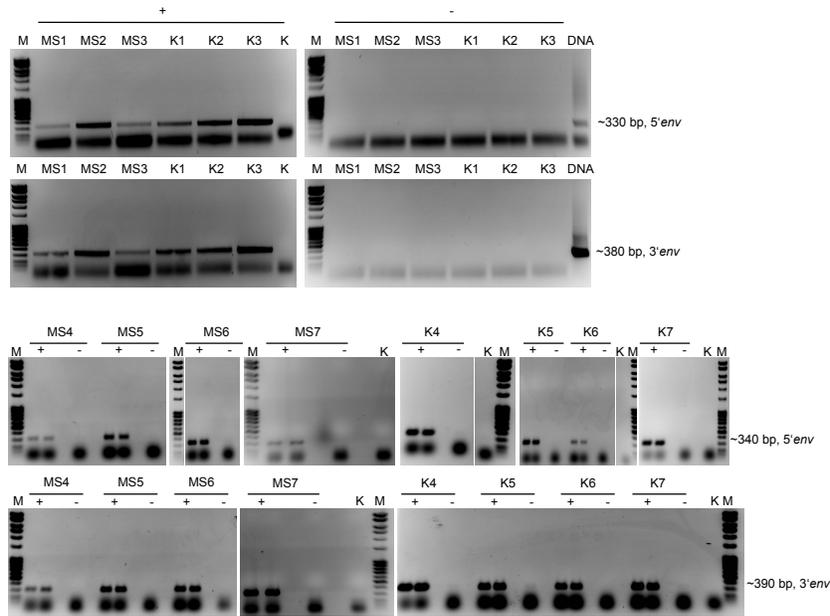
b)

```
>DNA14_07_T7
ATCACATGGATGGATAATCCTATAGAAATATGTTAAGGATAGTGTATGGGTACCTGGCCCCACAGATGATCTGCTGCC
CTGCCAAACCTGAGGAAGAGGGATGATGATAAATATTTCCATTGGGTATCGTTATCTCCTATTTGCTTAGGAGAGCA
CCAGGATGTTAATGCCTGCAATCCAAAATTTGGTTGGTAGAAGTACCTACTG
```

**Anhang 9:** Graphische Darstellung eines multiplen Sequenz-Alignments genomischer Loci der HERV-W-Gruppe. Die *env*-Bereiche genomischer HERV-W-Loci wurden, wie in Material und Methoden und Ergebnissen beschrieben (vgl. Abschnitte 3.2.3.1. und 4.2.2.), aus dem menschlichen hg18 Referenzgenom zusammengestellt und aliniert. Alinierbare Bereiche der Provirer sind als schwarze Balken dargestellt, horizontale Linien repräsentieren Lücken im Vergleich zur Sequenz anderer Loci. Über dem *Alignment* ist die Nukleotidposition relativ zum *env*-ORF der bei *Repb*ase hinterlegten HERV17- Konsensussequenz (entspricht HERV-W; gelb unterlegt) angegeben. Die Positionen der hier untersuchten Amplikons 5'*env* und 3'*env* sowie des in der Studie von Laufer *et al.* (2009) untersuchten Amplikons sind eingezeichnet.



**Anhang 10:** Ergebnis der gelelektrophoretischen Auftrennung der aus Hirnproben von MS-Patienten und gesunden Kontrollen amplifizierten HERV-W *env*-Amplikons. Die 280 bp und 330 bp großen Amplikons 5'*env* und 3'*env* wurden aus cDNA von sieben MS-Patienten (MS) und sieben Kontrollen (K) amplifiziert, die Produkte ausgeschnitten und aufgereinigt und anschließend mittels *454/FLX*-Technologie sequenziert (vgl. Abschnitt 4.2.3. der Ergebnisse). Die Proben MS1-3 und K1-3 sowie MS4-7 und K4-7 wurden in zwei separaten *454/FLX*-Sequenzier-Läufen analysiert. Zur PCR-Amplifikation wurden *Primer* mit entsprechenden *FLX*-Adaptoren und *Key*-Sequenzen an ihrem 5'-Ende verwendet (vgl. Tab. 6), die im Falle der Proben MS1-3 und K1-3 je 25 bp am *Forward*- und *Reverse-Primer* umfassten, im Falle der Proben MS4-7 und K4-7 je 30 bp, wodurch die amplifizierten Produkte entsprechend größer sind. Als Negativkontrolle diente für jede cDNA-Probe (+) der jeweilige RT(-) Ansatz ohne Enzym (-). K = PCR-Wasserkontrolle, DNA = Positivkontrolle mit genomischer DNA, M = 100 bp DNA-Leiter (Carl Roth, Karlsruhe, Deutschland).



**Anhang 11:** Übersicht über die mittels *454/FLX*-Sequenzierung generierten HERV-W 5'*env*- und 3'*env*-Sequenzen und deren Filterung. Angegeben sind die Gesamtanzahlen der für alle untersuchten MS- und Kontroll-Proben (K) generierten cDNA-Sequenzen, die durchschnittliche Sequenzlänge, die Anzahl manuell gelöschter Sequenzartefakte und eine detaillierte Übersicht, wie viele Sequenzen mit wie vielen Nukleotidunterschieden einem genomischen HERV-W-Locus zugeordnet werden konnten. Sequenzen mit mehr als drei Nukleotidunterschieden zu ihrem besten *BLAT-Match* wurden verworfen. Die jeweilige Anzahl an Sequenzen, die mit weniger als drei Nukleotidunterschieden einem genomischen HERV-W-Locus zugeordnet werden konnten und damit in die weitere Analyse eingingen, ist in der jeweils untersten Zeile angegeben. Zur Filterung, vgl. auch Material und Methoden, Abschnitt 3.2.3.2.2.

		5'env																											
		MS1	MS2	MS3	MS4	MS5	MS6	MS7	K1	K2	K3	K4	K5	K6	K7														
Gesamtanzahl Sequenzen		4297	2208	4304	2203	2116	1011	4010	5726	1622	1979	736	697	494	626														
durchschnittl. Sequenzlänge [bp]		238,5	245,4	239	215,6	225,3	216,1	215,1	242	242,4	241,2	236,3	236,5	236,4	236,8														
# verworfener Sequenzartefakte		155	50	111	69	45	27	144	114	42	74	13	13	8	6														
# Sequenzen mit x Nukleotidunterschieden		#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%												
0		3104	74,94	1606	74,42	2755	65,70	1596	78,58	1555	77,21	681	72,30	2665	71,99	4097	73,00	1250	79,11	1409	73,96	516	72,88	663	76,74	353	73,54	403	67,10
1		799	19,29	432	20,02	998	23,80	293	14,43	257	12,76	120	12,74	597	16,13	1195	21,29	272	17,22	417	21,89	120	16,95	127	14,70	65	13,54	129	21,46
2		181	4,37	97	4,50	346	8,25	84	4,12	151	7,50	104	11,04	305	8,24	232	4,13	46	2,91	63	3,31	53	7,49	49	5,67	47	9,79	46	7,65
3		39	0,94	21	0,97	75	1,79	32	1,58	25	1,24	23	2,44	70	1,89	48	0,86	8	0,51	13	0,68	6	0,85	9	1,04	11	2,29	13	2,16
4		16	0,39	1	0,05	16	0,38	14	0,69	22	1,10	4	0,42	42	1,13	19	0,34	3	0,19	3	0,16	9	1,27	11	1,27	2	0,41	6	0,10
5		3	0,07	0	0,00	3	0,07	7	0,34	4	0,20	5	0,53	14	0,38	11	0,20	1	0,06	1	0,04	1	0,14	2	0,23	1	0,21	4	0,67
>5		0	0,00	1	0,05	0	0,00	5	0,25	0	0,00	5	0,53	9	0,24	10	0,18	0	0,00	0	0,00	3	0,42	3	0,35	1	0,21	0	0,00
Summe		4142	2158	4193	2031	2014	942	3702	5612	1580	1905	708	864	480	601														
verworfen Sequenzen >3 nt-Unterschiede		19	2	19	26	26	14	65	40	4	3	13	16	4	10														
Sequenzen für weitere Analyse		4123	2156	4174	2108	2045	970	3801	5572	1576	1902	710	888	482	610														

		3'env																											
		MS1	MS2	MS3	MS4	MS5	MS6	MS7	K1	K2	K3	K4	K5	K6	K7														
Gesamtanzahl Sequenzen		1712	3077	6547	1741	2470	2169	2201	1989	2096	2288	754	680	786	950														
durchschnittl. Sequenzlänge [bp]		233,1	237,1	232,9	150,6	162,5	149,5	154,7	238,8	233,1	236,6	155,6	165,9	167	148,4														
# verworfener Sequenzartefakte		52	100	97	639	575	713	665	28	44	96	232	170	171	309														
# Sequenzen mit x Nukleotidunterschieden		#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%												
0		991	23,93	1853	85,87	3939	93,94	632	59,74	1327	73,49	1133	83,99	916	63,79	1247	22,22	1281	81,08	1472	77,27	378	75,60	380	78,35	447	76,67	455	74,84
1		495	11,95	743	34,43	1860	44,36	308	29,11	391	21,65	162	12,01	348	24,23	514	9,16	607	38,42	560	29,40	93	18,60	80	16,49	92	15,78	124	20,39
2		138	3,33	297	13,76	500	11,93	71	6,71	70	3,88	39	2,89	88	6,13	152	2,71	128	8,10	127	6,67	26	5,20	21	4,33	28	4,80	16	2,63
3		25	0,64	65	3,01	98	2,34	28	2,65	17	0,94	10	0,74	53	3,69	35	0,62	28	1,77	23	1,21	3	0,60	4	0,82	12	2,06	11	1,81
4		7	0,17	16	0,74	34	0,81	14	1,32	0	0,00	4	0,30	28	1,95	5	0,09	6	0,38	5	0,26	0	0,00	0	0,00	1	0,17	2	0,33
5		3	0,07	1	0,05	10	0,24	4	0,38	1	0,06	1	0,07	2	0,14	4	0,07	1	0,06	1	0,05	0	0,00	0	0,00	2	0,34	0	0,00
>5		1	0,02	2	0,09	9	0,21	1	0,09	0	0,00	0	0,00	1	0,07	4	0,07	1	0,06	4	0,21	0	0,00	0	0,00	1	0,17	0	0,00
Summe		1660	2977	6450	1058	1806	1349	1436	1961	2052	2192	500	485	583	608														
verworfen Sequenzen >3 nt-Unterschiede		11	19	53	19	1	5	31	13	8	10	4	4	4	2														
Sequenzen für weitere Analyse		1649	2958	6397	1083	1894	1451	1505	1948	2044	2182	522	510	611	639														

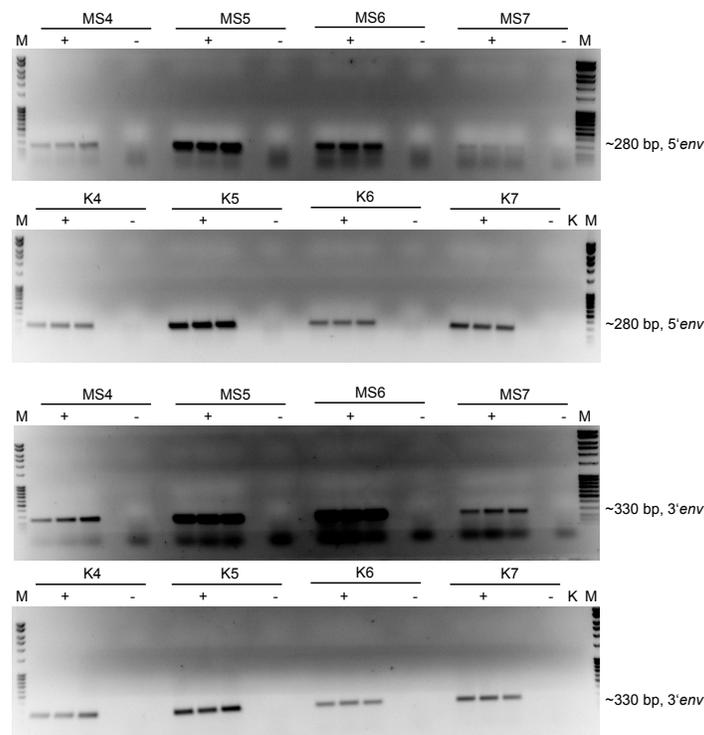


Anhang 12: Fortsetzung.

Start	Ende	HERVW-Locus	HGNC	Strang	MS1	MS2	MS3	MS4	MS5	MS6	MS7	K1	K2	K3	K4	K5	K6	K7	Mittewert [%] FLX & MSeq				
																				FLX	FLX	FLX	FLX
3*	164894958	3q28.1 <sup>b</sup>																					
3*	18025072	16026637	3q28.32	+	1,33	0,95	1,55	1,49	1,14	0,94	0,01	1,54	0,58	1,70	1,00	0,88	1,04	1,01	0,88	0,59	1,11	0,01	
3*	192860266	19286075	3q28	+							0,0002	0,54	0,23	0,54	0,20	0,11	0,02	0,02	0,02	0,02	0,83	0,01	
4	175551	180863	4p16.3	+	0,30	0,14	0,50	0,18	0,13	0,06	0,05	0,23	0,51	0,34	0,86	0,20	0,11	0,15	0,15	0,20	0,99	0,03	
4	3472391	8460392	4p16.1	+	1,15	1,62	1,06	0,79	0,94	0,89	0,60	1,25	1,79	1,60	0,40	0,60	0,83	0,90	0,93	0,62	1,19	0,08	
4	33419522	33421367	4p15.1	+							0,06	0,06			0,20	0,005				0,02	0,14	0,05	
4*	36118546	36122240	4p14	-							0,01					0,001						0,00	
4*	63522789	63526073	4q13.1	+																		0,01	
4	74010156	74017767	4q13.3	-																		0,01	
4	83613493	83620230	4q21.22	+		0,03					0,13	0,10	0,05	0,14	0,40	0,15	0,21	0,18	0,28	0,17	0,11	0,25	
4	86492302	86495571	4q21.23	+	0,55	0,47	0,77	0,35	0,39	0,17	0,42	1,39	0,21	0,10	0,19	0,27	0,21	0,54	0,54	0,45	0,36	0,53	
4*	106628394	106628806	4q24	+							0,0004	0,0004	0,0004	0,21	0,09	0,21	0,17	0,21	0,21	0,17	0,03	0,00	
4	115184984	115192421	4q26	+	0,12	0,07																0,16	0,11
4	134022904	134029743	4q28.3	+	0,36	1,18	1,39	1,67	1,01	2,27	0,003	0,56	0,56	1,45	1,89	0,60	0,99	0,21	1,13	0,69	0,83	1,05	
4*	139752390	139757603	4q31.1	+				0,09							0,004							0,98	1,01
4	153982430	153982932	4q31.3	+							0,02											0,00	
4	165796754	165796754	4q32.3	+							0,05											0,12	
4*	171347918	171358300	4q33	+													0,0003					0,04	
4*	184157539	184162547	4q35.1	+																		0,01	
4*	36364558	36368981	5p13.2	+	0,12	0,10	0,33	0,26	0,12	0,11	0,22	0,69	0,38	0,21	0,44	0,20	0,58	0,01	0,01	0,50	0,31	0,31	
5	44146946	44150936	5p12	+							0,01						0,41				0,50	0,02	
5	56851606	56854473	5q11.2	+							0,0002	0,50	0,51	0,87	0,52	0,40	0,86	0,66	0,66	0,17	0,38	0,02	
5	87508343	87509456	5q14.3 <sup>b</sup>	+	0,30	0,41	0,20	0,70	0,84	0,33	0,08											0,51	
5	89125459	89126610	5q14.3 <sup>b</sup>	+																		0,02	
5	107931705	107942182	5q21.3	+	0,30	0,17	0,42				0,10	0,07									0,04	0,24	
5	11814044	11814302	5q22.2	+							0,01										0,01	0,11	
6	1275281	1276849	6p25.3	+	0,36	0,20	0,02				0,0004	0,0004	0,0004	0,09	0,09	0,01	0,01	0,01	0,01	0,01	0,30	0,07	
6	13896545	13891534	6p23	+	0,49	0,14	0,72				0,02										0,02	0,28	
6*	52887632	52894671	6p12.1	-							0,02										0,02	0,18	
6*	6571605	65778581	6q12	+							0,02										0,03	0,03	
6*	82103381	82110363	6q14.1	+																	0,01	0,01	
6*	84210197	84216828	6q14.2	+																	0,03	0,03	
6*	85478138	85487587	6q14.3 <sup>b</sup>	+																	0,03	0,03	
6*	85710654	85711747	6q14.3 <sup>b</sup>	+																	0,02	0,02	
6*	89180630	89187945	6q15	+							0,04										0,04	0,04	
6*	10678204	106793640	6q21a	+	3,34	3,65	4,33	2,37	3,26	3,38	0,24	6,32	3,23	4,69	3,11	3,80	2,93	2,69	3,40	0,66	1,77	3,20	
6	107727309	107729640	6q21b	+	0,17	0,17	0,08				0,12	0,12	0,15	0,15	0,14	0,37	0,40	0,17	0,21	0,21	0,34	0,14	
6	111558913	111558969	6q21c	+	0,06		0,02				0,03	0,01				0,22	0,03	0,09	0,09	0,16	0,04	0,09	
6	138081757	138085532	6q21d	+																	0,07	0,04	
6	143442630	143450542	6q21e	+	0,91	1,69	2,34	2,02	1,23	0,50	1,39	0,62	2,82	1,11	2,17	1,40	2,14	1,66	3,13	2,27	0,99	2,43	
6	144479154	144484391	6q21f	+								0,07									0,86	1,60	
6	167224557	167231731	6q21g	+	2,25	2,60	0,72	0,26	0,58	0,78	0,86	0,49	1,26	1,18	0,53	0,47	0,60	0,83	0,85	0,30	0,33	0,72	
7	12859096	12859827	7p21.3	+	1,33	0,47	0,69	0,09	0,21	0,06	0,07	0,01	0,15	0,58	0,38	0,05	0,41	0,07	0,09	0,17	0,13	0,41	
7	16683219	16684043	7p21.1	-								0,004									0,05	0,05	
7	35702273	35703152	7p14.2	+						0,06											0,07	0,06	
7	40173581	40173987	7p14.1	+																	0,08	0,08	
7	91935248	91945442	7q21.2	+	3,22	4,63	5,78	6,77	10,98	11,80	13,51	2,91	5,74	8,61	9,34	17,60	14,74	16,36	18,60	19,52	11,88	11,98	
7	113806378	113813611	7q31.1	+	0,06	0,10	0,14				0,03	0,21	0,53	0,21	0,05	0,18	0,18	0,11	0,14	0,18	0,13	0,15	
7*	118983435	118988119	7q31.31	+																	0,0005	0,01	
7	121558944	121604895	7q31.32	+																	0,64	0,44	
7	131156140	131166223	7q32.3	+	0,12	0,14	0,63				0,68	0,41	0,97	0,58	0,33	0,65	0,41	0,41	0,39	0,22	0,42	0,44	
7	133920714	133925307	7q33	+	0,06			0,26	0,11	0,02	0,03	0,06									0,06	0,03	
7	146996620	146996245	7q36.1	+	0,12	0,14					0,001	0,0002	0,0002	0,10	0,42	0,10	0,26	0,69	0,06	0,13	0,38	0,23	
8	20066529	20061424	8p21.3	+																	0,03	0,03	
8	61494526	61500632	8q12.1	+							0,04										0,04	0,09	
8	65839066	65843472	8q12.3	+							0,04										0,05	0,05	
8*	68996751	68998420	8q13.2	+	0,67	0,14	0,19	0,09	0,41	0,06	0,42	0,36	0,21	0,48	0,14	0,21	0,71	0,70	0,70	0,17	0,30	0,32	
8	74896653	74897392	8q13.1	+	0,95	1,70	10,69	6,68	6,05	6,77	5,41	8,82	9,19	12,02	6,80	6,07	7,12	5,64	7,76	7,24	7,94	7,26	
8	81814505	81813982	8q21.13	+	0,87	0,76	2,13	4,66	4,77	3,43	5,95	3,47	3,55	2,67	2,66	2,55	3,60	4,97	3,28	6,60	4,13	3,55	
8	125981187	125986649	8q24.13	+	0,36	1,49	0,47	0,35	1,11	0,50	1,78	0,45	0,07	0,58	1,38	2,23	1,84	1,00	1,55	0,21	1,77	0,89	
9	8737223	8738523	9q24.1	+	0,36	1,08	0,98	0,06	0,33	0,11	0,15	0,16	0,97	0,05	0,24	0,07	0,21	0,13	0,35	0,04	0,33	0,41	
9	22813028	22813538	9q21.3	+	0,36	0,57	2,05	2,02	1,44	2,94	1,39	1,04	0,29	3,69	3,29	1,74	1,80	1,04	1,53	2,07	1,49	1,88	
9	50744745	50749159	9q22.1	+	0,42	0,24	0,69				0,04	0,14	0,31	0,25	0,05	0,05	0,01	0,40	0,40	0,001	0,33	0,66	
9	93762661	93764180	9q22.31	+								0,17	0,26	0,17	0,26	0,01	0,21	0,21	0,21	0,21	0,02	0,28	
9	113138813	113146280	9q31.3	+		0,41		0,33	0,06			0,21	0,05	0,88	0,40	0,27	0,31	0,35	0,27	0,17	0,47	0,27	



**Anhang 13:** Ergebnis der gelelektrophoretischen Auftrennung der aus Hirnproben von MS-Patienten und gesunden Kontrollen amplifizierten HERV-W *env*-Amplikons. Die 280 bp und 330 bp großen Amplikons 5'*env* und 3'*env* wurden aus cDNA von vier MS-Patienten (MS) und vier Kontrollen (K) amplifiziert, die Produkte ausgeschnitten und aufgereinigt und anschließend mittels *MiSeq/Illumina*-Technologie sequenziert (vgl. Abschnitt 4.2.6. der Ergebnisse). Als Negativkontrolle diente für jede cDNA-Probe (+) der jeweilige RT(-) Ansatz ohne Enzym (-). K = PCR-Wasserkontrolle, M = 100 bp DNA-Leiter (Carl Roth, Karlsruhe, Deutschland).



**Anhang 14:** Übersicht über die mittels *MiSeq/Illumina*-Sequenzierung generierten HERV-W cDNA-Sequenzen und deren Filterung. Angegeben sind die Gesamtanzahlen der für alle untersuchten MS- und Kontroll-Proben (K) generierten cDNA-Sequenzen, die Anzahl an Sequenzen, die aufgrund ihrer zu kurzen Länge (<185 bp) oder der nicht eindeutigen Rückkartierbarkeit (>1 bester *BLAT-Match*) verworfen wurden, sowie eine detaillierte Übersicht, wie viele Sequenzen mit wie vielen Nukleotidunterschieden im menschlichen Referenzgenom rückkartiert werden konnten. Sequenzen mit mehr als drei Nukleotidunterschieden zu ihrem besten *BLAT-Match* und solche, die keine HERV-Sequenz repräsentierten, wurden verworfen (zur Filterung der Sequenzen, vgl. auch Material und Methoden, Abschnitt 3.2.3.2.2.). Die jeweilige Anzahl an Sequenzen, die mit weniger als drei Nukleotidunterschieden einem genomischen HERV-W-Locus zugeordnet werden konnten und damit in die weitere Analyse gingen, ist in der untersten Zeile angegeben.

	K4	K5	K6	K7	MS4	MS5	MS6	MS7
Gesamtanzahl generierter Sequenzen	1134146	877248	1028010	1124659	1171497	963202	1453740	760466
>1 bester <i>BLAT-Match</i> oder <i>Match</i> <185 bp	107212	81021	77806	119896	115276	80011	83238	67669
nur 1 bester <i>BLAT-Match</i>	1026934	796227	950204	1004763	1056221	883191	1370502	692797
in %	90,55	90,76	92,43	89,34	90,16	91,69	94,27	91,10
# Sequenzen mit x Nukleotidunterschieden								
0	337277	287659	373024	376830	382577	338915	469821	230189
1	236466	185520	236888	268266	269981	225526	308496	182478
2	131043	97960	116222	129837	141457	113209	170553	102415
3	79721	56578	63798	67121	77828	61625	100496	54668
4	53489	37318	39965	41188	48244	37754	65272	33087
5	46459	76784	30464	29512	33818	28029	49367	22755
>5	142479	99408	89843	92009	102316	78133	206497	67205
verworfen Sequenzen >3 nt-Unterschiede in %	23,61	25,38	16,87	16,19	17,46	16,30	23,43	17,76
# Sequenzen zur weiteren Analyse	784507	627717	789932	842054	871843	739275	1049366	569750
annotiert als HERV17-int und für weitere Analyse verwendet	<b>778893</b>	<b>621236</b>	<b>785625</b>	<b>836399</b>	<b>865593</b>	<b>710593</b>	<b>1044013</b>	<b>565905</b>
% der Gesamtanzahl generierter Sequenzen	68,68	70,82	76,42	74,37	73,89	73,77	71,82	74,42

**Anhang 15:** cDNA-Frequenzen und Genompositionen der in Hirnproben von MS-Patienten und gesunden Kontrollen mit dem 5'env-Amplikon als transkribiert identifizierten HERV-W-Loci. Das 5'env-Amplikon wurde aus cDNA von vier MS- und Kontroll-Proben (K) amplifiziert und für alle Proben mittels 454/FLX- sowie MiSeq/Illumina-Technologie Sequenzen generiert und rückkartiert (vgl. Abschnitt 4.2.6. der Ergebnisse). Angegeben sind in den ersten fünf Spalten für jeden Locus die hg18 Chromosomenposition und -bande (Kleinbuchstaben kennzeichnen Banden, in denen mehr als ein Locus als transkribiert identifiziert wurde) sowie der offizielle HGNC-Name (vgl. Mayer et al., 2011). Fett geschriebene Loci wurden sowohl mit dem 5'env- als auch mit dem 3'env-Amplikon als transkribiert identifiziert. Für alle als transkribiert identifizierten Loci sind die relativen cDNA-Frequenzen in [%] in den verschiedenen Proben angegeben. Diese berechneten sich für jeden Locus in jeder Probe aus der Anzahl an eindeutig dem Locus zugeordneten Sequenzen dividiert durch die Gesamtanzahl an eindeutig rückkartierbaren Sequenzen, die für die Probe generiert wurden. Die cDNA-Frequenzen der 454/FLX-Sequenzierung finden sich bereits in Tabelle 20. Unter der Tabelle findet sich für jede Probe die Gesamtzahl an eindeutig rückkartierbaren Sequenzen sowie die Anzahl an transkribierten Loci. In den rechten Spalten sind die Gene in der Nachbarschaft der HERV-W-Loci aufgeführt, bzw. die Gene, in denen einige Loci lokalisiert sind. Kursiv geschriebene Gene werden in Richtung des HERV-Locus transkribiert. Krankheitsrelevante OMM-Gene sind orange hervorgehoben. #Loci, die zuvor in PBMCs von MS-Patienten und gesunden Kontrollen als transkribiert beschrieben wurden (Laufer et al., 2009).

Chr	Start	HERV-W-Locus	MS4		MS5		MS6		MS7		K4		K5		K6		K7		Interessiert in Gen		Interessiert Gen			
			FLX	MiSeq	FLX	MiSeq	FLX	MiSeq	FLX	MiSeq	FLX	MiSeq	FLX	MiSeq	FLX	MiSeq	FLX	MiSeq	FLX	MiSeq	FLX	MiSeq	FLX	MiSeq
1	421182714	1378567	0,05	0,13	0,34	0,37	0,39	0,38	0,39	0,38	0,39	0,38	0,39	0,38	0,39	0,38	0,39	0,38	0,39	HLX	HLX	HLX	HLX	
1	55145269	55157816	0,10		0,05	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	HLX	HLX	HLX	HLX	
2	210095762	210098886	1,32	2,22	1,32	2,22	1,32	2,22	1,32	2,22	1,32	2,22	1,32	2,22	1,32	2,22	1,32	2,22	1,32	2,22	1,32	2,22	1,32	2,22
2	136675526	136677399	2,81	1,1	0,05	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	HLX	HLX	HLX	HLX	
3	17589116	17589503	2,81	1,1	0,05	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	HLX	HLX	HLX	HLX	
3	17589116	17589503	2,81	1,1	0,05	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	HLX	HLX	HLX	HLX	
3	97685908	97685978	3,24	1,1	0,05	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	HLX	HLX	HLX	HLX	
3	143021324	14302045	3,23	1,1	0,05	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	HLX	HLX	HLX	HLX	
4	74010156	7401767	4,13	0,29	0,15	0,15	0,15	0,15	0,15	0,15	0,15	0,15	0,15	0,15	0,15	0,15	0,15	0,15	0,15	HLX	HLX	HLX	HLX	
4	139762390	139767803	4,31	0,14	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12	HLX	HLX	HLX	HLX	
4	165796233	165796754	4,32	0,14	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12	0,12	HLX	HLX	HLX	HLX	
5	56851646	56851646	5,11	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	HLX	HLX	HLX	HLX	
5	107937705	107947182	5,21	0,01	0,11	0,11	0,11	0,11	0,11	0,11	0,11	0,11	0,11	0,11	0,11	0,11	0,11	0,11	0,11	HLX	HLX	HLX	HLX	
6	106782704	106790382	6,21	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	HLX	HLX	HLX	HLX	
6	91935248	91945442	7,21	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	HLX	HLX	HLX	HLX	
9	93782681	93784180	9,22	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	HLX	HLX	HLX	HLX	
9	93782681	93784180	9,22	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	0,34	HLX	HLX	HLX	HLX	
11	82259293	82259293	11,14	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	HLX	HLX	HLX	HLX	
11	22259293	22259293	11,14	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	HLX	HLX	HLX	HLX	
12	49582525	49583411	12,13	0,81	0,81	0,81	0,81	0,81	0,81	0,81	0,81	0,81	0,81	0,81	0,81	0,81	0,81	0,81	0,81	HLX	HLX	HLX	HLX	
14	44558438	44562647	14,23	2,73	2,73	2,73	2,73	2,73	2,73	2,73	2,73	2,73	2,73	2,73	2,73	2,73	2,73	2,73	2,73	HLX	HLX	HLX	HLX	
15	53384371	53391866	15,06	12,01	9,66	12,06	12,99	19,52	15,62	16,18	16,88	19,64	20,75	19,73	20,38	18,87	17,94	19,88	19,88	FAM179B	FAM179B	FAM179B	FAM179B	
17	32784002	32787735	17,15	0,76	1,01	0,43	2,41	1,15	1,15	1,15	1,58	1,93	0,59	1,85	3,05	6,68	2,20	5,09	5,09	ACACA	ACACA	ACACA	ACACA	
20	7627209	7627209	20,15	0,05	0,02	0,07	0,05	0,02	0,12	0,03	0,02	0,12	0,03	0,05	0,05	16,92	16,02	16,02	HLX	HLX	HLX	HLX		
X	106184757	106184757	106,18	9,56	12,07	13,53	12,97	14,40	3,02	6,87	6,87	6,87	6,87	6,87	6,87	6,87	6,87	6,87	6,87	HLX	HLX	HLX	HLX	
		Gesamtzahl an Sequenzen	2005	2989/24	1988	291/293	924	483/869	3/87	47/52	695	417/859	848	3263/76	476	4895/79	591	432/646	432/646	HLX	HLX	HLX	HLX	
		Anzahl verschiedener Loci	14	16	12	16	10	23	13	17	11	20	12	20	8	21	9	18	18	HLX	HLX	HLX	HLX	

**Anhang 16:** Genomlokalisierung der in Hirnproben von MS-Patienten und gesunden Kontrollen mit dem 3'env-Amplicon als transkribiert identifizierten HERV-W-Loci sowie mit den Loci überlappende RNA-Seq-Sequenzen. Das 3'env-Amplicon wurde aus cDNA von je sieben MS- und Kontrollproben (K) amplifiziert und für alle Proben mittels 454/FLX-Technologie sowie für je vier MS- und Kontrollproben mittels MiSeq/Illumina-Technologie Sequenzen generiert und rückkartierte (vgl. Abschnitt 4.2.4. und 4.2.6. der Ergebnisse). Angegeben sind in den ersten fünf Spalten für jeden Locus die hg18 Chromosomenposition und -bande (Kleinbuchstaben kennzeichnen Banden, in denen mehr als ein Locus als transkribiert identifiziert wurde), der Strang sowie der offizielle HGNC-Name (vgl. Mayer et al., 2011). Fett geschriebene Loci wurden sowohl mit dem 5'env- als auch mit dem 3'env-Amplicon als transkribiert identifiziert. In den mittleren drei Spalten sind die Gene in der Nachbarschaft der HERV-W-Loci aufgeführt, bzw. die Gene, in denen einige Loci lokalisiert sind. Kursiv geschriebene Gene werden in Richtung des HERV-Locus transkribiert. Krankheitsrelevante OMIM-Gene sind orange hervorgehoben. In den rechten Spalten ist die Anzahl an ENCODE RNA-Seq-Sequenzen aus sieben verschiedenen Zelllinien angegeben, die mit den hier identifizierten Loci überlappen (zur weiteren Erklärung, vgl. Abschnitt 4.2.8. der Ergebnisse). Unter der Tabelle findet sich die Gesamtanzahl an mit den HERV-W-Loci überlappenden RNA-Seq-Sequenzen jeder Zelllinie sowie die Anzahl verschiedener Loci. Die gesamte Tabelle erstreckt sich über drei Seiten.

Chr	Start	Ende	HERV-W-Locus		HGNC	Intronisch in Gen	nächstes Gen stromauf	nächstes Gen stromab	ENCODE Caltech RNA-Seq Datensatz						
			Banden	Strang					GM12878	H1 hESC	HUVEC	HeLa-S3	Hep G2	K562	NIHEK
1	46624040	46629585	1p33 <sup>a</sup>	ERWV-7	ERWV-7	-	MSUN4	FAAH (FAAH)	73	2	2	2	2	3	11
1	47190149	47197166	1p33 <sup>a</sup>	ERWV-7	ERWV-7	-	CYP4A11 (CYP4A11)	CYP4A1	2	2	2	2	2	3	5
1	51465234	51468757	1p32.3 <sup>a</sup>	ERWV-7	ERWV-7	-	C1orf185	RNF11 (RNF11)	2	2	2	2	2	2	4
1	55149269	55157786	1p32.3 <sup>a</sup>	ERWV-7	ERWV-7	-	DHCR24 (DHCR24, KIAA0018)	RNF11 (RNF11)	5	10	1	3	1	4	7
1	56021444	5602729	1p32.3 <sup>a</sup>	ERWV-8	ERWV-8	-	AK12270	PPAP2B (PPAKA2, AMPK)	16	16	1	3	1	2	3
1	89162799	89170152	1p22.2 <sup>a</sup>	ERWV-8	ERWV-8	-	GFZB (GFZB, TFIB, TFZB)	CCBL2 (DKFZ5ANI117 (CCBL2, KAT3))	4	15	9	15	15	6	7
1	91417276	91419286	1p22.2 <sup>a</sup>	ERWV-8	ERWV-8	-	ZNF644 (ZNF644)	HFM1	1	1	1	1	1	1	4
1	119512444	119515509	1p12	ERWV-8	ERWV-8	-	WARS3 (WARS3)	HAO2 (HAO2, HAOA2)	2	2	1	4	7	1	4
1	15365804	153663079	1q22	ERWV-9	ERWV-9	-	GN598521	BCO4L646	1	6	3	9	2	1	11
1	17637975	176378917	1q25.2	ERWV-9	ERWV-9	-	RASAL2 (RASAL2, NGAP)	TEX35	2	3	5	23	1	5	4
1	204102082	20411558	1q32.1	ERWV-10	ERWV-10	-	LOC284581	LOC284581	2	3	5	23	1	5	4
1	210095762	210098866	1q42.13	ERWV-11	ERWV-11	-	ZNF678	INTS7 (DKFZp434B168 (611350))	2	2	1	2	2	5	5
1	225978837	225987883	1q42.13	ERWV-11	ERWV-11	-	LPGAT1 (610473)	INTS7 (DKFZp434B168 (611350))	2	1	8	2	1	9	7
2	143698	150782	2p25.3	ERWV-11	ERWV-11	-	BCD39356	ZNF847	6	5	1	8	6	5	4
2	17383688	17391462	2p24.2	ERWV-11	ERWV-11	-	FAM150B	MEM18 (TMEM18)	5	9	1	8	6	5	4
2	30592322	30596634	2p23.1 <sup>a</sup>	ERWV-11	ERWV-11	-	FAM49A	RADE51AP2	6	5	1	1	4	4	10
2	31707721	31712731	2p23.1 <sup>a</sup>	ERWV-11	ERWV-11	-	L07266	CAPN3	2	3	2	3	3	2	2
2	33735092	33743026	2p22.3	ERWV-12	ERWV-12	-	SRD5A2 (607306)	MEMO1, C2orf4 (MEMO1, MEMO, C2orf4)	7	57	1	2	9	4	11
2	53837283	53842086	2p16.2	ERWV-12	ERWV-12	-	FAM98A	CHAC2 (ASB3)	39	28	26	14	76	20	15
2	7592323	75960132	2p12 <sup>a</sup>	ERWV-12	ERWV-12	-	REG3A (PAL)	LRR7M4 (LRR7M4)	20	20	2	1	1	91	1
2	79243804	79250987	2p12 <sup>a</sup>	ERWV-12	ERWV-12	-	LOC285033	CTNNA2	2	2	2	4	3	2	3
2	96246608	96252790	2q11.2	ERWV-13	ERWV-13	-	LOC285033	TMEM87B	2	2	2	4	3	2	3
2	106192101	106198564	2q11.2	ERWV-13	ERWV-13	-	LOC285033	TMEM127 (TMEM127)	3	4	1	3	2	6	11
2	112513393	112519037	2q13	ERWV-13	ERWV-13	-	LOC285033	PLGLA (PLGLA, PRGA)	2	2	2	4	3	2	3
2	136675926	136677999	2q22.1	ERWV-13	ERWV-13	-	LOC285033	TMEM87B	2	3	1	1	3	1	7
2	143372719	143374723	2q22.2	ERWV-13	ERWV-13	-	LOC285033	THSD7B	6	1	1	4	1	2	1
2	165222666	165224367	2q24.3	ERWV-13	ERWV-13	-	LOC285033	ARHGAP15 (610578)	1	1	1	1	1	2	8
2	172124792	172126863	2q31.1 <sup>a</sup>	ERWV-13	ERWV-13	-	LOC285033	COBLL1 (610318)	1	1	2	4	1	2	1
2	175898116	175899503	2q31.1 <sup>a</sup>	ERWV-13	ERWV-13	-	LOC285033	DYNCL2 (DYNCL2, DNC2, IC2)	1	3	2	2	1	1	8
2	179085383	179086517	2q31.2 <sup>a</sup>	ERWV-14	ERWV-14	-	LOC285033	KIAA1715 (LNP, KIAA1715)	1	3	2	2	1	1	1
2	18004648	18012291	2q31.3 <sup>a</sup>	ERWV-14	ERWV-14	-	LOC285033	TTCC9B	12	12	1	3	1	2	1
2	195907645	195913044	2q32.3	ERWV-14	ERWV-14	-	LOC285033	FBP7 (607062)	3	1	3	3	1	3	11
2	218163813	218165733	2q35	ERWV-14	ERWV-14	-	LOC285033	UBREX3 (UBREX3, UBCH9)	3	1	3	3	3	3	3
2	239567147	239569039	2q37.3	ERWV-14	ERWV-14	-	LOC285033	SLC9A10 (608733)	5	5	1	3	3	3	3
3	38305912	38313820	3p22.2	ERWV-14	ERWV-14	-	LOC285033	TNS1 (TNS1, TNS)	1	6	1	1	1	5	38
3	48347019	48352735	3p21.31	ERWV-14	ERWV-14	-	LOC285033	HDAC4 (HDAC4, HDACA, BDMR, AH03)	2	3	1	5	1	4	4
3	75004673	75009927	3p12.3	ERWV-14	ERWV-14	-	LOC285033	FBXW12 (FBXW12, FBW12, FBXO35, FBXO12)	1	3	2	1	1	1	5
3	97868008	97876978	3q11.2	ERWV-14	ERWV-14	-	LOC285033	FAM86DP	1	3	2	1	2	1	4
3	116946111	116942229	3q13.31	ERWV-14	ERWV-14	-	LOC285033	EPHA6 (600066)	1	3	2	1	2	1	1
3	132408394	132409640	3q22.1	ERWV-14	ERWV-14	-	LOC285033	LSAMP (LSAMP, LAMP)	1	3	2	1	2	1	1
3	137136880	137140440	3q22.2	ERWV-14	ERWV-14	-	LOC285033	LOC3087A (NEK11)	2	1	1	1	1	1	1
3	143021321	143023045	3q23 <sup>a</sup>	ERWV-3	ERWV-3	-	LOC285033	PPP2R2A (609844)	1	1	1	4	2	2	1
3	143639710	143645186	3q23 <sup>a</sup>	ERWV-3	ERWV-3	-	LOC285033	ATP1B3 (ATP1B3)	1	1	1	1	1	1	1
3	150965938	150969502	3q25.1	ERWV-3	ERWV-3	-	LOC285033	ATR (ATR, FRP1, SOKL1, FCTCS)	2	4	3	2	1	6	2
3	156178558	156179784	3q25.2	ERWV-3	ERWV-3	-	LOC285033	ANKUB1	2	2	2	1	1	2	1
3	163523160	163528850	3q26.1 <sup>a</sup>	ERWV-3	ERWV-3	-	LOC285033	MME	1	1	1	3	1	1	3

Anhang 16: Fortsetzung.

Chr	Start	HERV-W1-Locus	IGNC	Intronsch in Gen	nächstes Gen stromauf	nächstes Gen stromab	GM12878 I H1 RESCI HUVEC HeLa-S3 Hep G2	K652	NHEK
3	164901958	3q26.1*	ERVW-5	ZMAT3 (WIG1)	LOC647107	SI (S)	1	2	5
3	180250072	18026037	ERVW-5	ZMAT3 (WIG1)	KCNMB2 (605214)	PIK3CA (PIK3CA, CLOCE, MCAIP, MCM, MCMWCT, CMV) FGF12 (FGF12, FHF1)	5	3	5
3	192859286	192866075	ERVW-5	ZMAT3 (WIG1)	PYDC2	ZNF876P	2	7	1
4	175551	180863	ERVW-15	ZNF732	AXT46755	ZNF141 (ZNF141, D4S590)	4	4	1
4	8472981	8480392	ERVW-15	ACOX3 (ACOX3)	PCDH7 (PCDH7, BHPCDH)	CAOP23 (MIETTL19)	3	2	1
4	334219622	33421967	ERVW-15	ACOX3 (ACOX3)	DRHD1	AK093205	1	2	1
4	36116846	36122240	ERVW-15	ACOX3 (ACOX3)	IPHM3	KIAA1239	1	1	5
4	63522769	63526073	ERVW-15	ACOX3 (ACOX3)	IPHM3	TECR	2	3	1
4	74010156	74017767	ERVW-15	ACOX3 (ACOX3)	ADAMT53 (ADAMT53)	COX18, COX18HS (s10429)	3	1	3
4	83613493	83620230	ERVW-15	ACOX3 (ACOX3)	ENOPHI	TMEM150C	6	9	3
4	86492302	86495571	ERVW-15	ACOX3 (ACOX3)	CDORF12	ARHGAP24 (ARHGAP24, RCGAP72, FILGAP)	1	2	2
4	106628394	106629806	ERVW-15	ACOX3 (ACOX3)	EEFJAL7	FU20184	2	1	2
4	115184984	115192421	ERVW-15	ACOX3 (ACOX3)	ARSI (ARSI)	UGT8 (UGT8, CGT)	3	7	1
4	134022904	134029743	ERVW-15	ACOX3 (ACOX3)	BC041865	PCDH10 (PCDH10, KIAA1400)	2	1	3
4	139762390	139767903	ERVW-15	ACOX3 (ACOX3)	UNICOM499	CCRN4L (CCRN4L, CCR4L, CCR4)	3	1	3
4	153982430	153982932	ERVW-15	ACOX3 (ACOX3)	TIGD4	FHDC1	1	1	1
4	165796233	165798754	ERVW-15	ACOX3 (ACOX3)	MARCH1 (MARCH1)	LOC100505989	1	6	1
4	184157539	184162547	ERVW-15	ACOX3 (ACOX3)	AADAT (AADAT, KATZ)	LOC100506122	1	1	2
4	36364586	36368961	ERVW-15	ACOX3 (ACOX3)	RANBP1	SIC1A3 (SIC1A3, EAAT1, EA6)	1	2	2
4	44146946	44150936	ERVW-15	ACOX3 (ACOX3)	NMT (NMT, GCCD4)	FGF10 (FGF10)	1	2	1
4	56851606	56854473	ERVW-15	ACOX3 (ACOX3)	ACTBL2 (ACTBL2)	PLK2 (PLK2, SNK)	1	5	1
4	59990078	59998037	ERVW-15	ACOX3 (ACOX3)	PART1 (PART1)	ELOV1 (ELOV1)	10	2	3
4	87508343	87509456	ERVW-15	ACOX3 (ACOX3)	CONH (CONH, CAK)	TMEM161B, UNQ679	1	10	2
4	89125459	89126610	ERVW-15	ACOX3 (ACOX3)	MEF2C (MEF2C, CSDE14.3)	CETN3 (602907)	2	2	1
4	107937705	107942182	ERVW-15	ACOX3 (ACOX3)	FBXL7 (FBXL7, FBXL7, FBXL13, FBXL15)	FER (FER, TYK)	2	1	3
4	11814302	11814302	ERVW-15	ACOX3 (ACOX3)	FU11235	APC (APC, GS, FPC, BTSP2)	2	2	1
4	1275281	1275849	ERVW-15	ACOX3 (ACOX3)	FOXJ1 (FOXJ1, HPH1)	FOX2 (FOX2, FRHL6, FREA2)	1	3	2
4	13986545	13991934	ERVW-15	ACOX3 (ACOX3)	CCO90A	RNF182	3	2	1
4	52897632	52894671	ERVW-15	ACOX3 (ACOX3)	GSTA3 (GSTA3)	GSTA4 (GSTA4)	1	1	2
4	65771605	65775681	ERVW-15	ACOX3 (ACOX3)	PHF3 (PHF3, KIAA0244)	SIC2AS1P1	6	1	2
4	82103381	82110363	ERVW-15	ACOX3 (ACOX3)	BCKDHB (BCKDHB, E1B)	FAM66A	3	5	2
4	84210197	84216828	ERVW-15	ACOX3 (ACOX3)	ME1 (ME1)	PRSS35	2	2	1
4	85478138	85487587	ERVW-15	ACOX3 (ACOX3)	TXL18 (604613)	NTSE (NTSE, NTF)	3	8	13
4	85710654	85711747	ERVW-15	ACOX3 (ACOX3)	TXL18 (604613)	NTSE (NTSE, NTF)	3	1	5
4	89180630	89187945	ERVW-15	ACOX3 (ACOX3)	CNR1 (CNR1)	RNGTT (RNGTT, HCE1, CAP1A)	3	11	3
4	106782704	106790382	ERVW-15	ACOX3 (ACOX3)	PRDM1 (PRDM1, BLIMP1)	SOBP (SOBP, JXCL, MRAMS)	3	8	2
4	107227309	10722840	ERVW-15	ACOX3 (ACOX3)	BEND3	KIAA1919	3	1	1
4	111568913	111565669	ERVW-15	ACOX3 (ACOX3)	OSTM2P1	LOC100130476	3	1	2
4	134085532	134085532	ERVW-15	ACOX3 (ACOX3)	OLG3 (OLG3, BHLHB7)	ADATZ	1	1	5
4	143442630	143460542	ERVW-15	ACOX3 (ACOX3)	LOC100507489	STXL1 (STXL1, FH14, HPLH4, HLH4)	5	1	1
4	144479154	144484391	ERVW-15	ACOX3 (ACOX3)	SPB5	RNASE2 (RNASE2, RNASE6P1)	11	3	1
4	167224557	167231731	ERVW-15	ACOX3 (ACOX3)	RP56A2 (RP56A2, RSIG)	ETV1 (ETV1)	1	2	2
4	12859096	12859827	ERVW-15	ACOX3 (ACOX3)	ANKK4 (ANKK4)	ETV1 (ETV1)	1	2	3
4	16663219	16664043	ERVW-15	ACOX3 (ACOX3)	ANKMY2	TSPAN13 (TSPAN13, NET6)	10	11	1
4	35702273	35703152	ERVW-15	ACOX3 (ACOX3)	HERPUD2	CR595224	1	8	4
4	40173561	40179887	ERVW-15	ACOX3 (ACOX3)	MPLKIP (MPLKIP, C7orf11, ABHS, TTDN1)	INHBA (INHBA)	1	2	4
4	91955248	91954442	ERVW-15	ACOX3 (ACOX3)	GATAD1 (GATAD1, ODAG, CMO2B)	PEXI (PEXI, ZWS1, PBD1A, PBD1B)	14	3	11
4	113806378	113813611	ERVW-15	ACOX3 (ACOX3)	PPP1R3A (PPP1R3A, PPP1R3)	MDFC (MDFC, HIC)	3	1	1
4	118993435	118998119	ERVW-15	ACOX3 (ACOX3)	ANKRD7 (ANKRD7)	KCNQ2 (KCNQ2, KIAA1044)	1	1	9
4	121598944	121604985	ERVW-15	ACOX3 (ACOX3)	AA5 (AA5)	FEZ1 (FEZ1, FEZ, ZNF328)	46	1	2
4	131168140	131160223	ERVW-15	ACOX3 (ACOX3)	PODXL (PODXL, PCLP)	PLNM4 (PLNM4)	1	2	3
4	13320714	13328307	ERVW-15	ACOX3 (ACOX3)	AKR1B5	PRGM (PRGM)	1	1	2
4	14889620	149005245	ERVW-15	ACOX3 (ACOX3)	ZNF767	KRBA1	1	2	4
4	20066529	20061424	ERVW-15	ACOX3 (ACOX3)	LPZ (LPZ, LPD, HDLC01)	ATP6V1B2 (ATP6V1B2, ATP6B2, VPP3)	1	5	3
4	61494526	61500632	ERVW-15	ACOX3 (ACOX3)	CA8 (CA8, CALS, CARP, CAMRQ3)	RAB2A (RAB2A, RAB2A)	1	5	3
4	68930066	68943472	ERVW-15	ACOX3 (ACOX3)	BHLHE22 (BHLHE22, BHLHB5, BETA3)	LINC00251	1	8	2
4	68984751	68988420	ERVW-15	ACOX3 (ACOX3)	CPAG (CPAG, CPAH, ETL5, FEB11)	PREX2 (PREX2, DEPDC2)	2	2	1
4	74886653	74897392	ERVW-15	ACOX3 (ACOX3)	STAU2 (STAU2)	TCEB1 (TCEB1)	3	11	1
4	81814505	81817982	ERVW-15	ACOX3 (ACOX3)	ZBTB10	PAG1 (PAG)	3	3	4
4	825981187	825986949	ERVW-15	ACOX3 (ACOX3)	MTSS1 (MTSS1, MIM, KIAA0429)	LOC157381	3	11	1
4	87327223	8738523	ERVW-15	ACOX3 (ACOX3)	C9orf123	TYRP1 (TYRP1, CAS2, GP75, SHEP11)	3	3	3
4	22813028	22813838	ERVW-15	ACOX3 (ACOX3)	DMRTA1 (DMRTA1, DMRT4)	EVAL2 (EVAL2, HELNI, HUB)	1	1	1
4	35630301	35632824	ERVW-15	ACOX3 (ACOX3)	TESK1 (TESK1)	SFTI (SFTI)	8	23	5
4	90749445	90749169	ERVW-15	ACOX3 (ACOX3)	LOC286238	Gnor47, SUPR3 (SUPR3, EDG3)	1	2	1
4	93762861	93784190	ERVW-15	ACOX3 (ACOX3)	ROR2 (ROR2, ROR2, ROR2, ROR2)	SPTLC1 (SPTLC1, LBCL1, SPTL1, HSN1, HSNAN)	2	1	1
4	113138513	113140280	ERVW-15	ACOX3 (ACOX3)	ROR2 (ROR2, ROR2, ROR2, ROR2)	KIAA0368	2	1	1

Anhang 16: Fortsetzung.

Chr	Start	Ende	HERV-WL-Locus	HGNC	intronisch in Gen	inächstes Gen stromauf	inächstes Gen stromab	ENCODE Calltech RNA-Seq Datensatz							
								GM12878	H1 hESC	HUVEC	HeLa-S3	Rep. G2	K562	NHEK	
10	23631096	10p12.2				PITFA (PITFA)	C10orf67	3	3	3	1	1	2	2	4
10	62463466	10q21.2				RHOBTB1 (RHOBTB1, KIAA0740)	TMEM236	2	4	4	1	1	1	1	4
10	65474511	10q21.3				REF3 (REF3, C10orf74)	CTNNA3 (CTNNA3)	3	3	3	3	1	4	1	4
10	86275106	10q23.33*				C5orf2	CP22C8 (CP22C8)	2	1	1	2	2	2	2	4
10	96584315	10q23.33*				CYP2C19 (CYP2C19, CYP2C19)	CP22C8 (CP22C8)	1	4	4	2	2	2	2	1
10	97447105	10q23.33*				ENTPD1 (ENTPD1, CD39)	TCTN3 (TCTN3, TECT3, C10orf61, ORFD4, IBT518)	1	4	4	1	2	2	2	1
11	9325928	11p15.4				ANO3 (ANO3, TMEM16C, C11orf75, DYT24)	MEM41B	1	4	4	1	2	2	1	1
11	26586553	11p14.2				AAADC	MUC15 (MUC15, ANO3)	1	3	3	6	28	12	1	1
11	38579714	11p12				RSF1 (RSF1, RNF101)	SLC5A12 (SLC5A12, SMC72)	1	7	7	1	6	28	1	1
11	77247020	11q14.2				RSF1 (RSF1, RNF101)	LRRGC (LRRGC, NGL1, KIAA1580)	1	3	3	1	3	3	1	1
11	86222185	11q14.2				PRSS23	INTS4 (INTS4)	6	5	5	15	24	13	17	16
11	107357020	11q23.3				IL13OR1	ORZEP2	1	1	1	13	1	1	3	16
11	117412710	11q23.3				TMPRSS4-AS1	TMPRSS4 (TMPRSS4)	3	8	8	1	1	1	4	4
11	117413759	11q23.3				IL13OR1	CUL5 (CUL5, VACM1)	3	8	8	1	1	1	1	1
12	7226009	12p13.31*				RAB39	PEXS (PEXS, PXR1, PTS1R, PBD2A, PBD2B)	2	11	11	1	4	2	1	1
12	8605276	12p13.31*				RIMKB (RIMKB, NAAGE)	AZMLL (AZMLL)	2	11	11	1	2	1	1	1
12	34144322	12p11.1				ALG10 (ALG10, KCR1)	ALG10B	1	1	1	1	1	1	1	1
12	37151649	12q12				ALG10B	CPNE8	1	1	1	1	1	1	1	1
12	37316569	12q12				ALG10B	CPNE8	1	1	1	1	1	1	1	1
12	39063704	12q12				LRRK2 (LRRK2, PARK8)	CPNE9	1	1	1	1	1	1	1	1
12	49582525	12q13.13				TMPRSS12	CPNE9	1	1	1	1	1	1	1	1
12	49582525	12q13.13				AKO93124	METTL7A	2	5	5	2	4	4	5	9
12	57531913	12q14.1				MIRS48T	LRRG3 (LRRG3, LIG3)	1	1	1	3	1	4	7	9
12	83711881	12q21.31				SILCA1A2 (SILCA1A2)	SLCGA15 (SLCGA15, SBAT1)	1	1	1	3	1	4	7	9
12	103861162	12q23.3				RPLP1 (RPLP1, RPL1, GOSPEL)	CL2orf45	1	1	1	1	1	1	1	1
12	122598917	12q24.33				MMP17 (MMP17)	TMED2	2	10	10	2	2	3	2	9
12	130823726	12q24.33				ALG5 (ALG5E)	UK1 (UK1, LINC51)	7	5	5	3	1	5	1	6
12	36428786	13q13.3				TCRA	EXOSC8 (EXOSC8, OIP2, RRP43)	7	5	5	3	1	5	1	6
13	54525766	13q21.1				STXBP6 (STXBP6, AMISYN, HSPCL56)	PRK20B	1	185	185	1	2	2	5	5
14	21741271	14q11.2				DDHD1 (DDHD1, PAPLAL, KIAA1705, SPQ28)	DAD1	1	1	1	1	2	4	6	5
14	25798768	14q12				DDHD1 (DDHD1, PAPLAL, KIAA1705, SPQ28)	NOVA1 (NOVA1)	1	1	1	1	1	1	1	1
14	52898110	14q22.2				C14orf159	BMP4 (BMP4, BMP2B1, BMP2B, MCOF56, OFC11)	1	14	14	1	1	2	1	1
14	57668751	14q23.1				SLC35F4	ACTR10	1	14	14	1	1	2	2	1
14	90761761	14q23.12				ERWW-27	GRP68 (GRP68, OGR1)	2	5	5	1	1	2	1	4
15	53384371	15q21.3				ERWW-4	PIGB (PIGB)	16	17	17	7	5	10	14	15
15	90642380	15q26.1				SLEFN14 (SLEFN14)	ST8SIA2 (STX)	1	4	4	1	1	2	3	3
17	30902419	17q12				ACACA (ACACA, ACAC, ACC1, ACACAD)	SNORD7	13	13	13	1	1	1	2	2
17	32767735	17q12				STXBP4 (STXBP4, G10415)	C17orf78 (ACACA, ACAC, ACC1, ACACAD)	1	1	1	1	1	1	2	2
17	50443884	17q22				ERWW-28	HLF (HLF)	4	6	6	6	1	5	2	9
18	4671679	18p11.31				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
18	13646680	18p11.31				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
18	56528820	18q21.32				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
18	58301303	18q21.33				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
19	20073271	19p12				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
19	21603630	19p12				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
19	22720178	19p12				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
19	23922792	19p12				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
19	45269687	19q13.2				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
19	46183987	19q13.2				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
20	53388912	20q13.2				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
21	19046932	21q21.1				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
21	27148626	21q21.1				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
21	40033060	21q22.2				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
22	32675530	22q12.3				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
X	7627209	Xp22.31				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
X	5741716	Xp11.3				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
X	75034856	Xq13.3				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
X	79080405	Xq21.1				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
X	10513459	Xq22.3				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
X	106182016	Xq22.3b				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
X	138967420	Xq27.1				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9
Y	19701211	Yq11.222				ERWW-29	HLF (HLF)	4	6	6	6	1	5	2	9

**Anhang 17:** Sequenzen der teilweise stark verkürzten HERV-W-LTRs, deren potentielle Promotoraktivität mittels Luciferase-Assay getestet wurde (vgl. Abschnitt 4.2.10. der Ergebnisse). Die LTR-Sequenzen wurden im Rahmen von Richter (2012) amplifiziert und stromauf des Luciferase-Gens in den *pGLuc-Basic*-Vektor kloniert. Die Sequenz des LTRs ist jeweils fett markiert, die interne provirale Sequenz ist grau unterlegt. Die Positionen der zur Amplifikation der LTR-Sequenzen verwendeten *Primer* sind blau markiert.

>1p32.3<sup>b</sup>\_5'-LTR

aatgacatcgaggcaatcctaccaaggaaatctcaactgcaccaccctactatgcccgaattcagcaggagcagcttagagggttgcggccaacc  
tcccaacagcacttgggttttctgttaaagagggggaactgagagacaggactagctggatttccaaagccaactaagaatccctagctagctg  
ggaaggtgaccgcatccacctttaaactggggcttgcacttagctcacaccaaccaatcaagtagtaaaagagagctcactaaaatgctaaatag  
gcaaaaacaggaggtaaagaatagccaatcatctacactgagagcacagcgggaggacaatgatcgggatataaaccaggcattcgagctga  
caatggctaccctcttgggtcccctcccttggatgggagctctgtttcactctataaatcttgaactgcaattctctggctccctgtttgtt  
acagctggagctgagctttcgctcgctgtgaccactgtgtttgcccctcacagaccgcaactgactttcctccaaatcctggcagggtg  
tccactgtgctcctgacgaaggcaccattgcccctctgctcctgatcgggctaaacgcttggcattgttctgcatggctgcccgggttcac  
ctaactgagctgaactagctactgggttccacggttctcttccatgaccacggcttctaatagagctataaactcaccgcatggcctaaagt  
ccatcccttggaaatccgtgaggccaagaaccccaggctcagagaacacaggcttgcaccattttggaagttagcccaccatctgggagctctg  
ggagaaacgacccccggtaacacccagctattgtgaagacttaagtaagaacatattggaacattcctaggcgtgtggcagctgataactaacg  
tgaataagtgccagcttggctgattaataataaaaaggcatccagaccaggctgacaggaggacaggagatc

>2p16.2 (ERVW-12) 5'-LTR

gaagatgtttcaggaagacattaccatttgaatcagcagactgaataaacagacttgcctccacaaatggggcaggcatcatgtaatccttcaag  
gacctaaagaaaaaaaggcagggaaccccccttgggtcccctccattgtatgggagctctgttttccactctataaatcttgcactgcacact  
ctctggctccgtgtttgtccagttcaactgagcttccactcaccctccaccactgctgatcaccgtctgagcagccgctgattgacttccacc  
tccgatctggcagggtgtcactgctctctgatccagcagggcggcccttgcactccagctcgggttagagctcaccattgttctctgcaca  
ggctaagtgccctgggttctacttaactgagctgaacagagctgtaacactcaccgcatgaccaagattccattccttggaaatccgtgaggtaag  
aacccgaggtcagagaacagaggcttgcaccatcttggaaagcagctaccaccatcttaggagctctaagaacaaggacccccagtaacattt  
ggtagccacaaaaggcactcaaaagcggtagtaaatattggaccactttcacttgcattctctgctctatcctt

>2q13 (ERVW-13) 5'-LTR

ccctagttcagaaggaagcagtttagagcagctgctgcccgcctccccaaagcacttgggttttctctgtttagagaggggactgagagacaggact  
agctgcttcttagcgggtaagaatccctaagcctagctgggaaggtgactgtatccactttaaaccagggcttgcacttagctcacacc  
gaccaatcaggtagtaaaagagagctcactaaaatgtaattaggcaaaaacaggaggtaaagaatagccaatctatctacactgagagtaacaggg  
gaggacaatgattgggataaaccaggcattcgagcggcaacggcaaccccccttgggtccttccattttatgggagctctgtttcactc  
tattaaatcttgcactgcaaaaaaaaaaaaaaaaaaaagcgttgacataaaaaatcttcaactacaatactataaccagcagaaatattctt  
ctaagcaaaaggagacagtgtgtaacaaaatggcagactagaaaactcgaaggattgttcttccacttcagaacattgaaaaaca

>6p25.3 3'-LTR

ccccattcagcaggaagcagtttagagcggctgctgcccacactccccaaacagcacttgggttttctctgtttagagagggggatgagagacaggact  
agctggatttcttagctgactaagaatccctaagcctagctgggaaggtgactgttccactttaaacaatggggcttgcacttagctcacacc  
aaccaatcggatagtaaaagagagctcactaaaatgtaattaggcaaaaacaggaggtaaagaatagccaatctatctatgacctgagagcacagg  
ggagagacaatgattgggataaaccaggcattttagcggcaacagcaaccccccttgggtcccctcccttggatgggagctctgtctcact  
ctattaaatcttgcactgcaaaaaaaaaaagaaagaaatgtagacatccattattgaaatacataaaggaactcaagcaactcaataacaagaaaac  
aataaccaatttaaaatgggcaaaagacatgaatagacatg

>6q21<sup>a</sup> (ERVW-17) 3'-LTR

ccccattcgcaggaagcagtttagactggtcgtcagccaacactccccaaagcacttgggttttctctgtttagaggggggactgagagacaggatt  
agctggatttcttagcggactaagaatcccaaagcctagctgggaaggtgaccacatccactttaaactgggcttgcacttagctcacacc  
gaccaatcaggtagtaaaagagagctcactaaaatgtaattaggcaaaaacaggaggtaaagaatagccaatctatctatgacctgagagcacagg  
ggaaggacaatgactgggataaaccaggcattcaagccggcaacggctaccttcttgggtcccctcccttggatgggagctctctctgtctt  
cactctattaaatattgcaactgcaaaaaaaaaaagcttaattgaaataaattaatacaataaaaggaatacattttaagatatacagttcaaac  
tgtaacagtggtacagttcaagaggaccccttcaacaagaattattggcatttccatcatgcccctaaaagtctcttctgtccctactggttgggt  
ccatctctactaca

>7q21.2 (ERVW-1) 5'-LTR

acatagacagaagtccctggggcgggttctcttctgggatgagggcaaaaagcctggagatacagcaattatcttgaactgagagacaggactag  
ctgatttcttagcggcactaagaatccctaagcctagctgggaaggtgaccacgtccactttaaaccagggcttgcacttagctcacaccctga  
ccaatcagagagctcactaaaatgtaattaggcaaaagacagggttaaagaatagccaatctatctatgacctgagagcacagcaggaggacaac  
aatcgggatataaaccaggcattcgagctggcaacagcagcccccttgggtcccctcccttggatgggagctgtttcatgctatttccactct  
atataatcttgaactgcactcttctggctcatgtttcttaccggctcgagctgagcttttctcaccgtccaccactgctgtttgcccaccaggcag  
acctgcccgtgactcccctccctctggatcctgcagggtgctcctgtgctcctgatccagcagggcggccattgcccctcccaattgggctaaagg  
cttgccattgttctgacaggctaaagtgcctgggttgttctaattgagctgaactagctcactgggttccatggttctctctgtgacccaggc  
ttctaataagaactataaacttaccacatggccaagattccattccttggaaatccgtgagccaagaactccaggtcagagaatacagaggcttgc  
accatcttgaagggcctgctaccaatcttgaagtggttaccaccatcttgggagctctgtgagcaaggacccccggtaacaatttggcaacca  
cgaacggacatcgaagtggtgagtaaatattggaccactttcacttct  
tcggccagttaaaacgattagtggtgcccaccggacttaagactcaggt

>13q13.3 (ERVW-25) 5'-LTR

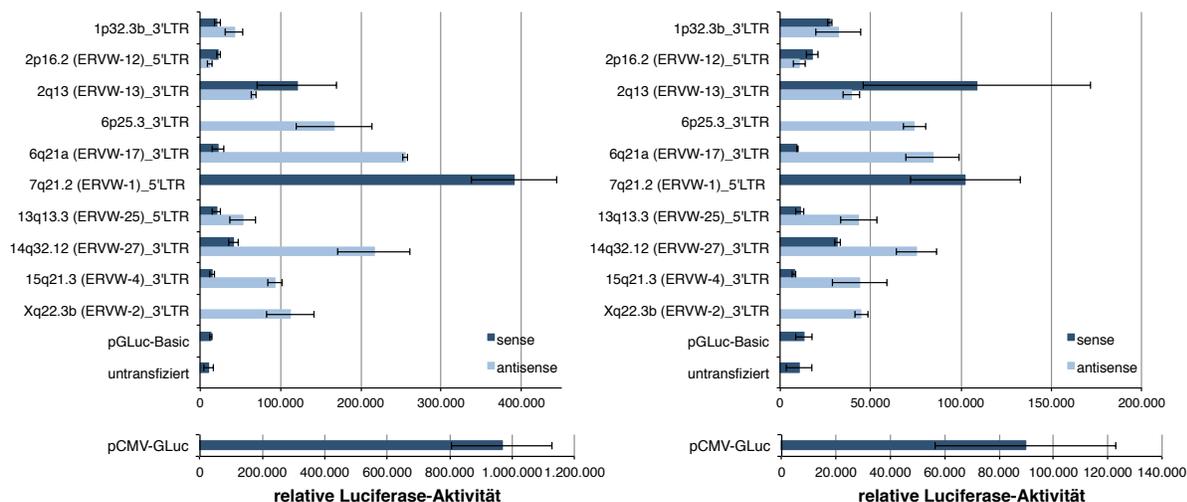
gttcacaacagccaagtatccaatcaacctaaaggctctgtcaacagaaaaactgggaaggaaaatgtggatattgtacacaatgaaatattattca  
gccttgggcaacctcttgggtcccctccccttggatgggaactctgtttcactctaatacatcttgaactgcacactcctctggctccgtgttt  
gttaccgctcagctgaagctttaaaggtggcagctcttaactgctgggatcatcagaaggaaagaaagaaagaaatagaagggaaccgcccagcag  
atattgaagcca

```
>14q32.12 (ERVW-27) 3'-LTR
aggaagcagtttagagcagttgtcagccaacctccccaacagcacttgggttttctctgttgagaggggtaetgagagacaggactagctggatttcc
tagggccaactaagaatccctaagcctagctgggaaggtgaccacatccacctttaaacaacagggcttgcaacttagctcacacccgaccaatcagag
agctcactaaaatgctaattagggcaaaaacaggaggtaaagaatagccaatcatctatcgctgagagcacagcaggagggaacaatgatcgggata
taaaccaggcattcgagccagcaatggctaccctctttgggtccctccctttgtatgggagctcagtttctactctattaaatcttgcaactgca
aaaacaacaacaaaaaaagactttcaccttttcaactaaaggaagccctttatggcctctttttggcatatccaatgcaagcactctctgct
cttgacctttgagga
```

```
>15q21.3 (ERVW-4) 3'-LTR
cccagttcagcaggaagcaggttagaattggttgcggccaacctccccaaaagcacttgggttttctctgttgagtagggactgagagacagggttagc
tggatttcttaggctgactaagcatccctaagcctagctgggaaggtgactgcatccacctttaaagcatggggcttgcaacttagctcacacctgac
caatcaggtagtaaagagatctcactaaaatgctaattagggcaaaaacaggaggaaagaatagccaatcatctatcgctgagagcacagtgaggga
aggacaatgattgggataataaactcaggcattccagccagcaactgctaccctctttgggtcccctccctttgtatgggagctccgttttctactcta
tgaatcttgcaactgcaaaaaacaaacaaacaaatacacacacacacatatataatataatataatggtcacactaaagctggtaaatgaatgttc
atcacatattcataatagccaaaaagtggaaaacacccaagtgtccattaactgataaggaataaataagtggtatttccgtataatggaatatt
attcagcca
```

```
>Xq22.3b (ERVW-2) 3'-LTR
ccccatttcagcaggaagcaggttagagcggtcagtcagcaacctccccagcagcacttgggttttctctgttgagaggggtaetgagagacaggact
agctggatttcttaggctgactaagaatccctaagcctagctgggaaggtgaccgctccacctttaaacaacggggcttgcaacttagctcacaccc
gaccaatcagagagctcactaaaatgctaattagcaaaaacaggaggtaaagaatagccagtcactatcgctgagagcacagcaagagggaca
atgatcgggataataaaccaggcattcgagccagcaacggctaccctctttgggtcccctccctttgtatgggagccctgttttctactctataaat
cttgcaactgcaaaaaataaataaataaagctagatgggtatggtttagttatgtgcattgtccacatttttaaaaaagcaaaccttcttata
gtctacaagccctataacagatctggctattgaccacctctctgacctctctactactctgccttacacatacagcagacatattggcat
```

**Anhang 18:** Promotoraktivität von 5'- und 3'-LTRs von als transkribiert identifizierten HERV-W-Loci. Zur Erklärung s. Text zu Abbildung 40. Gezeigt sind zwei Wiederholungen des in Abbildung 40 dargestellten Experimentes, die ein vergleichbares Bild lieferten. Man beachte, dass im dritten Experiment die relative Luciferase-Aktivität, verglichen mit den LTR-Konstrukten, deutlich niedrigere Werte zeigte als in den vorangehenden Experimenten.



## 8. Publikationen

### Journal-Beiträge

Schmitt K, Rösch A, Reichrath J, Meese E, Mayer J (2013) **Transcriptional profiling of human endogenous retrovirus group HERV-K(HML-2) loci in melanoma.** *Genome Biol Evol* 5(2):307-28.

Leidinger P, Backes C, Deutscher S, Schmitt K, Muller SC, Frese K, Haas J, Ruprecht K, Paul F, Stahler C, Lang CJ, Meder B, Bartfai T, Meese E, Keller A **A blood based 12-miRNA signature of Alzheimer disease patients.** *Genome Biol.* 2013 Jul 29;14(7):R78.

Schmitt K, Richter C, Backes C, Meese E, Ruprecht K, Mayer J (2013) **Comprehensive Analysis of Human Endogenous Retrovirus Group HERV-W Locus Transcription in Multiple Sclerosis Brain Lesions by High-Throughput Amplicon Sequencing.** *J Virol* 87(24):13837-52

Harz C, Ludwig N, Lang S, Werner TV, Galata V, Backes C, Schmitt K, Nickels R, Krause E, Jung M, Rettig J, Menger M, Zimmermann R, Meese E. **Preferential secretion of meningioma-associated antigen TXNDC16 from tumor cells may explain autoantibodies in meningioma patients.** *Manuskript in Revision*

### Posterpräsentationen

*International Congress on Transposable Elements*, 21.-24.04.2012, St. Malo, Frankreich:

Schmitt K, Rösch A, Reichrath J, Meese E, Mayer J. **Transcriptional profiles of loci of the human endogenous retrovirus group HERV-K(HML-2) in melanoma.**

Schmitt K, Meese E, Ruprecht K, Mayer J. **Comprehensive analysis of human endogenous retrovirus W (HERV-W) transcription patterns in multiple sclerosis lesions and control brain tissue using next generation sequencing.**

*23<sup>rd</sup> Annual Meeting of the Society for Virology*, 06.-09.03.2013, Kiel, Deutschland:

Schmitt K, Rösch A, Reichrath J, Meese E, Mayer J. **Transcriptional profiling of human endogenous retrovirus group HERV-K(HML-2) loci in melanoma.**

Schmitt K, Meese E, Ruprecht K, Mayer J. **Comprehensive analysis of human endogenous retrovirus W (HERV-W) transcription patterns in multiple sclerosis lesions and control brain tissue using next generation sequencing.**

## Danksagung

An dieser Stelle möchte ich mich zunächst ganz herzlich bei meinem Doktorvater, Herrn apl.-Prof. Dr. Jens Mayer, für die Bereitstellung des interessanten Promotionsthemas, die Betreuung der Arbeit sowie seine Unterstützung in den letzten drei Jahren bedanken. Seine zahlreichen Hilfestellungen sowie die Anmerkungen bei der Durchsicht der Arbeit haben maßgeblich zum Gelingen beigetragen.

Bei Herrn Prof. Dr. Friedrich Grässer bedanke ich mich herzlich für die Übernahme des Zweitgutachtens meiner Dissertation.

Ein großer Dank geht an Herrn Prof. Dr. Eckart Meese für die freundliche Aufnahme in seine Arbeitsgruppe sowie an meine Kollegen für die angenehme Arbeitsatmosphäre und die Hilfsbereitschaft bei allen auftretenden Problemen und Fragen. Ein besonderer Dank gilt hierbei Dr. Christina Backes für die bioinformatische Unterstützung bei der Rückkartierung meiner NGS-Daten und die statistische Auswertung sowie an meine Bachelor-Studentinnen Christin Richter und Jana Biermann für die Klonierung diverser Vektorkonstrukte, die ich im Rahmen der Arbeit weiter verwenden konnte. Danke auch an Esther Maldener für die nette Bürogesellschaft und viele aufmunternde Gespräche.

Ein großer Dank geht an Alexandra Stark und Heike Palm (Klinik für Dermatologie, Venerologie und Allergologie, Uniklinik Homburg), die mir bei der Zellkultur sowie UV-Bestrahlung der Melanom- und Melanozyten-Zelllinien eine große Hilfe waren.

Danke an Herrn apl.-Prof. Dr. Klaus Roemer (José Carreras Zentrum für Immun- und Gentherapie, Uniklinik Homburg) und Dr. Kristina Heyne, die mir die *rec/np9*-Klone zur Verfügung gestellt haben.

Ein Dank auch an Herrn Prof. Dr. Jörn Walter (Genetik/Epigenetik, Universität Saarbrücken), in dessen Arbeitsgruppe ich die *454/FLX*-Sequenzierungen durchführen konnte. Insbesondere bedanke ich mich hierbei ganz herzlich bei Jasmin Gries für die Durchführung der Experimente und die nette Gesellschaft an den langen Tagen, an denen ich hierbei helfen konnte. Danke auch an M.Sc. Pavlo Lutsik für die Filterung meiner Sequenzen.

Ein ganz großer Dank geht an meinen wunderbaren Bürokollegen, Schreibtischnachbarn und Rocky-Horror-Show-Partner ;-)) Christian Harz für die schönen drei Jahre und den Spaß im und außerhalb des Labors. Vielen Dank auch für das sehr kurzfristige Korrekturlesen der Arbeit.

Danke auch an meinen guten Freund und Kletter-Partner Timo Schumann für die nötige sportliche Ablenkung vom Schreiben und einige hilfreiche Kommentare zur Arbeit.

Ein großer Dank geht an meine Eltern, die mich immer unterstützt und an mich geglaubt haben und auch meine schlechte Laune in der Schreibphase geduldig ertragen haben. ;-)) Ein besonderer Dank auch an meine Schwester Anja für das Korrekturlesen der Arbeit.

Der größte und herzlichste Dank geht jedoch an meinen treuen Vierbeiner Floyd (1998-2013), der mich durch Schulzeit, Studium und einen Großteil der Doktorarbeit begleitet hat, und jeden Tag für die mehr als nötige Ablenkung, Aufmunterung und Bewegung an der frischen Luft gesorgt hat. Du fehlst, Großer...