From the Department of Neuroscience,
Karolinska Institutet, Stockholm, Sweden

CHARACTERIZATION OF TRANSCRIPTION OF GENOMIC REGIONS HARBORING HERV-W ELEMENTS

Fang Li

Stockholm 2014

Published by Karolinska Institutet. Printed by Åtta.45 Tryckeri AB

ISBN 978-91-7549-569-9

Karolinska Institutet

Institutionen för neurovetenskap

Characterization of transcription of genomic regions harboring HERV-W elements

AKADEMISK AVHANDLING
som för avläggande av medicine doktorsexamen vid Karolinska Institutet offentligen försvaras i Hillarp salen, Retzius väg 8

Fredag den 26 September, 2014, kl 9.00

av
Fang Li
MSc

Huvudhandledare:
Håkan Karlsson, PhD
Karolinska Institutet
Institutionen för Neurovetenskap

Bihandledare:
Prof. Robert H Yolken, MD, PhD
Johns Hopkins University School of Medicine, USA
Department of Pediatrics

Prof. Kristensson Krister, MD, PhD
Karolinska Institutet
Institutionen för Neurovetenskap

Christoffer Nellåker, PhD
Oxford University, UK
Department of Physiology, Anatomy & Genetics

Fakultetsopponent:
Ass. Prof. François Mallet
Lyon University
Joint Unit Hospices Civils de Lyon-bioMérieux

Betygsnämnd:
Ass. Prof. Torbjörn Ramqvist
Karolinska Institutet
Institutionen för onkologi-patologi

Ass. Prof. Afshin Ahmadian
Kungliga Tekniska högskolan (KTH)
Skolan för bioteknologi

Prof. Tomas Ekström
Karolinska Institutet
Institutionen för klinisk neurovetenskap

Stockholm 2014

ABSTRACT

Human endogenous retroviruses (HERV) comprise 8% of the human genome and can be classified into at least 31 families. A typical HERV provirus consists of internal *gag*, *pol* and *env* genes, flanked by two long terminal repeats (LTRs). HERV are by nature repetitive and have with few notable exceptions lost their protein-coding capacity. Therefore, HERV are consistently not included in large scale expression studies and hence little is known of their expression, regulation and potential functional significance. Aberrant expression of HERV-W has been associated with human diseases, such as multiple sclerosis and schizophrenia. It has been reported that HERV-W elements, including *ERVWE1* (the so far only known HERV-W gene functionally adopted by the human host), can be transactivated in a range of human non-placental cell-lines during influenza virus infections.

Applying a recently developed technique for obtaining high resolution melting temperature analyses, transcripts containing HERV-W *gag* sequences were found to be expressed in non-random patterns with extensive variations in the expression between tissues, brain regions and individuals. Furthermore, the patterns of such transcripts varied more between individuals in brain regions than other tissues (I). To determine the effect of genomic context, viral structure and orientation on the transcription of HERV-W, PCR directed at specific HERV-W loci were employed on panel of normal human tissues. HERV-W elements in intronic regions were found to be expressed at higher levels than elements in intergenic regions. With regard to intronic elements, proviruses were expressed at higher levels than pseudoelements or solo LTRs. Relative to their corresponding genes, intronic elements integrated on the sense strand appeared to be transcribed at higher levels than those integrated on the anti-sense strand. Furthermore, the expression of transcripts containing intronic proviral elements appeared to be independent from that of their corresponding genes (II). When addressing mechanisms underlying transactivation of HERV-W following virus infection we detected up-regulation of spliced *ERVWE1* transcripts and those encoding the transcription factor glial cells missing 1 (GCM1) which acts as an enhancer element upstream of *ERVWE1*. Knock-down of GCM1 by siRNA, followed by infection suppressed the transactivation of *ERVWE1*. In addition, ChIP assays detected decreased H3K9 trimethylation and histone methytransferase SETDB1 levels along with influenza viral proteins associated with *ERVWE1* and other HERV-W loci in infected CCF-STTG1 cells (III). Finally, by analyzing publicly available RNA sequencing datasets generated from three different regions of human brains of multiple individuals, a consistent expression (0.1-0.2% of mappable reads) of HERV families was observed across three regions of brains. Spearman correlations between tissues revealed highly correlated expression levels across 475 consensus sequences. By mapping sequences aligned to the consensus sequences of HERV-W and HERV-H families to individual loci on chromosome 7, more than 60 loci from each family were identified, some of which are being transcribed. Elevated expressions of overall HERV, as well as of HERV-W family were observed in samples from both schizophrenia and bipolar disorder patients (IV).

In conclusion, our studies show that 1) HERV-W gag transcripts appear to exhibit a highly diversified expression pattern across both tissues and individuals; 2) Both LTR directed and leaky transcription of HERV-W elements contribute to their tissue-specific expression pattern; 3) Chromatin modifications potentially mediate the effect of influenza A virus infection on HERV-W expression; 4) An independent method of RNA sequencing verifies expression of HERV in human brain regions.

LIST OF PUBLICATIONS

I. Nellåker C, Li F, Uhrzander F, Tyrcha J, Karlsson H. Expression profiling of repetitive elements by melting temperature analysis: variation in HERV-W gag expression across human individuals and tissues. *BMC genomic*. 2009 Nov 17; 10:532

II. Li F, Nellåker C, Yolken RH, Karlsson H. A systematic evaluation of expression of HERV-W element; influence of genomic context, viral structure and orientation. *BMC genomic*. 2011 Jan 12; 12:12

III. Li F, Nellåker C, Sabunciyan S, Yolken RH, Jones-Brando L, Johansson AS, Owe-Larsson B, Karlsson H. Transcriptional de-repression of ERVWE1 locus following influenza A virus infection. *J Virol*. 2014 Apr 88(8); 4328-4337 [Epub 2014 Jan 29]

IV. Li F, Sabunciyan S, Yolken RH, Hwang Y, Kim J, Lee D, Kim S, Karlsson H. Transcriptional expression of HERV in human brain by RNA-sequencing. Manuscript

CONTENTS

LIST OF ABBREVIATIONS

ChIP	chromatin immunoprecipitation
GCM1	glial cell missing 1
HERV	human endogenous retrovirus
H3K9me3	tri-methylation of histone H3 lysine 9
LTR	long terminal repeat
MS	multiple sclerosis
ORF	open reading frames
PBMC	peripheral blood mononuclear cell
qPCR	quantitative polymerase chase reaction
RT	reverse transcriptase
siRNA	small interfering RNA
Tm	melting temperature
TSE	trophoblast specific enhancer

1 INTRODUCTION

1.1 HUMAN ENDOGENOUS RETROVIRUSES

The human genome has a total of over 3 billion DNA base-pairs. Whereas only 2% of the genome is composed of protein-coding regions, approximately 45% of our DNA is recognized as containing transposable elements which have by and large been considered as "junk" (Lander et al., 2001; Venter et al., 2001). These elements consist of four types, i.e. long interspersed elements (LINE), short interspersed elements (SINE), DNA transposons and long terminal repeat (LTR) retrotransposons including human endogenous retroviruses (HERV). By using whole transcriptome sequencing, around 8% of mappable reads has been reported to originate from these repeat sequences in human cortex (Tyekucheva et al., 2011).

There are seven genera in the *retroviridae* family; alpharetroviruses, betaretroviruses, gammaretroviruses, deltaretroviruses, epsilonretrovirus, lentivirus and spumavirus, see chapter 57 in the text book (Knipe et al., 2001). HERVs constitute 8% of the human genome and have been broadly grouped into three classes based on sequence similarities. Class I contains HERVs related to gammaretroviruses (e.g. murine leukemia virus, MLV), including HERV-W and HERV-H; class II HERVs are related to betaretrovituses (e.g. mouse mammary tumor virus) and include several subgroups of HERV-K; while class III HERVs are related to spumaviruses (e.g. chimpanzee foamy virus), including HERV-L and HERV-S (Tristem, 2000), reviewed in (Bannert and Kurth, 2004; Griffiths, 2001). HERV sequences exist in multiple copies across different chromosomes and therefore are by nature repetitive. With a few notable exceptions, HERVs have lost their protein-coding capacity. For these reasons, HERV are consistently ignored in large scale expression studies and thus tremendous is not well understood regarding their expression, regulation and potential functional significance.

1.1.1 Evolution

ERVs are exogenous retroviruses that are believed to have integrated into the germ-line of the host, see chapter 9 in the text book (Boeke and Stoye, 1997). These sequences are widespread in many animal species such as mice; ERVs account for about 10% the mouse genome. The inserted ERVs, for instance, can detrimentally cause mutations in the host gene (e.g. supply alternative splicing or termination signals), leading to the loss of host from the population. On the other hand, a few selective advantages can be

provided by ERV insertions. For example, through evolution a positive pressure results in the functional preservation of *ERVWE1* locus of HERV-W family, which is critically involved in the morphogenesis of human placenta (Mallet et al., 2004), see section **1.4** below. Most of inserted ERV sequences are neutral and have persisted in the host genome.

HERV (coined "fossil viruses") are the ancient remnants of retroviral infections that integrated in the germ-line of human ancestors, reviewed in (Jern and Coffin, 2008; Kurth and Bannert, 2010; Mayer and Meese, 2005; Stoye, 2001). At least 31 HERV families have been identified, each derived from a single horizontal transmission event by an exogenous virus infection, followed by germ-line integration, inherited in a Mendelian fashion and have become fixed into the population during the evolution of the human lineage (Katzourakis and Tristem, 2005). In addition to reinfection, the mechanism of intracellular retrotransposition has been used by HERVs to expand their copy numbers within the genome (Belshaw et al., 2005). Intracellular retrotransposition is mediated by reverse transcriptase (RT) machinery, which takes place in a cytoplasmic virus-like particle primed by a cellular transfer RNA (tRNA). The nomenclature of HERV is primarily based on the tRNA specificity of the primer binding site (PBS), reviewed in, (Griffiths, 2001). For instance, the W family of HERV implies that it uses a tryptophan tRNA, which one letter code is W, for priming of reverse transcription. However, the great majority of HERV identified to date are defective and have lost the capacity for horizontal transmission. No single HERV element has been found to generate infectious particles, while reconstruction of infectious version of extinct HERV-K elements was reported from two independent groups (Dewannieux et al., 2006; Lee and Bieniasz, 2007).

HERV are detectable in the genomes of all *Catarrhini*, including humans, chimpanzees, gorillas, orangutans, gibbons and Old World monkeys (e.g. rhesus monkeys and baboons). The groups of class I and III HERVs are relatively old; HERV-H (class I) family, for example, is most abundant and inserted in genome prior to the divergence between Old and New World monkeys (e.g. squirrel monkeys and marmosets) over 40 million years ago (Mager and Freeman, 1995), while the W family of HERV has been considered to enter the human genome after the divergence of Old/ New World monkeys over 30 million years ago (Kim et al., 1999; Voisset et al., 1999). The youngest class II HERVs have been active (retain horizontal transmission) through

the last 5~25 million years, although several HERV-K (HML-5) elements are detected in New World monkeys, which integrated in the genome already 55 million years ago (Lavie et al., 2004).

1.1.2 Structure

HERVs are readily recognized in terms of their retroviral gene structure. According to the text book (Coffin et al., 1997), "a HERV provirus consists of internal group-specific antigen (*gag*), polymerase (*pol*) and envelope (*env*) genes flanked by more or less identical LTRs. These three genes: *gag* encodes important proteins to form structure of viral core, including capsid, matrix and nucleocapsid; *po*l encodes the viral enzymes, including protease (PR), reverse transcriptase (RT), RNAse H and integrase (IN); and *env* encodes the viral envelope glycoproteins. LTRs contain a range of regulatory sequences such as effective promoters, enhancers and transcription factor binding sites. The LTR is composed of U3, R and U5 regions. The U3 region of 5'LTR acts as proviral promoter, while the R region located in the 3'LTR serves as a polyadenylation signal. There are several regulatory elements located in the internal viral genome, including splice donor (**SD**) and acceptor (**SA**) sites which are involved in *env* expression (see section *splicing of ERVWE1 transcripts* below) and PBS for tRNA specific binding to initiate reverse transcription. Typical inactivating mutations originated from point mutations, deletions and frameshift retained in the three genes during the evolution. In addition, solitary LTRs, created by homologous recombination between the two LTRs generating to loss of the entire internal sequence, are frequently present in the human genome".

1.1.3 Distribution and expression

HERV elements are widely distributed in the human genome. Except for chromosomes 8, 13, 21 and 22, Yi and coworkers identified 70 *env* fragments of HERV-H across the other 20 chromosomes (Yi and Kim, 2004). A comprehensive analysis of HML-2 elements (HERV-K family), including 91 proviruses and 944 solo LTRs, has been reported recently (Subramanian et al., 2011). HERV insert in both genic and intergenic regions of the genome, but the majority of HERV families are more likely located in intergenic regions. From another aspect, HERV can be located in sense or antisense orientations of either introns or adjacent regions of annotated genes. Intronic HERV exhibit a significant antisense bias (van de Lagemaat et al., 2006).

Although all HERV elements are considered non-infectious and the vast majority is probably non-functional, there are a considerable number of studies that investigate their expression. In these studies, methods like microarray and quantitative polymerase chain reaction (qPCR) have most frequently been employed. HERV have been reported to be differentially expressed across human tissues and cell-types, with considerable expression in placenta but less expression in other tissues (Forsman et al., 2005; Pichon et al., 2006; Schon et al., 2001; Seifarth et al., 2005; Stauffer et al., 2004; Yi et al., 2004). Moreover, aberrant expression of HERV have been associated to human diseases (Frank et al., 2005; Frank et al., 2008; Hu et al., 2006; Johnston et al., 2001; Lower et al., 1993; Sekigawa et al., 2003). For example, HERV-K transcripts and proteins are expressed at higher levels in a series of cancers and autoimmune diseases. Increased expression of HERV *pol* RNA and proteins, most likely originating from ERV9 family, have been detected in blood samples from schizophrenia patients (Huang et al., 2006). HERV-H has been reported to be up-regulated in tissues from patients with multiple sclerosis (MS), reviewed in (Christensen, 2010).

1.2 HERV-W EXPRESSION

Following its integration into the human germ-line 30 million years ago, HERV-W is today found in many copies scattered throughout the human genome. Like other HERV elements, HERV-W is preferentially expressed in normal human placenta (Blond et al., 1999; Forsman et al., 2005; Stauffer et al., 2004). Differential expression of HERV-W has also been observed in various human tissues and cells. Hybridization-based studies detected HERV-W transcripts in a variety of human normal tissues, in particular, in placental and testis tissues (Seifarth et al., 2005). From results obtained by using a *pol*-probe, Schon and coworkers reported HERV-W transcripts in kidney and liver cells (Schon et al., 2001). Applying RT-PCR, the structural genes (*gag*, *pol* and *env*) of HERV-W were also found to be highly expressed in normal human placenta and testis out of 12 human normal tissues and 18 human cancer cells investigated. In addition, the *gag* gene was specifically transcribed in normal brain and spleen tissues but in none of the cancer cells. With the exception of normal heart and uterus tissues and three cell-lines, *pol* transcripts were widely observed in normal tissues and cancer cells. In contrast, *env* transcripts were expressed in all the tissues and cells tested (Yi et al., 2004). Expression of the HERV-W family has been reported in association with human diseases, such as cancer, MS and schizophrenia (see section **1.6** below).

Despite the differential detection of HERV-W in normal or pathological tissues, the repetitive nature of these sequences makes it difficult to identify expression of single elements for methodological reasons. Previous studies from our laboratory indicated a cell-specific transcription of a range of HERV-W elements in cell-lines of human origin. Such expression patterns of HERV-W *gag* and *env* genes were observed by using an analysis of amplicon melting temperatures (Tm). Since the unspecific designed primers were used to target toward HERV-W *gag* or *env* sequence, the different analyzed amplicons were grouped into 3~4 categories according to the discrete Tm ranges (Nellaker et al., 2006). This previously applied approach with low resolution limits the determination of sequence differences between amplified products. Furthermore, the extent of transcription of individual members of HERV-W across tissues and individuals has not been investigated.

1.3 HERV-W STRUCTURE

As for other HERVs, HERV-W in the human genome contains a number of elements with prototypical structure (denoted proviral elements, i.e. internal *gag*, *pol* and *env* genes flanked by two intact LTRs) but also elements bearing the hallmarks of processed pseudogenes (denoted pseudoelements) and elements that cannot be classified into either of these categories due to extensive deletions/truncations (denoted truncated elements), **Figure 1**.

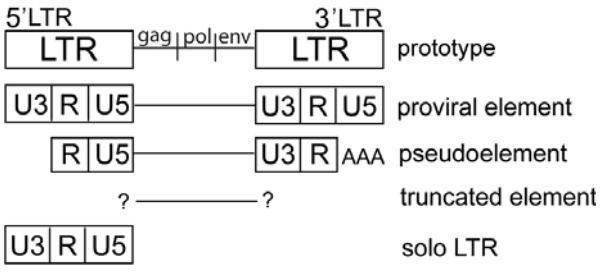

Figure 1 Schematic illustration of the structures of provirus, pseudoelement, truncated element and solo LTR. (Figure modified from "Li F, Nellåker C,Yolken RH and Karlsson H. A systematic evaluation of expression of HERV-W elements; influence of genomic context, viral structure and orientation. BMC Genomics. 2011")

The pseudoelements in common lack the U3 region of 5'LTR and the U5 region of 3'LTR, while truncated elements only displaying partial internal sequences. The U3 region of the LTR contains the regulatory motifs necessary for expression, therefore pseudoelements and truncated elements are presumably transcriptional silent (Costas, 2002). The HERV-W pseudoelements are typically characterized by their colinear structure with retroviral mRNA followed by poly(A) tails. Moreover, the majority of pseudoelements are flanked by direct repeats (i.e. a sequence is repeated with the same

pattern downstream) that are longer than those flanking proviruses (normally flanked by direct repeats of 4 bp) and displays a preference for TT|AAAA insertion motif similar to processed pseudogenes. These pseudoelements thus have been suggested to be generated by RT actively provided by LINE acting on the proviral transcripts. In contrast to very few detected pseudoelements in other HERV families (for instance less than 1% of HERV-H are pseudoelements), numerous pseudoelements (one fourth) have been observed in HERV-W which distinguishes it from other HERV families. In addition, the HERV-W family also contains a large number of solitary LTRs which account for at least half of the all elements in this family (Belshaw et al., 2005; Costas, 2002; Pavlicek et al., 2002).

As mentioned above, we previously observed transcription of HERV-W *gag* and *env* sequences in various cell-types. Surprisingly, cloning and sequencing data indicated that these detected transcripts could be mapped to three categories of HERV-W elements, including both proviral elements and elements lacking regulatory LTRs. In addition, mapped elements were located in both intronic and intergenic regions throughout the human genome (Nellaker et al., 2006; Yao et al., 2007). Whether transcription of HERV-W is initiated from the U3 region of the LTR or from promoters in flanking regions is therefore not clear.

1.4 SYNCYTIN-1 EXPRESSION

1.4.1 Biological role of syncytin-1

A group of studies have identified several *env*, with full-length open reading frames (ORF) able to produce potentially functional proteins, from different HERV families (de Parseval et al., 2003; Dewannieux et al., 2005; Villesen et al., 2004). *ERVWE1*, located on chromosome 7q21.2, is so far the only verified HERV-W proviral locus that has retained coding capacity. The product of the *env* glycoprotein gene, is termed syncytin-1 and appears to be exclusively expressed in the syncytiotrophoblast layer of the placenta. This gene, thus, appears to have been functionally adopted by the human host (Mallet et al., 2004; Mi et al., 2000; Voisset et al., 2000).

The syncytin-1 protein contains 538 amino-acids (AA), which structurally consists of a surface subunit (SU, AA1-317) and a transmembrane subunit (TM, AA318-538). SU is involved in receptor recognition and binding (Cheynet et al., 2006), while TM contains several fusion core peptides (Gong et al., 2005) and a putative immunosuppressive

domain (IM, AA377-396). In the placenta, syncytin-1 mediates trophoblast cell-cell fusion upon interaction with the D type mammalian retrovirus receptor (RDR, also known as neutral amino acid transporters, ASCT1 and ASCT2) (Blond et al., 2000; Lavillette et al., 2002). Other biological functions of syncytin-1 remain uncertain. A recent study reported that syncytin-1 serves the potential function of promoting trophoblast cell proliferation (Huang et al., 2013). Studies of The immune modulatory effect of syncytin-1 have provided different results. The tumor-rejection assay in mice model revealed that syncytin-1 is not immunosuppressive (Mangeney et al., 2007). However, syncytin-1 has recently been suggested to exert immunosuppressive function in lipopolysaccharide or phytohaemagglutinin stimulated human blood culture immune systems, which may play a role in maternal immune tolerance (Tolosa et al., 2012).

1.4.2 Regulation and expression of syncytin-1

ERVWE1 is composed of inactivated *gag* and *pol* genes and intact preserved *env* ORF (containing a specific 12 bp deletion which is located in the intracytoplasmic tail, AA471-538, and is pivotal for the constitutive fusogenic activity of syncytin-1), flanked by two intact LTRs. A 247 bp LTR retroviral promoter (U3 region of 5'LTR) cooperates with a 436 bp upstream regulatory element (URE) region, to regulate *ERVWE1* expression in the placenta. Whereas the U3 region from position +1 to +125 represents a basal placental promoter activity, remaining region of U3 in concert with part of R region (63 bp) suggests a cyclic adenosine monophosphate (cAMP)-responsive core promoter activity in all cell-types. Activation of cAMP/Protein Kinase A (PKA) pathway increase expression of *ERVWE1* and can induce syncytia formation in human trophoblast cells. The URE consist of a 208 bp non-retroviral sequence and a 228 bp mammalian apparent LTR-retrotransposon (MaLR) (Smit, 1993) LTR containing a 33 base-pair long trophoblast specific enhancer (TSE) region (-67~-35), in which a series of putative transcription binding sites are included, conferring a high and specific expression in the placenta (Bonnaud et al., 2005; Cheng et al., 2004; Mi et al., 2000; Prudhomme et al., 2004), see **Figure 2**.

Transcription factors

Several transcription factor binding sites, including Sp-1, GATA, Oct-1 and GCM1, have been identified in the TSE and promoter region (**Figure 2**). Transcription factor Sp-1, encoded by the *SP1* gene, can bind to GC-rich motifs of many promoters. It acts as an activator or repressor by post-translational modification such as phosphorylation

or acetylation, playing a critical role in regulating gene expression. GATA transcription factors represent a family of six closely related and evolutionarily conserved sequence-specific DNA-binding proteins, GATA1 to 6, contributing to gene regulation. Oct-1 belongs to the family of octamer transcription factors, which are characterized by their ability to bind to the "ATTTGCAT" sequence, which is important in the modulation of gene expression. These transcription factors have been shown to be involved in the *ERVWE1* promoter placental regulation (Cheng and Handwerger, 2005; Cheng et al., 2004), reviewed in (Loregger et al., 2003).

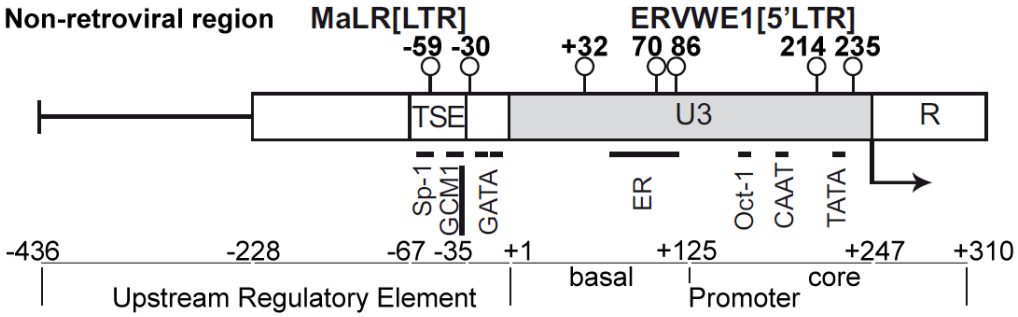

Figure 2 Schematic representation of the ERVWE1 transcriptional regulatory elements. An upstream regulatory element (URE) and LTR retroviral promoter synergistically regulate ERVWE1 transcription. The start and end bases of each constituent element are indicated at the bottom. Transcription starts at the U3/R boundary (arrow). CAAT and TATA boxes as well as effective transcription factors binding sites along the MaLR[LTR] and ERVWE1[5'LTR] are indicated. The GCM1 binding site is underlined. The top relative positions of the CpG sites located two LTR regions are illustrated by circles on vertical bars. (Figure modified from "Prudhomme S, Oriol G and Mallet F. A retroviral promoter and a cellular enhancer define a bipartite element which controls env ERVWE1 placental expression. J Virol, 2004")

GCM1, a highly conserved DNA-binding protein with a glial cells missing (gcm) motif, is a member of the Gcm protein family. The *gcm* gene was first identified in *Drosophila* as being involved in the glial cell fate specification by mediating neural stem cell (NSC) differentiation and promoting glial cell proliferation. In contrast, mammalian Gcm proteins are highly expressed in the placenta but expressed at low levels in the mature central nervous system (CNS), reviewed in (Mao et al., 2012). However, Iwasaki and coworkers has previously reported that mouse Gcm1 expression induce gliogenesis both *in vitro* and *in vivo* (Iwasaki et al., 2003). A previous study from our group reported that the levels of transcripts encoding Gcm1 are expressed highest at embryonal day 17 to postnatal day 7 and substantially decline at later time-points in mouse brain tissue (Asp et al., 2007). These findings suggest that Gcm1 also plays a role in mammalian CNS development. Otherwise, GCM1, is considered a

trophoblast-specific transcription factor and binds to the GCM1 binding site (GCM motif (A/G)CCCGCAT) located in the TSE region of *ERVWE1*, which greatly enhance promoter activity particularly in human placental cells (Nait-Oumesmar et al., 2000; Yu et al., 2002). Decreased transcription of *ERVWE1* has been observed under conditions of low oxygen in choriocarcinoma BeWo cells (Knerr et al., 2003). Induction of cAMP/PKA pathway compensates reduced *ERVWE1* expression by hypoxia since this pathway activation increases expression of GCM1 and subsequently stimulates *ERVWE1* transcription in trophoblast cells (Chang et al., 2005).

In addition, an oestrogen receptor (ER) binding site (i.e. oestrogen response element, **ERE**) and conventional CAAT and TATA boxes (overlapping with cAMP response element, **CRE**) located in the basal and core promoter respectively, which are essential to the promoter activation. The latter can be induced by the cAMP/PKA pathway, while ER bind specifically to the ERE to regulate *ERVWE1* transcription.

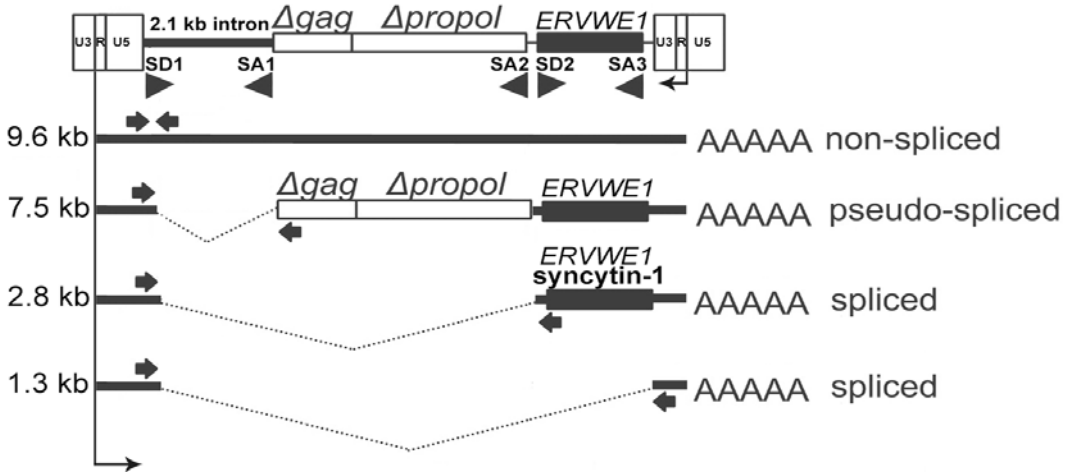

Figure 3 Schematic illustration of the *ERVWE1* splicing strategy. The transcription initiation site and the polyadenylation signal are indicated as thin right and left arrows, respectively. Splice donor (SD) and acceptor (SA) sites are indicated by thick right and left arrow heads, respectively. Non-spliced and spliced transcripts of *ERVWE1* can be detected by using the paired-assays which are indicated by thick right and left arrows. Only the 2.8-kb variant is responsible for syncytin-1 translation. (Figure modified from "Trejbalová K, Blazková J, Matousková M, Kucerová D, Pecnová L, Vernerová Z, Herácek J, Hirsch I, and Hejnar J. Epigenetic regulation of transcription and splicing of syncytins, fusogenic glycoproteins of retroviral origin. Nucleic Acids Res, 2011")

Splicing of ERVWE1 transcripts

According to Trejbalova et al (Trejbalova et al., 2011), the *ERVWE1* full-length pre-mRNA includes a 2.1 kb intron, which can be detected in BeWo cells. Two mRNA transcripts are mainly generated after the splicing process; a 7.5 kb spliced transcript

which contains the *gag*, *pol* and *env* frames and a 2.8 kb subgenomic-spliced *env* transcript which is responsible for syncytin-1 translation, **Figure 3**. The two spliced *ERVWE1* transcripts are rarely expressed in non-placental cells in contrast to their high expression levels in trophoblastic cells. Northern blot previously detected two spliced transcripts in human placenta but with low levels in other tissues (Mi et al., 2000). In addition, *ERVWE1* produces a small 1.3 kb spliced transcript. The 5'splice signal located at the end of U5 region produce these three major single-spliced transcripts, in combination with three different SA sites. Another SD site located at the 5'UTR of *env* ORF may give rise to presence of other splice variants (Blond et al., 1999). The splicing strategy of *ERVWE1* has been conserved among apes (Mallet et al., 2004).

Epigenetic and chromatin modification

DNA methylation is a biochemical process, implicated in the control of genomic imprinting, X chromosome inactivation, expression of endogenous retroviral sequences and tumorigenesis. The enzymes, DNA methyltransferases (Dnmts), catalyze the transfer of a methyl group to DNA during this process, leading to a general gene silencing. Three active Dnmts (Dnmt1, Dnmt3a, and Dnmt3b) that appear to establish and maintain DNA methylation patterns have been identified in the embryonic and neural development in mouse and human (Feng et al., 2010; Vassena et al., 2005), reviewed in (Kim et al., 2009). Gimenez and co-workers observed a hypomethylation of the *ERVWE1* promoter and TSE, along with high levels of mRNA encoding syncytin-1 in the placenta and BeWo cells (Gimenez et al., 2009). Methylation profiles from placentas of twins also indicated an inverse correlation between methylation levels of the *ERVWE1* promoter and its mRNA and protein levels (Gao et al., 2012). In contrast, suppression of expression of syncytin-1 stems from extensive 5' LTR methylation has been detected in non-placental cells such as peripheral blood mononuclear cells (PBMC), skin fibroblast and Hela cells (Matouskova et al., 2006).

In addition to DNA methylation, chromatin modification is another mechanism for gene regulation. The chromatin subunit, denoted as nucleosome, consists of 147 bp of DNA wrapped around eight histone protein cores (two copies of each of four core histones H2A, H2B, H3 and H4). Each core histone has a C-terminal histone-fold domain and a positively charged N-terminal, containing many lysine and arginine residues. Most post-translational covalent modifications occur within the N-terminal, which appears as an extensive tail protruding from the nucleosome. Histone

modifications such as acetylation and methylation are involved in the control of transcription. Whereas acetylation of histone H3 lysine 9 (H3K9Ac) is well-known to correlate with active transcription, tri-methylation of H3K9 (H3K9Me3) has been linked with transcriptional repression (Barski et al., 2007; Rosenfeld et al., 2009). Immunoprecipitation of *ERVWE1* promoter associated chromatin indicated high H3K9Ac but low H3K9me3 levels in BeWo cells, which corresponds to the active transcription of this locus. In contrary, transcriptional suppression of *ERVWE1* was detected in Hela cells, accompanied with high enrichment of H3K9me3 but low H3K9Ac at its 5'LTR region (Trejbalova et al., 2011). Hypermethylation of *ERVWE1* observed in non-placental cells (described above) is accompanied by H3K9 deacetylation, as well as H3K9me3 (Matouskova et al., 2006).

In contrast to genic promoter regions, class I and II ERVs are both highly DNA methylated and marked by H3K9me2 and H3k9me3 in mouse embryonic stem cells (mESCs) (Dong et al., 2008). A DNA methylation-independent pathway has been shown to be involved in the proviral silencing of ERVs in mESCs through H3K9me3 (Matsui et al., 2010). H3K9-specific lysine methyltransferases (KMTase) G9a and SETDB1, which are required for H3K9me2 and H3K9me3, respectively, are demonstrated to play crucial roles in the silencing of proviruses in mESCs (Leung et al., 2011; Matsui et al., 2010). Comparison of Dnmts- and SETDB1-mediated pathways in suppressing genes and retroelements in mESCs has been well characterized recently. Genome-wide profiling revealed that de-repression of class I and II ERVs with concomitant H3K9me3 loss is due to the depletion of SETDB1 rather than Dnmts (Karimi et al., 2011). These findings suggest that suppression of *ERVWE1* in non-placental cells is under the control of KMTases-mediated H3K9me2/3 independent of the methylation status of its promoter.

1.5 RELATIONSHIP BETWEEN HERV-W AND VIRUS

1.5.1 Transcription activation of HERV-W

In previous studies, dysregulation of *ERVWE1* has been shown to be associated with various exogenous stimuli. As mentioned above, oxygen deficiency gives rise to decreased expression levels of syncytin-1 (Knerr et al., 2003). Serum deprivation increase transcription of HERV-W elements *in vitro* (Nellaker et al., 2006). Viral infection can trigger transactivation of HERV-W. Induction of HERV-W by herpes virus simplex type I (HSV-1) infection has been observed in neuronal and brain

endothelial cells (Nellaker et al., 2006; Ruprecht et al., 2006). In addition, activation of HERV-W by Epstein-Barr virus (EBV) was recently observed in blood cells and astrocytes (Mameli et al., 2012).

1.5.2 Influenza A/WSN/33 virus infection

Influenza A viruses are negative-sense, single-stranded viruses with 8 RNA segments that encode 11 proteins, see **Table 1** (chapter 46 in the text book (Knipe et al., 2001)). Based on the type of two proteins on the surface of the viral envelope, hemagglutinin (HA) and neuraminidase (NA), 18 subtypes of HA and 11 subtypes of NA have been found so far. Influenza A/WSN/33 virus, of the H1N1 subtype, is a neurotropic mouse adapted strain which can be easily grown on Madin-Darby canine kidney (MDCK) cells.

Table 1 The segments of influenza A virus and their encoded proteins

Segment	Encoding protein (s)	Function
1	PB2 (RNA polymerase)	i) PB2 recognize mRNA cap, PB1 mediate RNA elongation, PB1-F2 mediate pro-apoptosis, PA cleaves the cap from the host's RNA ii) ssRNA is associated with PA&PB1&PB2 and coated with NPs to form ribonucleoprotein complex (vRNPs)
2	PB1 (RNA polymerase) & PB1-F2 & N40	
3	PA (RNA polymerase)	
4	HA (hemagglutinin)	Mediates binding of virions to the target cells and entry of the viral genome into the target cell
5	NP (nucleoprotein)	Bind to viral ssRNA; regulate nuclear import
6	NA (neuraminidase)	Mediates release of progeny virions from infected cells
7	M1(matrix proteins) &M2	Mediates attachment of vRNPs to lipid bilayer membrane through ion channel
8	NS1& NS2(NEP) (non-structure proteins)	Binds to host cell pre-mRNA transcripts to regulate host gene expression; regulate RNA export

An influenza A/WSN/33 virus infection induced transactivation of HERV-W elements, including *ERVWE1* locus, has been observed in a range of human cell-lines of non-placental origin (Nellaker et al., 2006). The mechanisms by which influenza A/WSN/33 activates HERV-W expression have not yet been resolved.

A previous study reported up-regulation of mouse Gcm1 and syncytin B *in vitro* in embryo fibroblast cell line (NIH-3T3) and primary cultures of neurons or glia cells, as well as increased levels of *Gcm1* transcripts *in vivo* in brains of mice, following influenza A virus infection (Asp et al., 2007). To what extent the expression of GCM1

is involved in the transactivation of *ERVWE1* in response to virus infection in human non-placental cells has not been investigated.

Efficient splicing of *ERVWE1* transcripts has been reported in placental cells, where the ratio of spliced to nonspliced transcripts is 6:1. In comparison, undetectable spliced transcripts and very low levels of nonspliced transcripts have been observed in non-placental cells (Trejbalova et al., 2011). The influence of virus infection on the *ERVWE1* splicing has not been investigated previously.

Hypermethylation and H3K9me3 on the *ERVWE1* promoter play critical roles in the suppression of *ERVWE1* as described above. Modification of the methylation status of host DNA has been observed in response to DNA viruses (Esteki-Zadeh et al., 2012; Macnab et al., 1988; Ryan et al., 2010; Vivekanandan et al., 2010), as well as in response to influenza A virus infection (Li et al., 2010; Tang et al., 2011) in different cell-lines. Histone modifications such as acetylation and methylation induced by DNA viruses have also been observed following infection of tumor cells (Iseki et al., 2005; Tian et al., 2013). If virus infection affect DNA methylation and/or chromatin modifications, and thereby induce up-regulation of HERV-W, especially the *ERVWE1* locus, in non-placental cells remain uncharacterized.

Virus encoded proteins could be potential factors involved in the gene regulation by interacting with histones. The influenza A virus-encoded protein, NS1, was reported to interfere with host gene transcription at many different levels (Hale et al., 2008). For example, NS1 was recently reported to affect transcription of antiviral genes selectively through a histone mimic (Marazzi et al., 2012). Viral proteins NP, M1 and the vRNP (see Table 1 mentioned above) have been found to bind to different histone regions (Zhirnov and Klenk, 1997). Interestingly, NP was recently observed to co-localize with H3K9me3 in transcriptionally repressed regions (Alfonso et al., 2011). If viral proteins play a role on the transactivation of HERV-W is not known.

1.6 ASSOCIATION BETWEEN HERV-W AND DISEASE

1.6.1 Cancer

An increasing body of evidence has demonstrated abnormal expression of HERV-W in cancerous cells. Up-regulation of HERV-W transcripts was detected in malignant tissues, such as testicular (Gimenez et al., 2010) and ovarian carcinoma (Menendez et

al., 2004). Correspondingly, global demethylation of HERV-W LTRs was observed in tumoral samples as compared to normal tissues, suggesting a correlation between HERV-W transcriptional capacity and CpG methylation control. Aberrant expression of syncytin-1 has also been observed in tumoral samples from breast cancer (Bjerregaard et al., 2006) and endometrial carcinoma (EnCa) (Strick et al., 2007). It has been reported that syncytin-1 play a functional role in the fusion process between breast cancer cell lines and endothelial cells (Bjerregaard et al., 2006). Furthermore, syncytin-1 is involved in the fusion or the proliferation of EnCa depending on cAMP pathway activation or steroid hormone stimulation, respectively. The switch of syncytin-1 function from fusion to cell differentiation appears to be regulated by transforming growth factor (TGF-β). Steroid hormone stimulates expression of syncytin-1 and TGF-β which leads to cell proliferation. Up-regulation of syncytin-1, however, induces cell fusion in the absence of TGF-β (Strick et al., 2007). The over-expression of syncytin-1 in EnCa could be a result of hypomethylation of 5' LTR, including the ERE and CRE (see *Transcription factors* mentioned above) binding sites (Strissel et al., 2012). Recent publications also indicates a significantly increased transcription of HERV-W in mycosis fungoides (cutaneous T-cell lymphoma) biopsies (Maliniemi et al., 2013) and specific detection of syncytin-1 in blood samples from patient with lymphoma and leukemia (Sun et al., 2010).

1.6.2 Abnormal placentation

Pre-eclampsia (PE), hemolysis elevated liver low platelets (HELLP) and intrauterine growth restriction (IUGR) are multisystem disorders during pregnancy associated with placental abnormalities. Hypoxia has been demonstrated to reduce the expression levels of syncytin-1 in PE (Kudo et al., 2003). Previous study reported decreased expression of GCM1 in PE (Chen et al., 2004). Hypoxia has also been considered to be involved in the down-regulation of GCM1 (Chiang et al., 2009). In all three syndromes, reduced expression of syncytin-1 (Ruebner et al., 2013) has been observed to coincide with increased apoptosis rate as compared with control placentas (Langbein et al., 2008). A DNMT3a-mediated hypermethylation of the promoter of *ERVW1* has been suggested to be responsible for its decreased activities (Ruebner et al., 2013).

1.6.3 Multiple sclerosis

Multiple sclerosis (MS) is a chronic, neuroinflammatory demyelinating disease of the CNS. The cause of MS is unknown; however, a leading hypothesis suggests that MS is

the result of exposure of genetically susceptible individuals to certain environmental factor(s) such as viral infections (Compston and Coles, 2002). MS-associated retrovirus (MSRV) has been found as retrovirus-like particles budding from leptomeningeal cells of MS patients (Perron et al., 1989). By using RT-PCR, the retroviral *pol* sequences, which resemble MSRV and exhibit highest RT activity, have been amplified and isolated from MS patients (Perron et al., 1997). The identification of MSRV subsequently allowed the identification of similar sequences, HERV-W, in the human genome. MSRV is proposed to be an exogenous replication-competent member of HERV-W (Perron et al., 1997), reviewed in (Dolei and Perron, 2009). A recent study however suggested that some of the MSRV sequences identified in patients are derived from template-switching between transcripts from dispersed genomic HERV-W loci during reverse transcription, i.e. an artifact occurring during complementary DNA (cDNA) synthesis (Laufer et al., 2009). Thus, the existence of an exogenous infectious retroviruses related to HERV-W remains to be firmly established.

Transcription of HERV-W has previously been investigated in many MS-related studies which suggest that HERV-W expression may play important roles in MS development. Using qRT-PCR, increased levels of HERV-W *env* transcripts were observed in PBMC (Mameli et al., 2007a) and brains from patients with MS relative to control individuals (Antony et al., 2006; Antony et al., 2007b; Mameli et al., 2007a). The fact that HERV-W GAG and ENV proteins are expressed in the neuronal and glial cells, respectively, in the normal brains suggests a physiological function of HERV-W in human brain, while macrophage-specific detection of ENV expression from MS lesions indicates HERV-W is involved in the pathogenesis of MS (Perron et al., 2005). Whereas enhanced expression of syncytin-1 was reported in glial cells of MS lesions (Antony et al., 2004), down-regulation of the receptor ASCT1 was observed in demyelinated regions in the brains of MS patients (Antony et al., 2007a). *In vitro* studies of astrocytes indicated that syncytin-1 expression can be activated by pro-inflammatory molecules such as tumor necrosis factor-alfa (TNF-α), interferon-gamma (IFN-γ), interleukin-6 (IL-6) and IL-1, but inhibited by anti-viral IFN-β (Mameli et al., 2007b). TNF-α-stimulated syncytin-1 expression through enhancing the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) binding to the NF-κB-responsive element located in the upstream enhancer region (MaLR LTR) of *ERVWE1* promoter (Mameli et al., 2007b), over-expression of syncytin-1 in turn resulted in up-regulation of pro-inflammatory cytokine IL-1β and oligodendrocyte cytotoxicity by

inducing release of redox reactants (Antony et al., 2004). Moreover, a distinct pathway implicated in MS pathogenesis was supported by the evidence that increased syncytin-1 levels cause increased inducible nitric oxide (NO) synthase and concurrent suppression of hASCT1 through induction of old astrocytes specifically induced substance (OASIS, an endoplasmic reticulum stress sensor), leading to oligodendrocyte damage and demyelination (Antony et al., 2007a).

1.6.4 Psychiatric disorders

Psychiatric disorders such as depression, bipolar disorder and schizophrenia are serious mental diseases. It is believed that both genetic components and environmental factors contribute to these disorders, reviewed in (Torrey et al., 1997; Yolken and Torrey, 1995). For example, exposures to various obstetric complications including pre-eclampsia increase the risk for later developing schizophrenia (Dalman et al., 1999). Accumulating evidence suggests that maternal infections such as prenatal influenza, reviewed in (Brown and Derkits, 2010), HSV-2 (Mortensen et al., 2010), *Toxoplasma gondii* and cytomegalovirus (Blomstrom et al., 2012) infections are risk factors for schizophrenia. Influenza viral infection has also been associated with bipolar disorders based on epidemiological and serological evidences (Parboosing et al., 2013). Retroviruses have been proposed as potential infectious agents involved in the pathogenesis of schizophrenia, reviewed in (O'Reilly and Singh, 1996; Yolken et al., 2000).

During a RT-PCR based study of potential retroviral activity in schizophrenia employing degenerate primers directed at a conserved region of the RT, HERV-W *pol* RNA was found in cerebrospinal fluids (CSF) of 30% of patients experiencing their first-episode of schizophrenia or schizoaffective disorder. Moreover, differential up-regulated transcription of HERV-W *pol* was detected in frontal cortex regions of *postmortem* brains from schizophrenic patients compared with controls (Karlsson et al., 2001). Similar up-regulation of HERV-W *env* transcripts were more recently reported from blood samples of patients with recent-onset schizophrenia (Huang et al., 2011) or bipolar disorder in comparison with healthy controls (Perron et al., 2012). In addition, elevated levels of HERV-W *gag* sequences were detected in plasma and PBMC from patients with a recent diagnosis of schizophrenia, as compared with individual controls (Karlsson et al., 2004; Yao et al., 2007). Intriguingly, Perron and coworkers reported that HERV-W GAG and ENV protein are predominantly expressed in serum of

schizophrenic patients (Perron et al., 2008). Employing microarray chip assay, in contrast, HERV-W *pol* transcripts were detected at similar levels in brains from patients with schizophrenia or bipolar disorder and control individuals (Frank et al., 2005).

The primary cause(s) of the aberrant expression of HERV-W members observed in these different patients, and their potential role in the pathogenesis of these diseases is unclear. These observations suggest a potential link between exogenous infections-driven, such as influenza virus infection-induced activation of HERV-W and psychiatric diseases. Further studies are required to improve our understanding of transcription and functional role of HERV-W play during these diseases which is important for the development of better treatment and prevention strategies.

2 AIMS

The general aim of this thesis project was to investigate the transcription of genomic regions harboring HERV-W elements. The major specific aims are;

1. To investigate the extent and variation of transcription of HERV-W elements in human tissues with a particular focus on the variation between different human brain regions.
2. To determine if transcription of HERV-W elements, in different tissues and cells, is a consequence of transcriptional leakage, or specifically directed from the retroviral LTR, or both.
3. To address the mechanism underlying transactivation of HERV-W elements, especially the *ERVWE1* locus, in response to exogenous influenza A/WSN/33 virus infection.
4. To characterize HERV expression in human brain regions using next generation sequencing.

3 METHODOLOGICAL CONSIDERATIONS

Detailed information of the material and methods are provided in the individual papers. Here, the advantages and limitations of the methods are briefly discussed.

3.1 CELL CULTURES

For the purpose of investigating the effects of exogenous/environment stimuli on the expression of HERV-W in human cells, primary fibroblast cells were prepared by cutaneous biopsy. Since details of growth media and conditions of fibroblast cultures were not provided in the published three papers, the description see below. Fibroblast cells were cultured in regular DMEM GlutaMAX media with Penicillin/Streptomycin and 15% fetal bovine serum (FBS), supplemental with 10mM HEPES, 1× MEM amino acids, 1× sodium pyruvate which is designed for the growth of these cells. Cell cultures were used after 2-3 passages. In addition, cell-lines originating from different tissue types were also used in this project. Human astrocytoma cell-line (CCF-STTG1), human neuroepithelioma cell-line (SK-N-MC, HTB-10), human lung epithelial cell-line (A549) and human embryonic kidney cell-line (293A) were employed to model extra-placental cells. By comparison, human choriocarcinoma cell-line (JEG-3) was applied to examine the properties of placental trophoblast cells.

Cell-lines are readily available and unlimited sources of biological material but are usually derived from tumor tissues and have adapted to growth in culture which may imply varied expression compared to a normal cell. In contrast, primary cells, such as dermal fibroblast, are better representing cells *in vivo* and thus able to generate more physiologically correct data, as well as addressing the role of genetic variation. The regional ethics committee has approved all the experiments regarding the primary cell culture (04-273/1, 2006/637-32 and 2009-06-12).

3.2 TREATMENTS

In order to study the expression alteration of HERV-W elements in response to stimuli treatment, here we investigate;

3.2.1 Serum deprivation

As an unspecific stressor, serum withdrawal commonly stimulates the cells to reduce their basal activity (Iyer et al., 1999). In **paper I and III**, human primary fibroblasts cultures and cell-line were subjected to serum starvation to examine the environmental

influence on expression pattern of HERV-W. Cells cultured in normal growth media were washed with corresponding growth media without fetal bovine serum (FBS) and incubated in serum deprived media for 24 hrs before harvest.

3.2.2 Poly (cytidylic-inosinic) acid (Poly(I:C)) treatment

The structure of poly (I:C) resembles double-stranded RNA (dsRNA). Therefore, in **paper III** poly (I:C) was used to elicit a cellular anti-viral response in CCF-STTG1 cultures by simulating viral infections. The outcome was compared to that influenced by influenza A/WSN/33 virus infection. The cultured cells were harvested after 6~24 hrs incubation.

3.2.3 Influenza A/WSN/33 virus infection

Influenza A/WSN/33 virus strain is a mouse adapted neurotropic strain and has been used to infect mouse brain cells, including both *in vitro* and *in vivo*, in our previous studies (Asp et al., 2007). HERV-W can be transactivated in human non-placental cells by influenza A/WSN/33 virus infection (Nellaker et al., 2006), in **paper II and III** this virus was applied to investigate the effects of virus infection on the expression and regulation of HERV-W and reveal the mechanism of transactivation of HERV-W following virus infection in cell-types mentioned in section **3.1** above. Cell cultures were washed with Minimum Essential Media and then infected with influenza virus at different multiples of infection (MOI) depending on the purpose of the experiment (usually at 0.5 MOI, i.e. ratio of virus particles to cells is 1:2). After 1 hour of virus adsorption in serum-free media, cells were rinsed and kept in complete growth media and harvested after 6~48 hrs incubation.

3.3 Q-RT-PCR

To investigate HERV-W transcription in the human genome, the amount of expressed element in a cell can be measured by the number of corresponding mRNA transcript copies present in a sample. For mRNA-based PCR the RNA population is first isolated and reverse transcribed to cDNA. Total RNA from most material such as cells and tissues were prepared with Qiagen RNeasy Mini Kit. To avoid the risk of the hemoglobin interfering with the affinity column, Qiazol Kit was used in combination with RNeasy Mini Kit to extract RNA from whole blood samples. Purified RNA was quantified by using a Picodrop Microliter UV/Vis spectrophotometer and a fixed amount of RNA was treated with DNase I (Invitrogen) to remove any residual genomic

DNA. RNA was then converted to cDNA using reverse transcriptase Superscript II and oligo (dT) primers (Invitrogen) primarily targeting poly-adenylated mRNAs.

The PCR, an indispensable tool commonly used in medical and biological research nowadays, was developed in 1980's (Mullis et al., 1986). This technique consists of a predetermined number of cycles with repeated temperature changes, each cycle normally consist of denaturation, annealing and extension steps. As PCR progresses, a small amount of specific DNA sequence is selectively amplified by using two primers that are complementary to the target sequence, deoxribonucleotide triphosphate (dNTP) and thermo stable *Taq* DNA polymerase. The amount of amplicon is doubled for every completed cycle and the final product at the end point can be visualized by agarose gel electrophoresis and staining with ethidium bromide, Gel-Red etc. The method was used in **paper III**.

Quantitative PCR (Q-PCR, also called real-time PCR) follows the general principle of conventional PCR but simultaneously detects and measures target sequence as the reaction progresses in "real time" which therefore provides more rapid and sensitive results over a far wider dynamic range than end-point analysis by electrophoresis. SYBR green and TaqMan, generating a fluorescence signal during the amplification, are the two most often used methods in qPCR. The latter requires the additional design and synthesis of a specific probe to monitor the amount of amplicons. To be more convenient and economic we applied SYBR green as an intercalating dye reporter in this study, bearing in mind its potential binding to non-specific dsDNA PCR products that may interfere with quantification of the target sequence. Application of SYBR green in qPCR requires identification of specific amplicons using analysis of their melting temperature (Tm) in a post-amplification dissociation step. Accurate results can be achieved by comparing the dissociation curves of the analyzed products.

Q-RT-PCR is widely used in the relative and absolute quantification of gene expression. Threshold, an arbitrary level of fluorescence, is chosen on the basis of the baseline variability. In this plot model, a signal detected above the threshold is defined as the threshold cycle (Ct) for a sample. The relative levels of target sequence between samples can then be acquired by comparing their Ct value to those of an endogenous reference. The expression of reference, often using "house keeping" gene such as GAPDH, β-actin, rRNA etc., is assumed to be constant across different tissues and

different experimental conditions. However, it is difficult to set an appropriate internal control when comparing between different tissues or treatments. For example, virus infections cause instability of cellular genomic material and altered expression of a large number of cellular genes. As the viral replication proceeding, total amount of RNA from viruses and host genes largely varies in the infected cells. Although no reference is optimal, in this thesis relative expression of HERV-W elements were arbitrarily normalized to β-actin transcripts in experiments encompassing different tissues and cell-lines. The ΔCt denoted as difference between target gene and β-actin Ct values in each sample is obtained by $Ct_{target} - Ct_{\beta\text{-actin}}$. The ΔΔCt denoted as difference between the ΔCts in treatment groups is then obtained by $\Delta Ct_{treated} - \Delta Ct_{untreated}$. Ultimately, fold differences in the levels of transcripts among groups is calculated by using the formula $2^{-\Delta\Delta Ct}$ (Livak and Schmittgen, 2001). In theory, this formula is applicable only when the reaction efficiency in an ideal PCR is 100%, which generates a doubling of the amount of amplicon every cycle. Differing efficiencies between the control and target gene leads to exponentially growing errors in the final fold calculation. In **paper II**, efficiency of each assay was determined by using $E=10^{-1/slope}-1$. Efficiencies of the assays designed for HERV-W elements and those of the assays designed for adjacent genes were equal. Hence, this equality enables the base "2" in the formula described above to suffice the calculation of fold change. The qPCR technology was used in **paper I-III**.

3.4 MELTING TEMPERATURE (TM) ANALYSIS

As described above, SYBR green chemistry is applied in qPCR based on the specific Tm of amplified fragment. The Tm measured in the dissociation stage is used to detect erroneous product and sequence variation between amplicons. Tm analysis has previously been used to identify the expression patterns of HERV-W *gag* and *env* elements (Nellaker et al., 2006). By this technique qPCR amplicons can be classified into distinct melting temperature ranges indicating sequence differences and thus transcripts from different genomic loci can be identified. A novel technique "high resolution Tm analysis" was subsequently developed by employing an additional molecular beacon as an internal control for temperature variations over the heat-block during the dissociation step (Nellaker et al., 2008; Nellaker et al., 2007). In denaturation step, the fluorescence of the molecular beacon increases on melting which generate dissociation curves distinguishable from that of SYBR products. In addition, the tool, Gaussian curve fit analysis of Tm (GcTm) program, is applied to determine

more precise and unbiased Tms of amplicons, **Figure 4**. By using the Tm-probe in combination with the curve-fit method, Tm resolution can be improved up to 3-fold and errors introduced by the instrument is eliminated to some extent. Furthermore, this technique allows additional melting temperature to be identified. In **paper I-III**, this approach was applied to examine the frequency distribution of HERV-W *gag* sequence across human tissues and individuals, along with the distribution variation in response to stress treatments including virus infection. The Tm analysis program presented in three papers was adapted for the ABI Prism 7000 SDS instrument but the principles can be adapted to other systems.

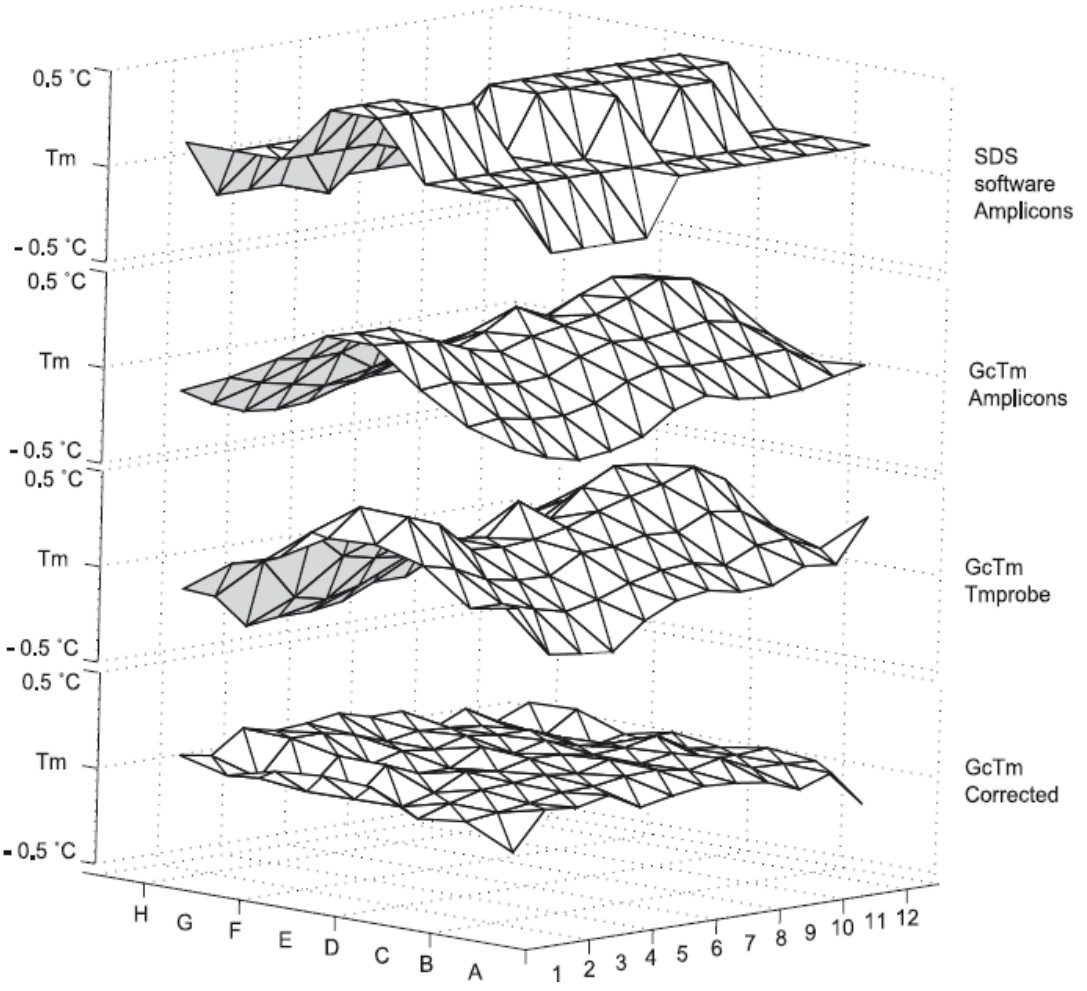

Figure 4 Distribution of reported temperatures over a 96-well plate in an ABI Prism 7000 for 1 template sequence. *Top*, one example of the Tms reported by the SDS software. *Upper middle*, indicates the amplicon Tms reported by SYBR and calculated by the Tm-analysis program. *Lower middle*, Tm-probe Tms as calculated by GcTm Tm-analysis program. *Bottom*, the normalized Tms of the amplicons, (i.e., calculated Tms corrected for temperature variations with the Tm-probe data). The lower 3 plots represent data averaged for 3 dissociation curves. (Christoffer Nellåker, Ulf Wållgren, and Håkan Karlsson. Molecular Beacon–Based Temperature Control and Automated Analyses for Improved Resolution of Melting Temperature Analysis Using SYBR I Green Chemistry. Clinical Chemistry 2007; v. 53, p.98-103. Figure reproduced with permission from the American Association for Clinical Chemistry).

3.5 PLASMID EXPRESSION SYSTEM AND TRANSFECTION

To investigate if over-expression of GCM1 regulates transcription of *ERVWE1* in **paper III**, CCF-STTG1 or JEG-3 cells were transfected with plasmid construct pcDNA3.1/V5-HIS-TOPO containing full length of ORF encoding human GCM1 using Lipofectamine 2000 in accordance with the manufacturer's instructions (Invitrogen). Transcription of insert GCM1 in the plasmid can be induced through the strong CMV promoter. Since an 80% similarity, specially a highly evolutionarily conserved "gcm motif", has been shared by homologous *GCM1* genes from human and mouse, this plasmid was previously used in the studies of examining the influence of Gcm1 on syncytin A and syncytin B, analogs for human syncytin-1, in murine cells (Asp et al., 2007).

3.6 RNA INTERFERENCE

RNA interference (RNAi) is terms describing a conserved biological response to complementary dsRNA which mediates resistance to parasitic nucleotide sequences and represses gene expression. In non-mammalian systems such as *C.elegans* and *Drosophila* long dsRNA is routinely used for the specific suppression of genes. However, in mammalian cells introduction of dsRNA molecules longer than 30 nt may trigger potent antiviral response, thereby leading to global changes of gene expression. In this thesis chemically synthesized small interfering RNA (siRNA), comprised of 21-mer dsRNAs which targeted to specific sequence of GCM1, was used to perform gene silencing experiment. Transfection was carried out in CCF-STTG1 or JEG-3 cells using HiPerFect (Qiagen) according to the manufacturer's instructions in **paper III**.

3.7 PYROSEQUENCING

The bisulfite reaction, deaminating cytosine residues through sodium bisulfite treatment, was discovered in 1970's (Hayatsu et al., 1970) and developed in 1990's to analyze 5-methylcytosine (5mC) in DNA (Frommer et al., 1992). For *ERVWE1* promoter-specific methylation analysis in **paper III,** cytosine was first converted to thymine in single strand DNA of genomic DNA from cell cultures after sodium bisulfite treatment, while 5mC remains unaffected. Using primers specific designed for bisulfite modified DNA, the DNA sequence of interest was amplified by conventional PCR. Unique amplicon was loaded in the following pyrosequencing reaction.

Developed in 1996, pyrosequencing is a technique of DNA sequencing based on the "sequencing by synthesis", (Ronaghi et al., 1998). In principle, the specific sequencing primer hybridizes with a single-stranded PCR product followed by incorporation of dNTP into DNA strand. Nucleotide incorporation generates light seen as a signal peak in the program trace. Unincorporated nucleotides are subsequently degraded and addition of dNTPs is performed sequentially which determine the strand sequence. Compared to conventional sequencing of plasmid clones with insert of interest, this technique is fast, simple to use to delivery real sequence data with high accuracy and quantity. This protocol has been reported to detect DNA methylation differences as small as 5% (Colella et al., 2003).

3.8 CHROMATIN IMMUNOPRECIPITATION

Chromatin immunoprecipitation (ChIP) assay is a powerful technique for characterizing the association of genomic DNA sequences with specific regulatory proteins. There are mainly two typical ChIP; cross-linked ChIP (XChIP) and native ChIP (NChIP). The former method was established in *Bacterial* research and developed in the field of *Drosophila* in the 1980's (Jackson, 1978; Solomon et al., 1988) and extensively refined thereafter. This method is widely used in living organisms and cell culture. In **paper III**, XChIP was used to investigate the levels of H3K9me3, SETDB1 and influenza A viral proteins in CCF-STTG1 cells. Typically, in ChIP DNA-protein complexes are cross-linked and then sheared. Associated DNA fragments are selectively immunoprecipitated by specific antibody, purified and amplified for the target gene analysis using real time PCR or next generation sequencing (NGS).

3.9 RNA SEQUENCING

By using NGS, RNA sequencing is a technology to quantitatively analyze entire transcriptomes. NGS is vigorously developed in recent years, generating high throughput sequencing data which can be used to identify expressed transcripts in a wide range of research (Chu and Corey, 2012). Basically, RNA is first extracted from samples, followed by reverse transcription. Optional fragmentation before reverse transcription may decrease the 5' bias of randomly primed-reverse transcription and the possibility of primer binding sites affected by secondary structures. Synthesized cDNA can be fragmented via size selection and then subject to the sequencing. For different purposes, mRNA can be preselected by poly-A enrichment or rRNA depletion during their isolation from total RNA. In **paper IV**, a publicly available RNA sequencing

datasets produced from human postmortem brain tissues (can be requested at http://sncid.stanleyresearch.org/) were used to investigate transcription of HERV repetitive sequences. 475 individual consensus sequences representing LTR repeats were downloaded from Repbase at www.girinst.org/repbase. These are the same sequences that are used by RepeatMasker to annotate repeats in sequenced genomes (Jurka, 2000). The consensus sequences were imported into CLC Genomics Workbench 5. This software were subsequently used for alignment of sequence reads to the consensus sequences allowing only unique matches using the following settings; similarity: 0.9, length fraction: 0.9, insertion cost: 3, deletion cost: 3, mismatch cost: 2. Reads were also aligned to the human genome (Hg19) using the same settings but including reads with multiple hits (randomly assigned) to obtain a measure of the proportion of mappable reads in each of the libraries.

4 RESULTS AND DISCUSSION

4.1 EXPRESSION PATTERN OF HERV-W GAG TRANSCTIPTS

The repetitive nature of HERV-W has lead to difficulties in identifying single elements due to methodological limitations, and hence little is known of the extent of expression of individual elements. Using high-resolution Tm analysis (see section **3.4** above) in **paper I** we examined the frequency distribution of transcripts containing HERV-W *gag* sequences in a wide range of human tissues and cells, including testis, ovary, thymus, PBMC, placenta, liver, blood, spleen and regions of the brain, as well as in human primary fibroblast cultures. A total 2775 indistinguishable Tm observations from these tissue samples were stratified into 13 categories to construct optimal mixture models according to Akaike's information criterion. From the profiles data, there were significant differing patterns of HERV-W *gag* transcripts between all the tissues investigated. Of these tissues, thymus and testis exhibited the most distinct expression patterns. Expression of such transcripts also differed between ten regions of the brain. The blood samples from different individuals, however, presented a far more homogeneous expression pattern of HERV-W *gag* transcripts. In addition to the tissue-specific expression pattern of HERV-W *gag*, expression of such transcripts varied more between individuals in cerebellum and four cortical regions of brain than in other tissues (spleen, blood and skin fibroblast).

Dendrograms based on gene expression patterns has been previously built from numerous human tissues (Shyamsundar et al., 2005). Similarly, our dendrogram, constructed on the basis of neighbor-joining of the Pearson correlation coefficients, grouped brain regions and gonads together and tissues rich in cells of the immune system together. Although insufficient data only provided inferring phylogenies, these results indicated that HERV-W *gag* transcripts vary between tissues in a similar way to that of coding transcripts and suggested an extensive and patterned transcription of HERV-W elements.

Analyses of primary fibroblast cultures from separate individuals avoid potential artifacts introduced by cell composition and dissection. In addition to determine tissues-specific and inter-individual varied patterns of expression, we used this approach to investigate the effect of exogenous stimuli on the expression pattern of HERV-W *gag* transcripts. Indeed, variation levels of expression patterns in fibroblast cultures were

reduced in response to serum deprivation. Further investigation, using the same Tm analysis, revealed significant qualitative variations of HERV-W *gag* transcripts in fibroblast cultures following influenza A/WSN/33 virus infection (**paper II and III**). In line with the changes in response to serum deprivation, the varied distribution pattern of HERV-W *gag* sequences detected in uninfected cultures became more homogenous in virus infected cultures.

Taken together, HERV-W gag transcripts appeared to exhibit a highly diversified expression pattern across both tissues and individuals. The Tm analysis provides a cheap and rapid approach to investigate changes of distribution patterns of HERV-W gag in response to environmental stimuli.

4.2 TRANSCRIPTION OF INDIVIDUAL HERV-W ELEMENTS

Although the method (applied in **paper I**) revealed a non-random expression pattern of HERV-W elements, the mechanisms underlying their transcription remains uncharacterized. Pseudoelements, LINE-1-mediated retrotransposition of proviral elements, and truncated elements are considered as transcriptionally silent due to loss of U3 region of LTR (Costas, 2002; Pavlicek et al., 2002). Our previous studies, however, indicated expressed sequences that mapped not only to proviral elements, but also to pseudoelements and truncated elements which raised questions regarding the transcriptional control of these loci (Nellaker et al., 2006; Yao et al., 2007). In **paper II**, we therefore investigated the transcriptional regulation of individual HERV-W elements representative of three classes; proviruses, pseudoelements and solo LTRs. Due to the absence of entire LTR, truncated elements were not included.

4.2.1 Distribution of HERV-W

Using the HERV17 internal sequence or the LTR17 sequence from Repbase Update (Jurka, 2000) followed by manual inspection of fragmental loci, we identified 295 HERV-W loci by BLAT searches (Kent, 2002) of the human genome (March 2006 assembly), see **Table 2**. Loci were defined as intronic if located within the boundaries of annotated RefSeq genes (Pruitt et al., 2007). Along with the gene distribution through the human genome, approximately 28% HERV-W elements are located in intronic regions of annotated genes which is consistent with previous finding that most HERV families are more likely to be found in intergenic than in intronic regions (van de Lagemaat et al., 2006). Similar percentage of integrations in intronic regions are

observed in all the proviral elements (contain the regulatory U3 region of LTR), pseudoelements and solo LTRs (formed by homologous recombination) studied, suggesting that machineries of both LINE retrotransposition and homologous recombinations are neutral with regard to genic or intergenic regions. Intronic HERV members exhibited an anti-sense insertion preference (van de Lagemaat et al., 2006). It appears that pseudoelements drive the anti-sense bias integration of HERV-W intronic elements since proviruses and solo LTRs were not significantly biased.

Table 2. Distribution of HERV-W elements

	Proviruses	Solo LTRs	Pseudoelements
Intergenic	37	110	66
Intronic	10	43	29
p-value		0.508 [a]	

	Total[1]	Expected*	Pro	Solo	Total[2]	Expected*	Pse	Expected*
Sense	21	43	3	16	19	28	2	14
Anti-sense	65	43	7	31	38	29	27	15
p-value	0.0005[a]		1.000[b]		0.087[a]		0.0008[b]	

Distribution of three categories of HERV-W elements in genic/intergenic regions and distribution of intronic HERV-W elements on sense and anti-sense strands, respectively. 1Total number of intronic HERV-W elements, 2 indicates number of proviruses and solo LTRs. * indicate expected number based on random integration (i.e. 50% on each strand). Pro, solo and pse indicate proviruses, solo LTRs and pseudoelements, respectively. a Chi-square test, b Fisher's exact test. P-values <0.05 were considered significant. (Table reproduced from "Li F, Nellåker C,Yolken RH and Karlsson H. A systematic evaluation of expression of HERV-W elements; influence of genomic context, viral structure and orientation. BMC Genomics. 2011")

4.2.2 Impacts of genomic context, viral structure and orientation on expression of HERV-W

Next we designed assays specific for a number of intronic and intergenic HERV-W elements with or without intact regulatory regions as well as for their corresponding flanking genes, as illustrated in **Figure 5**. By using qPCR, we measured the expression levels of these elements in various human tissues and subsequently examined the influence of genomic context, viral structure and orientation on their expression.

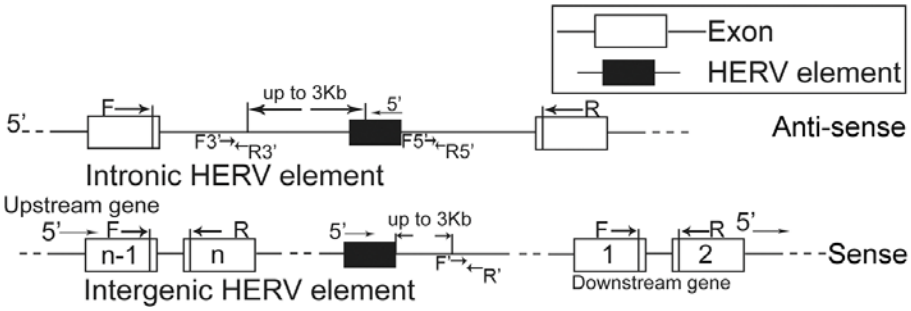

Figure 5 Strategy for design of assays specific for intronic and intergenic HERV-W elements and their corresponding genes. F3'/R3' and F5'/R5' denote forward and reverse primers in intronic regions 3' and 5' of HERV-W elements, respectively. (Figure modified from "Li F, Nellåker C,Yolken RH and Karlsson H. A systematic evaluation of expression of HERV-W elements; influence of genomic context, viral structure and orientation. BMC Genomics. 2011")

According to our data, all the HERV-W elements investigated are expressed in the tissues analyzed despite quantitative variation across individual tissues and elements, in line with previous studies of widespread expression of HERV-W elements in human cells (Nellaker et al., 2006; Yi et al., 2004) and tissues (Laufer et al., 2009; Yao et al., 2007). HERV-W elements in intronic regions were expressed at higher levels (i.e. lower Ct value) than elements in intergenic regions. It has also been observed that HERV-K LTRs in gene-rich regions exhibit significantly higher activities than those in gene-poor regions (Buzdin et al., 2006), suggesting an influence of genomic context on the transcriptional activity. Regarding the intronic elements, proviral elements exhibited higher expression levels than pseudoelements regardless of the transcriptional activities of the genes into which they are integrated, indicating the functional promoter activity of the U3 region in the proviral LTR. Indeed, such LTRs in the HERV-W and other families are active in different human cells to initiate transcription of HERV elements (Buzdin et al., 2006; Schon et al., 2001), or act as alternative promoters which modulate transcription of nearby genes (Conley et al., 2008; Gogvadze et al., 2009; Khodosevich et al., 2002), reviewed in(Griffiths, 2001). Proviral elements, particularly those with only one intact LTR, were detected at significantly higher levels than those of solo LTRs, suggests an involvement of internal sequence (i.e. gag, pol and/or env) in the expression of transcripts generated which might be due to an effect of regulatory element located in the internal sequence (Buzdin et al., 2006) (**Figure 6**).

Relative to their corresponding genes, intronic elements integrated on the sense strand appeared to be transcribed at higher levels than those integrated on the anti-sense strand. Furthermore, linear regression analyses suggest that expression of transcripts

containing intronic proviral elements is independent from that of their corresponding genes. In contrast, intronic pseudoelements are highly correlated to their corresponding gene transcripts and therefore appear to represent unprocessed pre-mRNAs or leaky transcription of coding genes.

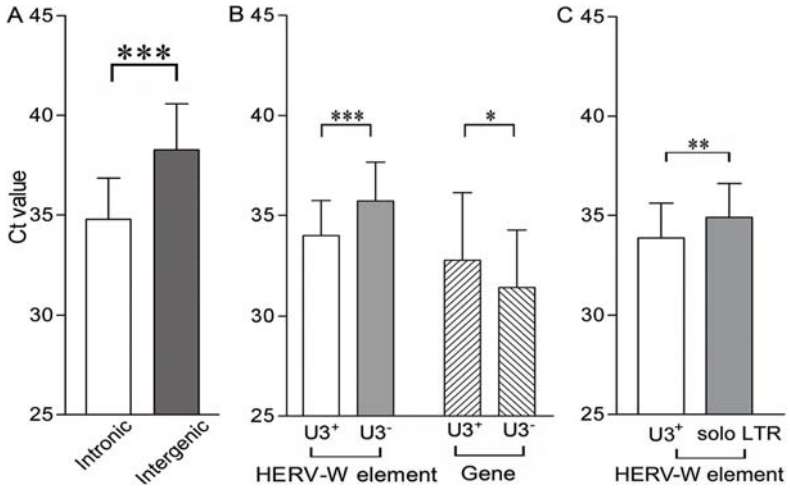

Figure 6 Comparison of average expression levels between intronic (n = 14) and intergenic (n = 10) HERV-W elements in eight human tissues (A). Comparison of expression levels between proviral (n = 7) and pseudoelements (n = 7), as well as comparison of expression levels between the genes into which these elements are integrated (B). Expression Levels of proviral elements and solo LTRs (n = 6) (C). Low ct value indicates high expression. Bars indicate means and error bars indicate SD. Statistical significance is indicated by *= $p<0.05$, **= $p<0.01$, ***= $p<0.001$. (Figure modified from "Li F, Nellåker C,Yolken RH and Karlsson H. A systematic evaluation of expression of HERV-W elements; influence of genomic context, viral structure and orientation. BMC Genomics. 2011")

Taken together, these findings suggest that our previously observed diversified and tissue-specific expression of elements in the HERV-W family is the result of both directed transcription (involving both the LTR and internal sequence) and leaky transcription of HERV-W elements in normal human tissues.

4.3 TRANSACTIVATION OF HERV-W

4.3.1 Effects of exogenous stimuli on the HERV-W promoter/LTR

In addition to the alterations in expression pattern of HERV-W *gag* sequence in primary fibroblast cultures in response to serum deprivation or influenza A/WSN/33 virus infection, significant quantitative up-regulation of HERV-W *gag* transcripts was also detected in virus-infected cultures (**paper II-III**), which expands our previous study of transactivation of HERV-W elements by cellular stress or virus infection in non-placental cell-lines (Nellaker et al., 2006). In **paper II**, we investigated effects of virus infection on the expression regulation of individual HERV-W elements in

fibroblast cells. Unlike pseudoelements, proviral elements displayed significantly elevated transcriptional activities following virus infection relative to their corresponding coding transcripts, further supporting independent and regulated expression of proviral elements.

Interestingly, the distribution pattern of HERV-W *gag* was very similar across individuals following virus infection, which exhibited an evident increase in one out of 13 categories (**paper III**). Specific HERV-W loci from this category were isolated and identified, one of which corresponds to a proviral elements investigated in **paper II**. This proviral element (chr6p22.2) is located in the intronic region of *ACOT13* and is expressed at significantly higher levels in virus infected fibroblast cells, suggesting an important and consistent role for the LTR of HERV-W in regulating transcription of proviral genome across multiple individuals.

Therefore, exogenous influenza A virus can promote transcription of members of the HERV-W family, including an intronic proviral element in *ACOT13*, in both non-placental cell-lines and primary fibroblast cultures.

4.3.2 Transactivation of *ERVWE1* by influenza A/WSN/33 virus infection

In contrast to most HERV-W loci, the promoter region in the *ERVWE1* locus has been investigated in great detail (Bonnaud et al., 2005; Cheng et al., 2004; Prudhomme et al., 2004). This locus encodes syncytin-1 (Voisset et al., 2000), the only known HERV-W gene known to be functionally adopted by the human host (Kim et al., 1999). We previously reported that *ERVWE1* can be induced in extra-placental cell-types in response to influenza A/WSN/33 virus infection by unidentified mechanisms (Nellaker et al., 2006). In **paper III**, we investigated potential factors that might contribute to the transactivation of *ERVWE1* in non-placental cells following virus infection.

Previous studies have revealed synergistic effect of 5'LTR and upstream TSE region on the promoter activity of *ERVWE1* (Prudhomme et al., 2004). The transcription factor GCM1, can bind to the TSE region and plays an important role in regulating its expression (Nait-Oumesmar et al., 2000; Yu et al., 2002). We detected inducible expression of both *ERVWE1* and *GCM1* in transformed malignant astrocytoma cell-line (CCF-STTG1 cells) and primary fibroblasts following influenza A/WSN/33 virus infection. These observations complement a previous report of up-regulation of mouse

Gcm1 following influenza A virus infection both *in vitro* and *in vivo* (Asp et al., 2007). However, studies by us and other indicate that over-expression of GCM1 is insufficient to induce transcription of *ERVWE1* in non-placenta cell-lines (Yu et al., 2002). On the other hand, siRNA analyses indicates that knock-down of GCM1 followed by virus infection suppress induction of transcription of *ERVWE1*, suggesting that GCM1 is required for the virus-driven transactivation of *ERVWE1*. Reductions of *GCM1* and *ERVWE1* transcripts do not affect levels of influenza RNA suggesting that these two genes are not critically involved in viral replication or in the host-cell defense to influenza A virus (**Figure 7**). GCM1 is thus involved in the physiological expression of *ERVWE1* in placenta-derived tissue, and the ectopic expression in other cell types.

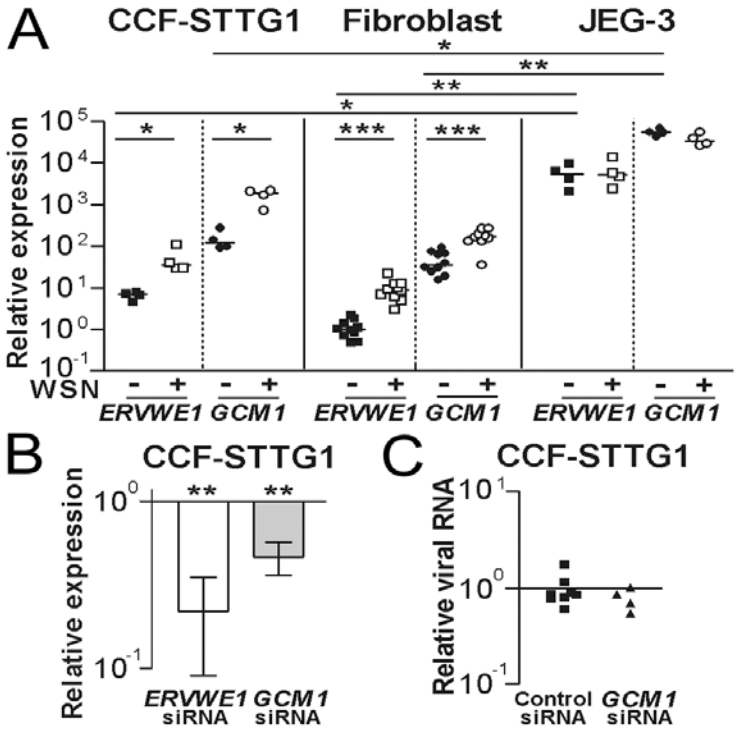

Figure 7 Transcription of *ERVWE1* and *GCM1* in response to virus infection. (A) Relative levels of *ERVWE1* transcripts (*env* exon transcripts) and transcripts encoding GCM1 were measured in uninfected and influenza A/WSN/33 virus-infected (MOI=0.5) CCF-STTG1 cells at 24 h postinfection, primary human fibroblasts at 48 h postinfection, and JEG-3 cells at 24 h postinfection. (B) Levels of *ERVWE1* and *GCM1* transcripts in CCF-STTG1 cells treated with siRNA targeted against transcripts encoding GCM1 and subsequently infected with influenza A/WSN/33 virus for 24 h relative to those in cells treated with control siRNA and infected with influenza A/WSN/33 virus. (C) Relative levels of transcripts encoding segment 8 of the influenza A/WSN/33 virus in CCF-STTG1 cells treated with scrambled control siRNA or no siRNA reagents and in cells treated with siRNA targeted against transcripts encoding GCM1. siRNA analyses data are presented as means ± SEM from four independent experiments. *, $p<0.05$; **, $p < 0.01$, ***, $p < 0.001$. (J Virol 2014; 88(8):4328-4337. doi: 10.1128/JVI.03628-13. FIG2. Figure reproduced with permission from American Society for Microbiology)

With regard to splicing of transcripts from *ERVWE1*, we demonstrated that the up-regulation of *ERVWE1* in response to a virus infection in non-placental cells appears to primarily result from increased generation of spliced transcripts. Kinetic analyses suggest that influenza A virus replication rather than the cellular stress response to starvation or the antiviral response to double-stranded RNA leads to increased levels of spliced transcript. The potential role of virus proteins in the transactivation of *ERVWE1* is further supported by the ChIP assays, see below. The fusogenic protein syncytin-1 encoded by *ERVWE1* spliced transcripts, exclusively expressed in the human placenta (Kim et al., 1999; Mi et al., 2000; Voisset et al., 2000), has been abnormally observed in diseases such as MS (Antony et al., 2007a; Antony et al., 2004) and schizophrenia (Perron et al., 2008). In addition to the detection of induced spliced *ERVWE1* transcripts by influenza A/WSN/33 virus infection in our study, several studies report transactivation of HERV-W, including syncytin-1, by EBV (Mameli et al., 2012) and HSV-1 infection (Nellaker et al., 2006; Ruprecht et al., 2006). These infections have been associated with both MS (Giovannoni et al., 2006) and schizophrenia (Karlsson, 2003), but if they causally contribute to these disorders remains to be established.

In addition to the impact of GCM1 on expression of *ERVWE1*, it has been reported that the high level of methylated CpGs in the 5'LTR and upstream TSE region of *ERVWE1* in extra-placental tissues contribute to its restricted expression in the human placenta (Gimenez et al., 2009; Matouskova et al., 2006). Interestingly, mammalian GCM proteins have been recently reported to directly be involved in active CpG demethylation (Hitoshi et al., 2011). In light of the findings that an influenza A virus infection enhances host genes expression by demethylating their promoters (Li et al., 2010; Tang et al., 2011), we next investigated the methylation status of 7 CpGs (see **Figure 3**) in the promoter and enhancer regions of *ERVWE1*. Pyrosequencing data indicate high methylation levels of these CpGs in a range of non-placenta cell-lines and fibroblast cultures, albeit with variation across 7 CpG sites. Comparably low methylation levels of two CpG sites (-32 or/and +86) of *ERVWE1* promoter are in line with previous reports of partial methylation of the *ERVWE1* enhancer-promoter region that did not completely silence promoter activity (Gimenez et al., 2009). However, no change in CpG methylation was observed in any of the cells following infection as compared to corresponding uninfected cells.

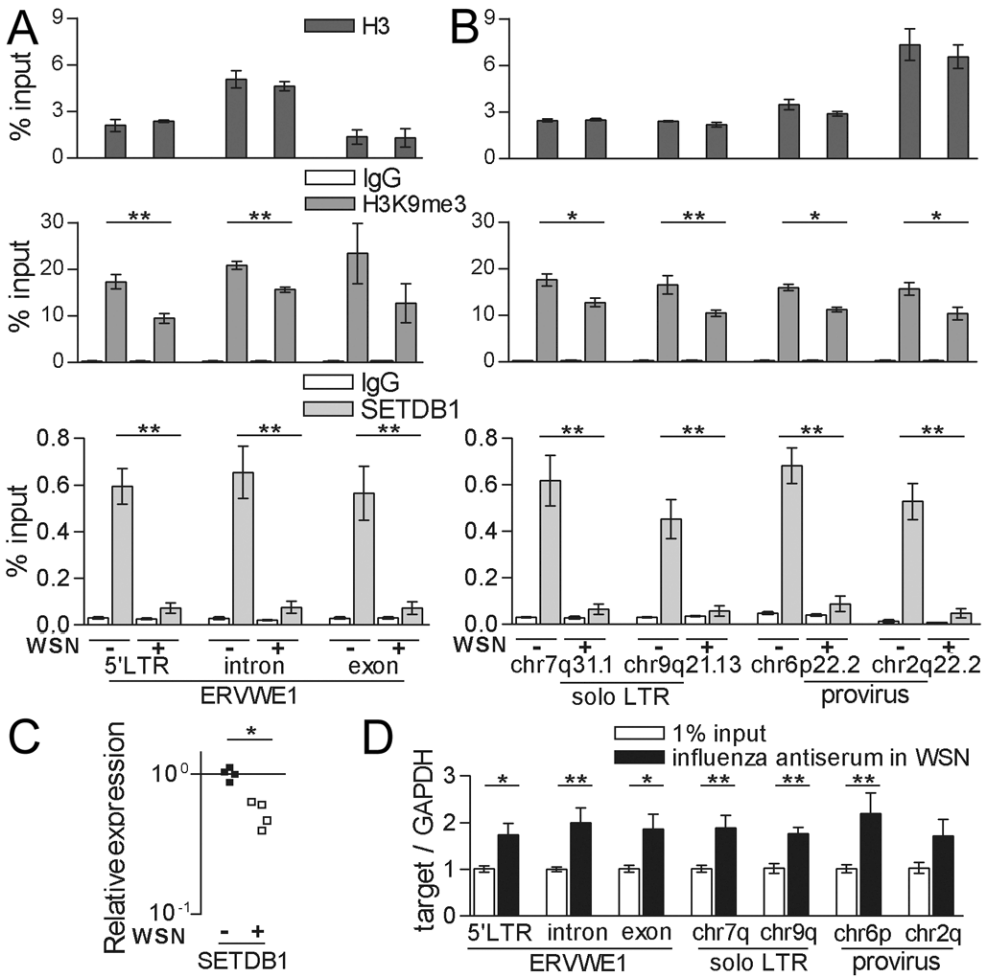

Figure 8. **Histone H3K9me3 associated with HERV-W elements during influenza A virus infection.** At 24 h postinfection, ChIP experiments were performed with rabbit IgG, antitrimethylated histone H3K9, H3 and SETDB1 antibodies on chromatin from uninfected and influenza A/WSN/33 virus-infected (MOI =0.5) CCF-STTG1 cells. Precipitated chromatin were determined by qPCR with specific assays for *ERVWE1* (A) and solo LTRs integrated in intronic regions of *ZNF277* and *C9orf85* genes and proviruses integrated in intronic regions of *KYNU* and *ACOT13* genes (B). (C) Relative levels of transcripts encoding SETDB1 were measured in uninfected and virus-infected (MOI=0.5) CCF-STTG1 cells at 24 h postinfection. (D) Levels of five HERV-W elements relative to levels of *GAPDH* in input chromatin and in chromatin immunoprecipitated with anti-influenza A/WSN/33 virus antiserum from virus-infected (MOI= 0.5) CCF-STTG1 cells. ChIP data are presented as means ± SEM from five independent experiments. *, p <0.05; ** p < 0.01. (J Virol 2014; 88(8):4328-4337. doi: 10.1128/JVI.03628-13. FIG5. Figure reproduced with permission from American Society for Microbiology)

Low transcriptional activity of *ERVWE1* in Hela cells has previously been reported to associate with enrichment of a repressive histone mark H3K9me3 in the 5' LTR region of this locus (Trejbalova et al., 2011). H3K9me2 and H3K9me3 have been reported to be critically involved in proviral silencing of ERVs in mESCs in a DNA methylation-independent manner, via H3K9-specific KMTase G9a and SETDB1 respectively (Leung et al., 2011; Matsui et al., 2010). We subsequently examined

H3K9me3 deposits and SETDB1 occupancy in *ERVWE1*, along with relative levels of transcripts encoding SETDB1, in uninfected and virus-infected CCF-STTG1 cells. In addition to the detection of decreased SETDB1 transcription by qPCR, our ChIP assays detected reduced H3K9me3 and SETDB1 levels associated with increased transcription of *ERVWE1* in infected cells, indicating that attenuation of SETDB1-mediated H3K9me3 play an important role in the transactivation of *ERVWE1* by influenza A virus infection. This correlation also applies to several other HERV-W loci including the intronic proviral element (chr6p22.2) studies in **paper II**, suggesting that the virus infection induces a global decrease in H3K9me3. Furthermore, influenza A viral protein are appeared to be more likely to bind to the HERV-W regions than to the *GAPDH* promoter, supporting previous finding that influenza encoded proteins co-localize with H3K9me3 tails during later stages of infection (Alfonso et al., 2011) (**Figure 8**).

Taken together, an influenza A virus infection can induce transcription of ERVWE1, particularly spliced transcripts encoding syncytin-1, via de-repression of SETDB1-mediated H3K9me3 on its LTR in concert with increased levels of GCM1 transcripts in CCF-STTG1 cells.

4.4 DETECTION OF HERV BY RNA SEQUENCING

Despite the observations of tissue-specific and patterned expression of HERV-W *gag* transcripts in **paper I** and regulated transcription of a number of individual HERV-W loci in **paper II**, the global expression of HERV is not well understood. In **paper IV**, we investigated the transcription of total HERV and the subfamilies HERV-W and HERV-H in particular, by analyzing publicly available RNA sequencing datasets generated from three regions of human brains; anterior cingulate cortex, hippocampus and orbitofrontal cortex, of multiple individuals.

4.4.1 Extent of HERV transcription in genome

In terms of the datasets, comparable 80~90% of total reads generated from each of three different brain regions aligned to the annotated human genome. Of all these reads, anterior cingulate cortex and hippocampus samples exhibited lower proportion of reads (25%) aligning to exon regions as compared to orbitofrontal cortex (75%) samples. The largest proportion of exons reads, concurrent with lowest proportion of intergenic reads

(13%) in the latter, is probably explained by differences in RNA and library preparation rather than biological differences between the three sequencing projects.

It has been reported that HERV sequences make up 8% of the human genome (Belshaw et al., 2005). According to our analyses, approximately 0.1~0.2% of reads in all three brain regions aligned to the 475 HERV consensus sequences. This finding is in general agreement with notion that majority of HERV elements are not being transcribed or are expressed at low levels. Despite the lowest proportion (0.1%) of HERV elements, the detection in orbitofrontal cortex samples indicates that HERV is also expressed in a highly poly-A enriched fraction of RNA. Surprisingly, the average normalized expression levels of 475 different consensus sequences were highly correlated across the three regions (range R^2 between 0.9472~0.9779) regardless of the differences observed between the libraries. The average normalized expression level for each consensus sequence as calculated based on a large numbers of reads obtained from multiple individuals for each dataset, the correlations therefore do not reflect intra-individual difference or inter-individual differences in terms of transcribed loci within different families.

4.4.2 Distribution of HERV in genome

Based on the overall HERV expression across three brain regions, we attempted to examine transcription of individual loci of interest by mapping reads which aligned to the consensus sequences of HERV17_int, LTR17 and HERVH to the human genome. The HERV-H (HERVH) family contains a larger numbers of members, which increases the chance of two elements being identical and probably explains why a smaller proportion (44~75%) of the reads to be uniquely mapped in comparison with HERV-W (HERV17_int, LTR17) family (75~98%). In addition, compared to anterior cingulate cortex (75 bp) and hippocampus (50 bp) read lengths, the orbitofrontal cortex library consisted of longer read lengths (100 bp) which also allows larger proportion of the reads to be uniquely mapped.

Next, we investigated the distribution of these repeat families across human chromosomes. From the mapping data, the proportion of reads which were uniquely mapped to each chromosome was not consistent with the proportion of integrated repeat sequences on each chromosome. This observation implies that some elements are relatively abundantly expressed whereas others are repressed.

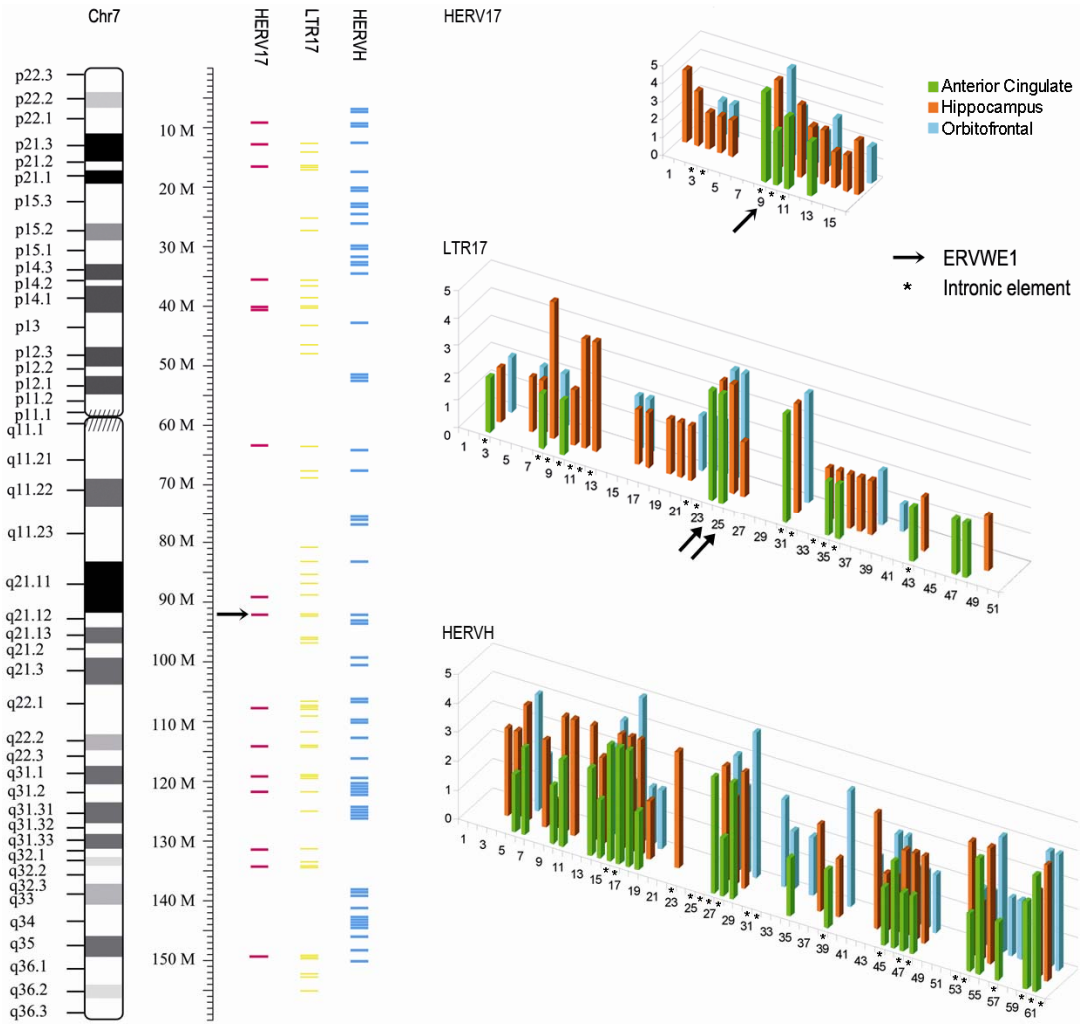

Figure 9 Relative transcript levels of repeat elements in brain samples on chromosome 7. Left, positions of individual HERV17_int, LTR17 and HERVH loci on chromosome 7. Right, relative transcription of each locus from anterior cingulate cortex, hippocampus and orbitofrontal cortex is indicated by the height of the bars. Individual loci located in introns of annotated genes are indicated by an asterisk (*) to distinguish them from loci located in intergenic regions.

To illustrate transcription of individual elements we mapped reads to individual loci along chromosome 7. Sixty-seven of HERV-W (i.e. 16 HERV17_int and 51 LTR loci) and sixty-two HERV-H loci, respectively, were identified to locate on chromosome 7. It is evident that not all elements in any one subfamily of repeats showed signs of expression. Moreover, the relative expression levels of certain loci might vary across three brain regions. In addition to intronic loci, several intergenic loci from both HERV-W and HERV-H families were expressed in three regions which in line with our previous study (Nellaker et al., 2006). Not only proviral elements but also psudoelements were expressed across different regions. This observation thus independently verifies our finding using locus-specific PCR in **paper**

II. A recent study using luciferase reporter assay reported promoter activity of HERV-W with incomplete LTRs (Schmitt et al., 2013). We detected expression from partially fragmented LTRs of two psudoelements, which correspond to those integrated in the intronic regions of genes *NRCAM* and *FOXP2* investigated in **paper II**, which further supports this finding. Of particular interest, the *ERVWE1* locus in the HERV-W family exhibited the highest levels of expression in all brain regions (**Figure 9**).

4.4.3 Transcription of HERV in disease states

The three brain regions (anterior cingulate cortex, hippocampus and orbitofrontal cortex) investigated were obtained from patients with psychiatric diagnoses and control individuals. Therefore we took this opportunity to examine the normalized expression levels of HERV across different diagnostic groups in each region separately. In the investigated RNA sequencing data, an overall increased expression levels of HERV (475 consensus sequences) were observed in autopsied anterior cingulate cortex from not only schizophrenia but also bipolar patients compared to control individuals.

Elevated levels of HERV-W transcripts were detected in anterior cingulate cortex from schizophrenia and in hippocampus from bipolar patients, confirming previous reports in postmortem brain tissues obtained by PCR (Karlsson et al., 2001) and extending these to include also bipolar disorder. Perron and coworkers reported similar findings in their recent studies conducted on blood samples (Perron et al., 2012). In addition to HERV-W, HERV-H was expressed at increased levels in anterior cingulate cortex and hippocampus of schizophrenia suggesting that other repeat families are differentially expressed in schizophrenia.

However, neither total HERV nor HERV-W and HERV-H families exhibited disease specific differences in orbitofrontal cortex libraries. Although these libraries were generated from the smallest number of individuals, they contain by far the largest number of reads per sample. Moreover, these libraries contain sequences representing exons mRNA which differ from anterior cingulate cortex and hippocampus libraries. Therefore, it is not clear if the lack of disease-related differences in the orbitofrontal cortex samples is caused by a lack of power or if these differences are detectable primarily in non-polyadenylated RNA.

Taken together, a consistent expression of HERV repetitive sequences can the detected by RNA sequencing in different brain regions regardless of the libraries preparation. This unbiased method can be used as a tool to investigate expression changes of HERV in disease states.

5 CONCLUDING REMARKS

In this thesis, application of Tm analysis, without having to rely on expensive and time-consuming sequencing, leads to a high resolution examination of the extent of transcripts originating from repetitive regions of HERV-W in various tissues and cell-types, and of changes of such transcripts in response to exogenous cues. The method can be potentially used for determining detailed expression patterns of repetitive sequences in disease states. As this approach allows detection of variations in the sequence of a limited number of products within a Tm category of interest, the PCR directed at specific loci is useful for detection of transcription of individual elements. Employing such specific PCR, we investigated the expression levels of a number of members in HERV-W family and thus concluded that diversified expression of HERV-W *gag* transcripts observed by Tm analysis comes from both LTR-directed transcription and transcriptional leakage. In light of the retained promoter activity in spite of U3 region lacking in the LTR, it appears that promoters from adjacent genes and also their own structurally incomplete LTRs contribute to the transcription initiation of psudoelements. Given the large number of HERV elements, RNA sequencing offers an unbiased technology to perform global analyses of HERV expression. A comprehensive understanding of extent of HERV repetitive sequences, especially of HERV-W and HERV-H families, in three different regions of human brains and disease states can be achieved from our analyzed sequencing data. Mapping of expressed sequences of HERV-W and HERV-H families to individual loci along chromosome 7 indicates active transcription of intronic and intergenic elements, as well as proviral elements and psudoelements, in general consistent with our observations investigated by PCR. Furthermore, elevated transcription of either total HERV or the W or the H family of HERV was observed in schizophrenia patients. Detection of HERV-W up-regulation in these patients supports previous similar findings based on PCR technique. Inasmuch as the three libraries were prepared under different protocols, these datasets might not be helpful for comparing expression levels of HERV sequence between brain regions.

In addition to characterize the extent of transcription of HERV-W, experimental studies were carried out to elucidate the mechanism by which influenza A/WSN/33 activates HERV-W elements, in particular the *ERVWE1* locus. Increased transcription levels of HERV-W members including *ERVWE1* were detected not only in immortalized cell-

lines, but also in primary fibroblast cultures of non-placental origin. In terms of our data, transactivation of *ERVWE1* in these cells in response to influenza A virus infection appears to be attributed to several mechanisms; increased levels of transcription factor *GCM1* transcripts, enhanced splicing strategy of *ERVWE1*, attenuation of H3K9me3 deposits on the *ERVWE1* promoter, yet the role of influenza viral proteins and other unknown factors is still unclear. H3K9me3 reduction is also associated with de-repression of other HERV-W elements during virus infections. Considering the fact that hypomethylation correlates to reactivation of a range of HERV members, involvement of demethylation in the virus-driven induction of HERV-W elements cannot be ruled out though this does not occur at the *ERVWE1* locus. Sequence differences in the U3 regions of individual HERV-W elements determine a series of transcription factor binding motif, still unidentified transcription factors, thus, might play an important role in the transactivation of HERV-W following virus infection.

Our observations based on the *in vitro* study are of potential relevance for understanding of conditions associated with aberrant transcription of HERV-W loci such as the psychiatric disorders; schizophrenia and bipolar disorder. The ChIP assays in this thesis reveal that chromatin changes potentially mediate the effects of environmental exposures (i.e. influenza A virus infection) on HERV-W. HERV-W expression may be a marker of chromatin changes, but also potentially directly involved in etiology or pathogenesis of human disorders. In addition to the indication of abnormal expression of HERV-W in patients, examining the differences of histone modifications between healthy and diseased individuals can help clinical researchers to address issues that are central for the development of better strategies for the prevention and treatment of these disorders.

6 ACKNOWLEDGEMENTS

Many people have contributed to this work in various ways and I want to express my deepest gratitude to all of you.

I would especially thank to my supervisor **Håkan Karlsson**. I am grateful for your patience, your earnest teaching, for giving me the opportunity to be your Ph.D student and for both inspiring research and warm living environment. As I left my far home country to pursue my educational career, I spent 6 years in this charming academic institution of Sweden. Your valuable supervision, support, encouragement and solicitude were great fortune to me. Under your advice and guidance, I am able to get through my studies and I really enjoyed the friendly atmosphere you created between us. During my whole doctorial studies periods, you remained your supervisor example ideal. I also thank to your lovely family members. Trust me, you are the best husband and father I have ever seen so far!

I would like to thank my co-supervisors:
Christoffer Nellåker, for providing a good foundation to this work. For most part of my project, Christoffer present excellent supervision. I thank you for discussions, laughs, instant solutions for tricky questions and for being great office mate. You pace me in studying industriously and exploring the science.
Robert H Yolken, for your valuable knowledge and comments on my project. The time I worked with you in US is precious and relatively short. It was fun to have those chats with you.
Krister Kristensson, for supervision in the neurovirology area. Thank you for your important suggestions in my study and for sharing your interesting leisure topics.
Katarina Luhr, for being my mentor.

All my friends and colleagues in;
Lab in the Dept. of Neuroscience at Karolinska Institutet, Sweden: Margareta Widing, Linnea Asp, Anne-Sofie Johansson, Elin Allard, Renee.Gardner, Gabriella Lundkvist, Gunnar Grant, Bo Sun, Mirjana Grujuic, Mikael Nygård, Daniel Amin, Björn Owe-Larsson, Evans, Monica, Claire, Yohannen. Thank you for working together and thanks for pleasant Fika time! I in particular thank **Margareta** for your attentively caring our lab and my daily lives.
Lab in the Dept. of Pediatrics, Stanley Division at Johns Hopkins University, US: Ou Chen, Emese O' Donnell, Sarven Sabunciyan, Ye Li, Shuojia Li Yang, Jianchun Xiao, Lorraine Jones-Brando, Emily Severance, Ann Cusic, Ruby Mason, Flora Leister, Nick Wohlgemuth, Elizabeth Rubalcaba, Kristin Gressitt, Claudia Bordon, Miranda Darby, Beyan Deuber. I am grateful to all the members for your cooperation and most grateful to **Ou** for your technical assistance and being a wonderful friend.

I want to thank our nice office neighbors: Tomas Hökfelt, Swapnali Barde, Mingdong Liu and Karin Lagerman for your friendly help. Thanks to all the members at the department.

Xiushan Wu, for introducing my interest in the bioscience research, who owns my definite appreciation and respect. With your recommendation, I had the chance to come to Sweden, followed by the fortunate encounter with Håkan and all the friends here.

My dear friends: Canhui Xu, Mingliang Wu, Citao Liu, Shuyuan Liang, Cuihua Lei, Man Zhang, Hong Cao, Sandro, Jie Lu, Zhihua Wu, Hua Xu, Cuiqin Li, Yongxing Zhao, Linzhan Wang, Xianghui Zeng, Weiwei Zhao, Nan Sheng, Chao Sun, Hidreg, Katarina, Yanling Wang, Jiang Chen, Guining Zhao, Yiran Jiao, Mingmei Shang, Shenqiu Wang, Qi Dai, Xinyi Yuan, Jia Liu, Chunyu Li, Ping, Yan, Lei, Xinhua, Jie, Kun, Kelvin. Thank you all for your friendship and all the fun times with you.

My parents, especially my mother, thank you for your constant understanding and support for my study. I love you! I always have and always will. My dear aunties, uncles, cousins and grandmother, thank you for caring and support.

This study was supported by;
The Stanley Medical Research Institute
The Swedish Brain Foundation
The Swedish Research Council (21X-20047)
The Karolinska Institutet Foundation for Virus Research
Karolinska Institutet travel allowance and Hirsch Fellowship for surgeons

7 REFERENCES

Alfonso, R., Lutz, T., Rodriguez, A., Chavez, J.P., Rodriguez, P., Gutierrez, S., and Nieto, A. (2011). CHD6 chromatin remodeler is a negative modulator of influenza virus replication that relocates to inactive chromatin upon infection. Cell Microbiol *13*, 1894-1906.

Antony, J.M., Ellestad, K.K., Hammond, R., Imaizumi, K., Mallet, F., Warren, K.G., and 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*, 1210-1224.

Antony, J.M., Izad, M., Bar-Or, A., Warren, K.G., Vodjgani, M., Mallet, F., and Power, C. (2006). Quantitative analysis of human endogenous retrovirus-W env in neuroinflammatory diseases. AIDS Res Hum Retroviruses *22*, 1253-1259.

Antony, J.M., van Marle, G., Opii, W., Butterfield, D.A., Mallet, F., Yong, V.W., Wallace, J.L., Deacon, R.M., Warren, K., and Power, C. (2004). Human endogenous retrovirus glycoprotein-mediated induction of redox reactants causes oligodendrocyte death and demyelination. Nat Neurosci *7*, 1088-1095.

Antony, J.M., Zhu, Y., Izad, M., Warren, K.G., Vodjgani, M., Mallet, F., and Power, C. (2007b). Comparative expression of human endogenous retrovirus-W genes in multiple sclerosis. Aids Res Hum Retrov *23*, 1251-1256.

Asp, L., Nellaker, C., and Karlsson, H. (2007). Influenza A virus transactivates the mouse envelope gene encoding syncytin B and its regulator, glial cells missing 1. J Neurovirol *13*, 29-37.

Bannert, N., and Kurth, R. (2004). Retroelements and the human genome: new perspectives on an old relation. Proc Natl Acad Sci U S A *101 Suppl 2*, 14572-14579.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Belshaw, R., Katzourakis, A., Paces, J., Burt, A., and Tristem, M. (2005). High Copy Number in Human Endogenous Retrovirus Families is Associated with Copying Mechanisms in Addition to Reinfection. Mol Biol Evol *22*, 814-817.

Bjerregaard, B., Holck, S., Christensen, I.J., and Larsson, L.I. (2006). Syncytin is involved in breast cancer-endothelial cell fusions. Cell Mol Life Sci *63*, 1906-1911.

Blomstrom, A., Karlsson, H., Wicks, S., Yang, S.J., Yolken, R.H., and Dalman, C. (2012). Maternal antibodies to infectious agents and risk for non-affective psychoses in the offspring-a matched case-control study. Schizophrenia Research *140*, 25-30.

Blond, J.L., Beseme, F., Duret, L., Bouton, O., Bedin, F., Perron, H., Mandrand, B., and Mallet, F. (1999). Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family. J Virol *73*, 1175-1185.

Blond, J.L., Lavillette, D., Cheynet, V., Bouton, O., Oriol, G., Chapel-Fernandes, S., Mandrand, B., Mallet, F., and Cosset, F.L. (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*, 3321-3329.

Boeke, J.D., and Stoye, J.P. (1997). Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements. Edited by Coffin JM HS, Varmus HE. Retroviruses. Cold Spring Harbor Laboratory Press (NY).

Bonnaud, B., Beliaeff, J., Bouton, O., Oriol, G., Duret, L., and Mallet, F. (2005). Natural history of the ERVWE1 endogenous retroviral locus. Retrovirology *2*, 57.

Brown, A.S., and Derkits, E.J. (2010). Prenatal infection and schizophrenia: a review of epidemiologic and translational studies. Am J Psychiatry *167*, 261-280.

Buzdin, A., Kovalskaya-Alexandrova, E., Gogvadze, E., and Sverdlov, E. (2006). At least 50% of human-specific HERV-K (HML-2) long terminal repeats serve in

vivo as active promoters for host nonrepetitive DNA transcription. J Virol *80*, 10752-10762.
Chang, C.W., Chuang, H.C., Yu, C.C., Yao, T.P., and Chen, H.W. (2005). Stimulation of GCMa transcriptional activity by cyclic AMP/protein kinase a signaling is attributed to CBP-Mediated acetylation of GCMa. Mol Cell Biol *25*, 8401-8414.
Chen, C.P., Chen, C.Y., Yang, Y.C., Su, T.H., and Chen, H. (2004). Decreased placental GCM1 (glial cells missing) gene expression in pre-eclampsia. Placenta *25*, 413-421.
Cheng, Y.H., and Handwerger, S. (2005). A Placenta-Specific Enhancer of the Human Syncytin Gene. Biol Reprod.
Cheng, Y.H., Richardson, B.D., Hubert, M.A., and Handwerger, S. (2004). Isolation and characterization of the human syncytin gene promoter. Biol Reprod *70*, 694-701.
Cheynet, V., Oriol, G., and Mallet, F. (2006). Identification of the hASCT2-binding domain of the Env ERVWE1/syncytin-1 fusogenic glycoprotein. Retrovirology *3*, 41.
Chiang, M.H., Liang, F.Y., Chen, C.P., Chang, C.W., Cheong, M.L., Wang, L.J., Liang, C.Y., Lin, F.Y., Chou, C.C., and Chen, H. (2009). Mechanism of hypoxia-induced GCM1 degradation: implications for the pathogenesis of preeclampsia. J Biol Chem *284*, 17411-17419.
Christensen, T. (2010). HERVs in neuropathogenesis. J Neuroimmune Pharmacol *5*, 326-335.
Chu, Y., and Corey, D.R. (2012). RNA sequencing: platform selection, experimental design, and data interpretation. Nucleic Acid Ther *22*, 271-274.
Coffin, J.M., Hughes, S.H., and Varmus, H.E. (1997). Retroviruses (New York, Cold Spring Harbor Laboratory Press).
Colella, S., Shen, L., Baggerly, K.A., Issa, J.P., and Krahe, R. (2003). Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites. Biotechniques *35*, 146-150.
Compston, A., and Coles, A. (2002). Multiple sclerosis. Lancet *359*, 1221-1231.
Conley, A.B., Piriyapongsa, J., and Jordan, I.K. (2008). Retroviral promoters in the human genome. Bioinformatics *24*, 1563-1567.
Costas, J. (2002). Characterization of the intragenomic spread of the human endogenous retrovirus family HERV-W. Mol Biol Evol *19*, 526-533.
Dalman, C., Allebeck, P., Cullberg, J., Grunewald, C., and Koster, M. (1999). Obstetric complications and the risk of schizophrenia: a longitudinal study of a national birth cohort. Arch Gen Psychiatry *56*, 234-240.
de Parseval, N., Lazar, V., Casella, J.F., Benit, L., and 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*, 10414-10422.
Dewannieux, M., Blaise, S., and Heidmann, T. (2005). Identification of a functional envelope protein from the HERV-K family of human endogenous retroviruses. J Virol *79*, 15573-15577.
Dewannieux, M., Harper, F., Richaud, A., Letzelter, C., Ribet, D., Pierron, G., and Heidmann, T. (2006). Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. Genome Res *16*, 1548-1556.
Dolei, A., and 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*, 4-13.
Dong, K.B., Maksakova, I.A., Mohn, F., Leung, D., Appanah, R., Lee, S., Yang, H.W., Lam, L.L., Mager, D.L., Schubeler, D., *et al.* (2008). DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. EMBO J *27*, 2691-2701.
Esteki-Zadeh, A., Karimi, M., Straat, K., Ammerpohl, O., Zeitelhofer, M., Jagodic, M., Mehrab-Mohseni, M., Sjoholm, L., Rahbar, A., Soderberg-Naucler, C., *et al.* (2012). Human cytomegalovirus infection is sensitive to the host cell DNA

methylation state and alters global DNA methylation capacity. Epigenetics *7*, 585-593.

Feng, J., Zhou, Y., Campbell, S.L., Le, T., Li, E., Sweatt, J.D., Silva, A.J., and Fan, G. (2010). Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. Nature Neuroscience *13*, 423-430.

Forsman, A., Yun, Z., Hu, L., Uzhameckis, D., Jern, P., and Blomberg, J. (2005). Development of broadly targeted human endogenous gammaretroviral pol-based real time PCRs Quantitation of RNA expression in human tissues. J Virol Methods *129*, 16-30.

Frank, O., Giehl, M., Zheng, C., Hehlmann, R., Leib-Mosch, C., and Seifarth, W. (2005). Human endogenous retrovirus expression profiles in samples from brains of patients with schizophrenia and bipolar disorders. J Virol *79*, 10890-10901.

Frank, O., Verbeke, C., Schwarz, N., Mayer, J., Fabarius, A., Hehlmann, R., Leib-Mosch, C., and Seifarth, W. (2008). Variable transcriptional activity of endogenous retroviruses in human breast cancer. J Virol *82*, 1808-1818.

Frommer, M., Mcdonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. (1992). A Genomic Sequencing Protocol That Yields a Positive Display of 5-Methylcytosine Residues in Individual DNA Strands. P Natl Acad Sci USA *89*, 1827-1831.

Gao, Y., He, Z.M., Wang, Z.L., Luo, Y.M., Sun, H.Y., Zhou, Y., Huang, L.H., Li, M.C., Fang, Q., and Jiang, S.W. (2012). Increased Expression and Altered Methylation of HERVWE1 in the Human Placentas of Smaller Fetuses from Monozygotic, Dichorionic, Discordant Twins. Plos One *7*.

Gimenez, J., Montgiraud, C., Oriol, G., Pichon, J.P., Ruel, K., Tsatsaris, V., Gerbaud, P., Frendo, J.L., Evain-Brion, D., and Mallet, F. (2009). Comparative methylation of ERVWE1/syncytin-1 and other human endogenous retrovirus LTRs in placenta tissues. DNA Res *16*, 195-211.

Gimenez, J., Montgiraud, C., Pichon, J.-P., Bonnaud, B., Arsac, M., Ruel, K., Bouton, O., and 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 research *38*, 2229-2246.

Giovannoni, G., Cutter, G.R., Lunemann, J., Martin, R., Munz, C., Sriram, S., Steiner, I., Hammerschlag, M.R., and Gaydos, C.A. (2006). Infectious causes of multiple sclerosis. Lancet Neurol *5*, 887-894.

Gogvadze, E., Stukacheva, E., Buzdin, A., and Sverdlov, E. (2009). Human-specific modulation of transcriptional activity provided by endogenous retroviral insertions. J Virol *83*, 6098-6105.

Gong, R., Peng, X., Kang, S., Feng, H., Huang, J., Zhang, W., Lin, D., Tien, P., and Xiao, G. (2005). Structural characterization of the fusion core in syncytin, envelope protein of human endogenous retrovirus family W. Biochem Biophys Res Commun *331*, 1193-1200.

Griffiths, D.J. (2001). Endogenous retroviruses in the human genome sequence. Genome Biol *2*, REVIEWS1017.

Hale, B.G., Randall, R.E., Ortin, J., and Jackson, D. (2008). The multifunctional NS1 protein of influenza A viruses. J Gen Virol *89*, 2359-2376.

Hayatsu, H., Wataya, Y., and Kazushige, K. (1970). The addition of sodium bisulfite to uracil and to cytosine. J Am Chem Soc *92*, 724-726.

Hitoshi, S., Ishino, Y., Kumar, A., Jasmine, S., Tanaka, K.F., Kondo, T., Kato, S., Hosoya, T., Hotta, Y., and Ikenaka, K. (2011). Mammalian Gcm genes induce Hes5 expression by active DNA demethylation and induce neural stem cells. Nature Neuroscience *14*, 957-U935.

Hu, L., Hornung, D., Kurek, R., Ostman, H., Blomberg, J., and Bergqvist, A. (2006). Expression of human endogenous gammaretroviral sequences in endometriosis and ovarian cancer. AIDS Res Hum Retroviruses *22*, 551-557.

Huang, Q., Li, J., Wang, F., Oliver, M.T., Tipton, T., Gao, Y., and Jiang, S.W. (2013). Syncytin-1 modulates placental trophoblast cell proliferation by promoting G1/S transition. Cell Signal *25*, 1027-1035.

Huang, W., Li, S., Hu, Y., Yu, H., Luo, F., Zhang, Q., and Zhu, F. (2011). Implication of the env gene of the human endogenous retrovirus W family in the expression of BDNF and DRD3 and development of recent-onset schizophrenia. Schizophr Bull *37*, 988-1000.

Huang, W.J., Liu, Z.C., Wei, W., Wang, G.H., Wu, J.G., and Zhu, F. (2006). Human endogenous retroviral pol RNA and protein detected and identified in the blood of individuals with schizophrenia. Schizophr Res *83*, 193-199.

Iseki, H., Shimizukawa, R., Sugiyama, F., Kunita, S., Iwama, A., Onodera, M., Nakauchi, H., and Yagami, K. (2005). Parvovirus nonstructural proteins induce an epigenetic modification through histone acetylation in host genes and revert tumor malignancy to benignancy. J Virol *79*, 8886-8893.

Iwasaki, Y., Hosoya, T., Takebayashi, H., Ogawa, Y., Hotta, Y., and Ikenaka, K. (2003). The potential to induce glial differentiation is conserved between Drosophila and mammalian glial cells missing genes. Development *130*, 6027-6035.

Iyer, V.R., Eisen, M.B., Ross, D.T., Schuler, G., Moore, T., Lee, J.C., Trent, J.M., Staudt, L.M., Hudson, J., Jr., Boguski, M.S., *et al.* (1999). The transcriptional program in the response of human fibroblasts to serum. Science *283*, 83-87.

Jackson, V. (1978). Studies on histone organization in the nucleosome using formaldehyde as a reversible cross-linking agent. Cell *15*, 945-954.

Jern, P., and Coffin, J.M. (2008). Effects of retroviruses on host genome function. Annu Rev Genet *42*, 709-732.

Johnston, J.B., Silva, C., Holden, J., Warren, K.G., Clark, A.W., and Power, C. (2001). Monocyte activation and differentiation augment human endogenous retrovirus expression: implications for inflammatory brain diseases. Ann Neurol *50*, 434-442.

Jurka, J. (2000). Repbase update: a database and an electronic journal of repetitive elements. Trends Genet *16*, 418-420.

Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M., *et al.* (2011). DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. Cell Stem Cell *8*, 676-687.

Karlsson, H. (2003). Viruses and schizophrenia, connection or coincidence? Neuroreport *14*, 535-542.

Karlsson, H., Bachmann, S., Schroder, J., McArthur, J., Torrey, E.F., and Yolken, R.H. (2001). Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia. Proc Natl Acad Sci U S A *98*, 4634-4639.

Karlsson, H., Schroder, J., Bachmann, S., Bottmer, C., and Yolken, R.H. (2004). HERV-W-related RNA detected in plasma from individuals with recent-onset schizophrenia or schizoaffective disorder. Mol Psychiatry *9*, 12-13.

Katzourakis, A., and Tristem, M. (2005). Phylogeny of human endogenous and exogenous retroviruses.), Retroviruses and Primate Genome Evolution Landes Bioscience, Georgetown, 186-203.

Kent, W.J. (2002). BLAT--the BLAST-like alignment tool. Genome Res *12*, 656-664.

Khodosevich, K., Lebedev, Y., and Sverdlov, E. (2002). Endogenous retroviruses and human evolution. Comp Funct Genomics *3*, 494-498.

Kim, H.S., Takenaka, O., and Crow, T.J. (1999). Isolation and phylogeny of endogenous retrovirus sequences belonging to the HERV-W family in primates. J Gen Virol *80 (Pt 10)*, 2613-2619.

Kim, K.C., Friso, S., and Choi, S.W. (2009). DNA methylation, an epigenetic mechanism connecting folate to healthy embryonic development and aging. Journal of Nutritional Biochemistry *20*, 917-926.

Knerr, I., Weigel, C., Linnemann, K., Dotsch, J., Meissner, U., Fusch, C., and Rascher, W. (2003). Transcriptional effects of hypoxia on fusiogenic syncytin and its receptor ASCT2 in human cytotrophoblast BeWo cells and in ex vivo perfused placental cotyledons. Am J Obstet Gynecol *189*, 583-588.

Knipe, D., Howley, P., Griffin, D., Lamb, R., Martin, M., and Roizman, B. (2001). Fields virology. Philadelphia (EUA): Lippincott Williams & Wilkins.

Kudo, Y., Boyd, C.A., Sargent, I.L., and Redman, C.W. (2003). Hypoxia alters expression and function of syncytin and its receptor during trophoblast cell fusion of human placental BeWo cells: implications for impaired trophoblast syncytialisation in pre-eclampsia. Biochim Biophys Acta *1638*, 63-71.

Kurth, R., and Bannert, N. (2010). Beneficial and detrimental effects of human endogenous retroviruses. Int J Cancer *126*, 306-314.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Langbein, M., Strick, R., Strissel, P.L., Vogt, N., Parsch, H., Beckmann, M.W., and Schild, R.L. (2008). Impaired cytotrophoblast cell–cell fusion is associated with reduced Syncytin and increased apoptosis in patients with placental dysfunction. Molecular reproduction and development *75*, 175-183.

Laufer, G., Mayer, J., Mueller, B.F., Mueller-Lantzsch, N., and 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., and 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*, 8788-8798.

Lavillette, D., Marin, M., Ruggieri, A., Mallet, F., Cosset, F.L., and Kabat, D. (2002). The envelope glycoprotein of human endogenous retrovirus type W uses a divergent family of amino acid transporters/cell surface receptors. J Virol *76*, 6442-6452.

Lee, Y.N., and Bieniasz, P.D. (2007). Reconstitution of an infectious human endogenous retrovirus. PLoS Pathog *3*, e10.

Leung, D.C., Dong, K.B., Maksakova, I.A., Goyal, P., Appanah, R., Lee, S., Tachibana, M., Shinkai, Y., Lehnertz, B., Mager, D.L., *et al.* (2011). Lysine methyltransferase G9a is required for de novo DNA methylation and the establishment, but not the maintenance, of proviral silencing. Proc Natl Acad Sci U S A *108*, 5718-5723.

Li, W., Sun, W., Liu, L., Yang, F., Li, Y.K., Chen, Y.N., Fang, J.L., Zhang, W.J., Wu, J.G., and Zhu, Y. (2010). IL-32: A Host Proinflammatory Factor against Influenza Viral Replication Is Upregulated by Aberrant Epigenetic Modifications during Influenza A Virus Infection. Journal of Immunology *185*, 5056-5065.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402-408.

Loregger, T., Pollheimer, J., and Knofler, M. (2003). Regulatory transcription factors controlling function and differentiation of human trophoblast--a review. Placenta *24 Suppl A*, S104-110.

Lower, R., Lower, J., Tondera-Koch, C., and Kurth, R. (1993). A general method for the identification of transcribed retrovirus sequences (R-U5 PCR) reveals the expression of the human endogenous retrovirus loci HERV-H and HERV-K in teratocarcinoma cells. Virology *192*, 501-511.

Macnab, J.C., Adams, R.L., Rinaldi, A., Orr, A., and Clark, L. (1988). Hypomethylation of host cell DNA synthesized after infection or transformation of cells by herpes simplex virus. Mol Cell Biol *8*, 1443-1448.

Mager, D.L., and Freeman, J.D. (1995). HERV-H endogenous retroviruses: presence in the New World branch but amplification in the Old World primate lineage. Virology *213*, 395-404.

Maliniemi, P., Vincendeau, M., Mayer, J., Frank, O., Hahtola, S., Karenko, L., Carlsson, E., Mallet, F., Seifarth, W., and Leib-Mösch, C. (2013). Expression of Human Endogenous Retrovirus-W Including Syncytin-1 in Cutaneous T-Cell Lymphoma. PloS one *8*, e76281.

Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G., Bonnaud, B., Lucotte, G., Duret, L., and 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*, 1731-1736.
Mameli, G., Astone, V., Arru, G., Marconi, S., Lovato, L., Serra, C., Sotgiu, S., Bonetti, B., and Dolei, A. (2007a). 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*, 264-274.
Mameli, G., Astone, V., Khalili, K., Serra, C., Sawaya, B.E., and Dolei, A. (2007b). Regulation of the syncytin-1 promoter in human astrocytes by multiple sclerosis-related cytokines. Virology *362*, 120-130.
Mameli, G., Poddighe, L., Mei, A., Uleri, E., Sotgiu, S., Serra, C., Manetti, R., and Dolei, A. (2012). Expression and Activation by Epstein Barr Virus of Human Endogenous Retroviruses-W in Blood Cells and Astrocytes: Inference for Multiple Sclerosis. Plos One *7*.
Mangeney, M., Renard, M., Schlecht-Louf, G., Bouallaga, I., Heidmann, O., Letzelter, C., Richaud, A., Ducos, B., and Heidmann, T. (2007). Placental syncytins: Genetic disjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins. P Natl Acad Sci USA *104*, 20534-20539.
Mao, H., Lv, Z., and Ho, M.S. (2012). Gcm proteins function in the developing nervous system. Dev Biol *370*, 63-70.
Marazzi, I., Ho, J.S.Y., Kim, J., Manicassamy, B., Dewell, S., Albrecht, R.A., Seibert, C.W., Schaefer, U., Jeffrey, K.L., Prinjha, R.K.*, et al.* (2012). Suppression of the antiviral response by an influenza histone mimic. Nature *483*, 428-433.
Matouskova, M., Blazkova, J., Pajer, P., Pavlicek, A., and Hejnar, J. (2006). CpG methylation suppresses transcriptional activity of human syncytin-1 in non-placental tissues. Exp Cell Res *312*, 1011-1020.
Matsui, T., Leung, D., Miyashita, H., Maksakova, I.A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M.C., and Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature *464*, 927-U149.
Mayer, J., and Meese, E. (2005). Human endogenous retroviruses in the primate lineage and their influence on host genomes. Cytogenet Genome Res *110*, 448-456.
Menendez, L., Benigno, B.B., and McDonald, J.F. (2004). L1 and HERV-W retrotransposons are hypomethylated in human ovarian carcinomas. Molecular cancer *3*, 12.
Mi, S., Lee, X., Li, X., Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.Y., Edouard, P., Howes, S.*, et al.* (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature *403*, 785-789.
Mortensen, P.B., Pedersen, C.B., Hougaard, D.M., Norgaard-Petersen, B., Mors, O., Borglum, A.D., and Yolken, R.H. (2010). A Danish National Birth Cohort study of maternal HSV-2 antibodies as a risk factor for schizophrenia in their offspring. Schizophrenia Research *122*, 257-263.
Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol *51 Pt 1*, 263-273.
Nait-Oumesmar, B., Copperman, A.B., and Lazzarini, R.A. (2000). Placental expression and chromosomal localization of the human Gcm 1 gene. J Histochem Cytochem *48*, 915-922.
Nellaker, C., Uhrzander, F., Tyrcha, J., and Karlsson, H. (2008). Mixture models for analysis of melting temperature data. BMC Bioinformatics *9*, 370.
Nellaker, C., Wallgren, U., and Karlsson, H. (2007). Molecular beacon-based temperature control and automated analyses for improved resolution of melting temperature analysis using SYBR I green chemistry. Clin Chem *53*, 98-103.
Nellaker, C., Yao, Y., Jones-Brando, L., Mallet, F., Yolken, R.H., and Karlsson, H. (2006). Transactivation of elements in the human endogenous retrovirus W family by viral infection. Retrovirology *3*, 44.
O'Reilly, R.L., and Singh, S.M. (1996). Retroviruses and schizophrenia revisited. Am J Med Genet *67*, 19-24.

Parboosing, R., Bao, Y., Shen, L., Schaefer, C.A., and Brown, A.S. (2013). Gestational influenza and bipolar disorder in adult offspring. JAMA Psychiatry *70*, 677-685.

Pavlicek, A., Paces, J., Elleder, D., and Hejnar, J. (2002). Processed pseudogenes of human endogenous retroviruses generated by LINEs: their integration, stability, and distribution. Genome Res *12*, 391-399.

Perron, H., Garson, J.A., Bedin, F., Beseme, F., Paranhos-Baccala, G., Komurian-Pradel, F., Mallet, F., Tuke, P.W., Voisset, C., Blond, J.L., *et al.* (1997). Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. The Collaborative Research Group on Multiple Sclerosis. Proc Natl Acad Sci U S A *94*, 7583-7588.

Perron, H., Geny, C., Laurent, A., Mouriquand, C., Pellat, J., Perret, J., and Seigneurin, J.M. (1989). Leptomeningeal cell line from multiple sclerosis with reverse transcriptase activity and viral particles. Res Virol *140*, 551-561.

Perron, H., Hamdani, N., Faucard, R., Lajnef, M., Jamain, S., Daban-Huard, C., Sarrazin, S., LeGuen, E., Houenou, J., Delavest, M., *et al.* (2012). Molecular characteristics of Human Endogenous Retrovirus type-W in schizophrenia and bipolar disorder. Transl Psychiatry *2*, e201.

Perron, H., Lazarini, F., Ruprecht, K., Péchoux-Longin, C., Seilhean, D., Sazdovitch, V., Créange, A., Battail-Poirot, N., Sibaï, G., Santoro, L., *et al.* (2005). Human endogenous retrovirus (HERV)-W ENV and GAG proteins: Physiological expression in human brain and pathophysiological modulation in multiple sclerosis lesions. J Neurovirol *11*, 23-33.

Perron, H., Mekaoui, L., Bernard, C., Veas, F., Stefas, I., and Leboyer, M. (2008). Endogenous Retrovirus Type W GAG and Envelope Protein Antigenemia in Serum of Schizophrenic Patients. Biological Psychiatry *64*, 1019-1023.

Pichon, J.P., Bonnaud, B., Cleuziat, P., and Mallet, F. (2006). Multiplex degenerate PCR coupled with an oligo sorbent array for human endogenous retrovirus expression profiling. Nucleic Acids Res *34*, e46.

Prudhomme, S., Oriol, G., and Mallet, F. (2004). A retroviral promoter and a cellular enhancer define a bipartite element which controls env ERVWE1 placental expression. J Virol *78*, 12157-12168.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res *35*, D61-65.

Ronaghi, M., Uhlen, M., and Nyren, P. (1998). A sequencing method based on real-time pyrophosphate. Science *281*, 363, 365.

Rosenfeld, J.A., Wang, Z., Schones, D.E., Zhao, K., DeSalle, R., and Zhang, M.Q. (2009). Determination of enriched histone modifications in non-genic portions of the human genome. BMC Genomics *10*, 143.

Ruebner, M., Strissel, P.L., Ekici, A.B., Stiegler, E., Dammer, U., Goecke, T.W., Faschingbauer, F., Fahlbusch, F.B., Beckmann, M.W., and Strick, R. (2013). Reduced syncytin-1 expression levels in placental syndromes correlates with epigenetic hypermethylation of the ERVW-1 promoter region. PLoS One *8*, e56145.

Ruprecht, K., Obojes, K., Wengel, V., Gronen, F., Kim, K.S., Perron, H., Schneider-Schaulies, J., and Rieckmann, P. (2006). Regulation of human endogenous retrovirus W protein expression by herpes simplex virus type 1: implications for multiple sclerosis. J Neurovirol *12*, 65-71.

Ryan, J.L., Jones, R.J., Kenney, S.C., Rivenbark, A.G., Tang, W., Knight, E.R., Coleman, W.B., and Gulley, M.L. (2010). Epstein-Barr virus-specific methylation of human genes in gastric cancer cells. Infect Agent Cancer *5*, 27.

Schmitt, K., Richter, C., Backes, C., Meese, E., Ruprecht, K., and 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*, 13837-13852.

Schon, U., Seifarth, W., Baust, C., Hohenadl, C., Erfle, V., and Leib-Mosch, C. (2001). Cell type-specific expression and promoter activity of human endogenous retroviral long terminal repeats. Virology *279*, 280-291.

Seifarth, W., Frank, O., Zeilfelder, U., Spiess, B., Greenwood, A.D., Hehlmann, R., and Leib-Mosch, C. (2005). Comprehensive analysis of human endogenous retrovirus transcriptional activity in human tissues with a retrovirus-specific microarray. J Virol *79*, 341-352.

Sekigawa, I., Ogasawara, H., Naito, T., Kaneko, H., Hishikawa, T., and Hashimoto, H. (2003). Systemic lupus erythematosus and human endogenous retroviruses. Mod Rheumatol *13*, 107-113.

Shyamsundar, R., Kim, Y.H., Higgins, J.P., Montgomery, K., Jorden, M., Sethuraman, A., van de Rijn, M., Botstein, D., Brown, P.O., and Pollack, J.R. (2005). A DNA microarray survey of gene expression in normal human tissues. Genome Biol *6*, R22.

Smit, A.F. (1993). Identification of a new, abundant superfamily of mammalian LTR-transposons. Nucleic Acids Res *21*, 1863-1872.

Solomon, M.J., Larsen, P.L., and Varshavsky, A. (1988). Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. Cell *53*, 937-947.

Stauffer, Y., Theiler, G., Sperisen, P., Lebedev, Y., and Jongeneel, C.V. (2004). Digital expression profiles of human endogenous retroviral families in normal and cancerous tissues. Cancer Immun *4*.

Stoye, J.P. (2001). Endogenous retroviruses: still active after all these years? Curr Biol *11*, R914-916.

Strick, R., Ackermann, S., Langbein, M., Swiatek, J., Schubert, S.W., Hashemolhosseini, S., Koscheck, T., Fasching, P.A., Schild, R.L., Beckmann, M.W., *et al.* (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 *85*, 23-38.

Strissel, P.L., Ruebner, M., Thiel, F., Wachter, D., Ekici, A.B., Wolf, F., Thieme, F., Ruprecht, K., Beckmann, M.W., and Strick, R. (2012). Reactivation of codogenic endogenous retroviral (ERV) envelope genes in human endometrial carcinoma and prestages: Emergence of new molecular targets. Oncotarget *3*, 1204-1219.

Subramanian, R.P., Wildschutte, J.H., Russo, C., and Coffin, J.M. (2011). Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. Retrovirology *8*, 90.

Sun, Y., Ouyang, D.Y., Pang, W., Tu, Y.Q., Li, Y.Y., Shen, X.M., Tam, S.C., Yang, H.Y., and Zheng, Y.T. (2010). Expression of syncytin in leukemia and lymphoma cells. Leuk Res *34*, 1195-1202.

Tang, B.K., Zhao, R.H., Sun, Y., Zhu, Y., Zhong, J.A., Zhao, G.P., and Zhu, N.S. (2011). Interleukin-6 expression was regulated by epigenetic mechanisms in response to influenza virus infection or dsRNA treatment. Molecular Immunology *48*, 1001-1008.

Tian, Y., Ni, D., Yang, W., Zhang, Y., Zhao, K., Song, J., Mao, Q., Tian, Z., van Velkinburgh, J.C., Yang, D., *et al.* (2013). Telbivudine treatment corrects HBV-induced epigenetic alterations in liver cells of patients with chronic hepatitis B. Carcinogenesis.

Tolosa, J.M., Schjenken, J.E., Clifton, V.L., Vargas, A., Barbeau, B., Lowry, P., Maiti, K., and Smith, R. (2012). The endogenous retroviral envelope protein syncytin-1 inhibits LPS/PHA-stimulated cytokine responses in human blood and is sorted into placental exosomes. Placenta *33*, 933-941.

Torrey, E.F., Miller, J., Rawlings, R., and Yolken, R.H. (1997). Seasonality of births in schizophrenia and bipolar disorder: a review of the literature. Shizophrenia Research *28*, 1-38.

Trejbalova, K., Blazkova, J., Matouskova, M., Kucerova, D., Pecnova, L., Vernerova, Z., Heracek, J., Hirsch, I., and Hejnar, J. (2011). Epigenetic regulation of transcription and splicing of syncytins, fusogenic glycoproteins of retroviral origin. Nucleic Acids Research *39*, 8728-8739.

Tristem, M. (2000). Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. J Virol *74*, 3715-3730.

Tyekucheva, S., Yolken, R.H., McCombie, W.R., Parla, J., Kramer, M., Wheelan, S.J., and Sabunciyan, S. (2011). Establishing the baseline level of repetitive element expression in the human cortex. BMC Genomics *12*, 495.

van de Lagemaat, L.N., Medstrand, P., and Mager, D.L. (2006). Multiple effects govern endogenous retrovirus survival patterns in human gene introns. Genome Biol *7*, R86.

Vassena, R., Schramm, R.D., and Latham, K.E. (2005). Species-dependent expression patterns of DNA methyltransferase genes in mammalian oocytes and preimplantation embryos. Molecular Reproduction and Development *72*, 430-436.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., *et al.* (2001). The sequence of the human genome. Science *291*, 1304-1351.

Villesen, P., Aagaard, L., Wiuf, C., and Pedersen, F.S. (2004). Identification of endogenous retroviral reading frames in the human genome. Retrovirology *1*, 32.

Vivekanandan, P., Daniel, H.D.J., Kannangai, R., Martinez-Murillo, F., and Torbenson, M. (2010). Hepatitis B Virus Replication Induces Methylation of both Host and Viral DNA. Journal of Virology *84*, 4321-4329.

Voisset, C., Blancher, A., Perron, H., Mandrand, B., Mallet, F., and Paranhos-Baccala, G. (1999). Phylogeny of a novel family of human endogenous retrovirus sequences, HERV-W, in humans and other primates. AIDS Res Hum Retroviruses *15*, 1529-1533.

Voisset, C., Bouton, O., Bedin, F., Duret, L., Mandrand, B., Mallet, F., and Paranhos-Baccala, G. (2000). Chromosomal distribution and coding capacity of the human endogenous retrovirus HERV-W family. AIDS Res Hum Retroviruses *16*, 731-740.

Yao, Y., Schröder, J., Nellåker, C., Bottmer, C., Bachmann, S., Yolken, R.H., and Karlsson, H. (2007). Elevated levels of human endogenous retrovirus-W transcripts in blood cells from patients with first episode schizophrenia. Genes Brain Behav *7*, 103-112.

Yi, J.M., Kim, H.M., and Kim, H.S. (2004). Expression of the human endogenous retrovirus HERV-W family in various human tissues and cancer cells. J Gen Virol *85*, 1203-1210.

Yi, J.M., and Kim, H.S. (2004). Evolutionary implication of human endogenous retrovirus HERV-H family. J Hum Genet *49*, 215-219.

Yolken, R.H., Karlsson, H., Yee, F., Johnston-Wilson, N.L., and Torrey, E.F. (2000). Endogenous retroviruses and schizophrenia. Brain Res Brain Res Rev *31*, 193-199.

Yolken, R.H., and Torrey, E.F. (1995). Viruses, schizophrenia and bipolar disorder. Clin Microbiol Rev *8*, 131-145.

Yu, C., Shen, K., Lin, M., Chen, P., Lin, C., Chang, G.D., and Chen, H. (2002). GCMa regulates the syncytin-mediated trophoblastic fusion. J Biol Chem *277*, 50062-50068.

Zhirnov, O.P., and Klenk, H.D. (1997). Histones as a target for influenza virus matrix protein M1. Virology *235*, 302-310.